

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

Τμήμα Μηχανικών Παραγωγής και Διοίκησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

“Εφαρμογή και σύγκριση μεθόδων
Πολυκριτήριας Ανάλυσης και Τεχνητής
Νοημοσύνης για την ανάλυση και
πρόβλεψη της συμπεριφοράς
καταναλωτών”

ΒΟΥΔΟΥΡΗΣ ΒΙΡΓΙΝΙΟΣ

Επιβλέπων: Νικ. Ματσατσίνης
Εξεταστική Επιτροπή: Αθ. Μυγδαλάς, Μιχ. Δούμπος

Χανιά 2003

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1^ο5

1.1 ΠΡΟΛΟΓΟΣ	6
1.2 ΕΙΣΑΓΩΓΗ ΣΤΗ ΣΥΜΠΕΡΙΦΟΡΑ ΚΑΤΑΝΑΛΩΤΩΝ	6
1.3 Η ΣΥΜΒΟΛΗ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ	8
1.4 ΠΛΗΡΟΦΟΡΙΕΣ ΓΙΑ ΤΗΝ ΕΡΕΥΝΑ ΑΓΟΡΑΣ	8
1.5 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ	10

ΚΕΦΑΛΑΙΟ 2^ο12

2.1 ΓΕΝΙΚΟ ΜΟΝΤΕΛΟ ΣΥΜΠΕΡΙΦΟΡΑΣ ΚΑΤΑΝΑΛΩΤΗ.....	13
2.2 ΔΙΑΔΙΚΑΣΙΑ ΙΚΑΝΟΠΟΙΗΣΗΣ ΠΕΛΑΤΩΝ.....	20
2.2.1 Η μέθοδος MUSA(Multicriteria satisfaction Analysis).....	22
2.3 ΑΠΟ ΠΟΥ ΑΝΑΚΥΠΤΕΙ Η ΑΝΑΓΚΗ ΕΡΕΥΝΑΣ ΑΓΟΡΑΣ.....	26
2.4 ΤΜΗΜΑΤΟΠΟΙΗΣΗ ΤΗΣ ΑΓΟΡΑΣ	27
2.5 ΈΡΕΥΝΑ.....	27
2.6 ΔΙΑΔΙΚΑΣΙΑ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ.....	30
2.7 ΠΡΟΤΕΙΝΟΜΕΝΗ ΜΕΘΟΔΟΛΟΓΙΑ.....	31

ΚΕΦΑΛΑΙΟ 3^ο33

3.1 WIZWHY	34
3.1.1 Γιατί είναι καλύτερο από άλλα data mining tools.	34
3.1.2 Πως λειτουργεί σε γενικές γραμμές.....	35
3.1.3 Τι ακριβώς είναι ένας κανόνας στο wizwhy	36
3.1.4 Που μπορεί να χρησιμοποιηθεί το wizwhy	37
3.1.5 Μαθηματικό υπόβαθρο του wizwhy.....	37
3.1.6 Πρόβλεψη.	41
3.1.7 Πόσοι κανόνες μπορούν να εξαχθούν απο μια βάση δεδομένων	41
3.2 NEUROSOLUTIONS	42
3.2.1 Νευρωνικά δίκτυα.....	42
3.2.2 Βασικές αρχές νευρωνικών δικτύων.....	43
3.2.3 Παραδείγματα εφαρμογών ΤΝΔ.....	45
3.2.4 Πλεονεκτήματα και μειονεκτήματα των ΤΝΔ.....	45
3.2.5 Επίπεδα χειρισμού του neurosolutions	46
3.2.6 Χρησιμοποιώντας το neurosolutions for excel.	47
3.2.7 Πως το neurosolutions υλοποιεί ένα νευρωνικό δίκτυο.	48
3.2.8 Αλγόριθμος back propagation.....	48
3.2.9 Multilayer perceptron.....	48
3.3 ROSETTA.....	50
3.3.1 Rough Sets.	51
3.3.2 Εφαρμογές των rough sets.	51
3.4 UTADIS.....	52

3.4.1 Βασικές αρχές του Utadis.....	52
3.4.2 Διαδικασία ανάπτυξης του προτύπου ταξινόμησης.....	55
3.4.3 Τα μαθηματικά πίσω από το Utadis.....	58

ΚΕΦΑΛΑΙΟ 4^ο65

4.1 ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ.....	66
4.2 ΕΦΑΡΜΟΓΗ WIZWHY	88
4.3 ΕΦΑΡΜΟΓΗ NEUROSOLUTIONS.....	102
4.4 ΕΦΑΡΜΟΓΗ ROSETTA.....	119
4.5 ΕΦΑΡΜΟΓΗ UTADIS	125

ΚΕΦΑΛΑΙΟ 5^ο129

5.1 ΣΥΜΠΕΡΑΣΜΑΤΑ	130
------------------------	-----

ΒΙΒΛΙΟΓΡΑΦΙΑ137

ΠΑΡΑΡΤΗΜΑ.....139

Στην οικογένεια μου.

Ευχαριστώ τον κύριο Ματσατσίνη Νικόλαο, τον κύριο Δούμπο Μιχάλη και την κυρία Κρασσαδάκη Λία, για την πολύτιμη βοήθεια που μου παρείχαν κατά την εκπόνηση της διπλωματικής μου εργασίας.

Ευχαριστώ, επίσης, όλους τους καθηγητές μου και την οικογένεια μου για την εκπαιδευτική και κοινωνική μου ανάπτυξη, καθώς και τους φίλους μου για τη στήριξη που μου προσέφεραν.

ΚΕΦΑΛΑΙΟ 1^ο

ΕΙΣΑΓΩΓΗ

1.1 Πρόλογος

Η διπλωματική αυτή έχει σαν σκοπό της την ανάλυση και πρόβλεψη της συμπεριφοράς καταναλωτών μέσω εφαρμογής και σύγκρισης μεθόδων πολυκριτήριας ανάλυσης και τεχνητής νοημοσύνης.

Δεν είναι λίγες οι φορές που γίνεται μια επιχειρηματική κίνηση και αποτυγχάνει. Αυτή μπορεί να είναι είτε το λανσάρισμα ενός καινούριου προϊόντος σε μια αγορά, είτε η ανάπτυξη μιας καινούριας αγοράς, ή ο σχεδιασμός ενός νέου προϊόντος και άλλες παρόμοιες κινήσεις που μπορεί να κάνει μια εταιρία προκειμένου να εξαπλωθεί στον καταναλωτικό χώρο. Για να υποστηριχθούν οι παραπάνω ενέργειες χρειάζεται να γίνει, αρχικά, μια σωστή και εμπειριστατωμένη μελέτη ή όπως λέμε μια έρευνα αγοράς.

Η έρευνα αγοράς γίνεται με σκοπό να διερευνηθούν ορισμένες παράμετροι της αγοράς και να βγουν ορισμένα συμπεράσματα για αυτήν. Δηλαδή πρέπει να μελετηθεί τι ενδεχομένως λείπει από τη συγκεκριμένη αγορά, τι είδους πελάτες-καταναλωτές διαθέτει αυτή, την οικονομική τους κατάσταση, την κοινωνική τους θέση, την ηλικία τους κτλ, σε ποιους απευθύνεται το προϊόν, τι θέση κατέχει στον ανταγωνισμό το τελευταίο και κάποια άλλα παρόμοια χαρακτηριστικά τα οποία αποτελούν πολύτιμες πληροφορίες για την εκάστοτε εταιρία, ώστε να σχεδιάσει κατάλληλα το προϊόν της, να κάνει σωστή διαφήμιση αυτού, να το κοστολογήσει ορθά, και να το παράγει-διανέμει σε σωστές ποσότητες. Αυτό θα αποτελέσει ένα ισχυρό έρεισμα για αυτήν ώστε η επένδυσή της να αποφέρει κέρδος.

Μια τέτοια έρευνα αγοράς έχει γίνει και στη δική μας περίπτωση, σε αγορά του εξωτερικού (Ολλανδία) το 1997-1998, και έχουν προκύψει τα δεδομένα που θα παρουσιαστούν παρακάτω. Πιο συγκεκριμένα, η έρευνα αφορά τρία προϊόντα έξτρα παρθένου ελαιολάδου και ένα οργανικό. Τα τρία είναι ιταλικά και είναι: **“LORENA”**, **“AH ORGANIC”**, **“BERTOLLI”**. Το τέταρτο είναι ελληνικό και μάλιστα από την Κρήτη και λέγεται **“SITIA”**. Τα δεδομένα μας αποτελούν οι απαντήσεις 553 ατόμων σε ένα σύνολο ερωτήσεων, που παρουσιάζονται στο παράρτημα και σχετίζονται με οικογενειακά χαρακτηριστικά, με γενικές ερωτήσεις περί ελαιολάδου καθώς και συγκεκριμένες γύρω από τα προαναφερθέντα προϊόντα λαδιού.

Αυτό που έχει σημασία σε αυτές τις περιπτώσεις είναι να μπορέσουμε με κατάλληλες μεθόδους να αναλύσουμε και να ερμηνεύσουμε τα δεδομένα μας. Εδώ λοιπόν κάνει την είσοδό της η επιστήμη της Τεχνητής Νοημοσύνης και της Πολυκριτήριας ανάλυσης και στόχος της εργασίας είναι η μελέτη και η σύγκριση των αποτελεσμάτων της επιμέρους εφαρμογής των μεθοδολογιών στο ίδιο σύνολο δεδομένων και η διερεύνηση της δυνατότητάς τους να αναλύουν και να προβλέπουν με ακρίβεια τη συμπεριφορά των καταναλωτών.

Επίσης έμφαση δίνεται στη μελέτη του χειρισμού των ποιοτικών δεδομένων και ιδιαίτερα του χειρισμού των δεδομένων της ποιοτικής κλίμακας (καλό, πολύ καλό, ...) και των προδιατάξεων στις οποίες αντιστοιχούν, μέσω της εφαρμογής των μεθοδολογιών.

1.2 Εισαγωγή στη Συμπεριφορά Καταναλωτών

Λέγοντας συμπεριφορά καταναλωτών εννοούμε τη μελέτη ενός συνόλου ατόμων(καταναλωτών) ή οργανισμών και των διαδικασιών που αυτοί χρησιμοποιούν για να επιλέξουν, να παράγουν και να διαθέσουν προϊόντα, υπηρεσίες και ιδέες για

την ικανοποίηση των αναγκών τους. Επιπρόσθετα μελετά το πώς επιδρά αυτό στον καταναλωτή[Del.I.Hawkins, Roger G.Best, Kenneth A.Coney, 2001].

Απαιτείται οι διάφορες επιχειρήσεις να μελετούν και να συλλέγουν πληροφορίες για την καταναλωτική συμπεριφορά σε καθημερινή βάση. Αυτό γίνεται διότι η γνώση της συμπεριφοράς των καταναλωτών και η συνεχής έρευνα και μελέτη της είναι πάρα πολύ σημαντική και επηρεάζει ολόκληρη τη παραγωγική διαδικασία του κάθε προϊόντος καθώς και τη δυνατότητα ανακύκλωσής του.

Στην εποχή μας η θεωρία της συμπεριφοράς του καταναλωτή δημιουργεί ερωτήματα στον παραγωγό σχετικά με την αποδοχή του εκάστοτε προϊόντος του στην αγορά, ο οποίος είναι σε θέση να απαντήσει μέσω ερευνών που κατά διαστήματα πρέπει να κάνει σχετικά με την πορεία του προϊόντος αυτού. Η σημασία των ερευνών της συμπεριφοράς του καταναλωτή είναι πολύ σημαντική και σύμφωνα με τα λεγόμενα ενός στελέχους επιχειρήσεων, «κάθε εβδομάδα ερευνητές της εταιρίας μιλούν με περισσότερους από 4000 καταναλωτές θέτοντάς τους τα εξής ερωτήματα:

- Τι πιστεύουν για τα προϊόντα της εταιρίας και τους ανταγωνιστές της.
- Ποιες βελτιώσεις στα προϊόντα αυτά θα μπορούσαν να γίνουν.
- Πως χρησιμοποιούν τα προϊόντα αυτά
- Τι πιστεύουν για το ρόλο τους στην οικογένεια και τη θέση τους στην κοινωνία και ποιες είναι οι ελπίδες τους και τα όνειρά τους για αυτούς και την οικογένεια τους»

Σήμερα περισσότερο παρά ποτέ η παραγωγή ενός προϊόντος δε μπορεί να θεωρηθεί δεδομένη, μιλώντας για μια συνεχώς μεταβαλλόμενη αγορά. Για αυτό η γνώση της συμπεριφοράς και η μελέτη της ενδεχόμενης αντίδρασης του καταναλωτή στο προϊόν είναι η βάση για το προγραμματισμό της παραγωγής του τελευταίου.

Η συμπεριφορά του καταναλωτή είναι μια πολύπλοκη και πολυδιάστατη διαδικασία και οι πρακτικές της αγοράς που εφαρμόζονται για τον επηρεασμό της συμπεριφοράς αυτής συχνά επηρεάζουν την ίδια την εταιρία παραγωγής των προϊόντων, τα άτομα και την κοινωνία.

Πρέπει να αναφερθεί ότι όλες οι έρευνες αγοράς βασίζονται στη μελέτη της συμπεριφοράς καταναλωτή για κάθε προϊόν. Αυτό σημαίνει πως για να επιβιώσει μια εταιρία σε αυτό το ιδιαίτερα ανταγωνιστικό περιβάλλον, πρέπει να προσφέρει στους καταναλωτές στους οποίους απευθύνεται, καλύτερα προϊόντα από αυτά των ανταγωνιστών. Το καλύτερο προϊόν για τον πελάτη, είναι αυτό που το σύνολο των πλεονεκτημάτων του είναι πιο ισχυρό από το κόστος απόκτησής του. Η στρατηγική της προσφοράς καλύτερου προϊόντος απαιτεί από τον παραγωγό πολύ καλή γνώση της αναμενόμενης αντίδρασης του ενδιαφερόμενου για το προϊόν αυτό, και άμεσο χρόνο δραστηριοποίησης σε σχέση με τον ανταγωνισμό. Τα παραπάνω είναι αυτά που τελικά προσδιορίζουν τη επιτυχία ή αποτυχία του προϊόντος στην αγορά.

Οι εταιρίες πρέπει να προσδιορίζουν τις ομάδες ατόμων, ή και τα νοικοκυριά που θα απευθυνθούν με βάση:

- Δημογραφικά στοιχεία
- Προτίμηση σε MME
- Γεωγραφικές τοποθετήσεις και άλλα ειδικότερα κριτήρια.

Ο ανωτέρω προσδιορισμός γίνεται πάντα έχοντας υπόψη τις δυνατότητες της εταιρίας σε σχέση με τον ανταγωνισμό, με δεδομένη την παρούσα και μελλοντική οικονομική και τεχνολογική κατάσταση αυτής.

1.3 Η συμβολή της Τεχνητής Νοημοσύνης

Είναι γνωστό πλέον ότι μπορούμε να μελετάμε τον τρόπο σκέψης των ανθρώπων και την διαδικασία απόκτησης της εμπειρίας τους. Επιπλέον, όμως προσπαθούμε να εμφυτεύσουμε στους υπολογιστές ένα είδος εξυπνάδας ώστε να «συμπεριφέρονται» σαν νοήμονα όντα και να βασίζονται στην αντίληψη. Αυτές οι λειτουργίες χαρακτηρίζουν αυτό που από την δεκαετία του 60 ονομάζουμε Τεχνητή Νοημοσύνη (artificial intelligence).

Η Τ.Ν είναι μια επιστήμη που μας δίνει τη δυνατότητα να εξάγουμε συμπεράσματα μέσα από ένα πλήθος δεδομένων, που μπορεί να αφορούν οποιοδήποτε τομέα. Με χρηματοοικονομικά μεγέθη ενδεχομένως κρίνουμε την βιωσιμότητα της εν λόγω επιχείρησης ή αν τα μεγέθη μας είναι από τον χώρο του marketing τότε μπορούμε να κρίνουμε για παράδειγμα αν θα πρέπει να λανσάρουμε το εν λόγω προϊόν κτλ. Όμως δεν είναι πάντα σίγουρη 100% η αξιοπιστία των συμπερασμάτων μας, οπότε πρέπει συγκρινόμενα και με άλλες μεθόδους να κρίνουμε την επιτυχία του συστήματός μας. Η συγκεκριμένη εργασία σκοπό έχει να γίνει μια τέτοια σύγκριση μεθόδων Τεχνητής Νοημοσύνης και μεθόδων πολυκριτήριας ανάλυσης πάνω στα συγκεκριμένα δεδομένα που μας έχουν δοθεί. Το ερώτημα είναι: «λειτουργούν σωστά και αξιόπιστα τα συστήματά μας ώστε να εξαχθούν ασφαλή συμπεράσματα;»

1.4 Πληροφορίες για την έρευνα αγοράς

Η έρευνα αγοράς έχει πραγματοποιηθεί στην Ολλανδία μέσω του project DIMITRA (1997-1998) στα πλαίσια ενός διατμηματικού προγράμματος που ονομάζεται “FAIR-PL 95-844” και οργανώθηκε από το Wageningen Agricultural University σε συνεργασία με το Εργαστήριο Συστημάτων Υποστήριξης Αποφάσεων του Πολυτεχνείου Κρήτης. Στην έρευνα παίρνουν μέρος 553 άτομα-ερωτηθέντες, και αυτοί αποτελούν το δείγμα μας. Το ερωτηματολόγιο που έχουν κληθεί να συμπληρώσουν παρατίθεται στο παράρτημα.

Υπάρχουν 22 ερωτήσεις προς απάντηση στο ερωτηματολόγιο μας. Πρέπει να σημειωθεί πως οι απαντήσεις που δίνουν οι καταναλωτές κωδικοποιήθηκαν όλες με ποιοτικές μεταβλητές για να μην χρησιμοποιηθούν περιφραστικές απαντήσεις, που θα μας δυσκόλευαν στη εφαρμογή των εργαλείων ανάλυσης των δεδομένων μας. Έτσι οι μεταβλητές αυτές αποτελούν ποιοτικά δεδομένα και όχι ποσοτικά. Οι επεξηγήσεις τους δίνονται παρακάτω:

Code	Question Text	Type	Multicriteria	Alternatives
1	Do you use Extra Virgin Olive oil for frying?	0	0	2

Συνέχεια:

Min	Max	Alt1	Text_alt1	Alt2	Text_alt2
1	2	1	Yes	2	No

- Το πρώτο κελί είναι ο αύξων αριθμός της ερώτησης
- Το δεύτερο είναι η ίδια η ερώτηση.

- Το τρίτο είναι ο τύπος της ερώτησης, δηλαδή αν είναι ποιοτική το δηλώνουμε με 0, αν είναι ποσοτική το δηλώνουμε με 2, και αν είναι ερώτηση κατάταξης(ranking) βάζουμε 3.
- Το τέταρτο κελί είναι η δήλωση αν είναι πολυκριτήρια(1) ή όχι (0)
- Το πέμπτο είναι οι ενδεχόμενες εναλλακτικές που έχουμε για την απάντηση ως προς το πλήθος.
- Το έκτο και έβδομο είναι η ελάχιστη και μέγιστη κωδικοποιημένη τιμή που μπορεί να έχει η απάντηση.
- Οι τελευταίες στήλες έπειτα είναι το πώς έχει γίνει η κωδικοποίηση, δηλαδή τι συμβολίζει το κάθε νούμερο στην περιγραφική μέθοδο.

Επόμενο παράδειγμα είναι :

44	What is your opinion about the Color of "Lorena" olive oil?	0	1	6
----	---	---	---	---

Συνέχεια:

1	6	1	very good	2	good	3	indifferent
---	---	---	-----------	---	------	---	-------------

Συνέχεια:

4	bad	5	very bad	6	I don't know
---	-----	---	----------	---	--------------

Αυτό είναι ένα παράδειγμα πολυκριτήριας ερώτησης και τα νούμερα εξηγούνται βάση των παραπάνω. Το επόμενο πλαίσιο είναι για ερώτηση κατάταξης:

64	Rank "Lorena"	3	1	4	1	4	1	first	2	second	3	third	4	fourth
----	---------------	---	---	---	---	---	---	-------	---	--------	---	-------	---	--------

Και τέλος οι γενικές ερωτήσεις κωδικοποιούνται :

Are you male or female?	0	0	2	1	2	1	Male	2	Female	0	0	0	0	0	0
What is your age?	0	0	3	1	3	1	17-28	2	29-45	3	46-65	4	66-90	0	0
level of education	0	0	3	1	3	1	primary school	2	high school	3	college-university	0	0	0	0
How many adults live in your household?	0	0	4	1	4	1	one	2	two	3	three	4	More than three	0	0
How many children live in your household?	0	0	6	1	6	0	0	1	one	2	two	3	Three	4	four
														5	more

Όπως βλέπουμε και οι ηλικίες έχουν ομαδοποιηθεί σε τέσσερις κατηγορίες. Άρα τελικά όλες οι απαντήσεις έχουν αριθμητική μορφή και συμβολίζουν ενδεχομένως μια περιγραφική απάντηση.

Δυστυχώς η ποιότητα των δεδομένων που συλλέχθηκαν δεν είναι και τόσο καλή επειδή υπήρχαν αρκετά κενά ή ορισμένες απαντήσεις χωρίς λογική. Μπορεί δηλαδή κάποιος να απάντησε με θετικό τρόπο για κάποιο προϊόν και ύστερα στις προτιμήσεις να μην του δίνει την αντίστοιχη αξία. Κάποιες τέτοιες αντιφατικές καταστάσεις δεν θα μπορούσαν σε καμιά περίπτωση να βοηθήσουν εμάς ή το πρόγραμμα να βγάλουμε ορθά συμπεράσματα. Σχεδόν όλες όμως οι εγγραφές, οι οποίες δεν θα μας έδιναν κανένα στοιχείο σωστής πληροφόρησης απομακρύνθηκαν από τη βάση δεδομένων μας και τελικά προέκυψαν 553 εγγραφές.

Μπορούμε να υποθέσουμε για τα δεδομένα μας ότι τα πέντε πρώτα πεδία αποτελούν κριτήρια επιλογής τιμής της κατάταξης(ranking). Δηλαδή είναι λογικό για τον καταναλωτή όταν δίνει απαντήσεις για το πώς βαθμολογεί το χρώμα, τη συσκευασία, τη φήμη, τη ποιότητα, και τη τιμή του κάθε προϊόντος λαδιού ανάλογα να κατατάσσει και το προϊόν αυτό. Αν, για παράδειγμα, κάποιος καταναλωτής υποστηρίζει ότι η ποιότητα του lorena είναι πολύ καλή(τιμή 1), είναι αναμενόμενο πως το κατατάσσει στις πρώτες επιλογές ανάμεσα στα άλλα τρία. Ομοίως ισχύει και για τα υπόλοιπα πεδία. Μπορούμε να υπολογίσουμε ωστόσο, τη συσχέτιση μεταξύ του καθενός πεδίου(από τα πέντε των ανεξάρτητων μεταβλητών) σε σχέση με αυτό της εξαρτημένης μεταβλητής(ranking) για να δούμε κατά πόσο είναι σε θέση η καθεμία από αυτές να επηρεάσει την τελευταία. Η συσχέτιση των πέντε πεδίων με το έκτο για το lorena, το ah organic, το sitia και το bertolli αντίστοιχα παρουσιάζεται παρακάτω:

συσχέτιση	χρώμα-κατάταξη	συσκευασία-κατάταξη	φήμη-κατάταξη	ποιότητα-κατάταξη	τιμή-κατάταξη
lorena	0,289590404	0,193953238	0,186303584	0,21482261	0,17179104
ah organic	0,062291174	0,106532424	0,162047589	0,210329415	0,225309256
sitia	0,20535309	0,169122042	0,243877857	0,271414225	0,082911244
bertolli	0,220001248	0,202414124	0,280242864	0,329127608	0,170914259

Πίνακας 1.1-Συσχετίσεις μεταβλητών

Είναι φανερό από τον παραπάνω πίνακα πως οι τιμές συσχέτισης είναι σχετικά χαμηλές και σε μερικές περιπτώσεις πλησιάζουν το μηδέν. Αυτό σημαίνει πως τα αντίστοιχα πεδία δεν έχουν σημαντική συσχέτιση, ενώ θα έπρεπε, και αυτό ενδεχομένως οφείλεται στη χαμηλή ποιότητα δεδομένων.

1.5 Δομή της εργασίας

- Αρχικά η εργασία ξεκινάει με μια μικρή αναφορά για το ποιος είναι ο σκοπός της και ποιες είναι οι βασικές θεωρίες που χρησιμοποιεί. Ύστερα σχολιάζονται τα δεδομένα που μας έχουν δοθεί από την έρευνα αγοράς. Τέλος παρουσιάζεται η δομή της εργασίας επιγραμματικά. Σχολιάζονται κάποια γενικά στατιστικά φαινόμενα όπως ποσοστά προτίμησης και οι συσχετίσεις τους με κάποια γενικά χαρακτηριστικά των ερωτηθέντων κτλ.
- Το δεύτερο κεφάλαιο αναλύει τη θεωρία της Συμπεριφοράς Καταναλωτών, αναφέρει σύγχρονες μεθοδολογίες που χρησιμοποιούνται και παρουσιάζεται η μέθοδος επίλυσης που προτείνουμε, δηλαδή τα επιστημονικά εργαλεία, με τα οποία επεξεργαζόμαστε τα δεδομένα μας.
- Το επόμενο κεφάλαιο(3^ο) αναφέρεται κυρίως στην ουσία των προγραμμάτων και τον τρόπο λειτουργίας τους, καθώς επίσης και στη θεωρία που κρύβεται πίσω από αυτά. Μεταξύ άλλων αναλύεται τι είναι το data mining και πως βοηθά στην ανάλυση των δεδομένων μας, τι είναι τα νευρωνικά δίκτυα κτλ, καθώς επίσης και το μαθηματικό υπόβαθρο πίσω από αυτά.

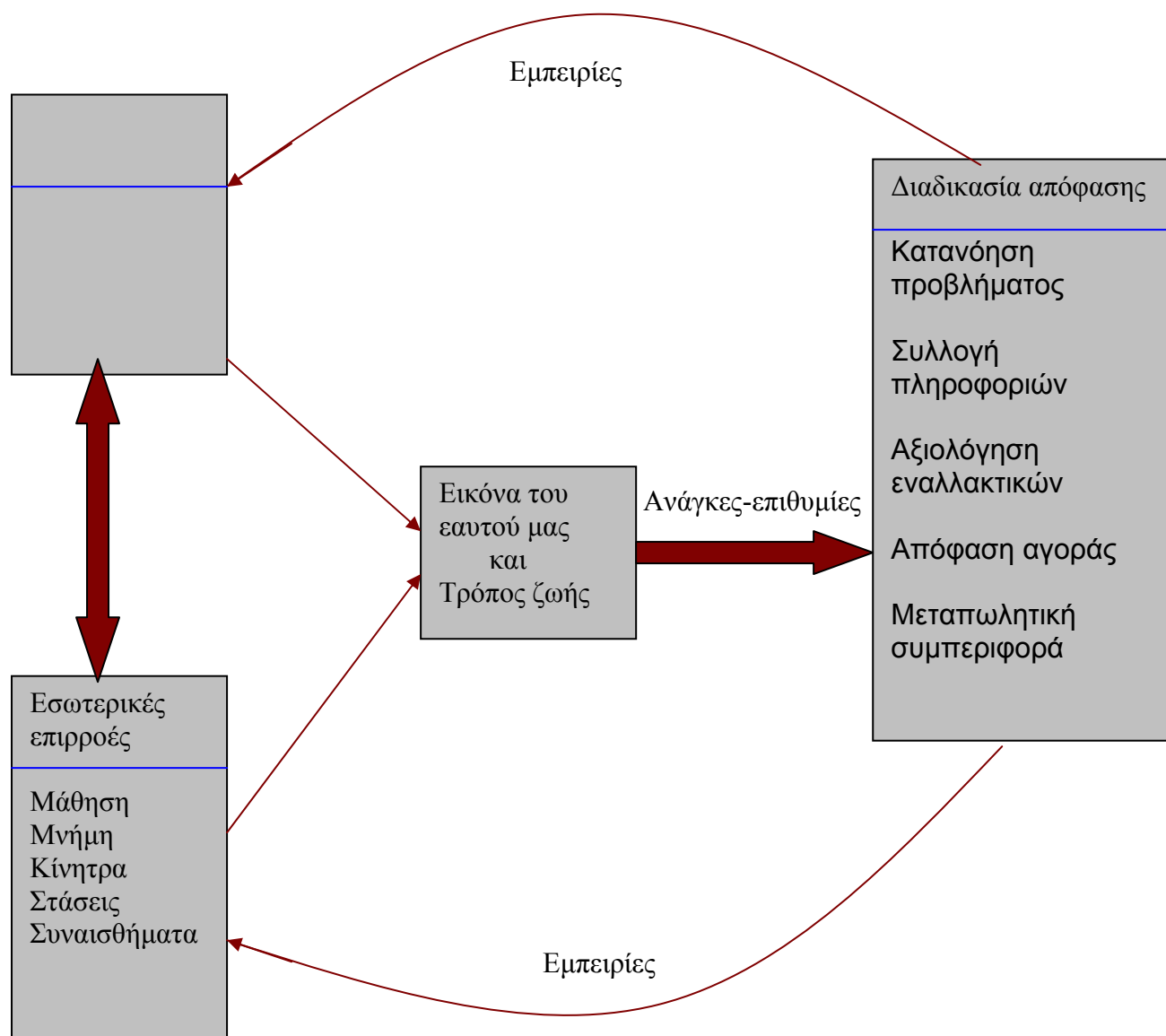
- Το κεφάλαιο τέσσερα περιέχει κάποιες πληροφορίες για το interface και την εφαρμογή των εργαλείων που περιγράφηκαν παραπάνω και γίνεται η στατιστική παρουσίαση των αποτελεσμάτων που πήραμε, η δημιουργία των πινάκων σφαλμάτων και τέλος παρατίθενται τα επιμέρους συμπεράσματα που προκύπτουν.
- Το πέμπτο κεφάλαιο περιέχει τα γενικά συμπεράσματα της εργασίας αυτής πάνω στην εφαρμογή των τεσσάρων προϊόντων λογισμικού, σύγκριση των τελευταίων, προβλήματα που προέκυψαν και γενικές παρατηρήσεις για τη χρησιμότητά τους στην πρόβλεψη της Συμπεριφοράς Καταναλωτών .

ΚΕΦΑΛΑΙΟ 2^ο

***Η ΕΠΙΣΤΗΜΗ ΤΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ
ΚΑΤΑΝΑΛΩΤΩΝ, Η ΧΡΗΣΙΜΟΤΗΤΑ
ΤΗΣ ΚΑΙ ΠΡΟΤΕΙΝΟΜΕΝΗ
ΜΕΘΟΔΟΛΟΓΙΑ***

2.1 Γενικό μοντέλο Συμπεριφοράς Καταναλωτή

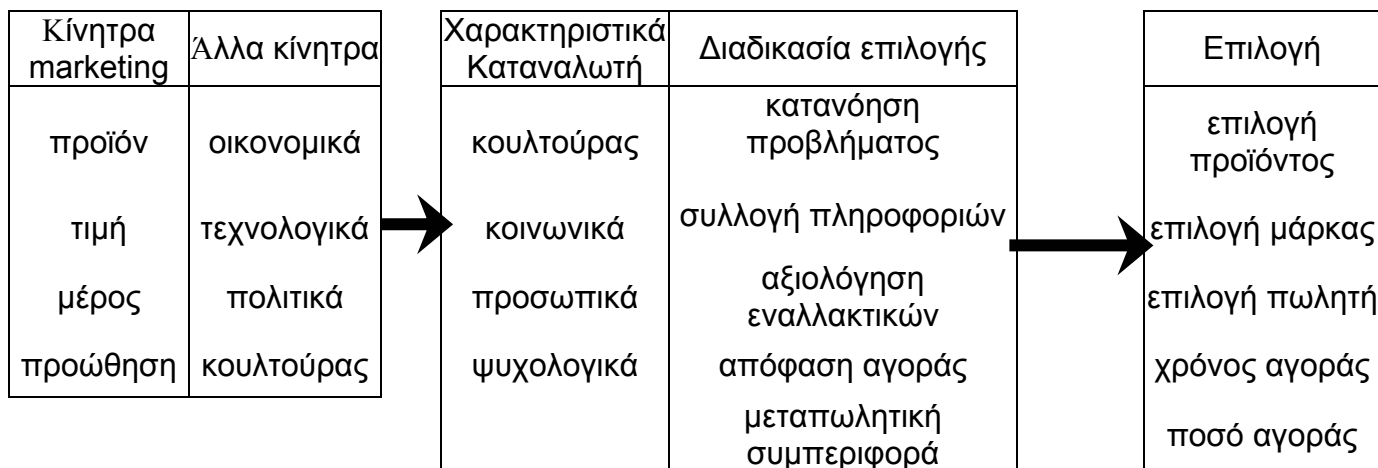
Η Συμπεριφορά Καταναλωτή είναι μια πολυσύνθετη διαδικασία που περιλαμβάνει τη ροή της αγοράς από τη σύλληψη της ύπαρξης της ανάγκης ή της επιθυμίας μέχρι και την χρησιμοποίηση των προϊόντων μετά την αγορά τους. Παραστατικά περιγράφεται:



Σχήμα 2α-Γενικό μοντέλο Συμπεριφοράς Καταναλωτή [Del.I.Hawkins, Roger G.Best, Kenneth A.Coney, 2001]

Τα κίνητρα του marketing και του περιβάλλοντος καθορίζουν τη συνείδηση του καταναλωτή. Τα χαρακτηριστικά αυτού και οι διαδικασίες απόφασης οδηγούν σε συγκεκριμένες κινήσεις αγοράς. Το έργο του ανθρώπου που κάνει το marketing (marketeer) είναι να καταλάβει τι ακριβώς συμβαίνει στη συνείδηση του καταναλωτή μεταξύ της εμφάνισης των εξωτερικών κινήτρων και της απόφασης αγοράς.

Η βάση για να καταλάβουμε τον τρόπο λειτουργίας της Συμπεριφοράς Καταναλωτή είναι το παρακάτω σχήμα:



Σχήμα 2β-Τρόπος λειτουργίας Συμπεριφοράς Καταναλωτή [Philip Cotler, 2003]

Η Συμπεριφορά Καταναλωτή επηρεάζεται από παράγοντες κουλτούρας , κοινωνικούς , προσωπικούς και ψυχολογικούς για να μπει τελικά στη διαδικασία της επιλογής[Philip Cotler, 2003].

Παράγοντες κουλτούρας.

Κουλτούρα, υποκουλτούρα, και κοινωνική θέση είναι σημαντικές παράμετροι της Συμπεριφοράς Καταναλωτή και πολλές φορές ασκούν τη μεγαλύτερη επιρροή. Η κουλτούρα είναι αυτή που καθορίζει σε πολύ σημαντικό βαθμό τα θέλω και τη συμπεριφορά ενός ατόμου. Κάθε παιδί καθώς μεγαλώνει υιοθετεί ένα σύνολο αξιών, προτιμήσεων, αντιλήψεων, και συμπεριφορών διαμέσου της οικογένειάς του και των κατεστημένων και κάθε κουλτούρα σχηματίζεται από μικρότερες υποκουλτούρες, που μπορεί να περιλαμβάνουν εθνικότητες, θρήσκευμα, φυλετικούς διαχωρισμούς και γεωγραφικές ιδιαιτερότητες. Όταν οι υποκουλτούρες είναι ισχυροποιημένες σε μια κοινωνία τότε συχνά οι εταιρίες σχεδιάζουν ειδικά πλάνα marketing για να συμβαδίσουν με αυτές.

Κάθε κοινωνία παρουσιάζει κοινωνική διαστρωμάτωση μέσα στην οποία τα μέλη ανάλογα με το που ανήκουν έχουν τους ρόλους τους. Τις περισσότερες φορές έχουμε την καθαρή μορφή των κοινωνικών τάξεων που είναι ιεραρχικά διατεταγμένες και τα μέλη της καθεμίας μοιράζονται παρόμοιες αξίες, ενδιαφέροντα και συμπεριφορές. Οι κοινωνικές τάξεις διαφέρουν όχι μόνο στο εισόδημα αλλά και στο επάγγελμα, στη μόρφωση, στη περιοχή διαμονής , στον τρόπο ενδυμασίας, στη προφορά και σε άλλα χαρακτηριστικά. Ξεχωριστά όμως ο καθένας μπορεί κατά τη διάρκεια της ζωής του να ανέβει ή να κατέβει στη κοινωνική δομή.

Κάθε κοινωνική τάξη δηλώνει ενδεχομένως προτιμήσεις σε αγαθά και μάρκες όπως για παράδειγμα την ενδυμασία, τα αυτοκίνητα την επίπλωση του σπιτιού κτλ. Διαφορές εντοπίζονται και στην προτίμηση των Μέσων Μαζικής Επικοινωνίας αφού, για παράδειγμα, οι υψηλές κοινωνικές τάξεις προτιμούν περιοδικά και βιβλία ή θέματα ειδησιογραφίας ενώ οι χαμηλότερες, τηλεόραση ή αθλητικά προγράμματα κτλ. Η έρευνα λοιπόν πρέπει να στρέφεται στην κατάλληλη κοινωνική τάξη.

Κοινωνικοί παράγοντες.

Η Συμπεριφορά Καταναλωτή επηρεάζεται από κοινωνικούς παράγοντες όπως είναι τα γκρουπ πληροφόρησης, η οικογένεια, οι κοινωνικοί ρόλοι και τα κατεστημένα.

- Τα γκρουπ πληροφόρησης ενός ατόμου αποτελούν όλοι εκείνοι που έχουν έμμεση ή άμεση επιρροή στη στάση και στη συμπεριφορά του ατόμου. Αυτοί μπορεί να είναι η οικογένεια, οι φίλοι, οι γείτονες κτλ με τους οποίους το άτομο είναι σε διαρκή επαφή και ανταλλαγή απόψεων. Οι άνθρωποι επηρεάζονται σημαντικά από τα γκρουπ πληροφόρησης, τα οποία οδηγούν σε νέες συμπεριφορές και τρόπους ζωής. Επίσης, επηρεάζονται και από γκρουπ, πολλές φορές, στα οποία δεν ανήκουν ούτε καν οι ίδιοι και τα οποία μπορεί να πρεσβεύουν μη αποδεκτές αντιλήψεις για αυτούς. Οι κατασκευαστές των προϊόντων σε περιπτώσεις όπου η επιρροή τέτοιων γκρουπ είναι πολύ σημαντική πρέπει να σχεδιάζουν και να προσεγγίζουν την άποψη των «ηγετών» αυτών.
- Η οικογένεια είναι ο πιο σημαντικός και θεμελιώδης λίθος της κοινωνίας. Τα μέλη της οικογένειας αποτελούν και αυτά ένα γκρουπ πληροφόρησης τρομερά σημαντικό, πόσο μάλλον σε περιπτώσεις όπου μιλάμε για οικογένειες με ένα γονιό και παιδί που ο επηρεασμός είναι καθημερινός. Οι marketers ασχολούνται με τους ρόλους και τις επιδράσεις που έχει το κάθε μέλος στο άλλο, αναφορικά με το ποιος επιθυμεί το προϊόν και ποιος πραγματοποιεί την αγορά του προϊόντος και πως αυτά ποικίλουν ανάλογα με τη χώρα και τη κοινωνική τάξη.
- Κάθε άτομο σήμερα έχει πολλούς ρόλους και κάθε ένας καθορίζεται από το κατεστημένο. Οι ρόλοι αποτελούν τις δραστηριότητες που ο άνθρωπος έχει να εκτελέσει και ρυθμίζονται αναλογικά με μια σχετική βαρύτητα στη ζωή του. Για παράδειγμα ένας δικαστικός έχει ρόλους πολύ πιο βαρείς από έναν διευθυντή πωλήσεων και η επιλογή αγοράς προϊόντων σχετίζεται άμεσα με αυτό. Δηλαδή ένας πρόεδρος εταιρίας συνήθως θα οδηγάει Mercedes και θα φοράει ακριβά ρούχα. Τέτοιου είδους φαινόμενα χρήζουν της προσοχής του marketer.

Προσωπικοί παράγοντες.

Η Συμπεριφορά Καταναλωτή επηρεάζεται επίσης από προσωπικά χαρακτηριστικά του όπως η ηλικία και η θέση στον κύκλο της ζωής του καταναλωτή, η εργασία του, οι οικονομικές συνθήκες, ο τρόπος ζωής, η προσωπικότητα και η αυτοεκτίμηση που αυτός έχει.

- Οι άνθρωποι αγοράζουν διαφορετικά αγαθά και υπηρεσίες κατά τη διάρκεια της ζωής τους. Το τι θα φάνε, πως θα ντυθούν, τι έπιπλα θα αγοράσουν κτλ έχει άμεση σχέση με την ηλικία. Οι marketers δίνουν μεγάλη προσοχή στην αλλαγή των συνθηκών ζωής ανάλογα με την ηλικία και στην επίδρασή τους στην κατανάλωση.
- Η εργασία σαφώς και επηρεάζει την καταναλωτική κατάσταση. Άλλα ρούχα αγοράζει ένας εργάτης και άλλα ο πρόεδρος μιας επιχείρησης, οπότε ο marketer αναγνωρίζει και στοχεύει στο καταναλωτικό κοινό που ενδιαφέρονται τα προϊόντα και οι υπηρεσίες του.
- Υπάρχουν άνθρωποι που έχουν την ίδια κουλτούρα, κοινωνική θέση και θέση εργασίας αλλά έχουν τελείως διαφορετικό τρόπο ζωής. Ο τρόπος ζωής χαρακτηρίζει για ένα άτομο τις δραστηριότητες, τα ενδιαφέροντα και τις

απόψεις του. Οι marketers ερευνούν τη σχέση μεταξύ των προϊόντων τους και των τρόπων ζωής και στοχεύουν καλύτερα σε αυτούς που πρέπει.

- Η προσωπικότητα μπορεί να αποτελέσει μια πολύ σημαντική παράμετρο στη ανάλυση επιλογής προϊόντων από τον καταναλωτή. Η ιδέα βρίσκεται στο γεγονός ότι και τα προϊόντα(μάρκες) έχουν την προσωπικότητά τους, και πως αρέσει στον καταναλωτή η άποψη να ταυτίζεται η μια προσωπικότητα με την άλλη. Οι marketers προσπαθούν να αναπτύξουν προσωπικότητες στις μάρκες που θα μπορούν να προσελκύσουν προσωπικότητες καταναλωτών.

Ψυχολογικοί παράγοντες.

Όταν ο καταναλωτής κάνει καταναλωτικές κινήσεις είναι επηρεασμένος από τέσσερις βασικούς ψυχολογικούς παράγοντες: τα κίνητρα, τη διαίσθηση, τη μάθηση και τα πιστεύω.

- Το άτομο γενικά έχει πολλές ανάγκες ανά πάσα στιγμή, κάποιες από αυτές είναι βιολογικές. Οι υπόλοιπες είναι ψυχογενείς και προέρχονται από την ανάγκη για αναγνώριση, σεβασμό, κτητικότητα κτλ. Μια ανάγκη γίνεται κίνητρο όταν είναι πολύ έντονη η επιθυμία της. Ένα κίνητρο είναι ανάγκη όταν σαφώς πιέζει το άτομο να ενεργήσει.

Οι καταναλωτές δεν αγοράζουν προϊόντα αλλά στην πραγματικότητα αγοράζουν λύσεις για τα προβλήματά τους ή ικανοποίηση των κινήτρων τους. Για παράδειγμα δεν αγοράζουν άρωμα ή aftersave αλλά αγοράζουν ρομαντικότητα, ευχαρίστηση για τις αισθήσεις, κουλτούρα και άλλα συναισθηματικά και ψυχολογικά οφέλη. Οι marketers πρέπει να ανακαλύπτουν τα κίνητρα, τα οποία τα προϊόντα ή οι υπηρεσίες τους καλούνται να ικανοποιήσουν, και να αναπτύσσουν τεχνάσματα γύρω από την ικανοποίηση αυτών.

Πολλές φορές οι καταναλωτές αγοράζουν κάποιο προϊόν και όταν ερωτηθούν γιατί το αγόρασαν απαντούν ότι ενδεχομένως «ταιριάζει στη μόδα, ταιριάζει πάνω μου» κτλ αλλά στην πραγματικότητα υπάρχουν άλλοι λόγοι που οι καταναλωτές δεν παραδέχονται ή δεν συνειδητοποιούν, όπως για παράδειγμα «φαίνομαι ότι έχω χρήματα» ή «με κάνουν σεξουαλικό» κτλ. Τα κίνητρα της πρώτης κατηγορίας, που είναι γνωστά και είναι εύκολο να τα παραδεχτεί ο καταναλωτής, ονομάζονται *φανερά κίνητρα*. Τα δεύτερου τύπου κίνητρα είναι είτε άγνωστα στον καταναλωτή είτε είναι απρόθυμος να τα παραδεχτεί και λέγονται *κρυμμένα κίνητρα* [Del.I.Hawkins, Roger G.Best, Kenneth A.Coney, 2001]. Στόχος του marketer είναι να καθορίσει τον συνδυασμό των κινήτρων που επηρεάζουν το τμήμα της αγοράς που τον ενδιαφέρει. Τα μεν φανερά τα συγκεντρώνει με απλές ερωτήσεις, τα δε κρυμμένα τα εκμαιεύει με πιο πολύπλοκες μεθόδους.

- Ένα άτομο που έχει κίνητρα είναι έτοιμο να δράσει. Το πώς πραγματικά θα δράσει εξαρτάται από την αντίληψη που έχει για τη συγκεκριμένη κατάσταση. Η διαίσθηση δεν εξαρτάται μόνο από τα φυσικά ερεθίσματα αλλά και από τα ατομικά σε σχέση με τα περιβάλλοντα. Πολλές φορές στο marketing η διαίσθηση είναι πολύ σημαντική, όπου για παράδειγμα ένας πελάτης κρίνει έναν πωλητή που μιλάει γρήγορα είτε ως επιθετικό και υποκριτή, είτε ως ευφυή και συνεργάσιμο.
- Όταν οι άνθρωποι δραστηριοποιούνται μαθαίνουν. Η μάθηση σχετίζεται με αλλαγές ατομικής συμπεριφοράς που προέρχονται από την εμπειρία. Για παράδειγμα αν αγοράσουμε ένα αγαθό και μείνουμε ευχαριστημένοι τότε θα

αγοράσουμε και κάποιο άλλο από την ίδια εταιρία εμπιστευόμενοι την πρώτη μας αγορά.

- Όταν οι άνθρωποι δημιουργούν και μαθαίνουν αποκτούν πιστεύω και νοοτροπία. Αυτά με τη σειρά τους καθορίζουν τη Συμπεριφορά Καταναλωτή. Έτσι οι marketeers ενδιαφέρονται να μάθουν για τα παραπάνω που κρύβει ο καθένας σε σχέση με τα προϊόντα και τις μάρκες που προσφέρονται. Οι άνθρωποι έχουν νοοτροπίες και στάσεις σχεδόν για τα πάντα και αυτά είναι που χαρακτηρίζουν την συμπάθεια ή την αντιπάθεια προς ένα προϊόν. Δεν αντιδρούν για κάθε ένα αγαθό από μηδενική βάση αλλά χειρίζονται ήδη υπάρχουσες σκέψεις.

Οι marketeers επεκτείνονται και πέρα από τις διάφορες επιρροές της αγοράς και αναπτύσσουν ένα επίπεδο αντίληψης για το πως πραγματικά οι καταναλωτές πραγματοποιούν αποφάσεις αγοράς. Συγκεκριμένα προσδιορίζουν ποιος ακριβώς παίρνει την απόφαση, τι τύπου απόφαση είναι και ποια είναι τα βήματα στη διαδικασία αγοράς.

Ρόλος του καταναλωτή.

Είναι εύκολο να προσδιορίσει κανείς ποιος είναι αυτός που κάνει τελικά την αγορά και τι ρόλο παίζουν οι υπόλοιποι στη διαδικασία απόφασης. Συνήθως, για παράδειγμα, οι άντρες αγοράζουν τον εξοπλισμό για το ξύρισμα ενώ οι γυναίκες επιλέγουν το άρωμά τους. Πολλές φορές όμως και αυτοί οι ρόλοι αλλάζουν. Γενικά μπορούμε να ξεχωρίσουμε πέντε ρόλους που οι άνθρωποι κατέχουν στη διαδικασία αγοράς[Philip Cotler, 2003]:

- Μνητής: αυτός που πρώτος προτείνει την ιδέα αγοράς ενός προϊόντος ή υπηρεσίας.
- Επιδρών: εκείνος του οποίου η άποψη ή η συμβουλή επηρεάζει την τελική απόφαση.
- Αποφασίζων: το άτομο που αποφασίζει όλες τις παραμέτρους της αγοράς, που, πως, τι κτλ.
- Αγοραστής: αυτός που πραγματικά κάνει την αγορά.
- Χρήστης: αυτός που τελικά χρησιμοποιεί το προϊόν.

Η διαδικασία απόφασης του καταναλωτή ποικίλει ανάλογα με το προϊόν που θα αγοραστεί. Επειδή δεν είναι ίδια η βαρύτητα του κάθε προϊόντος, οι αποφάσεις για παράδειγμα αγοράς μιας ρακέτας, ενός υπολογιστή, ενός αυτοκινήτου κτλ διαφέρουν πάρα πολύ ως προς τα άτομα που θα εμπλακούν σε αυτές, ως προς τη χρονική στιγμή που θα πραγματοποιηθεί, ως προς τον τρόπο πληρωμής κτλ.

Στάδια της διαδικασίας απόφασης αγοράς.

Για να καταλάβουν οι marketeers τα στάδια της διαδικασίας αγοράς των προϊόντων τους πρέπει να μουν οι ίδιοι στη θέση του καταναλωτή και να μελετήσουν πως εκείνοι θα αντιδρούσαν στην αντίστοιχη περίπτωση. Μπορούν επίσης να κάνουν ερωτήσεις σε ένα σύνολο καταναλωτών ώστε αυτοί να τους περιγράψουν τα βήματα που τους οδήγησαν στην αγορά του αγαθού. Άλλος τρόπος είναι να εντοπίσουν κάποιους καταναλωτές που είναι έτοιμοι να αγοράσουν το προϊόν και να τους ζητήσουν να περιγράψουν τις σκέψεις τους τη στιγμή που τις κάνουν προκειμένου να πραγματοποιήσουν την αγορά. Τέλος μια άλλη επιλογή είναι να

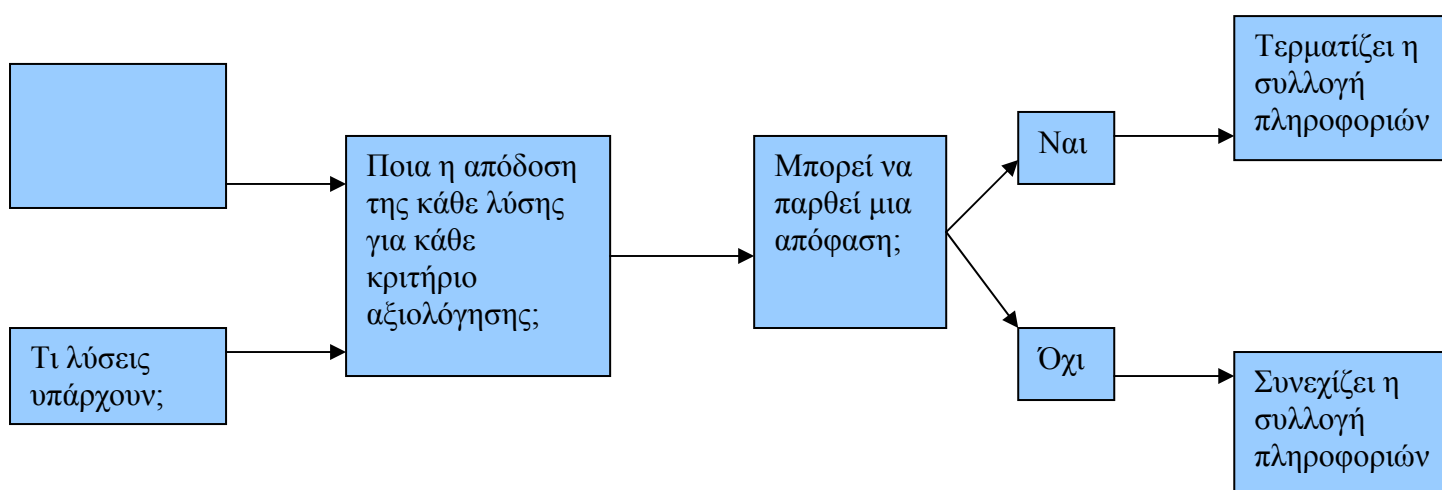
ζητήσουν από τους καταναλωτές να περιγράψουν τον ιδανικό τρόπο αγοράς του αγαθού. Κάθε μέθοδος θα φανερώσει τα βήματα της διαδικασίας αγοράς.

Προκειμένου ο καταναλωτής να αγοράσει ένα αγαθό, η σκέψη του θα περάσει από πέντε επίπεδα: κατανόηση προβλήματος, συλλογή πληροφοριών, αξιολόγηση εναλλακτικών, απόφαση αγοράς, μεταπωλητική συμπεριφορά. Ουσιαστικά δηλαδή η διαδικασία αγοράς ξεκινά πολύ πριν από την πραγματική αγορά και έχει συνέπειες και πολύ μετά από αυτήν.

➤ Κατανόηση του προβλήματος.

Η διαδικασία αγοράς ξεκινά όταν ο καταναλωτής αντιλαμβάνεται ένα πρόβλημα ή ανάγκη, που προέρχεται από εσωτερικά (όπως η πείνα, η δίψα κτλ) ή εξωτερικά ερεθίσματα(βλέποντας το αυτοκίνητο του γείτονα γεννιέται η επιθυμία απόκτησης του κτλ). Οι marketers χρειάζεται να μελετήσουν τις συνθήκες που δημιουργούν την εκάστοτε ανάγκη και αφού συγκεντρώσουν αρκετό όγκο πληροφόρησης τότε μπορούν να δημιουργήσουν στρατηγικές που θα ερεθίζουν το ενδιαφέρον του καταναλωτή.

➤ Συλλογή πληροφοριών.



Σχήμα 2γ-Συλλογή πληροφοριών για απόφαση καταναλωτή[Del.I.Hawkins, Roger G.Best, Kenneth A.Coney, 2001]

Όταν ο καταναλωτής προτίθεται να κάνει μια αγορά τότε συλλέγει συνεχώς πληροφορίες συνήθως κατά δύο τρόπους. Ο πρώτος τρόπος είναι απλά ακούγοντας και αποθηκεύοντας οποιαδήποτε πληροφορία αφορά το συγκεκριμένο προϊόν και ο δεύτερος είναι πιο ουσιαστικός, όπου ο ενδιαφερόμενος ψάχνει σε βάθος να βρει υλικό πληροφόρησης. Πηγές πληροφόρησης μπορεί να είναι είτε προσωπικές, από την οικογένεια, τους φίλους κτλ, είτε από πηγές εμπορίου(διαφημίσεις, πωλητές κτλ), είτε από κοινά μέσα ενημέρωσης(MME κτλ), είτε τέλος από προσωπική εμπειρία. Γενικά ο περισσότερος όγκος πληροφόρησης προέρχεται από εμπορικές πηγές ενώ η πιο καθοριστική μορφή είναι αυτή που προέρχεται από προσωπικές πηγές. Διαμέσου της πληροφόρησης μαθαίνει ο καταναλωτής για νέα

προϊόντα, για ανταγωνιστικά προϊόντα και για πληθώρα προδιαγραφών που αυτά έχουν.

Οι διάφορες εναλλακτικές που υπάρχουν σε μια διαδικασία απόφασης μπορούν να χωριστούν σε κάποιες κατηγορίες ανάλογα με τη προτίμηση που τους δείχνει ο καταναλωτής. Το συνολικό σετ είναι αυτό που περιέχει όλες ανεξαιρέτως τις εναλλακτικές. Από αυτές(διάφορες μάρκες προϊόντων για παράδειγμα) ο καταναλωτής θα έχει ακουστά ένα υποσύνολο αυτών που λέγεται γνωστό σετ. Κάποια προϊόντα από τα παραπάνω θα πληρούν τα βασικά κριτήρια του καταναλωτή και αποτελούν το υπό μελέτη σετ. Καθώς όμως ο καταναλωτής συλλέγει πληροφορίες, ορισμένες εναλλακτικές μένουν τελικά να «αναμετρηθούν» και το νέο υποσύνολο λέγεται σετ επιλογής, από το οποίο θα προκύψει και η τελική.

Κάθε επιχείρηση οφείλει να προσπαθεί ώστε τα προϊόντα της να μην αποτελούν απλές εναλλακτικές για τον καταναλωτή(να μην ανήκουν στο συνολικό σετ) αλλά να ανήκουν τουλάχιστον στο σύνολο προϊόντων λίγο πριν την τελική απόφαση(υπό μελέτη σετ τουλάχιστον). Οφείλει, επίσης, να μελετά τις ανταγωνιστικές εναλλακτικές ώστε να εισχωρεί με αξιώσεις στον ανταγωνισμό. Βασικό επίσης είναι να εξετάζονται οι πηγές πληροφόρησης και η βαρύτητα αυτών γιατί έτσι η εταιρία μπορεί να δημιουργήσει ενεργή και ουσιαστική επικοινωνία με τη συγκεκριμένη αγορά.

➤ Αξιολόγηση των εναλλακτικών.

Δεν υπάρχει μια τακτική αξιολόγησης των εναλλακτικών αλλά ποικίλει ανάλογα με τον αγοραστή ή/και το αγαθό. Αρχικά ο καταναλωτής προσπαθεί να ικανοποιήσει μια ανάγκη, ύστερα ψάχνει να επωφεληθεί όσο το δυνατόν περισσότερο από την αγορά που έκανε. Το αγαθό για αυτόν είναι ένα σύνολο χαρακτηριστικών που το καθένα θα ικανοποιεί και στους επιμέρους τομείς την συγκεκριμένη ανάγκη. Άρα η αξιολόγηση των εναλλακτικών γίνεται πρώτα αναλύοντας τα χαρακτηριστικά καθεμίας και τα οφέλη αυτών και έπειτα εκτιμώνται συνολικά τα ανταγωνιστικά προϊόντά.

➤ Απόφαση αγοράς.

Για να παρθεί μια απόφαση αγοράς από το σύνολο των εναλλακτικών που αποτελούν το σετ επιλογής υπάρχουν δυο παράγοντες που παρεμβαίνουν. Ο πρώτος είναι η στάση ή η άποψη των άλλων και αναφέρεται στην άποψη κάποιου τρίτου που μπορεί να είναι αρνητική για την εναλλακτική του καταναλωτή ή μπορεί τα κίνητρα του τελευταίου να μην συμβαδίζουν με την επιθυμία του πρώτου. Όσο πιο σημαντικό άτομο είναι αυτό για τον αποφασίζοντα τόσο πιο πολύ επηρεάζεται η άποψή του τελευταίου. Είναι αναμενόμενο ότι όσοι πιο πολλοί είναι αυτοί οι άλλοι τόσο πιο πολύπλοκο γίνεται το πρόβλημα της απόφασης. Ο δεύτερος παράγοντας είναι κάποιες αναπάντεχες συνθήκες που μπορεί να προκύψουν και να επηρεάσουν την διαδικασία της τελικής επιλογής(για παράδειγμα κάποια άλλη αγορά να προκύψει που είναι πιο σημαντική και επείγουσα).

Η επιλογή ενός καταναλωτή στο να αποδεχτεί ή να απορρίψει μια εναλλακτική δημιουργεί ένα ρίσκο, του οποίου η μείωση αποτελεί στόχο του καταναλωτή. Αυτό μπορεί να γίνει μέσω κατασκευής κατάλληλων αλγορίθμων μείωσης ρίσκου, μέσω περαιτέρω συλλογής πληροφοριών κτλ. Οι marketers πρέπει να μελετήσουν τους παράγοντες που προκαλούν την αίσθηση του ρίσκου στους καταναλωτές και να βρουν τρόπους μείωσης αυτού, ώστε να κερδίσουν την εμπιστοσύνη του πελάτη.

➤ Μεταπωλητική συμπεριφορά.

Μετά την αγορά ενός προϊόντος ο καταναλωτής θα βιώσει την ικανοποίηση ή την απογοήτευση της απόφασής του. Η θεωρία του marketing δεν τελειώνει όταν η αγορά γίνει αλλά συνεχίζει στοχεύοντας στη μεταπωλητική ικανοποίηση του πελάτη, στις μεταπωλητικές ενέργειες καθώς και στη μεταπωλητική χρησιμότητα των προϊόντων[Philip Cotler, 2003].

Μεταπωλητική ικανοποίηση.

Το πόσο ικανοποιημένος είναι ένας πελάτης σχετίζεται με το πόσο αυτό που αγόρασε ανταποκρίνεται στις προσδοκίες του. Πόσο ταυτίζεται αυτό που επιθυμούσε με αυτό που τελικά αγόρασε. Αν είναι χαμηλότερα από τις προσδοκίες του τότε είναι απογοητευμένος, αν τις έχει φτάσει είναι ικανοποιημένος και αν τις ξεπερνά τότε είναι υπερικανοποιημένος και θα ξαναπροτιμήσει το προϊόν ή την εταιρία και θα κάνει και καλή διαφήμιση στο περιβάλλον του.

Μεταπωλητικές ενέργειες.

Επειδή όπως λένε οι marketeers «η καλύτερη διαφήμισή μας είναι ένας ικανοποιημένος πελάτης», για αυτό οι κινήσεις που αυτοί κάνουν μεταπωλητικά είναι να ενισχύουν την ιδέα ότι οι πελάτες τους έκαναν μια πολύ καλή αγορά. Έστω η περίπτωση που οι πελάτες δεν είναι ικανοποιημένοι. Οι πωλητές θα πρέπει να μην δίνουν περιθώρια να εδραιωθεί η άσχημη άποψη που έχει δημιουργηθεί αλλά να είναι συνεχώς δίπλα στον πελάτη, να ακούν τα προβλήματά του και να του παρέχουν ότι χρειαστεί, όπως για παράδειγμα καλά εγχειρίδια χρήσης ή πληροφορίες για διαθέσιμο δίκτυο σέρβις κτλ.

Μεταπωλητική χρήση και διάθεση των προϊόντων.

Οι marketeers πρέπει να προσέχουν πως οι καταναλωτές χρησιμοποιούν τα αγαθά και τι χρήση τους κάνουν. Αν ένας πελάτης έχει το προϊόν σε ντουλάπι για παράδειγμα τότε αυτό σημαίνει προφανώς ότι δεν είναι ιδιαίτερα ικανοποιημένος από την αγορά του, ενώ αν πληρεί τις προδιαγραφές που αυτός έχει θέσει τότε εκμεταλλεύεται πλήρως τη χρησιμότητά του και είναι ικανός να βρει και νέους τρόπους χρησιμοποίησής τους.

2.2 Διαδικασία ικανοποίησης πελατών

Η βάση για τη Συμπεριφορά Καταναλωτή είναι το κατά πόσο κάποιο προϊόν θα μπορέσει να ικανοποιήσει το σύνολο των αναγκών του. Από αυτό θα εξαρτηθεί εν τέλει η αγορά ή όχι του προϊόντος. Η ικανοποίηση μπορεί αν κυμαίνεται από μηδενική ως πλήρης σε δύο επίπεδα: αυτό των πραγματικών αναγκών και αυτό των αντιλαμβανόμενων αναγκών. Άλλες φορές αυτά τα δύο συμπίπτουν και άλλες διαφέρουν.

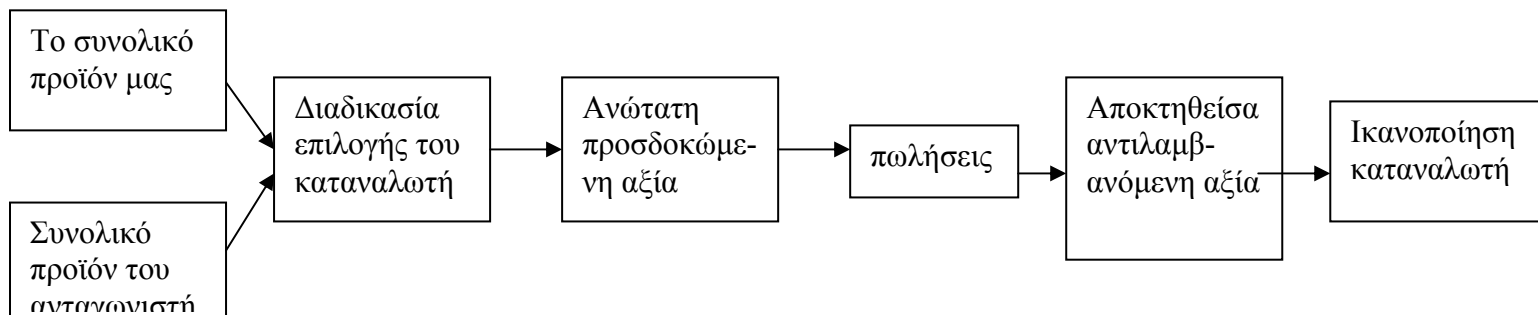
Η ικανοποίηση του πελάτη (customer satisfaction) ορίζεται με δύο βασικούς τρόπους[Γρηγορούδης Β. , Σίσκος Ι. , 2000]:

- αποτέλεσμα της εμπειρίας κατανάλωσης ή χρήσης ενός προϊόντος ή μιας υπηρεσίας (outcome).
- διαδικασία αξιολόγησης σύμφωνα με τις προσδοκίες του πελάτη (process).

Ο πρώτος τρόπος καθορίζει την ικανοποίηση ως μια τελική κατάσταση ή ως ένα αποτέλεσμα της εμπειρίας κατανάλωσης ή χρήσης ενός προϊόντος ή μιας υπηρεσίας.

Ο δεύτερος θεωρεί την ικανοποίηση ως μια διαδικασία, με έμφαση στους ψυχολογικούς παράγοντες αντίληψης και αξιολόγησης που την επηρεάζουν.

Οι marketers έχουν ανακαλύψει ότι είναι πιο σημαντικό να διατηρείς τους ήδη υπάρχοντες πελάτες από το να δημιουργείς ή να αντικαθιστάς με καινούριους. Για να διατηρήσει κανείς όμως τους πελάτες του πρέπει να μπορεί να ικανοποιεί τις ανάγκες τους στο μέγιστο βαθμό. Το βασικό συστατικό για να γίνει η πρώτη αγορά ενός προϊόντος είναι να μπορεί να πείσει η αντίστοιχη εταιρία για την αξία του αγαθού της. Από εκεί και ύστερα όμως έχει πολύ δουλειά για να πετύχει την πλήρη ικανοποίηση του πελάτη. Στο παρακάτω σχήμα φαίνεται παραστατικά το πώς δημιουργούνται ικανοποιημένοι πελάτες.



Σχήμα 2δ-Διαδικασία ικανοποίησης πελατών[Del.I.Hawkins, Roger G.Best, Kenneth A.Coney, 2001]

Έτσι το να δημιουργήσει μια εταιρία ικανοποιημένους πελάτες, οπότε και μελλοντικές πωλήσεις, απαιτεί οι πελάτες αυτοί να μείνουν ικανοποιημένοι και μετά την αγορά και χρήση του αγαθού. Για αυτό πρέπει ο πελάτης να αντικρίξει μεγαλύτερη αξία από αυτή που περίμενε και μεγαλύτερη ικανοποίηση αναγκών για το προϊόν που αγόρασε.

Η μέτρηση της ικανοποίησης του πελάτη μπορεί να βοηθήσει στην κατανόηση των γενικότερων αντιλήψεων του και πιο συγκεκριμένα στον προσδιορισμό και την ανάλυση των αναγκών, των προσδοκιών και των επιθυμιών του. Όσον αφορά το πρόβλημα της ύπαρξης διαφορετικής αντίληψης της ικανοποίησης ανάμεσα στον πελάτη και τη διοίκηση της εταιρείας, αυτό μπορεί να αμβλυνθεί με την υλοποίηση ενός προγράμματος μέτρησης ποιότητας.

Η σύγκριση της απόδοσης ενός προϊόντος ή υπηρεσίας θα πρέπει να γίνεται με τις επιθυμίες και όχι με τις προσδοκίες των πελατών. Η απόδοση των επιμέρους χαρακτηριστικών των προϊόντων, κυρίως στην περίπτωση διαρκών αγαθών, είναι ένας πολύ σημαντικός δείκτης της ικανοποίησης των πελατών. Η σύγκριση θα πρέπει να βασίζεται σε συγκεκριμένα πρότυπα απόδοσης που βασίζονται στην εμπειρία από τη χρήση (experience- based norms) όχι μόνο του προϊόντος/υπηρεσίας, αλλά και άλλων ανταγωνιστικών. Μερικά συστήματα μέτρησης ικανοποίησης του πελάτη είναι τα παρακάτω:

- Άμεσα συστήματα μέτρησης
- Έμμεσα συστήματα μέτρησης
- Στατιστικά δείγματα
- Συνεντεύξεις και έρευνες

2.2.1 Η μέθοδος MUSA(Multicriteria satisfaction Analysis)

Η μέθοδος MUSA[Γρηγορούδης Β. , Σίσκος Ι. , 2000] είναι ένα μοντέλο πολυκριτήριας μέτρησης της ικανοποίησης πελατών. Στηρίζεται στην ανάλυση της ικανοποίησης επιμέρους κριτηρίων, που απαρτίζουν τη συνολική ικανοποίηση του πελάτη, όπως σχηματικά φαίνεται παρακάτω:

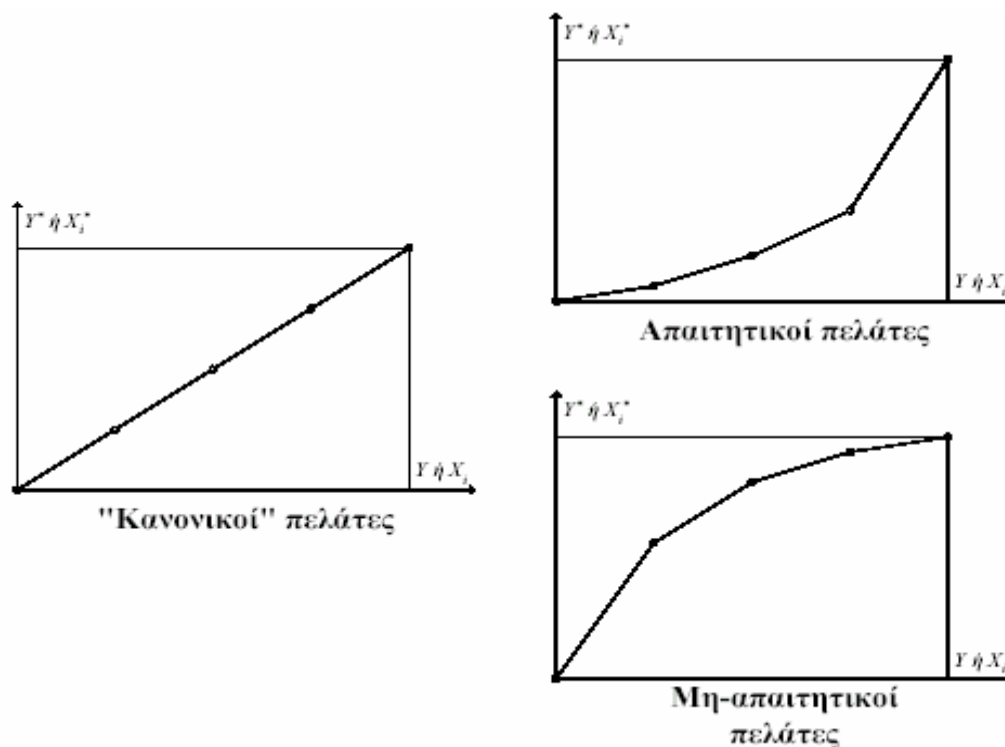


Σχήμα 2ε-Συνολική ικανοποίηση πελάτη[Γρηγορούδης Β. , Σίσκος Ι. , 2000]

Η συνολική ικανοποίηση του πελάτη εξαρτάται από ένα σύνολο μεταβλητών που αντιπροσωπεύουν χαρακτηριστικά του προσφερόμενου προϊόντος ή υπηρεσίας. Βασικός σκοπός της μεθόδου MUSA είναι η σύνθεση των προτιμήσεων ενός συνόλου πελατών σε μια μαθηματική συνάρτηση αξιών. Η μέθοδος ακολουθεί τις γενικές αρχές της ποιοτικής ανάλυσης παλινδρόμησης υπό περιορισμούς, χρησιμοποιώντας τεχνικές γ.π. Στόχος είναι η ελαχιστοποίηση του αθροίσματος των σφαλμάτων υπό τους περιορισμούς:

- εξίσωση παλινδρόμησης για κάθε πελάτη
- κανονικοποίησης των συναρτήσεων ικανοποίησης
- μονοτονίας των συναρτήσεων ικανοποίησης

Οι εκτιμώμενες συναρτήσεις ικανοποίησης εκφράζουν την πραγματική αξία που προσδίδει το σύνολο των πελατών σε ένα καθορισμένο ποιοτικό επίπεδο ικανοποίησης. Η μορφή των συναρτήσεων αυτών είναι σε θέση να προσδιορίσει το βαθμό απαιτητικότητας των πελατών.



Σχήμα 2στ-Συναρτήσεις ικανοποίησης[Γρηγορούδης Β. , Σίσκος Ι. , 2000]

Ως βάρη των κριτηρίων ικανοποίησης ορίζουμε το σχετικό βαθμό σπουδαιότητας που δίνει το σύνολο των πελατών στις αξίες των διαστάσεων ικανοποίησης που έχουν καθοριστεί. Οι μέσοι δείκτες ικανοποίησης και οι μέσοι δείκτες απαιτητικότητας είναι οι μεταβλητές που μας δείχνουν κατά πόσο οι πελάτες είναι ικανοποιημένοι και απαιτητικοί αντίστοιχα και υπολογίζονται από κάποιες εξισώσεις. Οι μέσοι δείκτες απαιτητικότητας είναι κανονικοποιημένοι στο διάστημα $[-1, 1]$ και ισχύει ότι για[Γρηγορούδης Β. , Σίσκος Ι. , 2000]:

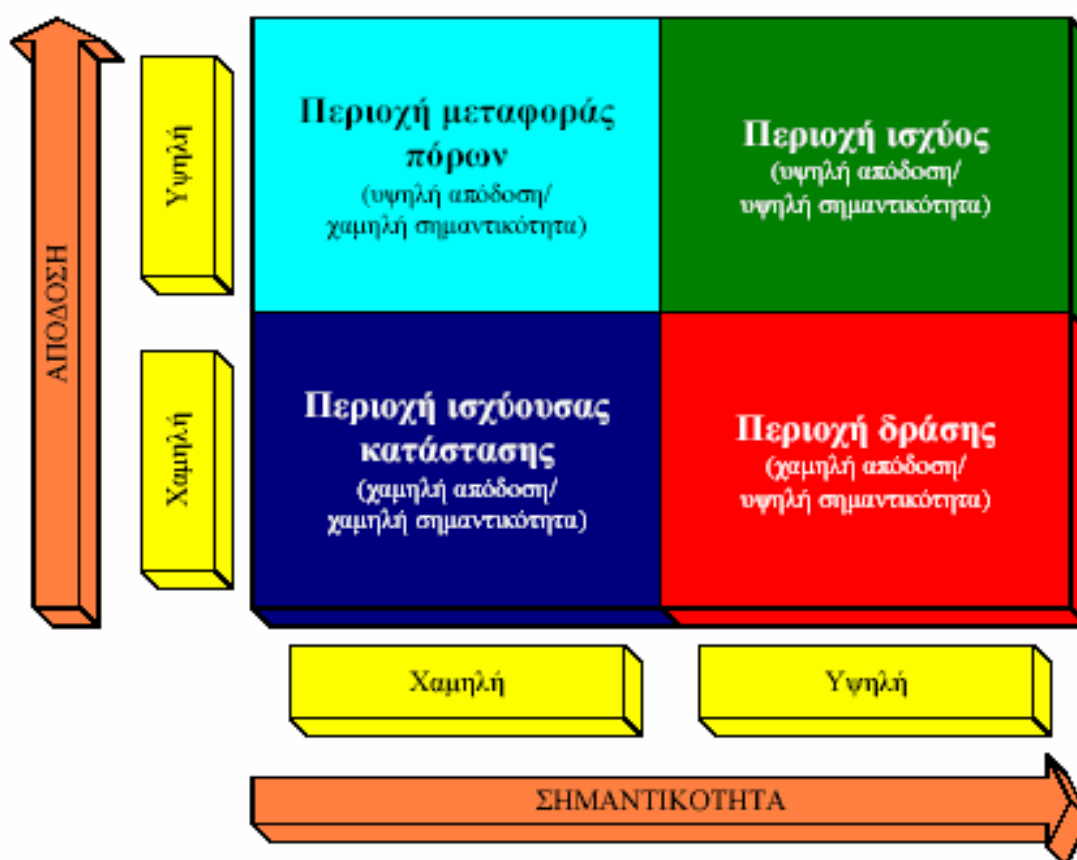
- **Ουδέτερους πελάτες:** η συνάρτηση ικανοποίησης έχει γραμμική μορφή, γεγονός που σημαίνει ότι οι συγκεκριμένοι πελάτες όσο περισσότερο ικανοποιημένοι δηλώνουν ότι είναι, τόσο μεγαλύτερο είναι το ποσοστό των προσδοκιών τους που εκπληρώνεται(Δείκτης 0).
- **Απαιτητικούς πελάτες:** η συνάρτηση ικανοποίησης έχει κυρτή μορφή, δεδομένου ότι η ομάδα αυτή των πελατών δεν είναι ικανοποιημένη παρά μόνο αν τους προσφέρεται το βέλτιστο επίπεδο υπηρεσιών(Δείκτης 1).
- **Μη απαιτητικούς πελάτες:** η συνάρτηση ικανοποίησης έχει κοίλη μορφή, γεγονός που υποδηλώνει ότι οι συγκεκριμένοι πελάτες δηλώνουν ότι είναι ικανοποιημένοι παρόλο που ένα μικρό ποσοστό των προσδοκιών τους εκπληρώνεται(Δείκτης -1).

Το αποτέλεσμα των ενεργειών βελτίωσης εξαρτάται τόσο από τη σημαντικότητα του κριτηρίου, όσο και από τη συνεισφορά του στη μη-ικανοποίηση (δυσανεπεία) των πελατών. Για το λόγο αυτό, ορίζεται ένα σύνολο μέσων δεικτών αποτελεσματικότητας. Οι συγκεκριμένοι δείκτες ορίζονται στο διάστημα $[0, 1]$ και δείχνουν τα περιθώρια βελτίωσης σε ένα συγκεκριμένο κριτήριο ικανοποίησης, λαμβάνοντας υπόψη και τη σπουδαιότητά του.

Συνδυάζοντας τα βάρη των κριτηρίων ικανοποίησης με τους μέσους δείκτες ικανοποίησης είναι δυνατός ο υπολογισμός μιας σειράς διαγραμμάτων δράσης

(action diagrams). Τα διαγράμματα δράσης μπορούν να προσδιορίσουν ποια είναι τα δυνατά και τα αδύνατα σημεία της ικανοποίησης των πελατών, καθώς και το που πρέπει να στραφούν οι προσπάθειες βελτίωσης. Κάθε διάγραμμα δράσης χωρίζεται σε τεταρτημόρια ανάλογα με την απόδοση (μέσοι δείκτες ικανοποίησης) και τη σημαντικότητα (βάρη) των κριτηρίων:

- **Περιοχή ισχύουσας κατάστασης** (χαμηλή απόδοση και χαμηλή σημαντικότητα)
- **Περιοχή ισχύος** (υψηλή απόδοση και υψηλή σημαντικότητα)
- **Περιοχή δράσης** (χαμηλή απόδοση και υψηλή σημαντικότητα)
- **Περιοχή μεταφοράς πόρων** (υψηλή απόδοση και χαμηλή σημαντικότητα)

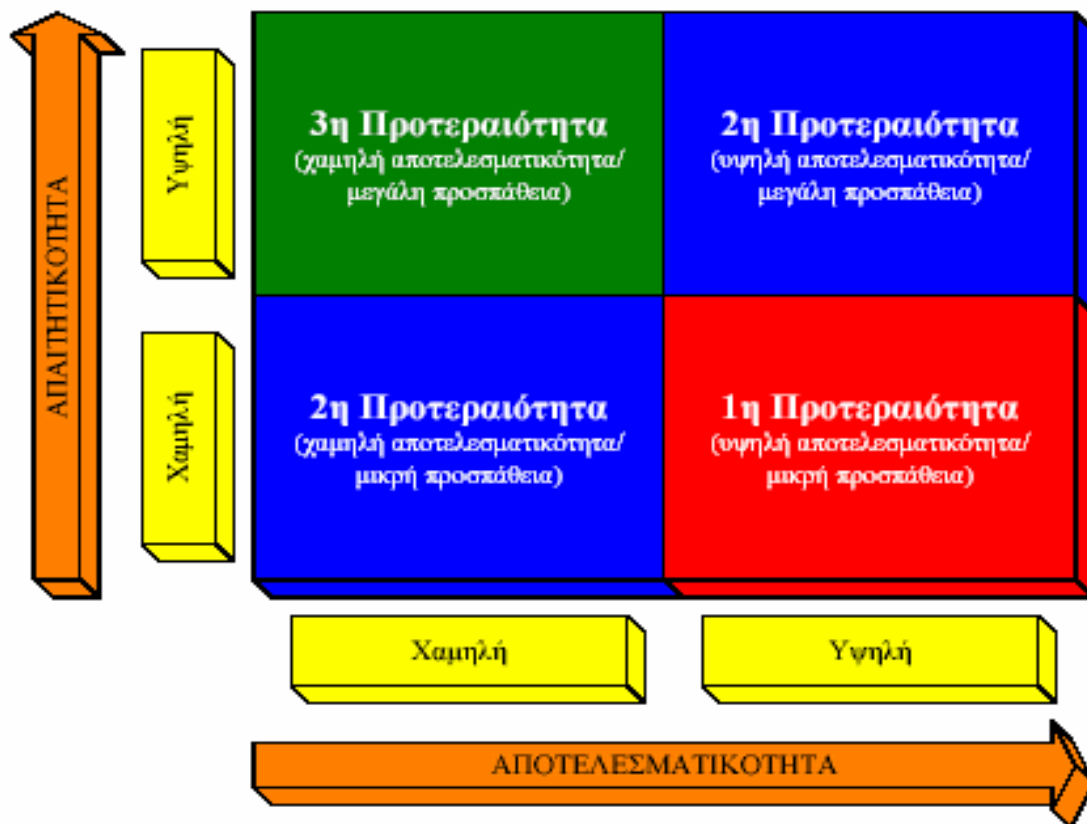


Σχήμα 2η-Διάγραμμα δράσης[Γρηγορούδης Β. , Σίσκος Ι. , 2000]

Όμοια, συνδυάζοντας τους μέσους δείκτες απαιτητικότητας και αποτελεσματικότητας κατασκευάζεται μια σειρά διαγραμμάτων βελτίωσης (improvement diagrams). Κάθε διάγραμμα βελτίωσης χωρίζεται σε τεταρτημόρια ανάλογα με την απαιτητικότητα και την αποτελεσματικότητα των διαστάσεων ικανοποίησης, με αποτέλεσμα τον προσδιορισμό των προτεραιοτήτων βελτίωσης:

- Η επιχείρηση θα πρέπει να επικεντρώσει τις προσπάθειες βελτίωσης στις διαστάσεις ικανοποίησης που έχουν μεγάλη αποτελεσματικότητα ενώ οι πελάτες δεν εμφανίζονται ιδιαίτερα απαιτητικοί.
- Η δεύτερη προτεραιότητα των ενεργειών βελτίωσης αποτελούν τα κριτήρια που είτε παρουσιάζουν μεγάλη αποτελεσματικότητα και μεγάλο βαθμό απαιτητικότητας, είτε εμφανίζουν μικρή αποτελεσματικότητα, ενώ οι πελάτες δε φαίνονται ιδιαίτερα απαιτητικοί.

- Τέλος, τα χαρακτηριστικά που παρουσιάζουν μικρή αποτελεσματικότητα και μεγάλη απαιτητικότητα αποτελούν την τελευταία προτεραιότητα βελτίωσης.



Σχήμα 2θ-Διάγραμμα βελτίωσης[Γρηγορούδης Β. , Σίσκος Ι. , 2000]

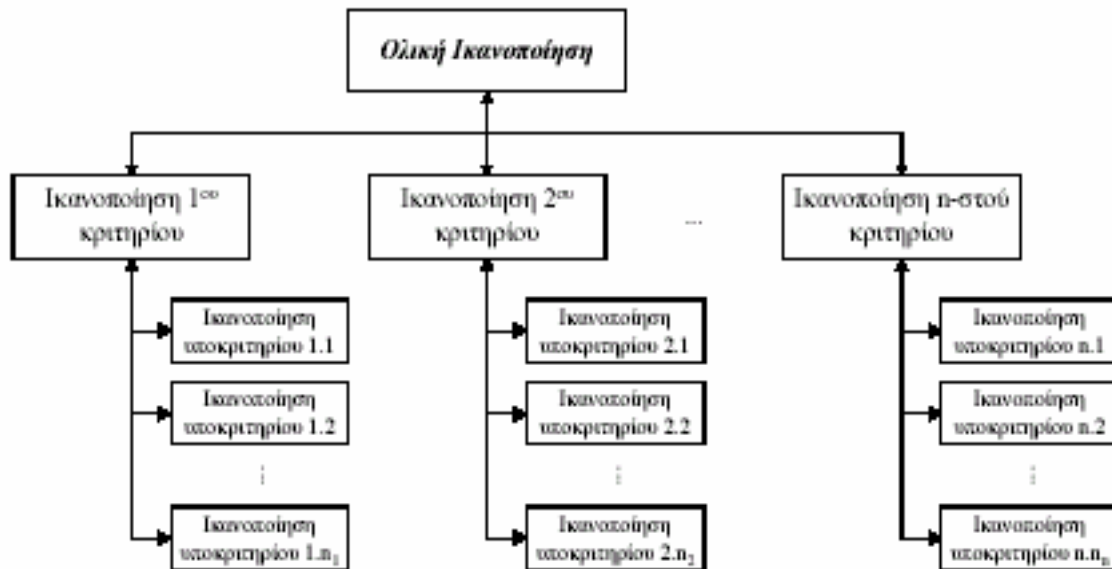
Η λογική ασυνέπεια των δεδομένων του προβλήματος εκτίμησης της ικανοποίησης επηρεάζει άμεσα την αξιοπιστία και την ευστάθεια της εκάστοτε μεθόδου. Για αυτό κατά τη διαδικασία εφαρμογής της μεθόδου MUSA θα πρέπει να υπάρχει ένα προκαταρκτικό στάδιο ελέγχου της λογικής συνέπειας των δεδομένων.

Η βασική μέθοδος MUSA εξετάζει το πρόβλημα της σύνθεσης της ικανοποίησης ενός συνόλου βασικών κριτηρίων. Όμως σε αρκετές περιπτώσεις, κρίνεται ιδιαίτερα χρήσιμος ο καθορισμός ενός επιπλέον επιπέδου ικανοποίησης όπου:

- Το πρώτο επίπεδο των κριτηρίων περιλαμβάνει τις βασικές διαστάσεις της ικανοποίησης των πελατών σε μια γενική μορφή (π.χ. προσωπικό).
- Το δεύτερο επίπεδο των κριτηρίων (υποκριτήρια) αφορά λεπτομερείς διαστάσεις των βασικών κριτηρίων ικανοποίησης (π.χ. συμπεριφορά, γνώσεις προσωπικού, κλπ).



(α) 1 επίπεδο κριτηρίων ικανοποίησης



(β) 2 επίπεδα κριτηρίων ικανοποίησης

Σχήμα 2ι-Πολλαπλά επίπεδα κριτηρίων[Γρηγορούδης Β. , Σίσκος Ι. , 2000]

2.3 Από πού ανακύπτει η ανάγκη έρευνας αγοράς.

Περίπου 126 εκατομμύρια αμερικανών έχουν τεθεί υπό συνέντευξη για οποιοδήποτε λόγο κάποια στιγμή στη ζωή τους. Κάθε χρόνο παραχωρείται συνέντευξη από περίπου 70 εκατομμύρια ανθρώπων στην Αμερική, που ισοδύναμα σημαίνει διάθεση τουλάχιστον δεκαπέντε λεπτών για κάθε ενήλικο το χρόνο. Οι συνεντεύξεις αυτές έχουν πολύ μεγάλη χρησιμότητα για την έρευνα αγοράς για τους παρακάτω λόγους[Carl McDaniel and Roger Gates, 2002]:

- Χρειάζεται να ξέρουμε το γιατί. Στην έρευνα αγοράς πρέπει να ξέρουμε γιατί οι άνθρωποι κάνουν ή δεν κάνουν κάτι. Για παράδειγμα γιατί αγοράζουμε ή δεν αγοράζουμε κάποια συγκεκριμένη μάρκα προϊόντος; Ή ποιος μας επηρέασε σε αυτήν μας την απόφαση;
- Χρειάζεται να ξέρουμε το πώς. Παράλληλα ο marketer πρέπει να ξέρει μέσα από ποια στάδια περνάει η σκέψη του καταναλωτή πριν καταλήξει στην τελική του απόφαση. Πως πήραν την απόφαση; Πόσο χρονικό διάστημα χρειάστηκε για αυτήν την διαδικασία; Πότε και που πάρθηκε η απόφαση;
- Χρειάζεται να ξέρουμε ποιος. Ο marketer μελετά και γνωρίζει το εκάστοτε άτομο ως προς τα δημογραφικά και προσωπικά του στοιχεία, μέσω ερωτήσεων για ηλικία, εισόδημα, επάγγελμα, μόρφωση κτλ.

Πρόβλεψη[Philip Cotler, 2003].

Πρόβλεψη είναι η τέχνη του να κάνει κανείς πρόγνωση στο πως θα ενεργήσουν οι καταναλωτές όταν τους δοθεί ένα πακέτο συνθηκών. Επειδή η Συμπεριφορά Καταναλωτή είναι πολύ σημαντική οι καταναλωτές πρέπει να ερευνώνται σε βάθος. Ένας πολύ σημαντικός παράγοντας για να γίνονται έρευνες αγοράς είναι το να καθοριστούν οι δυνατότητες της αγοράς. Όταν η έρευνα τελειώνει η εταιρία πρέπει να υπολογίζει και να προβλέπει το μέγεθος, την ανάπτυξη και το ενδεχόμενο κέρδος κάθε δυνατότητας της αγοράς. Η πρόβλεψη στις πωλήσεις, για παράδειγμα, πραγματοποιείται σε οικονομικά μεγέθη ώστε να αυξηθούν τα απαιτούμενα μετρητά για έρευνα και λειτουργία. Για τον τομέα των κατασκευών δύναται να υπολογιστεί η δυναμικότητα και η παραγωγή. Όσον αφορά την αγορά υπολογίζεται η σωστή ποσότητα αποθεμάτων, ενώ αναφερόμενοι σε ανθρώπινο δυναμικό υπολογίζονται, ενδεχομένως, οι απαιτούμενοι εργάτες.

2.4 Τμηματοποίηση της αγοράς

Ίσως η πιο σημαντική απόφαση που έχει να πάρει μια εταιρία είναι το που θα στοχεύσει για να προωθήσει τα προϊόντα της. Δηλαδή πρέπει να κάνει αυτό που λέμε τμηματοποίηση της αγοράς και να στραφεί σε εκείνους τους καταναλωτές που ενδιαφέρονται άμεσα για τα προϊόντα της. Μόνο έτσι μπορεί να ικανοποιήσει στο μέγιστο βαθμό τις επιθυμίες των πελατών της και δεν πρέπει να προσπαθεί να προσεγγίσει τα θέλω πολλαπλών τμημάτων της αγοράς. Όσο πιο μεγάλη είναι η τμηματοποίηση της αγοράς, άρα και πιο μικρά τα τμήματά της τόσο πιο καλά μπορεί μια εταιρία να ακολουθήσει τις επιθυμίες των πελατών της. Βέβαια η ιστορία μας δείχνει ότι όσο πιο μικρό τμήμα της αγοράς καλούμαστε να εξυπηρετήσουμε τόσο ανεβαίνει το κόστος, όπως για παράδειγμα τα ρούχα μαζικής παραγωγής είναι πιο φθηνά από αυτά που κατασκευάζονται ατομικά.

Για να τμηματοποιηθεί η αγορά πρέπει να γίνουν τα ακόλουθα τέσσερα βήματα[Del.I.Hawkins, Roger G.Best, Kenneth A.Coney, 2001]:

1. καθορισμός των αναγκών που ικανοποιεί το προϊόν.
2. ομαδοποίηση πελατών με παραπλήσιες ανάγκες.
3. περιγραφή κάθε ομάδας.
4. επιλογή ενός ελκυστικού τμήματος για εξυπηρέτηση.

Πρέπει να σημειωθεί πως γίνεται λόγος για σύνολο αναγκών διότι σε όλες πλέον τις αναπτυγμένες οικονομίες η αγορά ενός προϊόντος καλύπτει πάνω από μία ανάγκη. Ας πάρουμε για παράδειγμα το αυτοκίνητο. Μπορεί να είναι μέσο μεταφοράς, μέσο ευχαρίστησης και διασκέδασης, μέσο πρόσδωσης κύρους κτλ. Πολλές φορές ο καταναλωτής σταματά να είναι πελάτης μιας εταιρίας, παρόλο που είναι ικανοποιημένος από το προϊόν της, για λόγους μη ικανοποίησης της περιβαλλοντικής του ευαισθησίας, επί παραδείγματι.

2.5 Έρευνα.

Ανάλογα με τα δεδομένα που έχουμε να επεξεργαστούμε και το σκοπό που την κάνουμε διακρίνουμε την έρευνα σε ποιοτική και ποσοτική[Philip Cotler, 2003].

Ποσοτική: είναι συνήθως η έρευνα η οποία χρησιμοποιεί μαθηματική ανάλυση των δεδομένων για να συσχετίσει κάποιες ποσότητες ή να αποδώσει επίπεδα σημαντικότητας σε κάποιες μεταβλητές.

Ποιοτική: είναι συνήθως η έρευνα η οποία εξετάζει συμπεριφορές, αισθήματα κίνητρα κτλ όσον αφορά τους μετέχοντες στην έρευνα και είναι μια ανάλυση που γίνεται σε βάθος και με λίγα δεδομένα που όμως συνήθως περιέχουν και περιφραστικές απαντήσεις(ελεύθερες απαντήσεις). Αυτή η έρευνα είναι πιο ουσιαστικής και απαιτεί από τον «ερευνητή» να έχει ιδιαίτερες ικανότητες και γνώσεις πάνω σε ψυχολογία, κοινωνιολογία, marketing κτλ.

Μέθοδοι ποιοτικής έρευνας.

Focus Groups: αυτή η μορφή έρευνας πραγματοποιείται σε οκτώ με δώδεκα άτομα συνήθως που κατευθύνονται από τον «συντονιστή-ερευνητή» σε μια ουσιαστική, περιεκτική και πυκνή συζήτηση για ένα συγκεκριμένο θέμα ή θεωρία. Ο σκοπός αυτής της έρευνας είναι να μάθουμε και να καταλάβουμε τις απόψεις των ερωτηθέντων αλλά και το γιατί απάντησαν κατά αυτό τον τρόπο. Μας ενδιαφέρει το πώς νιώθουν για ένα προϊόν, ιδέα, οργανισμό κτλ, καθώς επίσης και πώς αυτά όλα εντάσσονται στη ζωή τους.

Depth interviews: είναι τύπος έρευνας όπου γίνεται συνέντευξη ένας με έναν και ο αναλυτής προσπαθεί να εκμαιεύσει και να αποσπάσει λεπτομερείς απαντήσεις πάνω στα θέματα προς έρευνα, χρησιμοποιώντας πολλές φορές εναλλακτικές μεθόδους για να αποκαλύψει κρυμμένα κίνητρα. Κατά τη διάρκεια της συνέντευξης οι απαντήσεις που δίνονται κάθε φορά αποτελούν τη βάση για καινούριες ερωτήσεις πιο ουσιαστικές και πιο βαθιές.

Projective tests: πολλές φορές οι τεχνικές αυτές είναι ενσωματωμένες σε Depth interviews. Οι βάσεις των projective τεχνικών προέρχονται από την κλινική ψυχολογία. Στη πραγματικότητα το αντικείμενο των τεστ αυτών είναι να τρυπώνουν κάτω από επιφανειακές απαντήσεις για να παρατηρήσουν πραγματικά συναισθήματα, νοήματα και κίνητρα, γιατί πολλές φορές οι άνθρωποι δεν θέλουν ή δεν μπορούν να εξωτερικεύσουν τα βαθύτερα συναισθήματά τους.

Οι πιο συνηθισμένες φόρμες των projective techniques που χρησιμοποιούνται στην έρευνα αγοράς είναι τα:

- word association tests, όπου ο αναλυτής λέει μια λέξη και ο ερωτώμενος πρέπει να πει την πρώτη σκέψη που έρχεται στο μυαλό του ,
- sentence and story completion tests, όπου οι ερωτηθέντες συμπληρώνουν προτάσεις ή ιστορίες με δικά τους λόγια,
- cartoon tests, όπου ο ερωτηθείς συμπληρώνει τα κείμενα διαλόγου ενός ή δύο προσώπων σε ένα καρτούν ,
- photo sorts, όπου ο εξεταζόμενος διαχωρίζει κάποιες φωτογραφίες ατόμων ανάλογα με το αν θα χρησιμοποιούσαν το υπό εξέταση προϊόν ή υπηρεσία,
- consumer drawings, οι ερωτηθέντες σχεδιάζουν πώς αισθάνονται ή πώς αντιλαμβάνονται ένα αντικείμενο,
- storytelling, όπου το εξεταζόμενο άτομο καλείται να αφηγηθεί ιστορίες για τις εμπειρίες του σε σχέση με ένα προϊόν ή μια εταιρία,
- και οι third person techniques, όπου ο αναλυτής μαθαίνει για τα συναισθήματα του ατόμου ρωτώντας τον διαμέσου τρίτων προσώπων,

δηλαδή, χωρίς να θέτει τον ίδιο μέσα στο πρόβλημα αλλά κάποιο υποθετικό πρόσωπο.

Τύποι έρευνας και ανάλυσης[Carl McDaniel and Roger Gates, 2002].

Για να προσεγγίσουμε τον καταναλωτή και να αντλήσουμε πληροφορίες για αυτό που μας ενδιαφέρει χρειάζεται να κάνουμε κάποιες ερωτήσεις, δηλαδή μια έρευνα. Τι τύπο έρευνας θα χρησιμοποιήσουμε όμως ποικίλει ανάλογα με την κάθε περίπτωση. Παρακάτω εξετάζονται έξι τρόποι έρευνας:

- Ο «Πόρτα-πόρτα συνέντευξη», στην οποία γίνονται ερωτήσεις ανά άτομο στο σπίτι τους. Ενδεχομένως να είναι η πιο παλιά μέθοδος συλλογής πληροφοριών και η οποία παρουσιάζει το πλεονέκτημα της άμεσης απάντησης ή της εξήγησης αν κάτι δεν είναι κατανοητό κτλ. Επίσης ο εξεταζόμενος βρίσκεται στο σπίτι του χωρίς άγχος έχοντας όλες του τις ανέσεις.
- Οι «Ανώτερες συνεντεύξεις» οι οποίες χρησιμοποιούνται αντίστοιχα με τις προηγούμενες αλλά σε βιομηχανικό περιβάλλον. Δηλαδή εδώ πλέον μιλάμε για συνεντεύξεις σε ανώτερα στελέχη επιχειρήσεων στα γραφεία τους για προϊόντα και υπηρεσίες που αφορούν τη βιομηχανία. Αυτός ο τύπος έρευνας κοστίζει περισσότερο, και αυτό γιατί οι ερωτηθέντες πρέπει να εντοπισθούν και να αναγνωριστούν και συνήθως αυτό γίνεται μέσω τηλεφώνου. Αφού εντοπισθούν τα κατάλληλα άτομα πρέπει να ρυθμιστούν τα αντίστοιχα ραντεβού που σε πολλές περιπτώσεις είτε καθυστερούν, είτε ακυρώνονται.
- «Συνεντεύξεις σε εμπορικά κέντρα» που είναι από τις πολύ γνωστές μεθόδους απόκτησης πληροφορίας και αποτελούν το ένα τρίτο του συνόλου των προσωπικών συνεντεύξεων. Γίνονται σε άτομα που έχουν βγει να ψωνίσουν σε εμπορικά κέντρα κτλ και είτε ερωτώνται επί τόπου, είτε σε κάποιο διαμορφωμένο χώρο της επιχείρησης που κάνει την έρευνα. Πολλές φορές δεν επιτρέπεται σε κάποια τέτοια κέντρα να γίνονται αυτές οι έρευνες γιατί ενοχλούνται κατά κάποιο τρόπο οι πελάτες. Αυτή η μέθοδος είναι περισσότερο οικονομική γιατί άλλωστε ο αναλυτής δεν χρειάζεται να μεταφέρεται συνεχώς από σπίτι σε σπίτι, ούτε να ψάχνει για καταναλωτές και συνήθως έρχονται μόνοι τους να ερωτηθούν. Κάποια μειονεκτήματα της μεθόδου είναι ότι πολλές φορές δεν είναι αντιπροσωπευτικό το δείγμα των ερωτηθέντων μέσα σε ένα εμπορικό κέντρο ή ότι άλλες φορές αρνούνται να συνεργαστούν λόγω πίεσης χρόνου κτλ.
- «Τηλεφωνικές συνεντεύξεις». Είναι από τις πολύ διαδεδομένες μεθόδους συλλογής πληροφοριών και συνήθως επιλέγονται τυχαία νούμερα μέσω αλγορίθμων γέννησης τυχαίων αριθμών και γίνονται τηλεφωνικές συνομιλίες που είναι και σχετικά ανέξοδες. Επίσης παρέχεται η δυνατότητα να ρυθμιστεί το δείγμα μας όπως επιθυμούμε. Παρόλα αυτά όμως υποστηρίζεται ότι μέσω τηλεφώνου δεν μπορεί να γίνει σωστή κρίση και αξιολόγηση των απαντήσεων που δίνονται ή ότι δεν μπορεί να γίνει μια βαθιά συζήτηση ιδίως σε ερωτήσεις που χρειάζονται περιφραστική απάντηση. Τέλος και σε αυτήν την περίπτωση δύναται να σπαταλείται πολύς χρόνος για να βρεθούν και να δεχτούν τα διάφορα άτομα τη συνέντευξη.
- Υπάρχει επίσης η μέθοδος των «Αυτοδιαχειριζόμενων ερωτηματολογίων». Σε αυτή τη μέθοδο δεν υπάρχει ερευνητής παρών αλλά τα άτομα συμπληρώνουν μόνα τους τις απαντήσεις που επιθυμούν, συνήθως μέσω του διαδικτύου ή σε κάποια ειδικά διαμορφωμένα κιόσκια που παρέχουν σύνδεση στο δίκτυο. Είναι σαφές ότι σε ενδεχόμενες ερωτήσεις των καταναλωτών δεν υπάρχει κανείς εκεί για να τους βοηθήσει. Αντίθετα πολλές φορές είναι καλό για να

μην πιέζεται ο ερωτηθείς και έτσι οι απαντήσεις του να είναι πιο ειλικρινείς και αυθόρμητες.

- Εναλλακτικός τρόπος έρευνας είναι και η «Δια αλληλογραφίας έρευνα», όπου χρησιμοποιούνται κυρίως δύο τύποι: ο «Μιας φορές αλληλογραφία» που στέλνονται ερωτηματολόγια σε επιλεγμένα άτομα ή διευθύνσεις χωρίς να γίνει κάποια προσυνεννόηση, και ο «αλληλογραφία με κατάλογο», όπου εκεί υπάρχει μια συνεχής συνεργασία της εταιρίας που κάνει την έρευνα με τους συμμετέχοντες, που έχουν επιλεγεί. Αυτοί ειδοποιούνται νωρίτερα και τους στέλνονται περιοδικά τα ερωτηματολόγια. Ο κατάλογος αυτός των ατόμων δημιουργείται αρχικά μέσω αλληλογραφίας όπου εξηγείται ο σκοπός της έρευνας και αυτοί ενδεχομένως δέχονται δωρεάν να συμμετέχουν. Ύστερα συμπληρώνουν κάποια ερωτηματολόγια με προσωπικές τους ερωτήσεις και τα οποία παρέχουν την απαραίτητη πληροφορία στην εταιρία της έρευνας ώστε να στέλνει όταν χρειάζεται τα κατάλληλα ερωτηματολόγια. Για παράδειγμα αν γίνεται μια έρευνα για σκυλοτροφή πρέπει να ερωτηθούν μόνο οι κάτοχοι σκύλων.

2.6 Διαδικασία ανάλυσης δεδομένων.

Όταν γίνει η συμπλήρωση των ερωτηματολογίων και συγκεντρωθούν όλα τα δεδομένα των απαντήσεων που έχουν δοθεί, ξεκινάει η διαδικασία ανάλυσής τους. Αυτό όμως αποτελεί μια πολύ δύσκολη διαδικασία λόγω του εξαιρετικά μεγάλου όγκου πληροφοριών σε πολλές περιπτώσεις. Αν φανταστούμε μια έρευνα αγοράς με 1300 ερωτηματολόγια των 10 σελίδων το καθένα καταλαβαίνουμε πως η μέθοδος της απλής ανάγνωσής και μελέτης τους είναι μάλλον αδύνατη. Έτσι οι επαγγελματίες αναλυτές ακολουθούν μια διαδικασία πέντε βημάτων για να αξιοποιήσουν τα δεδομένα που τους έχουν δοθεί[Carl McDaniel and Roger Gates, 2002]:

Βήμα πρώτο: έλεγχος αξιοπιστίας και διόρθωση.

Βήμα δεύτερο: κωδικοποίηση.

Βήμα τρίτο: είσοδος δεδομένων.

Βήμα τέταρτο: καθαρισμός δεδομένων από πρόγραμμα.

Βήμα πέμπτο: ταξινόμηση σε πίνακα και στατιστική ανάλυση.

1. Σκοπός μας εδώ είναι να σιγουρευτούμε ότι τα ερωτηματολόγια έχουν συμπληρωθεί σωστά, είναι πλήρη και όπως έχει οριστεί.

Πρέπει να εξακριβωθεί αν οι συνεντεύξεις συντάχθηκαν όπως έπρεπε, γιατί σε πολλές περιπτώσεις δεν μπορεί η εταιρία να ελέγξει το που και πως έγιναν οι συνεντεύξεις, για αυτό άλλωστε συμπληρώνονται πάντα τα προσωπικά στοιχεία του ερωτηθέντος. Είναι σύνηθες για τις εταιρίες που κάνουν τις έρευνες να ελέγχουν την αξιοπιστία των ερευνητών τους μέσω δειγματοληπτικού ελέγχου των ερωτηθέντων. Συνήθως δέκα με είκοσι τοις εκατό αυτών καλείται μέσω τηλεφώνου από την εταιρία για να επιβεβαιωθεί ότι έχουν συμμετάσχει στην έρευνα και ότι είχαν καλή συνεργασία με τον ερευνητή.

Η διόρθωση περιλαμβάνει έλεγχο των λαθών των απαντήσεων του ερωτηθέντα. Συνήθως αυτή η διαδικασία γίνεται εις διπλούν, μια φορά από την εταιρία που συνέταξε τα ερωτηματολόγια και μια από την εταιρία που κάνει την ανάλυση. Η διόρθωση έχει να κάνει με περιπτώσεις όπου δεν ερωτήθηκαν όλες οι ερωτήσεις ή μερικές απαντήσεις αντιγράφηκαν, με κομμάτια που ανάλογα την απάντηση παραπέμπουν σε άλλες ερωτήσεις ή με

περιπτώσεις που παραφράστηκαν τα λόγια του καταναλωτή σε ελεύθερες ερωτήσεις.

Είναι μια αρκετά επίπονη και χρονοβόρα διαδικασία να γίνει το πρώτο βήμα, ωστόσο είναι βασικό για τη σωστή προσέγγιση των απαντήσεων.

2. Εδώ γίνεται η διαδικασία της ομαδοποίησης και του προσδιορισμού αριθμητικής κωδικοποίησης των εναλλακτικών απαντήσεων σε μια ερώτηση. Πολλές απαντήσεις είναι κωδικοποιημένες εξ αρχής σε αριθμητικές απαντήσεις που είναι εύκολο σχετικά να συγκεντρωθούν και να μελετηθούν, όπως για παράδειγμα «πολύ καλό-1, καλό-2» κτλ. Άλλες ερωτήσεις όμως είναι οι λεγόμενες ελεύθερες όπου εκεί ο καθένας απαντά με τον δικό του τρόπο. Αυτές αποτελούν τις πιο δύσκολες περιπτώσεις και για να κωδικοποιηθούν πρέπει να ομαδοποιηθούν ενδεχομένως οι απαντήσεις, που έχουν δοθεί, όσο γίνεται καλύτερα και να αποδοθούν τελικά και σε αυτές αριθμητικές αντιστοιχίες. Αξίζει να σημειωθεί ότι σήμερα είναι δυνατό να γίνει η παραπάνω κωδικοποίηση μέσω υπολογιστή.
3. Σε αυτό το βήμα πρέπει να περαστούν τα δεδομένα μας, που είναι πλέον ελεγμένα και κωδικοποιημένα, σε ηλεκτρονική μορφή στον υπολογιστή, ώστε να γίνει η περαιτέρω επεξεργασία τους. Οι απαντήσεις που έχουν δοθεί χρειάζονται προσοχή στη μεταφορά στον υπολογιστή στα κατάλληλα πεδία που υπάρχουν για κάθε απάντηση. Είναι συχνό φαινόμενο να γίνονται λάθη στη μεταφορά των δεδομένων.
4. Σε αυτό το βήμα αφού έχει γίνει η καταχώρηση των δεδομένων στον υπολογιστή μένει να γίνει κάποια μορφή επεξεργασίας από ειδικά προγράμματα, ώστε αυτά να καθαριστούν, όπως λέμε. Δίνεται η δυνατότητα να κατασκευαστούν αλγόριθμοι που θα κάνουν αυτή τη δουλειά και θα ελέγχουν τη συμβατότητα των αποτελεσμάτων, καθώς επίσης είναι δυνατό να παρουσιαστεί και μια αναφορά με το τι λάθη εντοπίστηκαν και που. Αυτό είναι το τελικό βήμα πριν να γίνει η επεξεργασία των αποτελεσμάτων.
5. Είμαστε πλέον έτοιμοι να κάνουμε την στατιστική ανάλυση και να δούμε τα αποτελέσματα της έρευνάς μας συγκεντρωτικά σε πίνακες με τις απαντήσεις των καταναλωτών. Μπορούν να κατασκευαστούν πίνακες, όπου για κάθε ερώτηση υπάρχει ο αριθμός των απαντήσεων, ή πίνακες που να συσχετίζουν τις απαντήσεις μιας ερώτησης με τις απαντήσεις άλλων ερωτήσεων κτλ. Μπορούν επίσης να κατασκευαστούν διαγράμματα (γραφικές παραστάσεις, πίτες, ράβδοι κτλ) που μπορούν παραστατικά, να μας δώσουν την πληροφόρηση που θέλουμε.

2.7 Προτεινόμενη μεθοδολογία

Τα εργαλεία που θα χρησιμοποιήσουμε είναι από το χώρο της Τεχνητής Νοημοσύνης και της Πολυκριτήριας Ανάλυσης και ο τρόπος λειτουργίας τους στηρίζεται στην ίδια βάση.

Αντικείμενο της εργασίας αυτής είναι να προβλέψουμε την προτίμηση καταναλωτών όσον αφορά την κατάταξη των τεσσάρων προϊόντων μας. Από τη στιγμή που έχουμε 553 εγγραφές δεδομένων στη διάθεσή μας, μπορούμε να στηριχτούμε στη «γνώση» και την «εμπειρία» που μας προσφέρει η μελέτη αυτών και

να κάνουμε πρόγνωση των απόψεων κάποιων νέων καταναλωτών. Πρέπει να έχουμε πάντα υπόψιν μας βέβαια ότι στηρίζομαστε σε πιθανότητες που απλώς προσπαθούμε να μεγιστοποιήσουμε με όσο το δυνατόν περισσότερη εκπαίδευση, άρα πλήθος δεδομένων.

Η βάση, λοιπόν, στην οποία στηρίζονται τα εργαλεία που θα χρησιμοποιήσουμε είναι το γεγονός της εκπαίδευσής τους. Δηλαδή όλα τα συστήματα που θα χρησιμοποιήσουμε εκπαιδεύονται, παρέχοντάς τους ένα πλήθος δεδομένων, τα οποία τα χρησιμοποιούν το καθένα με το δικό του τρόπο, ώστε δίνοντας τους νέα δεδομένα να μπορούν να παράγουν τη ζητούμενη πρόβλεψη. Ο τρόπος με τον οποίο αξιοποιεί το καθένα τα δεδομένα εκπαίδευσης αποτελεί και τη διαφορετικότητα του κάθε εργαλείου. Άλλα μπορεί να στηρίζονται στη δημιουργία σχέσεων μεταξύ των ποσοτήτων που παίρνουν μέρος στην έρευνα, άλλα μπορεί να παράγουν λογικές συσχετίσεις μεταξύ των ποσοτήτων, άλλα δημιουργώντας νευρωνικά δίκτυα κτλ.

Τα εργαλεία που θα χρησιμοποιήσουμε είναι το WIZWHY της Wizsoft και χρησιμοποιεί την τεχνική του data mining , το NEUROSOLUTIONS της Neurodimension, που στηρίζεται στα νευρωνικά δίκτυα , το ROSETTA από το Norwegian University of Science&Technology και βασίζεται στη θεωρία των rough sets και τέλος το UTADIS του Devaud et al. (1980), των Jacquet–Lagrèze και Siskos (1982), Zorounidhs και Doubos, χρησιμοποιώντας μεθοδολογία πολυκριτήριας ανάλυσης. Χρησιμοποιούμε τις παραπάνω μεθοδολογίες χειρισμού των δεδομένων μας ώστε να αποφανθούμε ποια λειτουργεί καλύτερα, δηλαδή ποια μας δίνει ποιοτικότερα αποτελέσματα στη πρόβλεψη των τελικών επιλογών των καταναλωτών.

ΚΕΦΑΛΑΙΟ 3^ο

ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

3.1 Wizwhy

Ένα από τα πακέτα λογισμικού που θα χρησιμοποιηθούν και θα μας βοηθήσουν στην πρόβλεψη της συμπεριφοράς καταναλωτή είναι το wizwhy της Wizsoft[www.wizsoft.com]. Αυτό στηρίζεται γενικά στη μεθοδολογία που είναι γνωστή ως «data mining»[Wizsoft, 1996]. Πιο συγκεκριμένα, η δουλειά αυτού του προγράμματος είναι να μελετά τα δεδομένα μας και μετά να προσπαθεί να δημιουργήσει κανόνες και σχέσεις μεταξύ των μεγεθών που περιγράφουν αυτά ούτως ώστε να μπορεί ύστερα να προβλέψει μελλοντικές τιμές όταν του ζητηθεί.

Πολλές βάσεις δεδομένων περιέχουν πολύτιμες πληροφορίες που δεν είναι προφανείς οπτικά. Για παράδειγμα κρυφές πληροφορίες μπορεί να έχουμε στις παρακάτω περιπτώσεις:

- Στη σχέση μεταξύ του ιατρικού ιστορικού ασθενών και τη διάγνωσή της υγείας τους
- Μέσα στις βάσεις δεδομένων των οικονομικών των επιχειρήσεων, όπου παρατηρούνται κάποιες επιχειρήσεις να είναι υψηλού ρίσκου όπως λέγονται.
- Τύπους καταναλωτών οι οποίοι πιθανώς να ανήκουν σε κατηγορία υψηλών λογαριασμών κατανάλωσης .

Ψάχνοντας λοιπόν γι' αυτές τις πολύτιμες πληροφορίες , που είναι ωστόσο «κρυμμένες» στις βάσεις δεδομένων, λέμε ότι κάνουμε «data mining». Μέσω αυτής της μελέτης δύναται να δημιουργούνται κανόνες που μπορούν να μας βοηθήσουν να κάνουμε προβλεψη μεγεθών ή να προγνώσουμε πιθανά αποτελέσματα των αποφάσεων μας.

Σήμερα χρησιμοποιούνται πολλά εργαλεία που βασίζονται στο data mining και τα οποία μπορεί να είναι διαφορετικού τρόπου λειτουργίας ως προς τη φιλοσοφία τους όπως για παράδειγμα :

- Data visualization
- Neural networks
- Decision trees
- Rule induction programs

Τελικά το wizwhy είναι ένα λογισμικό που παράγει κατά κάποιον τρόπο κανόνες ώστε να λειτουργήσει η λογική του data mining σε μεγάλες και πολύπλοκες βάσεις δεδομένων. Σύμφωνα με τον αλγόριθμο που χρησιμοποιεί σε αρκετά μικρό χρονικό διάστημα μπορεί να εμφανίσει κανόνες που θα μας δώσουν τις απαραίτητες πληροφορίες που θέλουμε από το δείγμα.

3.1.1 Γιατί είναι καλύτερο από άλλα data mining tools.

Η διαφορά του wizwhy από άλλα εργαλεία που χρησιμοποιούν τη διαδικασία του data mining, όπως για παράδειγμα είναι τα νευρωνικά δίκτυα ή τα δένδρα απόφασης έγκειται στο ό,τι βρίσκει περιεκτικούς και εύχρηστους κανόνες σε πολύ λίγο χρόνο. Πιο συγκεκριμένα μπορούμε να κάνουμε τις εξής παρατηρήσεις[Wizsoft, 1996]:

- Το πρόγραμμα αυτό όταν πραγματοποιεί μια πρόβλεψη παρουσιάζει τους κανόνες που σχετίζονται συγκεκριμένα με τη πρόβλεψη σε αντίθεση με τα νευρωνικά δίκτυα που δεν κάνουν τέτοιους διαχωρισμούς.

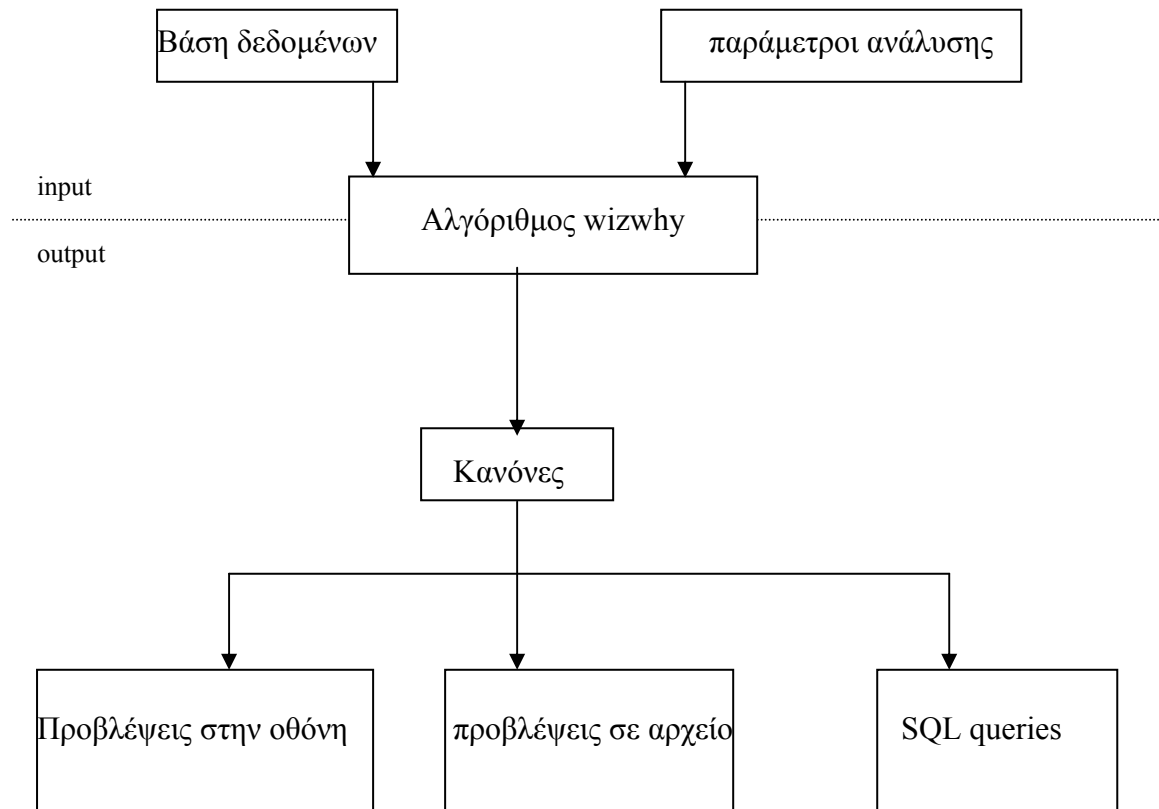
- Τα νευρωνικά δίκτυα συνήθως επηρεάζονται από κάποιες παραμέτρους όπως είναι η τοπολογία του δικτύου για παράδειγμα ή το μέγεθος του δείγματος, τα επίπεδα που θα χρησιμοποιηθούν και άλλα και τα οποία βεβαίως ορίζονται από το χρήστη, οπότε μιλάμε για υποκειμενικότητα αποτελεσμάτων. Αντίθετα τώρα το wizwhy ,ανεξαρτήτως χρήστη, δεν επηρεάζεται από υποκειμενικούς παράγοντες.
- Μια επίσης μεγάλη διαφορά είναι στο πόσο αξιόπιστα είναι τα αποτελέσματα που δίνει το κάθε πρόγραμμα. Πιο συγκεκριμένα το wizwhy είναι ικανό να μας δίνει ποσοστά αξιοπιστίας της πρόβλεψης που εκτελεί και επίσης να μας δείχνει και τη πιθανότητα λάθους των αποτελεσμάτων του, ενώ συνήθως τα νευρωνικά δίκτυα και τα δένδρα απόφασης δεν έχουν ανεπτυγμένους τέτοιους μηχανισμούς .
- Σε πολλές περιπτώσεις τα νευρωνικά δίκτυα αντιμετωπίζουν πρόβλημα απέναντι σε ποιοτικά δεδομένα γιατί λειτουργούν με δυαδικό τρόπο αντίληψης. Άρα εκεί θα πρέπει το πρόγραμμα νευρωνικών δικτύων να κάνει κωδικοποίηση των δεδομένων σε πολλά καινούρια που θα συμβαδίζουν με τη δικιά του αντίληψη, οπότε όπως είναι αναμενόμενο χρειάζεται και πολύ περισσότερο χρόνο επεξεργασίας όλων αυτών. Εν αντιθέσει το wizwhy αναγνωρίζει τα ποιοτικά δεδομένα και τα χειρίζεται άμεσα αρκεί να το κατευθύνει ο χρήστης.
- Ούτε τα νευρωνικά δίκτυα , ούτε τα δένδρα απόφασης ανακαλύπτουν-δημιουργούν μαθηματικές εξισώσεις ανάμεσα στα πεδία έστω και αν είναι ποσοτικά, ενώ το wizwhy δημιουργεί και τις εφαρμόζει κιόλας στη διενέργεια της πρόβλεψης.

3.1.2 Πως λειτουργεί σε γενικές γραμμές

Όταν το wizwhy αναλύει μια βάση δεδομένων σε γενικές γραμμές εκτελεί τις παρακάτω λειτουργίες:

- Αρχικά διαβάζει τη βάση δεδομένων και μετά ζητάει να κατευθυνθεί από το χρήστη ώστε να ρυθμιστούν κάποιες παράμετροι όπως είναι «η ελάχιστη πιθανότητα των κανόνων» ή «ο ελάχιστος αριθμός περιπτώσεων που επιβεβαιώνουν τους κανόνες». Έτσι τώρα το wizwhy είναι έτοιμο να κάνει τη μελέτη και παρουσίαση των κανόνων(rule report).
- Μέσα σε λίγο χρονικό διάστημα πραγματοποιεί τη παρουσίαση των κανόνων που έχουν να κάνουν μεταξύ των πεδίων πληροφόρησης και του πεδίου πρόβλεψης. Η μορφή αυτών είναι είτε μαθηματικές εξισώσεις είτε προτάσεις «if-then».
- Έχοντας βρει τους κανόνες μπορεί πλέον να βρει και τιμές του πεδίου που είναι προς πρόβλεψη. Συγκεκριμένα δίνοντάς του νέες τιμές των πεδίων πληροφόρησης μπορεί να προβλέψει το πεδίο που μας ενδιαφέρει και να μας δώσει και τις τιμές σε καινούριο αρχείο.
- Τέλος το wizwhy μπορεί να μεταφράσει τους κανόνες που βρίσκει σε SQL. Έτσι μπορεί να χρησιμοποιεί κανείς δεδομένα σε οποιαδήποτε βάση δεδομένων αν ταιριάζουν στις συνθήκες και τις παραμέτρους των κανόνων.

Κάπως έτσι έχουν να κάνουν τα παραπάνω παραστατικά:



Σχήμα 3α-Wizwhy Operations Flowchart[wizwhy-User's guide, 1996]

3.1.3 Τι ακριβώς είναι ένας κανόνας στο wizwhy

Όπως έχουμε πει το πρόγραμμα αυτό παράγει δυο ειδών κανόνες, τους if-then και των μαθηματικών εξισώσεων.

Για παράδειγμα ένας if-then κανόνας είναι:

If πεδίο της επιχείρησης είναι «υψηλό»
 Και αριθμός υπαλλήλων είναι « 200 ± 30 »
 Και ετήσιες πωλήσεις είναι « 38000 ± 1000 »
 Και το κεφάλαιο είναι « 1500 ± 500 »
 Τότε
 Η αξία αποθέματος είναι « 460000 ± 5000 »
 Πιθανότητα κανόνα: 0,90
 Ο κανόνας επαληθεύεται από 370 δεδομένα
 Επίπεδο σημαντικότητας:πιθανότητα λάθους<0,01

Ένα παράδειγμα κανόνων που βασίζονται σε μαθηματικές εξισώσεις είναι:

$A=5*B$
 Όπου: A είναι η αξία αποθέματος
 B είναι καθαρό κέρδος
 Επίπεδο ακριβείας του κανόνα: 0,95

Ο κανόνας επαληθεύεται από 1890 δεδομένα

Η πιθανότητα για τους κανόνες if-then αποτελεί το πηλίκο μεταξύ του πλήθους των δεδομένων για τα οποία οι συνθήκες και τα αποτελέσματα ισχύουν, δια του πλήθους των δεδομένων για τα οποία ισχύουν μόνο οι συνθήκες.

Το επίπεδο σημαντικότητας μας δείχνει το βαθμό στον οποίο μπορούμε να στηριχτούμε στον αντίστοιχο κανόνα, ώστε να κάνουμε μια πρόβλεψη. Ισούται με τη μονάδα μείον την πιθανότητα λάθους και συμβολίζει τη πιθανότητα ο αντίστοιχος κανόνας να υφίσταται στα συγκεκριμένα δεδομένα που αναλύουμε αλλά όχι και σε ολόκληρο το δείγμα. Στο παραπάνω παράδειγμα το επίπεδο σημαντικότητας θα ήταν $1-0,01=0,99$ και δείχνει ότι ο κανόνας μας είναι μια καλή βάση για προβλέψεις.

Επίπεδο ακριβείας στους κανόνες μαθηματικών εξισώσεων ονομάζουμε την διαφορά μεταξύ των τιμών των δεδομένων και των τιμών των εξισώσεων.

3.1.4 Που μπορεί να χρησιμοποιηθεί το wizwhy

Γενικά η μέθοδος του data mining και πιο συγκεκριμένα το wizwhy έχει πολλές πρακτικές εφαρμογές. Μπορεί να χρησιμοποιηθεί σε αναλύσεις και προβλέψεις δεδομένων και απλών και πολύπλοκων, όπως:

- Σε δεδομένα από τον ιατρικό χώρο ή τις κοινωνικές επιστήμες, όπου θα βοηθήσει τους αντίστοιχους επαγγελματίες να πραγματοποιήσουν καλές διαγνώσεις ή να επεκτείνουν την έρευνά τους.
- Σε τραπεζικά και γενικότερα οικονομικά δεδομένα, όπου το wizwhy μπορεί να επιδείξει «επικίνδυνους» πελάτες.
- Σε περιπτώσεις που κάποια επιχείρηση χρησιμοποιεί τη μέθοδο του direct marketing μπορεί να επιτύχει καλύτερα αποτελέσματα(αποστέλλοντας αλληλογραφία) χρησιμοποιώντας το wizwhy.

3.1.5 Μαθηματικό υπόβαθρο του wizwhy

Παρακάτω παρουσιάζεται το μαθηματικό υπόβαθρο το οποίο καλύπτει τη λειτουργία του wizwhy.

Καταρχήν το wizwhy μελετά τα δεδομένα μας που αποτελούν το τμήμα της εκπαίδευσης ή το training set, όπως λέγεται. Σκοπός σ' αυτή τη φάση είναι να βρεθούν από το πρόγραμμα κάποια βασικά πρότυπα μέσα στο αρχείο προς εξέταση, δηλαδή αυτές οι πληροφορίες οι οποίες θα αποτελέσουν το υλικό για να σχηματιστούν οι κανόνες, οι οποίοι στη συνέχεια θα μας βοηθήσουν να κάνουμε την πρόβλεψη.

Ύστερα από αυτό το wizwhy πραγματοποιεί τη διαδικασία της πρόβλεψης σύμφωνα με αυτά που του έχει ορίσει ο χρήστης ως ζητούμενα για πρόβλεψη. Επιπλέον για κάθε προβλεπόμενη τιμή που θα προκύψει θα μας δοθεί και η πιθανότητά της καθώς και το επίπεδο σημαντικότητας. Το wizwhy σε αντίθεση με άλλα εργαλεία data mining, εκτός από τη τιμή της πρόβλεψης, μας παρουσιάζει και τους αντίστοιχους κανόνες που πλαισιώνουν τη διαδικασία της συγκεκριμένης πρόβλεψης.

3.1.5.1 Κανόνες if then

Ο συμβολισμός R χρησιμοποιείται για να δηλώσει το πεδίο προς πρόβλεψη. Όλοι οι κανόνες που θα βρεθούν έχουν να κάνουν με το R.

Το wizwhy ψάχνει για κανόνες if then που να έχουν σχέση με το prediction field. Αυτοί οι κανόνες έχουν την ακόλουθη δομή:

If (συνθήκη)

Then $R=r$

Πιθανότητα κανόνα= p

Ο κανόνας επαληθεύεται από m δεδομένα

Επίπεδο σημαντικότητας : πιθανότητα λάθους= a

Η συνθήκη if μπορεί να είναι είτε ενός ορίσματος είτε περισσοτέρων. Η συνθήκη ενός ορίσματος είναι του ακόλουθου τύπου:

η τιμή του πεδίου X είναι a (αν το πεδίο είναι ποιοτικό) ή $x \pm D$ (αν το πεδίο είναι ποσοτικό).

Μια συνθήκη πολλών ορισμάτων είναι ένας συνδυασμός συνθηκών ενός ορίσματος. Πρέπει φυσικά το πλήθος των ορισμάτων να είναι μικρότερο/ίσο του πλήθους των πεδίων του αρχείου που μελετάμε. Για παράδειγμα μια συνθήκη τριών ορισμάτων περιγράφεται ως εξής:

Η τιμή του ποιοτικού πεδίου X είναι a

Και η τιμή του ποιοτικού πεδίου Y είναι β

Και η τιμή του ποσοτικού πεδίου Z είναι $z \pm D$.

Το συμπέρασμα του κανόνα είναι του παρακάτω τύπου:

$R=r$ (αν το πεδίο είναι ποιοτικό) ή $R=r \pm D$ (αν το πεδίο είναι ποσοτικό).

Επίσης πρέπει να πούμε πως μπορεί το D να είναι 0 εν γένει είτε αναφερόμαστε στο if κομμάτι είτε στο then ενός κανόνα.

3.1.5.2 Κανόνες μαθηματικών εξισώσεων

Όταν το prediction field R είναι ποσοτικό το wizwhy ψάχνει για εξισώσεις που συσχετίζουν το R με τα άλλα ποσοτικά πεδία και είναι του τύπου:

Πεδίο R είναι συνάρτηση των άλλων πεδίων

Επίπεδο ακρίβειας= p

Ο κανόνας επαληθεύεται από m δεδομένα

Αριθμητικές σχέσεις μεταξύ πεδίων

Το wizwhy ψάχνει για συναρτήσεις του τύπου:

$$R = (x_1 \oplus x_2) \otimes (x_3 \oplus x_4)$$

Όπου:

\oplus δηλώνει πρόσθεση ή πολλαπλασιασμό

\otimes δηλώνει ή πρόσθεση ή αφαίρεση ή πολλαπλασιασμό ή διαίρεση

Κανόνες που αναφέρονται σε ποσοστά.

Έχουμε για παράδειγμα το $(x_3 \oplus x_4)$ το οποίο μπορεί να αντικατασταθεί με μια εκ των δύο παραστάσεων:

- $(1 - x_3 / 100)$

- $(1 + x_3 / 100)$

έτσι η εξίσωση θα γίνει για παράδειγμα

$$R = x_1 \square x_2 \square (1 + x_3 / 100)$$

Έτσι λειτουργούν οι συναρτήσεις όταν η τιμή του x_3 είναι ποσοστό.

Χρησιμοποιούνται οι παρακάτω συναρτήσεις:

Γραμμικές συναρτήσεις

$$R = a \square x_1 + b$$

Παραβολικές συναρτήσεις

$$R = a \square x_1^2 + b \square x_1 + c$$

Υπερβολικές συναρτήσεις

$$R = a / x_1.$$

3.1.5.3 Καθορισμός της a priori πιθανότητας ενός ενδεχομένου όταν το prediction field είναι μη δυαδικό.

Ορίζουμε τα παρακάτω:

m είναι το πλήθος των δεδομένων ενός κανόνα

p είναι η πιθανότητα του κανόνα

n=m/p είναι ο αριθμός των δεδομένων που ικανοποιούν τη συνθήκη του κανόνα

a είναι η a priori πιθανότητα (πιθανότητα λάθους) του κανόνα.

Έστω r_1, \dots, r_k είναι οι τιμές του prediction field. Εδώ $r_i, i=1, \dots, k$ είναι οι καθορισμένες τιμές του R, αν το πεδίο είναι ποιοτικό. r_i είναι η καθορισμένη τιμή της προσθαφαιρούμενης ποσότητας που έχουμε βρει, της ποσοτικής μεταβλητής, αν το R είναι ποσοτικό.

Έστω q_i είναι η συχνότητα της τιμής r_i του R, στο αρχείο προς εξέταση. Στον if then κανόνα είναι $R = r_l$ όταν αναφερόμαστε στο then κομμάτι, όπου $l \in \{1, \dots, k\}$.

Το wizwhy καθορίζει την a priori πιθανότητα a ενός κανόνα σύμφωνα με τους παραπάνω ορισμούς. Έστω x_i είναι η συχνότητα της τιμής r_i του R, υπό τη συνθήκη του κανόνα. Το wizwhy υπολογίζει ότι όλα τα διανύσματα (x_1, \dots, x_k) είναι ισοπίθανα. Η μαθηματική μοντελοποίηση γίνεται ως εξής:

$$a = \frac{Q_1}{Q_0}$$

όπου:

Q_0 είναι ο αριθμός των διαφορετικών λύσεων της ισότητας:

$$\sum_{i=1}^k x_i = n, 0 \leq x_i \leq q_i, i = 1, \dots, k, \quad x_i \text{ είναι ακέραιος.}$$

Q_1 είναι ο αριθμός των διαφορετικών λύσεων της ισότητας:

$$m \leq x_l \leq n, \\ \sum_{i=1}^k x_i = n, \quad 0 \leq x_i \leq q_i, \quad , x_i \text{ είναι ακέραιος} \\ i \in \{1, \dots, k\} / \{l\}$$

για τη συγκεκριμένη περίπτωση και αν $n < q_{\min}$, όπου $q_{\min} = \min q_i$. Αν η συνάρτηση που υπολογίζει τον αριθμό των διαφορετικών λύσεων της ισότητας

$$\sum_{i=1}^k X_i = n, 0 \leq X_i < n, X_i \in \mathbb{Z}, \text{ είναι η } Q(n, k).$$

Η συνάρτηση αυτή υπολογίζεται ως εξής:

if $n \geq k$ τότε

$$Q(n, k) = \sum_{i=0}^{k-1} C_{n-1}^{k-1-i} \square C_k^i$$

ή

$$Q(n, k) = \sum_{i=k-n}^{k-1} C_{n-1}^{k-1-i} \square C_k^i$$

βάση αυτής της συνάρτησης η a priori πιθανότητα a υπολογίζεται ως:

$$a = \frac{Q_1}{Q_0}, \text{ όπου}$$

$$Q_0 = Q(n, k)$$

$$Q_1 = 1 + \sum_{j=m}^{n-1} Q(n-j, k-1)$$

3.1.5.4 Βαθμός ακριβείας για τους κανόνες μαθηματικών εξισώσεων.

Αν το R είναι ποσοτικό, το wizwhy ψάχνει για τη συνάρτηση που συνδέει το πεδίο R με τα άλλα ποσοτικά πεδία. Για τις παραμετρικές συναρτήσεις μιας μεταβλητής, το wizwhy καθορίζει την καλύτερη συνάρτηση κατά προσέγγιση των δεδομένων που περιέχονται στα πεδία. Για τις συναρτήσεις πολλαπλών μεταβλητών και χωρίς παραμέτρους, το wizwhy ελέγχει τη ακρίβεια της προσέγγισης.

Το επίπεδο ακρίβειας ενός τέτοιου κανόνα καταδεικνύει την ακρίβεια της προσέγγισης βάση της ευρεθείσας συνάρτησης. Φυσικά το μέγιστο επίπεδο που μπορεί να φτάσει η ακρίβεια είναι η μονάδα. Αυτή αποτελεί και την πιο ακριβή προσέγγιση άλλωστε. Όταν η ακρίβεια μιας συνάρτησης είναι πάνω από ένα

κατώτερο επίπεδο ακριβείας τότε αυτή αποτελεί ένα νέο κανόνα. Αυτό σημαίνει ότι ο κανόνας αυτός μπορεί να λειτουργήσει βάση των προδιαγραφών που εμείς έχουμε θέσει. Τελικά το επίπεδο ακριβείας ρ μπορεί να υπολογιστεί για μια συνάρτηση

$$R = f(x, y, z), \text{ ως εξής:}$$

Η τιμή του πεδίου R σε μια εγγραφή δεδομένων i δηλώνεται ως r_i , $i=1, \dots, N$, όπου N είναι ο αριθμός των δεδομένων του προς εξέταση αρχείου. Αν X_i , Y_i , Z_i είναι οι τιμές των μεταβλητών x, y, z για το δεδομένο i , τότε:

$$\rho = 1 - d, \text{ όπου}$$

$$d = \frac{1}{N} \sum_{i=1}^N d_i,$$

$$d_i = \frac{|r_i - f(x_i, y_i, z_i)|}{|r_i|}.$$

Το d_i εδώ είναι θετικό και συμβολίζει τη σχετική τιμή του σφάλματος της συνάρτησης $R=f(x,y,z)$ στο δεδομένο i , d είναι μια μέση τιμή του d_i για όλα τα δεδομένα. Αν η συνάρτηση R έχει λειτουργήσει για όλα τα δεδομένα που μας ενδιαφέρουν ακριβώς τότε $d=0$ και $\rho=1$. Όσο μεγαλώνει η τιμή του d τόσο μικραίνει η τιμή του ρ . Επίσης αν $d>1$ τότε $\rho<0$. Ωστόσο το *wizwhy* δεν μπορεί να παράγει κανόνα με αρνητικό ρ από τη στιγμή που ισχύει η σύμβαση για το επίπεδο ακριβείας

$$\rho \geq \rho_{\min}.$$

3.1.6 Πρόβλεψη.

Μια πρόβλεψη μπορεί να γίνει αφού πρώτα ευρεθούν όλοι οι κανόνες που πλαισιώνουν τα δεδομένα μας. Η πρόβλεψη που κάνει το *wizwhy* βασίζεται στις παραμέτρους που του ρυθμίζει ο χρήστης, δηλαδή υπάρχουν οι λεγόμενες συνθήκες πρόβλεψης που ανάλογα με το *prediction field* ο χρήστης βγαίνει και τα ανάλογα συμπεράσματα των τιμών που τον ενδιαφέρουν.

Αν το *prediction field* δεν είναι δυαδικό τότε η πρόβλεψη παράγεται, παρουσιάζοντας μαζί και τους κανόνες τους σχετικούς με τις τιμές του πεδίου που μας ενδιαφέρει.

3.1.7 Πόσοι κανόνες μπορούν να εξαχθούν από μια βάση δεδομένων

Αν σκεφτούμε το πόσο πολύπλοκο είναι το πρόβλημα που καλείται να λύσει το *wizwhy* τότε θα καταλάβουμε γιατί κάνει και τόσο χρόνο να εξαγάγει τους κανόνες. Ας υπολογίσουμε παρακάτω τον αριθμό των πιθανών κανόνων που μπορεί να σχετίζονται με τη βάση δεδομένων μας.

Ας υποθέσουμε ότι το prediction field περιέχει κ διαφορετικές τιμές. Επίσης έστω ότι το αρχείο μας περιέχει η πεδία προς εξέταση με μια μέση τιμή m για κάθε ποιοτικό πεδίο και m προκύπτουσες διαφορές για κάθε ποσοτικό πεδίο. Ο αριθμός των διαφορετικών πιθανοτήτων να παράγουμε κανόνες if-then με l πεδία στο if τμήμα του κανόνα είναι

$$k \square m^l \square C_n^l.$$

Ο συνολικός αριθμός των πιθανών κανόνων if-then είναι

$$k \square \sum_{l=1}^n m^l \square C_n^l.$$

Είναι φανερό ότι ο αριθμός των πιθανών κανόνων είναι τρομακτικά μεγάλος. Για παράδειγμα αν ένα αρχείο περιέχει 20 πεδία με κατά μέσο όρο 10 τιμές για καθένα, τότε ο αριθμός των πιθανών κανόνων ξεπερνά τους 10^{20} ! Αν υποθέσουμε πως ένας υπολογιστής χρειαστεί να μελετήσει τους κανόνες αυτούς έναν προς έναν, αυτό θα έπαιρνε τρομερά πολύ χρόνο να γίνει. Αν υποθέσουμε ότι υπάρχουν 10000 δεδομένα στο αρχείο τότε απ'τη στιγμή που κατα μέσο όρο πρέπει να εκτελεστούν 10 λειτουργίες για να ελεγχθεί εάν ο συγκεκριμένος κανόνας επαληθεύεται για το συγκεκριμένο δεδομένο, τότε ο συνολικός χρόνος του υπολογιστή για αυτόν τον έλεγχο είναι 10^{25} . Οι υπολογισμοί αυτοί ακόμα και στον καλύτερο σύγχρονο H/Y θα απαιτούσαν χιλιάδες χρόνια για να ολοκληρωθούν.

Βασισμένο σε έναν «έξυπνο» αλγόριθμο το wizwhy κατασκευάζει αυτούς τους κανόνες σε πολύ μικρό χρονικό διάστημα. Ο αλγόριθμος από μόνος του είναι ιδιαίτερα πολύπλοκος για να παρουσιαστεί εδώ, αλλά αξίζει να σημειωθεί ότι δεν βασίζεται σε ευριστικές μεθόδους.

3.2 Neurosolutions

Το neurosolutions είναι ένα πρόγραμμα το οποίο μπορεί να κάνει επεξεργασία και ανάλυση δεδομένων αλλά και πρόβλεψη κάποιας ποσότητας μέσω της θεωρίας των νευρωνικών δικτύων. Είναι προϊόν της NeuroDimension [www.neurosolutions.com] και είναι η έκδοση 4.22 στη μορφή κοινής χρήσης, όπου περιορίζεται στα 1000 δεδομένα εισόδου.

3.2.1 Νευρωνικά δίκτυα

Το τεχνητό νευρωνικό δίκτυο (ΤΝΔ) [Α.Λύκας, 1999] είναι μια αρχιτεκτονική δομή αποτελούμενη από πλήθος διασυνδεδεμένων μονάδων (τεχνητοί νευρώνες). Κάθε μονάδα χαρακτηρίζεται από εισόδους και εξόδους και υλοποιεί τοπικά έναν απλό υπολογισμό. Κάθε σύνδεση μεταξύ δυο μονάδων χαρακτηρίζεται από μια τιμή βάρους. Οι τιμές των βαρών των συνδέσεων αποτελούν τη γνώση που είναι αποθηκευμένη στο δίκτυο και αποτελούν τη λειτουργικότητά του. Η έξοδος κάθε μονάδας καθορίζεται από τον τύπο της μονάδας, τη διασύνδεση με τις υπόλοιπες μονάδες και πιθανώς κάποιες εξωτερικές εισόδους. Πέρα από μια πιθανή δεδομένη (εκ κατασκευής) λειτουργική ικανότητα ενός δικτύου, συνήθως ένα δίκτυο αναπτύσσει μια συνολική λειτουργικότητα μέσω μιας μορφής εκπαίδευσης.

Η συνολική λειτουργικότητα ενός νευρωνικού δικτύου καθορίζεται από τη τοπολογία του δικτύου, τα χαρακτηριστικά των νευρώνων, τη μέθοδο εκπαίδευσης και τα δεδομένα με τα οποία γίνεται η εκπαίδευση. Παρότι δεν το απαιτεί ο προηγούμενος ορισμός, ο υπολογισμός που εκτελεί κάθε νευρώνας είναι απλός και

κοινός για όλους τους νευρώνες. Επειδή οι νευρώνες λειτουργούν παράλληλα(ταυτόχρονα) και ο αριθμός τους μπορεί να είναι πολύ μεγάλος, τα τεχνητά νευρωνικά δίκτυα αποτελούν χαρακτηριστικό παράδειγμα μαζικά παράλληλου υπολογισμού.

3.2.2 Βασικές αρχές νευρωνικών δικτύων

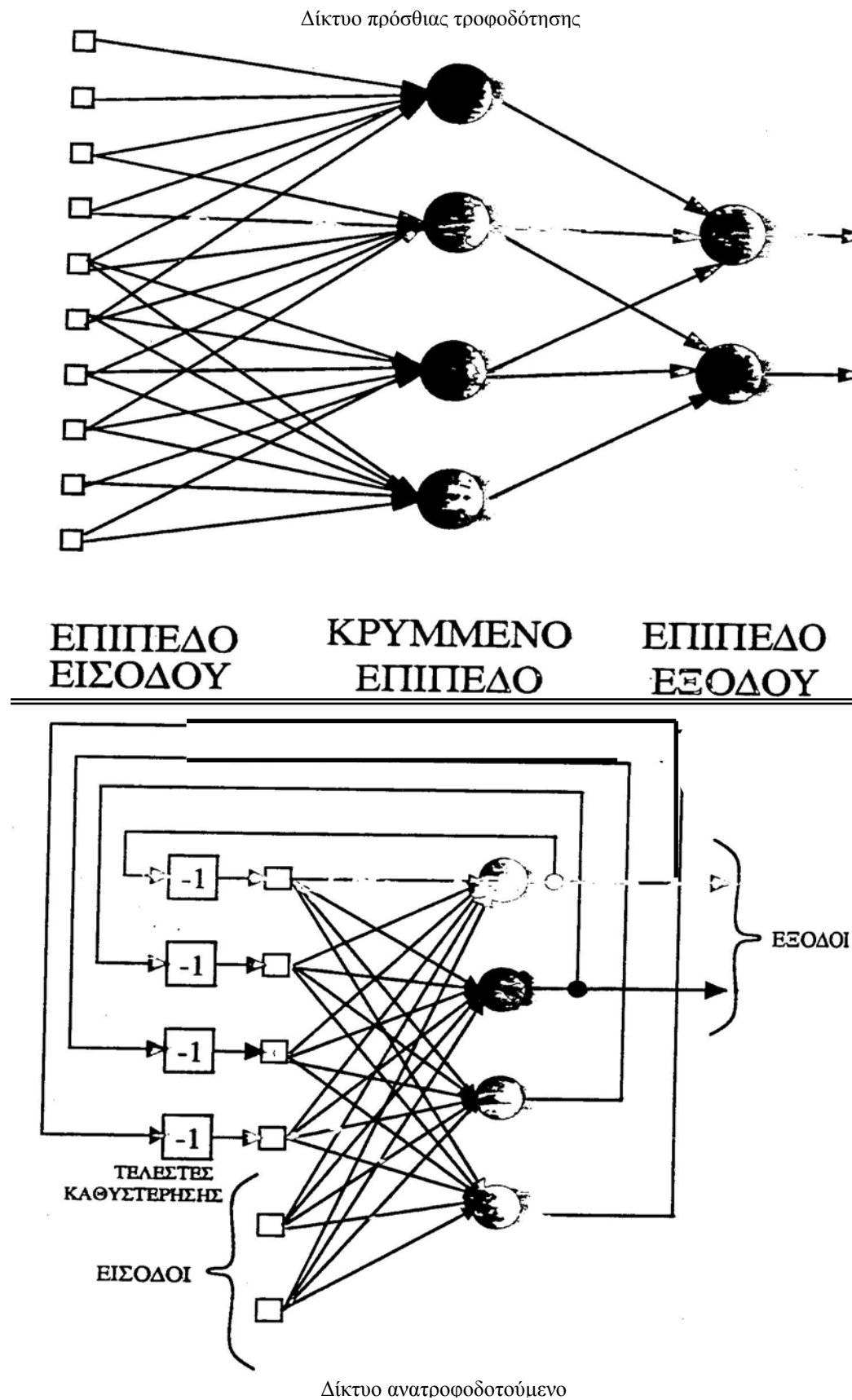
Τα νευρωνικά δίκτυα[A.Λύκας, 1999] είναι κάποια κατάλληλα συστήματα, τα οποία μπορούν να δημιουργούν συσχετίσεις μεταξύ δεδομένων και να γενικεύουν κανόνες σε νέα δεδομένα, οι οποίοι δεν είναι προφανείς. Υπάρχει η δυνατότητα να οδηγούνται από τον χειριστή του προγράμματος για το τι πρέπει να αφομοιώνουν από τα δεδομένα ενώ άλλα είναι έτσι εξελιγμένα ώστε από μόνα τους να αποθηκεύουν πληροφορίες μέσα στη δομή τους.

Όπως προκύπτει από τον ορισμό του ΤΝΔ, το υπολογιστικό μοντέλο αποτελείται από συνδέσεις απλών στοιχείων ή μονάδων. Οι συνδέσεις αυτές χαρακτηρίζονται από βάρη. Στα σχήματα 3.β απεικονίζονται δυο παραδείγματα δικτύων μικρής κλίμακας στα οποία οι μονάδες παρουσιάζονται με κύκλους και οι συνδέσεις με βέλη. Στο πρώτο απεικονίζεται ένα μη ανατροφοδοτούμενο δίκτυο, το οποίο δεν περιέχει κλειστά μονοπάτια από συνδέσεις. Όπως φαίνεται οι μονάδες ομαδοποιούνται σε επίπεδα (layers). Αντίθετα, στο δεύτερο παρουσιάζεται η αρχιτεκτονική της ανατροφοδότησης η οποία επιτρέπει την ύπαρξη κύκλων από συνδέσεις μεταξύ μονάδων. Αυτή η δεύτερη αρχιτεκτονική προσδίδει στο δίκτυο πολύ περισσότερες δυνατότητες, αλλά είναι πιο δύσκολο να αντιμετωπισθεί μαθηματικά. Οι μονάδες που δεν έχουν συνδέσεις με τον εξωτερικό κόσμο λέμε ότι είναι *κρυμμένες*(hidden) ή εσωτερικές.

Κάθε μονάδα υλοποιεί μια συνάρτηση τοπικά και ολόκληρο το δίκτυο υλοποιεί μια συγκεκριμένη λειτουργία. Στις περισσότερες περιπτώσεις η εφαρμογή είναι αυτή που μέσω των προδιαγραφών καθορίζει την απαιτούμενη λειτουργία. Ο καθορισμός των παραμέτρων του δικτύου(τιμών των βαρών) που θα ικανοποιούν αυτές τις προδιαγραφές επιτυγχάνεται μέσω της διαδικασίας μάθησης.

Η γνώση, η εμπειρία και η εκπαίδευση του δικτύου αποθηκεύεται στις διασυνδέσεις των μονάδων και στις τιμές των βαρών. Η εκπαίδευση επιτυγχάνεται με τη συνεχή τροποποίηση των τιμών των βαρών.

Στη πλειοψηφία τους τα ΤΝΔ εκπαιδεύονται με την ελπίδα ότι θα παρουσιάσουν καλή γενικευτική ικανότητα όταν θα τους ζητηθεί να αναγνωρίσουν ή να κατηγοριοποιήσουν καινούρια(άγνωστα) δεδομένα(πρότυπα). Αυτός είναι ο αντικειμενικός στόχος της διαδικασίας εκπαίδευσης, να αναπτύξει δηλαδή το ΤΝΔ κατάλληλη εσωτερική δομή ώστε να αναγνωρίζει πρότυπα που θα μοιάζουν με αυτά με τα οποία εκπαιδεύτηκε. Τα ΤΝΔ εκπαιδεύονται τόσο με τεχνικές μάθησης με επίβλεψη όσο και με τεχνικές μάθησης χωρίς επίβλεψη.



Σχήμα 3β- Μη Επαναληπτικό και Επαναληπτικό δίκτυο [Α.Λύκας, 1999]

3.2.3 Παραδείγματα εφαρμογών ΤΝΔ

Χρησιμοποιούνται σε [Α.Λύκας, 1999]:

- επεξεργασία εικόνας και μηχανική όραση
- επεξεργασία σήματος
- αναγνώριση προτύπων
- ιατρική
- αμυντικά συστήματα
- οικονομία
- σχεδιασμός, έλεγχος και αναζήτηση (robotics)
- τεχνητή νοημοσύνη (expert systems)
- δυναμικά εξελισσόμενα συστήματα, πρόβλεψη χρονοσειρών
- επικοινωνία ανθρώπου υπολογιστή.

3.2.4 Πλεονεκτήματα και μειονεκτήματα των ΤΝΔ.

Επειδή τα ΤΝΔ αποτελούν ένα σχετικό πρόσφατο μοντέλο υπολογισμού, θα πρέπει να πούμε ότι τα πλεονεκτήματα, τα μειονεκτήματα και η σχέση τους με τις κλασικές μεθόδους υπολογισμού δεν είναι πλήρως ξεκαθαρισμένα. Οι προσδοκίες σε αυτόν τον τομέα είναι μεγάλες. Σε γενικές γραμμές τα ΤΝΔ είναι κατάλληλα να αντιμετωπίσουν συγκεκριμένες κατηγορίες προβλημάτων και ειδικότερα προβλήματα εκπαίδευσης που αντιμετωπίζονται με συσχέτιση προτύπων. Η αντίληψη ότι τα ΤΝΔ μπορούν να λύσουν όλα τα προβλήματα αυτόματης απόκτησης γνώσης είναι μη ρεαλιστική.

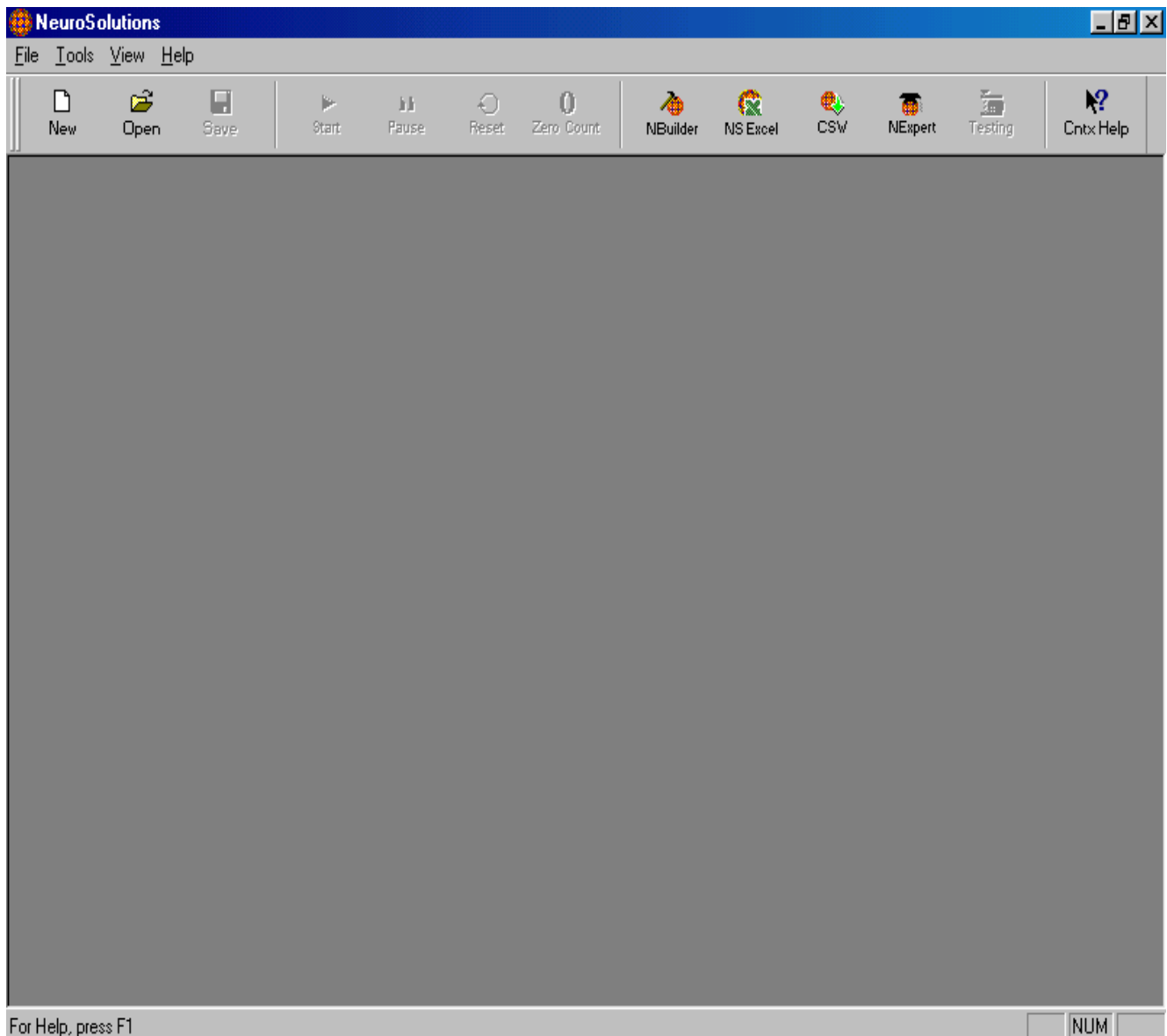
Ως πλεονεκτήματα μπορούμε να αναφέρουμε τα εξής:

- έμφυτος παραλληλισμός
- ανοχή σε βλάβες (λόγω του παραλληλισμού)
- σχεδιασμένα να είναι προσαρμοζόμενα
- δεν υπάρχει η ανάγκη για χαρακτηρισμό του προβλήματος πέρα από το σύνολο των προτύπων εκπαίδευσης.

Ως μειονεκτήματα θεωρούμε τα παρακάτω:

- δεν υπάρχουν σαφείς κανόνες για την ανάπτυξη ΤΝΔ για οποιαδήποτε εφαρμογή
- δεν υπάρχει γενικός τρόπος για την ερμηνεία της εσωτερικής λειτουργίας του δικτύου
- η εκπαίδευση μπορεί να είναι δύσκολη ή αδύνατη
- η ικανότητα γενίκευσης είναι δύσκολα προβλέψιμη

3.2.5 Επίπεδα χειρισμού του neurosolutions



Σχήμα 3γ-Παράθυρο χειρισμού neurosolutions

Υπάρχουν έξι επίπεδα του neurosolutions καθένα από τα οποία σου επιτρέπουν να αναπτύξεις το δικό σου νευρωνικό δίκτυο[NeuroDimension, 2003]. Το EDUCATOR είναι το πρώτο επίπεδο και το οποίο σου δίνει τη δυνατότητα να αρχίσεις να γνωρίζεις τα νευρωνικά δίκτυα. Επόμενη μορφή είναι το USERS το οποίο σου δίνει τη δυνατότητα να επιλέξεις ανάλογα με τις απαιτήσεις σου ανάμεσα σε πληθώρα τύπων νευρωνικών δικτύων. Ύστερα υπάρχουν και πιο αναπτυγμένα μοντέλα για πιο εξειδικευμένες εργασίες όπως το CONSULTANTS, PROFESSIONAL, και τέλος το DEVELOPER, το οποίο παρέχει τη δυνατότητα στον χρήστη να επεκτείνει τη λειτουργικότητα του neurosolutions δημιουργώντας τα δικά του νευρωνικά δίκτυα, κάνοντας προεπεξεργασία και αλγορίθμους εισόδου/εξόδου. Το DEVELOPERS LITE είναι των ίδιων δυνατοτήτων με το παραπάνω με τη διαφορά ότι δεν περιέχει τη λειτουργία υλοποίησης κώδικα ANSI C++.

Το NeuroSolutions for Excel είναι ένα ειδικό πρόσθετο για το excel το οποίο συνεργάζεται με οποιοδήποτε από τα παραπάνω επίπεδα λειτουργίας και δημιουργεί ένα πολύ δυνατό περιβάλλον λειτουργίας για να επεξεργαστούμε τα δεδομένα μας. Ο χειριστής απλά διαλέγει στήλες δεδομένων από τη βάση του ως είσοδο στο σύστημά του και γραμμές δεδομένων για την εκπαίδευσή του συστήματος, το τεστάρισμα κτλ και ύστερα το neurosolutions αντίστοιχα εκπαιδεύει ή τεστάρει το δίκτυο κτλ.

Τα νευρωνικά δίκτυα είναι φτιαγμένα για να δημιουργούν και εμπειρικές σχέσεις μεταξύ ποσοτήτων(μεταβλητών), «regression» αλλά και για να κάνουν ταξινόμηση αυτών, «classification». Στη πρώτη περίπτωση του regression οι τιμές που έχουμε στην έξοδο είναι αποτέλεσμα μιας συνεχούς μεταβαλλόμενης εκτίμησης των «ερεθισμάτων» της εισόδου. Στη δεύτερη περίπτωση του classification ,το ζητούμενο είναι να συσχετίσουμε τα σήματα εισόδου με κάποιο από τη πληθώρα κατηγοριών ή τάξεων, που συνήθως αναπαριστώνται από εξόδους που παίρνουν τιμές από 0 σε 1 και ουσιαστικά δίνουν τη πιθανότητα να βρίσκονται μέσα στην αντίστοιχη κατηγορία.

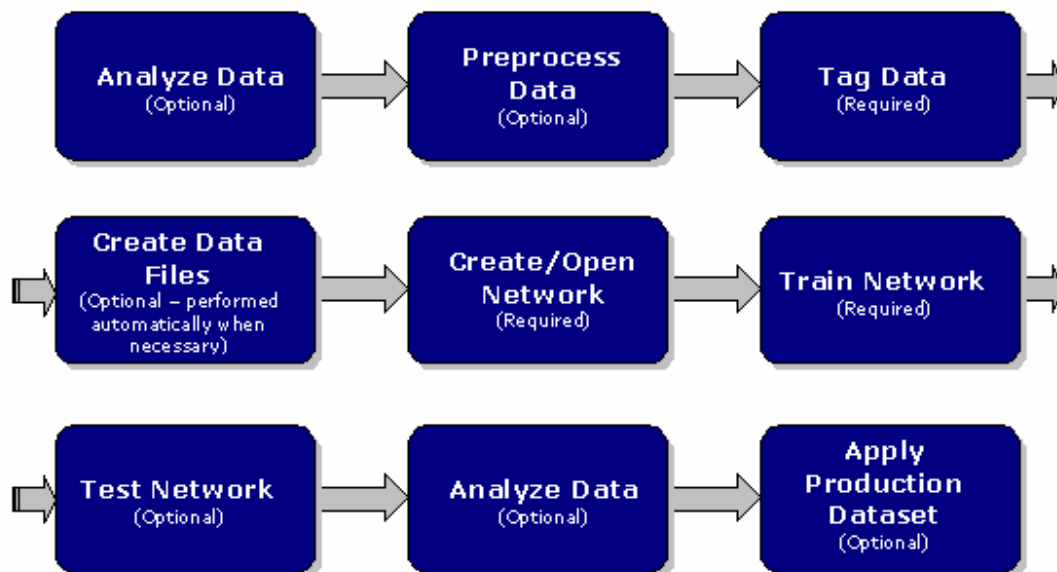
Η δυνατότητα neuralbuilder είναι μια πρόσθετη εφαρμογή που δίνει τη δυνατότητα στο χρήστη να σχεδιάσει και να σετάρει το δικό του νευρωνικό δίκτυο. Αυτόματα χτίζει μια από τις οκτώ πιο γνωστές νευρωνικές δομές.

3.2.6 Χρησιμοποιώντας το neurosolutions for excel.

Αυτό το εργαλείο έχει σχεδιαστεί για να μπορούμε να αναπτύξουμε μια ολοκληρωμένη λύση στο πρόβλημά μας παρέχοντας παράλληλα τη ελαστικότητα να χρησιμοποιήσουμε και τη visual basic για κάποιες αλλαγές αν θέλουμε. Το επόμενο σχήμα αποτελεί ένα flow chart και παρουσιάζει τη σειρά των κινήσεων που μπορούν να γίνουν ώστε να λύσουμε το εκάστοτε πρόβλημα. Αν παρατηρήσουμε, θα καταλάβουμε ότι το ελάχιστο αυτών για να λειτουργήσει το neurosolutions for excel είναι τρείς:

- ο χαρακτηρισμός των δεδομένων
- η δημιουργία του δικτύου και
- η εκπαίδευση του δικτύου.

Συνήθως χρησιμοποιούμε και την επιλογή της δοκιμής(test) της απόδοσης του μοντέλου στην τροφοδοσία νέων δεδομένων εισόδου όπου η έξοδος είναι άγνωστη.



Flow diagram illustrating the suggested usage of NeuroSolutions for Excel

Σχήμα 3δ-Διάγραμμα ροής επίλυσης[NeuroDimension, 2003]

3.2.7 Πως το neurosolutions υλοποιεί ένα νευρωνικό δίκτυο.

Το neurosolutions οδηγεί στο λεγόμενο «local additive model». Σύμφωνα με αυτό το μοντέλο κάθε τμήμα του δικτύου μπορεί να ενεργοποιηθεί και να ξεκινήσει τη διαδικασία της μάθησης βάση των δικών του βαρών και λειτουργιών ή και των λειτουργιών των «γειτόνων» του. Αυτό φυσικά αποτελεί σημαντικό πλεονέκτημα αφού κάθε τμήμα του δικτύου μπορεί να στέλνει και να δέχεται μηνύματα.

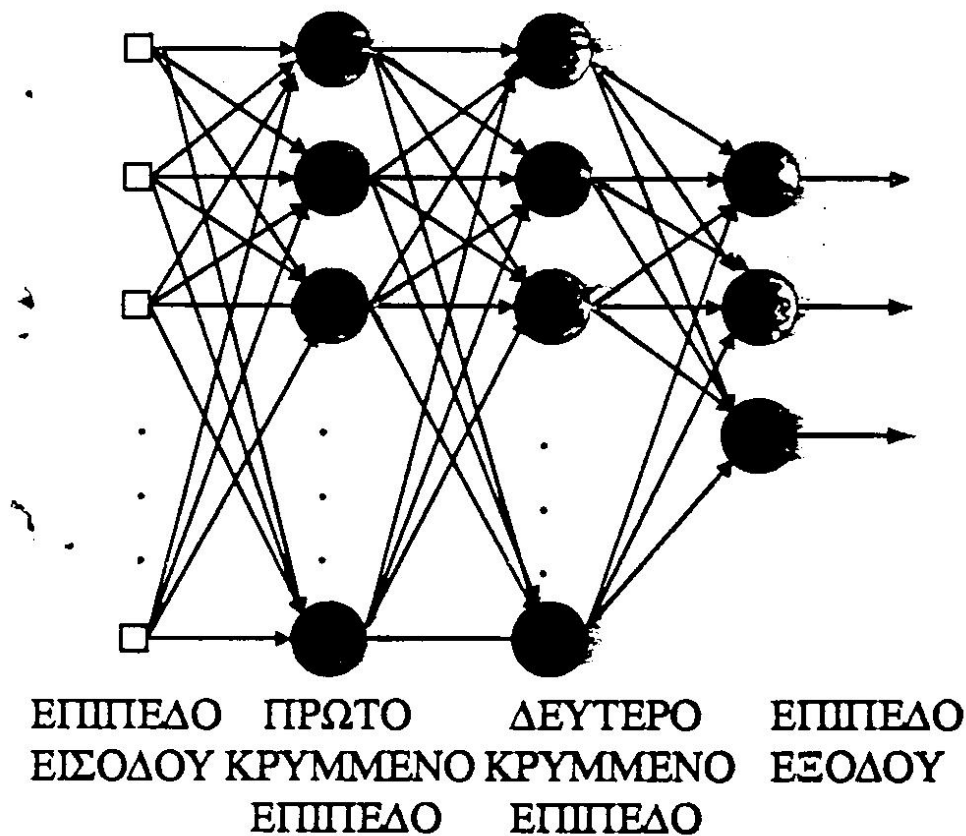
3.2.8 Αλγόριθμος back propagation.

Το neurosolutions χρησιμοποιεί για να εκπαιδεύσει το δίκτυο τον αλγόριθμο back propagation[A.Λύκας, 1999] διαμέσου χρόνου. Ο αλγόριθμος back propagation εμφανίστηκε μετά το 1985 αλλά η ιδέα του περιγράφηκε για πρώτη φορά στη διδακτορική διατριβή του Werbos το 1974. Αποτελεί ορόσημο για τα νευρωνικά δίκτυα διότι παρέχει μια υπολογιστικά αποτελεσματική μέθοδο εκπαίδευσης των πολυεπίπεδων perceptrons.

3.2.9 Multilayer perceptron.

Το multilayer perceptron (MLP)[A.Λύκας, 1999] είναι η ευρύτερα χρησιμοποιούμενη κατηγορία νευρωνικών δικτύων. Τα δίκτυα αυτά είναι πρόσθιας τροφοδότησης(feed forward), και περιλαμβάνουν ένα επίπεδο εισόδου, ένα ή περισσότερα κρυμμένα επίπεδα και ένα επίπεδο εξόδου. Τα multilayer perceptron εφαρμόζονται σε πολλά και δύσκολα προβλήματα και εκπαιδεύονται με επίβλεψη με το γνωστό αλγόριθμο της ανάστροφης διάδοσης του σφάλματος(error back propagation) και μπορεί να θεωρηθεί ως γενίκευση του αλγορίθμου ελαχίστων τετραγώνων. Η διαδικασία εκπαίδευσης στον back propagation περιλαμβάνει υπολογισμούς που υλοποιούνται σε δυο περάσματα μέσω των επιπέδων του δικτύου : ένα πέρασμα κατά την ευθεία φορά(από είσοδο προς έξοδο) και ένα πέρασμα κατά

την ανάστροφη φορά(από έξοδο προς είσοδο). Στο πρώτο πέρασμα τα βάρη των συνδέσεων του δικτύου είναι σταθερά ενώ κατά το δεύτερο (ανάστροφο) τα βάρη προσαρμόζονται σύμφωνα με τον κανόνα διόρθωσης σφάλματος. Πιο συγκεκριμένα η πραγματική τιμή εξόδου αφαιρείται από την αντίστοιχη επιθυμητή και παράγεται το σήμα σφάλματος, το οποίο στη συνέχεια προωθείται στο δίκτυο κατά την ανάστροφη πορεία. Σημειώνουμε πως τα βάρη των συνδέσεων αναπροσαρμόζονται ώστε να μετατοπισθεί η απόκριση του δικτύου πλησιέστερα στην επιθυμητή. Τα multilayer perceptron έχουν πολύ μεγάλη υπολογιστική ισχύ αλλά παρουσιάζουν και πολλές ασάφειες στη θεωρητική ανάλυσή τους λόγω παρουσίας κατανεμημένης μορφής μη γραμμικότητας και μεγάλης συνδεσιμότητας. Επίσης η χρήση κρυμμένων νευρώνων δυσκολεύει την οπτικοποίηση της διαδικασίας μάθησης. Άποψη του multilayer perceptron φαίνεται παρακάτω:



Σχήμα 3ε- Πολυεπίπεδο perceptron [Α.Λύκας, 1999]

3.2.9.1 Εφαρμογές multilayer perceptron.

Η βασική λειτουργία που επιτελεί ένα πολυεπίπεδο perceptron είναι η υλοποίηση απεικόνισης(mapping) από το χώρο των εισόδων στο χώρο των εξόδων, χρησιμοποιώντας τα ζεύγη εκπαίδευσης και τους κατάλληλους αλγορίθμους εκπαίδευσης(back propagation).

Μάλιστα έχει αποδειχθεί θεωρητικά ότι το MLP έχει αυξημένες δυνατότητες απεικόνισης και συγκεκριμένα χαρακτηρίζεται από την ιδιότητα της παγκόσμιας προσέγγισης. Παραλείποντας τον αυστηρό μαθηματικό ορισμό, η ιδιότητα αυτή με απλά λόγια μας λέει το εξής: ένα multilayer perceptron με τουλάχιστον ένα κρυμμένο επίπεδο με μη γραμμικές μονάδες μπορεί να προσεγγίσει μια οποιαδήποτε συνάρτηση με οποιαδήποτε ακρίβεια, αυξάνοντας κατάλληλα τον αριθμό των κρυμμένων μονάδων.

Η ιδιότητα αυτή είναι τρομερά σημαντική διότι μας εξασφαλίζει ότι το multilayer perceptron μπορεί να υλοποιήσει οποιαδήποτε απεικόνιση, αλλά δεν είναι πρακτικά χρήσιμη, διότι δε μας λέει τίποτα για το πώς θα υλοποιήσουμε την απεικόνιση. Το πρόβλημα του καθορισμού του αριθμού των κρυμμένων μονάδων που απαιτούνται για ένα δεδομένο σύνολο εκπαίδευσης αποτελεί σήμερα βασικό ερευνητικό ζήτημα σχετικά με το multilayer perceptron.

Από τη στιγμή που ένα MLP έχει την ικανότητα να απεικονίζει ένα διάνυσμα πραγματικών εισόδων, έχει χρησιμοποιηθεί με ιδιαίτερη επιτυχία στη κατασκευή συστημάτων πρόβλεψης(prediction), για την κατασκευή μοντέλων από δεδομένα, για τον έλεγχο συστημάτων, μέχρι και για την επίλυση διαφορικών εξισώσεων.

Η βασικότερη όμως εφαρμογή του multilayer perceptron είναι σε προβλήματα ταξινόμησης(classification). Στη περίπτωση αυτή τα δεδομένα είναι της μορφής (πρότυπο, κατηγορία) και προκειμένου να εκπαιδευτεί το MLP απαιτείται μια διαδικασία κωδικοποίησης των κατηγοριών. Σκοπός της διαδικασίας αυτής είναι η μετατροπή του προβλήματος ταξινόμησης σε πρόβλημα απεικόνισης, μέσω της αντιστοίχισης κάθε κατηγορίας σε κάποιο διάνυσμα εξόδου. Με τον τρόπο αυτό το αρχικό σύνολο εκπαίδευσης μετασχηματίζεται ώστε να περιέχει ζεύγη της μορφής (πρότυπο, διάνυσμα εξόδου) και να μπορεί να χρησιμοποιηθεί το multilayer perceptron για την υλοποίηση της απεικόνισης.

Ο ευρύτερα χρησιμοποιούμενος τρόπος κωδικοποίησης των κατηγοριών είναι η κωδικοποίηση 1-από-κ για ένα πρόβλημα κατηγοριών. Στη κωδικοποίηση αυτή το διάνυσμα εξόδου έχει συνιστώσες (t_1, \dots, t_k) και η κατηγορία C_k κωδικοποιείται θέτοντας $t_k=1$ και $t_i=0$ για $i \neq k$. Για παράδειγμα σε ένα πρόβλημα με τρεις κατηγορίες τα αντίστοιχα 3 διανύσματα εξόδου είναι τα $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ και φυσικά απαιτείται ένας MLP με τρεις εξόδους. Η ταξινόμηση ενός προτύπου γίνεται εφαρμόζοντας το πρότυπο ως είσοδο στο δίκτυο και επιλέγοντας την κατηγορία που αντιστοιχεί στην έξοδο με τη μεγαλύτερη τιμή. Όσο πιο κοντά στο 1 είναι αυτή η έξοδος και κοντά στο 0 οι υπόλοιπες εξοδοί, τόσο πιο αξιόπιστη είναι η ταξινόμηση. Ειδικά για την περίπτωση δυο κατηγοριών, χρησιμοποιείται και η κωδικοποίηση με μια έξοδο: αντιστοιχίζουμε την έξοδο $t=0$ στη μια κατηγορία (C_1) και την έξοδο $t=1$ στην άλλη κατηγορία(C_2). Στην περίπτωση αυτή, η ταξινόμηση ενός προτύπου(αφού έχει γίνει η εκπαίδευση) γίνεται ως εξής: αν η έξοδος είναι μεγαλύτερη του 0.5 τότε το πρότυπο ταξινομείται στη κατηγορία C_2 αλλιώς στη C_1 .

3.3 Rosetta.

Το rosetta είναι ένα εργαλείο τεχνητής νοημοσύνης που στηρίζεται στη λεγόμενη θεωρία των rough sets. Είναι προϊόν της διδακτορικής διατριβής του Aleksander Ohn, φοιτητή του τμήματος Computer and Information science του Norwegian University of science and technology, Trondheim, Norway[Alexander Ohn, 2001]. Εξαιτίας του γεγονότος ότι δεν αποτελεί προϊόν λογισμικού κάποιας εμπορικής εταιρίας, τα εγχειρίδια χρήσης του και ανάλυσής του είναι πολύ «φτώχα» σε πληροφόρηση και πολύ συνοπτικά. Έτσι δεν είναι πολύ εύκολο να αναλυθούν οι

μεθοδολογίες που χρησιμοποιεί ο συγκεκριμένος φοιτητής, ο οποίος άλλωστε δηλώνει πως ο σκοπός που γράφτηκε το εγχειρίδιο αυτό είναι απλά να κατευθύνει αυτούς που είναι γνώστες του αντικειμένου, και πως κάποιος για να έχει συνολική εικόνα πρέπει να κατέχει τη γνώση της βιβλιογραφίας του.

3.3.1 Rough Sets.

Η θεωρία των rough sets είναι υπο συνεχή ανάπτυξη για πάνω από 20 χρόνια τώρα, και πολλοί ερευνητές ασχολούνται με αυτή τη μεθοδολογία. Παρουσιάστηκε από τον Zdzislaw Pawlak στις αρχές της δεκαετίας του 80 και αποτελεί ένα μαθηματικό εργαλείο που ασχολείται με περιπτώσεις αβεβαιότητας και πολυπλοκότητας. Αυτή η προσέγγιση είναι ουσιώδους σημασίας για την επιστήμη της τεχνητής νοημοσύνης ειδικά σε πεδία μηχανικής μάθησης, εκμαίευσης της γνώσης, ανάλυση αποφάσεων, ταίριασμα προτύπων κτλ.

Η θεωρία των rough sets στηρίζεται στην υπόθεση πως για οτιδήποτε και αν συζητάμε στο σύμπαν, του προσδίδουμε παράλληλα και ένα πλήθος πληροφοριών που το χαρακτηρίζουν. Κάθε σετ δυσδιάκριτων ή παρόμοιων αντικειμένων ονομάζεται elementary set, και δημιουργεί ατομική(κοκκώδη) μορφή της γνώσης για το σύμπαν. Κάθε ένωση μερικών τέτοιων σετ αποτελεί ένα διαυγές σετ αλλιώς είναι rough, δηλαδή πολύπλοκο και ανακριβές.

Υποθέτουμε πως στη προσέγγισή μας κάθε πολύπλοκο σύνολο αποτελείται και αντικαθιστάται από ένα ζευγάρι από ακριβή υποσύνολα, που λέγονται ανώτατη και κατώτατη προσέγγιση του πολύπλοκου συνόλου. Η κατώτατη αποτελείται από όλα τα αντικείμενα τα οποία σίγουρα ανήκουν στο σύνολο, και η ανώτατη αποτελείται από εκείνα τα αντικείμενα τα οποία ενδεχομένως να ανήκουν στο σύνολο. Προφανώς η διαφορά μεταξύ της ανώτατης και της κατώτατης προσέγγισης ορίζει τη συνοριακή περιοχή του πολύπλοκου συνόλου.

Οι βασικές λειτουργίες της θεωρίας των rough sets ανάγονται στην εύρεση θεμελιωδών προτύπων στα δεδομένα που διαθέτουμε. Έτσι τα rough sets σχετίζονται με τη μηχανική μάθηση, την εκμαίευση γνώσης, τη στατιστική και τη επαγωγική εξαγωγή συμπεράσματος. Ωστόσο η ερμηνεία των συμπερασμάτων κινείται εκτός της θεωρίας των rough sets και μπορεί να χρησιμοποιηθεί κατά ποικίλους τρόπους.

Είναι άξιο να σημειωθεί πως τα rough sets υπερκαλύπτουν σε σημαντικό βαθμό πολλές άλλες μαθηματικές θεωρίες. Ιδιαίτερα ενδιαφέροντα είναι η σχέση με τη θεωρία των ασαφών συστημάτων, γιατί οι δυο θεωρίες αλληλοσυμπληρώνονται κατά μια έννοια αλλά είναι θεμελιωδώς διαφορετικές, μιας και αναφέρονται σε ποικίλες όψεις της ανακρίβειας.

3.3.2 Εφαρμογές των rough sets.

Τα πιο σημαντικά πεδία εφαρμογής των rough sets είναι:

- ιατρική διάγνωση
 - φαρμακολογία
 - πρόβλεψη αποθέματος και ανάλυση οικονομικών δεδομένων
 - τραπεζικά
 - αναγνώριση προτύπων, περιλαμβανομένων αναγνώρισης γραφής και ομιλίας.
 - Σχεδιασμός συστημάτων ελέγχου
 - Σχεδίαση ψηφιακής λογικής
- και πολλά άλλα.

3.4 Utadis.

Η μέθοδος UTADIS (UTilités Additives DIScriminantes) βασίζεται στις αρχές της αναλυτικής-συνθετικής προσέγγισης. Η συγκεκριμένη μέθοδος δεν είναι καινούργια στο χώρο της πολυκριτήριας ανάλυσης. Η πρώτη παρουσίασή της πραγματοποιήθηκε ήδη στις αρχές της δεκαετίας του 1980 από τους Devaud et al. (1980), καθώς και τους Jacquet-Lagrèze και Siskos (1982)[M.Doumpos and C.Zorounidis, 2002]. Το Utadis αποτελούσε ένα εργαλείο που ήταν πολύ ευέλικτο στη διαδικασία ανάπτυξης προτύπων ταξινόμησης αλλά και στη μοντελοποίηση του συστήματος αξιών του αποφασίζοντος, αλλά παρόλα αυτά δεν έτυχε μεγάλης αποδοχής για περισσότερο από μια δεκαετία. Μοναδικές μέχρι το 1997 εφαρμογές της μεθόδου ήταν αυτές που παρουσιάστηκαν στην εργασία των Devaud et al. (1980) για την αξιολόγηση των υποκαταστημάτων μιας εμπορικής επιχείρησης, καθώς στην εργασία του Jacquet-Lagrèze (1995) για την αξιολόγηση προγραμμάτων έρευνας και ανάπτυξης. Μετά το 1997 παρουσιάστηκαν διάφορες εφαρμογές της μεθόδου, κυρίως στο χώρο της χρηματοοικονομικής διοίκησης [Zorounidis και Doumpos (1997, 1998, 1999α-β), Doumpos και Zorounidis (1998), Zorounidis et al. (1999)] με την παράλληλη υλοποίηση της μεθόδου σε πολυκριτήρια συστήματα υποστήριξης αποφάσεων, όπως στα συστήματα FINCLAS (Zorounidis και Doumpos, 1998) και PREFDIS (Zorounidis και Doumpos, 2000α).

3.4.1 Βασικές αρχές του Utadis.

Μια από τις πλέον γνωστές μεθόδους στο χώρο της αναλυτικής-συνθετικής προσέγγισης της πολυκριτήριας ανάλυσης αποφάσεων είναι η μέθοδος UTA (Utilités Additives)[M.Doumpos and C.Zorounidis, 2002]. Η μέθοδος UTA αναπτύχθηκε από τους Jacquet-Lagrèze και Siskos (1982) ως μια διαδικασία μονότονης παλινδρόμησης με σκοπό την ανάπτυξη προσθετικών συναρτήσεων χρησιμότητας για την κατάταξη ενός συνόλου εναλλακτικών δραστηριοτήτων από τις καλύτερες προς τις χειρότερες, βάσει μιας δεδομένης προδιάταξης των δραστηριοτήτων, η οποία καθορίζεται από τον αποφασίζοντα. Η μέθοδος UTADIS αποτελεί μια προσαρμογή της μεθόδου UTA στην περίπτωση όπου σκοπός δεν είναι η κατάταξη των εναλλακτικών δραστηριοτήτων, αλλά η ταξινόμησή τους σε προκαθορισμένες ομοιογενείς κατηγορίες. Οι κατηγορίες αυτές είναι διατεταγμένες από τις καλύτερες προς τις χειρότερες ως εξής:

$$C_1 \succ C_2 \succ \dots \succ C_q$$

Ως C_1 συμβολίζεται η κατηγορία που αποτελείται από τις καλύτερες εναλλακτικές δραστηριότητες. Οι δραστηριότητες που ανήκουν στην κατηγορία C_1 προτιμώνται έναντι των δραστηριοτήτων των υπόλοιπων κατηγοριών. Αντίστοιχα, η τελευταία κατηγορία C_q αποτελείται από τις χειρότερες εναλλακτικές δραστηριότητες. Προβλήματα ταξινόμησης στα οποία οι κατηγορίες ορίζονται κατά διατεταγμένο (ordinal) και όχι ονομαστικό (nominal) τρόπο εμφανίζονται ιδιαίτερα συχνά σε προβλήματα λήψης αποφάσεων. Αποτέλεσμα της βασικής αυτής υπόθεσης της μεθόδου όσον αφορά τη διάταξη των προκαθορισμένων κατηγοριών είναι ότι τα

χαρακτηριστικά που περιγράφουν την κάθε εξεταζόμενη εναλλακτική δραστηριότητα έχουν τη μορφή κριτηρίων αξιολόγησης.

Σκοπός της μεθόδου είναι η ανάπτυξη ενός υποδείγματος σύνθεσης των κριτηρίων αξιολόγησης έτσι ώστε το αποτέλεσμα της σύνθεσης αυτής να αποδίδει υψηλά σκορ στις εναλλακτικές δραστηριότητες της κατηγορίας C_1 και σταδιακά χαμηλότερα σκορ στις δραστηριότητες που ανήκουν στις χαμηλότερες κατηγορίες. Το υπόδειγμα σύνθεσης των κριτηρίων που χρησιμοποιείται στα πλαίσια της μεθόδου UTADIS, έχει τη μορφή μιας προσθετικής συνάρτησης χρησιμότητας:

$$U(g) = \sum_{i=1}^n p_i u_i(g_i) \quad (\alpha)$$

όπου

$g = (g_1, g_2, \dots, g_n)$ είναι το διάνυσμα n κριτηρίων αξιολόγησης

p_i είναι το βάρος(σημαντικότητα) του κριτηρίου

$$g_i \left(\sum_{i=1}^n p_i = 1 \right)$$

$u_i(g_i)$ είναι η συνάρτηση μερικής χρησιμότητας του κριτηρίου g_i

Οι συναρτήσεις μερικών χρησιμοτήτων (marginal utility functions) είναι μονότονες συναρτήσεις οριζόμενες στην κλίμακα του κάθε κριτηρίου αξιολόγησης. Οι συναρτήσεις αυτές δύνανται να έχουν οποιαδήποτε μορφή, γραμμική ή μη γραμμική και ικανοποιούν τις ακόλουθες δύο βασικές συνθήκες:

$$\left. \begin{array}{l} u_i(g_{i^*}) = 0 \\ u_i(g_i^*) = 1 \end{array} \right\}$$

Όπου, ως g_{i^*} και g_i^* ορίζονται, αντίστοιχα, η λιγότερο και η περισσότερο προτιμητέα τιμή του κριτηρίου g_i . Ουσιαστικά μέσω των συναρτήσεων μερικών χρησιμοτήτων πραγματοποιείται ένας μετασχηματισμός της κλίμακας του κάθε κριτηρίου αξιολόγησης σε μια νέα κλίμακα στο διάστημα $[0, 1]$. Η νέα αυτή κλίμακα αναπαριστά τη χρησιμότητα / αξία της κάθε τιμής του κριτηρίου.

Η αναγωγή των επιδόσεων των εναλλακτικών δραστηριοτήτων στα κριτήρια

αξιολόγησης σε όρους χρησιμότητας, μέσω του ορισμού των κατάλληλων συναρτήσεων μερικών χρησιμοτήτων παρέχει τα ακόλουθα δύο βασικά πλεονεκτήματα:

1. Επιτρέπει τη μοντελοποίηση και αναπαράσταση στο αναπτυσσόμενο υπόδειγμα της μη γραμμικής συμπεριφοράς του αποφασίζοντος κατά την αξιολόγηση των εναλλακτικών δραστηριοτήτων.
2. Επιτρέπει την αξιοποίηση ποιοτικών κριτηρίων αξιολόγησης χωρίς να απαιτείται η ποσοτικοποίησή τους μέσω του ορισμού μιας ποιοτικής κλίμακας. Για παράδειγμα, έστω ένα ποιοτικό κριτήριο σχετικό με το βαθμό οργάνωσης μιας επιχείρησης, το οποίο έχει τρεις διαβαθμίσεις: (α) καλή οργάνωση, (β) μέτρια οργάνωση, (γ) κακή οργάνωση. Η χρησιμοποίηση ενός τέτοιου ποιοτικού κριτηρίου σε ένα απλοϊκό υπόδειγμα που έχει τη μορφή ενός σταθμισμένου μέσου, απαιτεί την αντιστοίχιση μιας ποσοτικής κλίμακας στις τρεις αυτές διαβαθμίσεις, ώστε να είναι δυνατή η στάθμιση των τιμών του κριτηρίου με το βάρος του κριτηρίου. Ένα παράδειγμα μιας τέτοιας αντιστοίχισης είναι το ακόλουθο: κακή οργάνωση=1, μέτρια Οργάνωση=2, καλή οργάνωση=3. Αντίθετα, η αναγωγή της ποιοτικής κλίμακας σε όρους χρησιμότητας δεν απαιτεί την ποσοτικοποίηση του κριτηρίου, αλλά μόνο τον προσδιορισμό των μερικών χρησιμοτήτων των τριών διαβαθμίσεων της ποιοτικής κλίμακας: $u(\text{κακή οργάνωση})$, $u(\text{μέτρια οργάνωση})$ και $u(\text{καλή οργάνωση})$. Στο συγκεκριμένο παράδειγμα και δεδομένου ότι η κακή και η καλή οργάνωση αποτελούν τις λιγότερο και περισσότερο προτιμητέες τιμές του κριτηρίου αντίστοιχα g_i^* και g_i^* απαιτείται μόνο ο καθορισμός της μερικής χρησιμότητας $u(\text{μέτρια οργάνωση})$, καθώς εξ'όρισμού $u(\text{κακή οργάνωση})=0$ και $u(\text{καλή οργάνωση})=1$.

Πολλαπλασιάζοντας τις μερικές χρησιμότητες μιας εναλλακτικής δραστηριότητας X_i σε καθένα από τα κριτήρια αξιολόγησης, με τα αντίστοιχα βάρη των κριτηρίων (από σχέση α) υπολογίζεται η ολική χρησιμότητα (αξία) της δραστηριότητας. Οι ολικές χρησιμότητες κυμαίνονται στο διάστημα $[0, 1]$ και αποτελούν το συνολικό δείκτη αξιολόγησης των εναλλακτικών δραστηριοτήτων λαμβάνοντας υπόψη όλα τα κριτήρια αξιολόγησης. Οι ολικές χρησιμότητες αποτελούν και το κριτήριο βάσει του οποίου λαμβάνεται η απόφαση ταξινόμησης των εναλλακτικών δραστηριοτήτων στις προκαθορισμένες κατηγορίες.

Γενικά στην περίπτωση q κατηγοριών, η ταξινόμηση των εναλλακτικών δραστηριοτήτων πραγματοποιείται βάση των ακόλουθων κανόνων:

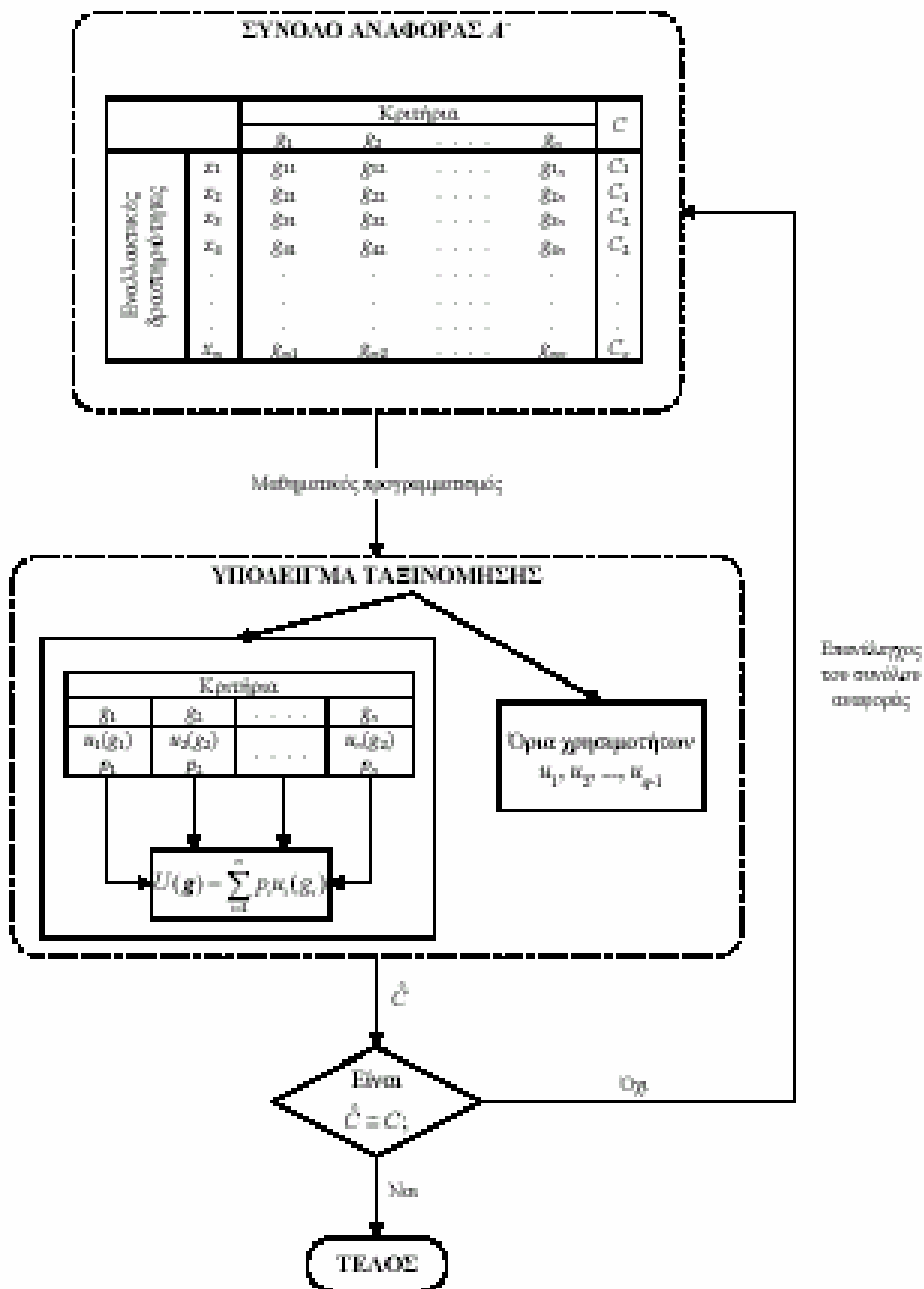
$$\left. \begin{array}{l} U(g_j) \geq u_1 \Rightarrow x_j \in C_1 \\ u_2 \leq U(g_j) < u_1 \Rightarrow x_j \in C_2 \\ u_k \leq U(g_j) < u_{k-1} \Rightarrow x_j \in C_k \\ U(g_j) < u_{q-1} \Rightarrow x_j \in C_q \end{array} \right\} (\beta)$$

3.4.2 Διαδικασία ανάπτυξης του προτύπου ταξινόμησης.

Όπως παρουσιάστηκε στην προηγούμενη παράγραφο, τα βασικά συστατικά στοιχεία του υποδείγματος ταξινόμησης που αναπτύσσεται μέσω της μεθόδου UTADIS περιλαμβάνουν τα βάρη των κριτηρίων αξιολόγησης και τη μορφή των μερικών συναρτήσεων χρησιμότητας. Τα δύο αυτά στοιχεία καθορίζουν τη μορφή της αναπτυσσόμενης προσθετικής συνάρτησης χρησιμότητας. Παράλληλα όμως, βασικό στοιχείο του αναπτυσσόμενου υποδείγματος ταξινόμησης, αποτελούν και τα όρια χρησιμότητας βάσει των οποίων λαμβάνεται η απόφαση για την ταξινόμηση των εναλλακτικών δραστηριοτήτων.

Ο καθορισμός αυτών των συστατικών στοιχείων του αναπτυσσόμενου υποδείγματος ταξινόμησης πραγματοποιείται στα γενικά πλαίσια που διέπουν την αναλυτική-συνθετική προσέγγιση. Πιο συγκεκριμένα, η ανάπτυξη του υποδείγματος ταξινόμησης επιτυγχάνεται μέσω μιας διαδικασίας, η οποία βασίζεται στη φιλοσοφία της γνωστής παλινδρόμησης (Σχήμα α). Αρχικά, χρησιμοποιείται ένα σύνολο αναφοράς A' (reference set), αποτελούμενο από m εναλλακτικές δραστηριότητες, οι επιδόσεις των οποίων περιγράφονται βάσει των n προεπιλεγμένων κριτηρίων αξιολόγησης. Η ταξινόμηση $C = \{C_1, C_2, \dots, C_q\}$ των εναλλακτικών δραστηριοτήτων του συνόλου A' στις προκαθορισμένες q κατηγορίες θεωρείται γνωστή. Για την κάθε κατηγορία το σύνολο αναφοράς θα πρέπει να περιλαμβάνει επαρκή αριθμό εναλλακτικών δραστηριοτήτων που εντάσσονται στην κατηγορία αυτή. Το πλήθος των εναλλακτικών δραστηριοτήτων του συνόλου αναφοράς που εντάσσεται στην κατηγορία C_1 θα συμβολίζεται στο εξής ως m_1 . Αντίστοιχα, ως m_2, m_3, \dots, m_q θα συμβολίζεται το πλήθος των εναλλακτικών δραστηριοτήτων του συνόλου αναφοράς που ανήκουν στις κατηγορίες C_2, C_3, \dots, C_q .

Ουσιαστικά, το σύνολο αναφοράς είναι το αντίστοιχο του δείγματος εκμάθησης (training sample), όρος ο οποίος χρησιμοποιείται στο χώρο της στατιστικής, της οικονομετρίας, και της τεχνητής νοημοσύνης (νευρωνικά δίκτυα, μηχανική μάθηση) υποδεικνύοντας το δείγμα στο οποίο βασίζεται η ανάπτυξη των υποδειγμάτων ταξινόμησης (σχήμα 3ζ).



Σχήμα 3στ-[Doumpos, M. and Zopounidis, 2002]

Έχοντας ως δεδομένη την ταξινόμηση C των εναλλακτικών δραστηριοτήτων του συνόλου αναφοράς, σκοπός της μεθόδου είναι η ανάπτυξη ενός υποδείγματος σύνθεσης των κριτηρίων αξιολόγησης το οποίο, σε συνδυασμό με τον καθορισμό των ορίων χρησιμότητας, θα ταξινομεί τις εναλλακτικές δραστηριότητες του συνόλου αναφοράς στις προκαθορισμένες κατηγορίες με τη μεγαλύτερη δυνατή ακρίβεια, ή ανάλογα με το μικρότερο δυνατό σφάλμα ταξινόμησης. Ο όρος «σφάλμα»

αναφέρεται στις διαφορές που εντοπίζονται μεταξύ της ταξινόμησης \hat{C} που επιτυγχάνει το αναπτυσσόμενο υπόδειγμα σύνθεσης των κριτηρίων, σε σχέση με τη δεδομένη ταξινόμηση C των εναλλακτικών δραστηριοτήτων που εμπεριέχονται στο σύνολο αναφοράς. Οι διαφορές αυτές μπορούν να αποδοθούν μέσω του ορισμού μιας δυαδικής μεταβλητής E , η οποία ορίζεται για κάθε εναλλακτική δραστηριότητα X_j ως εξής:

$$E_j = \begin{cases} 0, & \text{αν } \hat{C}_j = C_j \\ 1, & \text{αν } \hat{C}_j \neq C_j \end{cases}$$

Βάσει αυτής της δυαδικής μεταβλητής, το σφάλμα γ της ταξινόμησης μπορεί να υπολογιστεί ως το ποσοστό των εναλλακτικών δραστηριοτήτων του συνόλου αναφοράς που ταξινομούνται κατά εσφαλμένο τρόπο:

$$\gamma = \frac{\sum_{j=1}^m E_j}{m} \in [0,1]$$

ή ακόμα καλύτερα ως:

$$\gamma = \frac{1}{q} \sum_{k=1}^q \left\{ \frac{\sum_{\forall x_j \in C_k} E_j}{m_k} \in [0,1] \right\} \quad (\gamma)$$

Εφόσον το σφάλμα του υποδείγματος στην ταξινόμηση των εναλλακτικών δραστηριοτήτων του συνόλου αναφοράς, όπως πλέον αυτό ορίζεται από τη σχέση (γ) , κριθεί ικανοποιητικό, τότε το υπόδειγμα είναι δυνατόν να χρησιμοποιηθεί για την ταξινόμηση οποιασδήποτε άλλης εναλλακτικής δραστηριότητας που δεν ανήκει στο

σύνολο αναφοράς. Διαφορετικά, εάν η ταξινόμηση \hat{C} που υποδεικνύει το αναπτυσσόμενο υπόδειγμα ταξινόμησης διαφέρει σημαντικά από την προκαθορισμένη ταξινόμηση C τότε ο αποφασίζων θα πρέπει να επανελέγξει το σύνολο αναφοράς σχετικά με την πληρότητα και επάρκεια των κριτηρίων αξιολόγησης, καθώς και την επάρκεια των εναλλακτικών δραστηριοτήτων που περιλαμβάνει για την περιγραφή των προκαθορισμένων κατηγοριών. Εναλλακτικά, βέβαια, είναι πιθανό η μορφή του υποδείγματος όπως αυτή διατυπώνεται μέσω της προσθετικής συνάρτησης χρησιμότητας να μην είναι σε θέση να αναπαραστήσει το σύστημα προτιμήσεων και αξιών του αποφασίζοντος, γεγονός το οποίο υποδεικνύει την ανάγκη διερεύνησης μιας διαφορετικής μορφής σύνθεσης των κριτηρίων αξιολόγησης.

Θα πρέπει βέβαια να τονιστεί ότι το χαμηλό σφάλμα κατά την ταξινόμηση των εναλλακτικών δραστηριοτήτων του συνόλου αναφοράς δεν συνεπάγεται και την υψηλή δυνατότητα γενίκευσης του υποδείγματος ταξινόμησης, αν και αποτελεί μια ισχυρή ένδειξη προς την κατεύθυνση αυτή. Αντίθετα, υψηλό σφάλμα κατά την ταξινόμηση των εναλλακτικών δραστηριοτήτων του συνόλου αναφοράς οδηγεί, με βεβαιότητα, στο συμπέρασμα ότι το υπόδειγμα ταξινόμησης είναι ανεπαρκές.

3.4.3 Τα μαθηματικά πίσω από το Utadis

Η επίτευξη του στόχου της ανάπτυξης ενός υποδείγματος το οποίο μεγιστοποιεί τη συμφωνία μεταξύ της εκτιμώμενης ταξινόμησης \hat{C} και της δεδομένης ταξινόμησης C των εναλλακτικών δραστηριοτήτων του συνόλου αναφοράς, επιτυγχάνεται στη μέθοδο UTADIS μέσω της χρησιμοποίησης τεχνικών μαθηματικού προγραμματισμού.

Ουσιαστικά, η επίλυση του προβλήματος αυτού αφορά την ελαχιστοποίηση της συνάρτησης (γ) . Καθώς οι μεταβλητές E_j της συνάρτησης αυτής είναι δυαδικές 0–1 μεταβλητές, είναι προφανές ότι η άμεση ελαχιστοποίηση της συνάρτησης (γ) μπορεί να επιτευχθεί μόνο μέσω τεχνικών ακέραιου μαθηματικού προγραμματισμού. Η επίλυση όμως προβλημάτων ακέραιου μαθηματικού προγραμματισμού απαιτεί ιδιαίτερα αυξημένο υπολογιστικό φόρτο. Παρά τις μέχρι σήμερα ερευνητικές προσπάθειες που έχουν πραγματοποιηθεί με σκοπό την ανάπτυξη ευρεστικών αλγορίθμων και τη χρησιμοποίηση εξελιγμένων τεχνικών βελτιστοποίησης για την επίλυση προβλημάτων ακέραιου μαθηματικού προγραμματισμού στο χώρο της ταξινόμησης, ο απαιτούμενος υπολογιστικός φόρτος για την επίλυση τέτοιων προβλημάτων, παραμένει ακόμα και σήμερα ιδιαίτερα αυξημένος.

Για το λόγο αυτό προτείνεται η προσέγγιση της συνάρτησης (γ) η οποία διατυπώνεται ως εξής:

$$\gamma' = \frac{1}{q} \sum_{k=1}^q \left\{ \frac{\sum_{\forall x_j \in C_k} \sigma_j}{m_k} \right\} \quad (\delta)$$

Σε αυτή τη συνάρτηση μέτρησης του σφάλματος ταξινόμησης, οι μεταβλητές σ_j ανήκουν στο σύνολο των θετικών πραγματικών αριθμών, και ορίζονται ως εξής:

$$\sigma_j = \begin{cases} > 0, \text{αν } \hat{C}_j \neq C_j \\ 0, \text{αν } \hat{C}_j = C_j \end{cases} \quad (\varepsilon)$$

Ουσιαστικά οι μεταβλητές σ_j αναπαριστούν το βαθμό του σφάλματος που πραγματοποιείται κατά την ταξινόμηση της κάθε εναλλακτικής δραστηριότητας x_j . Το σφάλμα αυτό ορίζεται ως η απόλυτη διαφορά μεταξύ της ολικής χρησιμότητας κάθε εσφαλμένως ταξινομημένης εναλλακτικής δραστηριότητας x_j και του ορίου χρησιμότητας το οποίο παραβιάζεται. Έτσι, βάσει του κανόνα ταξινόμησης (β) για τις εναλλακτικές δραστηριότητες της κατηγορίας C_1 , το σφάλμα αφορά την παραβίαση του ορίου χρησιμότητας u_1 , το οποίο αποτελεί το κάτω όριο της κατηγορίας C_1 διαχωρίζοντάς την από τις υπόλοιπες κατηγορίες. Για τις εναλλακτικές δραστηριότητες της τελευταίας κατηγορίας C_q , το σφάλμα αφορά την παραβίαση του ορίου χρησιμότητας u_{q-1} , το οποίο αποτελεί το άνω όριο της αντίστοιχης κατηγορίας. Για όλες τις υπόλοιπες εναλλακτικές δραστηριότητες που ανήκουν σε οποιαδήποτε ενδιάμεση κατηγορία C_k ($1 < k < q-1$) το σφάλμα ταξινόμησης μπορεί να αφορά είτε την παραβίαση του ορίου χρησιμότητας u_{k-1} (άνω όριο της κατηγορίας C_k), είτε την παραβίαση του ορίου χρησιμότητας u_k (κάτω όριο της κατηγορίας C_k).

Στην περίπτωση όπου το σφάλμα αφορά την παραβίαση του κάτω ορίου μιας κατηγορίας, τότε αυτό θα συμβολίζεται εφεξής ως σ^+ , ενώ αντίστοιχα ως σ^- θα συμβολίζεται το σφάλμα που αφορά την παραβίαση του άνω ορίου μιας κατηγορίας. Θα πρέπει βέβαια να τονιστεί ότι οι συναρτήσεις σφάλματος (γ) και (δ) δεν είναι απολύτως ισοδύναμες.

Η συνάρτηση σφάλματος της σχέσης (δ) αποτελεί μια ικανοποιητική προσέγγιση του πραγματικού σφάλματος της ταξινόμησης που ορίζει η συνάρτηση (γ), αποφεύγοντας παράλληλα τον αυξημένο υπολογιστικό φόρτο που απαιτεί η ελαχιστοποίηση του σφάλματος ταξινόμησης (γ.).

Σχετίζοντας τις δύο μορφές σφαλμάτων με τον αντίστοιχο κανόνα ταξινόμησης της σχέσης (β), αυτά μπορούν να διατυπωθούν μαθηματικά ως εξής:

$$\sigma_j^+ = \max \{0, u_k - U(g_j)\}, \forall x_j \in C_k, k = 1, 2, \dots, q-1$$

$$\sigma_j^- = \max \{0, U(g_j) - u_{k-1}\}, \forall x_j \in C_k, k = 2, \dots, q$$

Η ελαχιστοποίηση του σφάλματος της ταξινόμησης μπορεί να πραγματοποιηθεί μέσω της επίλυσης ενός προβλήματος μαθηματικού προγραμματισμού, το οποίο έχει την ακόλουθη γενική μορφή (πρόβλημα ΓΠ1):

$$\min \gamma' = \frac{1}{q} \sum_{k=1}^q \left\{ \frac{\sum_{\forall x_j \in C_k} \sigma_j}{m_k} \Leftrightarrow \min \left\{ \sum_{k=1}^q \left\{ \frac{\sum_{\forall x_j \in C_k} (\sigma_j^+ + \sigma_j^-)}{m_k} \right\} \right\} \right\}$$

υπό τους περιορισμούς:

$$U(g_j) - u_1 + \sigma_j^+ \geq \delta_1, \forall x_j \in C_1$$

$$\left. \begin{array}{l} U(g_j) - u_k + \sigma_j^+ \geq \delta_1 \\ U(g_j) - u_{k-1} - \sigma_j^- \leq -\delta_2 \end{array} \right\}, \forall x_j \in C_k$$

$$U(g_j) - u_{q-1} - \sigma_j^- \leq -\delta_2, \forall x_j \in C_q$$

$$U(g^*) = 1$$

$$U(g_*) = 0$$

$$u_k - u_{k+1} \geq s, \forall k = 1, 2, \dots, q-2$$

$$u_i(g_i) \text{ αυξουσες συναρτησεις}$$

$$\sigma_j^+ \geq 0, \sigma_j^- \geq 0, \forall j = 1, 2, \dots, m$$

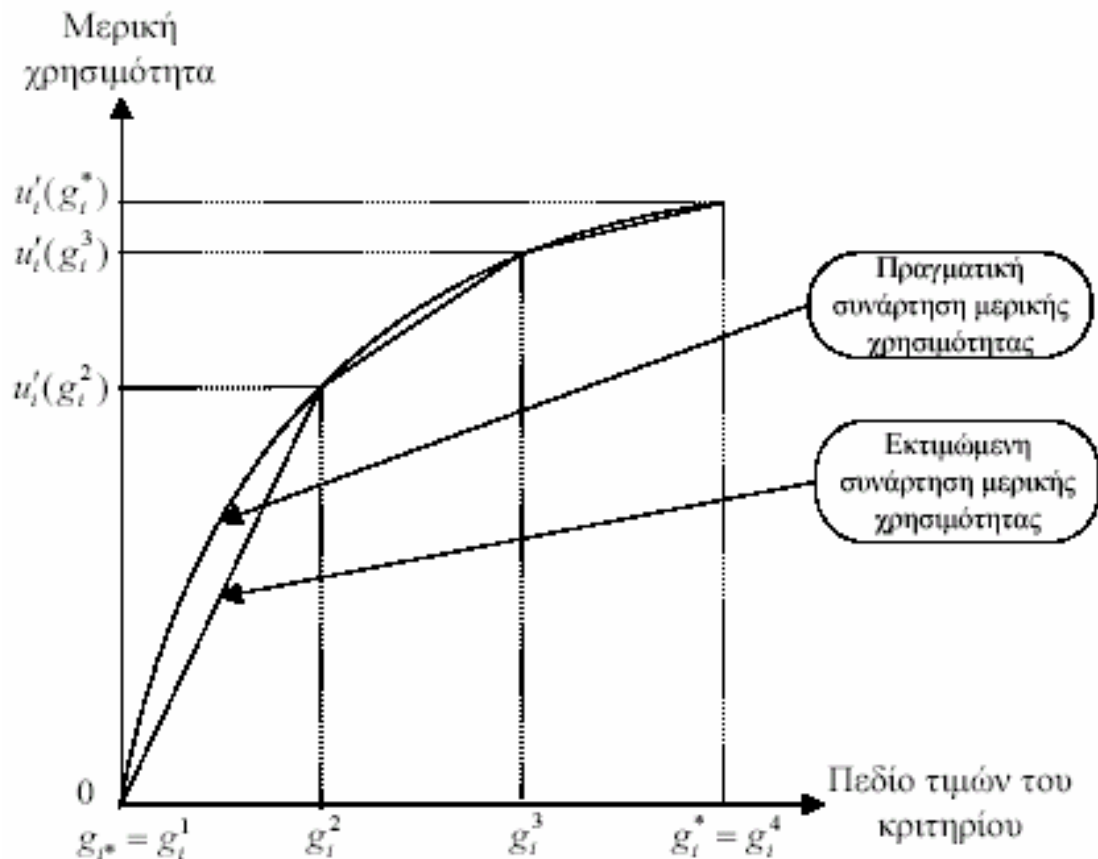
Η χρησιμοποίηση της μορφής της συνάρτησης χρησιμότητας όπως αυτή ορίστηκε από τη σχέση (α) οδηγεί μέσω του προβλήματος ΓΠ1 την επίλυση ενός μαθηματικού προγράμματος με μη γραμμικούς περιορισμούς, καθώς τόσο τα βάρη των κριτηρίων αξιολόγησης όσο και οι συναρτήσεις μερικών χρησιμοτήτων είναι άγνωστες μεταβλητές του προβλήματος. Προκειμένου να αποφευχθεί ο αυξημένος υπολογιστικός φόρτος που συνεπάγεται η επίλυση ενός προβλήματος βελτιστοποίησης υπό μη γραμμικούς περιορισμούς, στη μέθοδο UTADIS χρησιμοποιείται η ακόλουθη μορφή της προσθετικής συνάρτησης χρησιμότητας, η οποία είναι ισοδύναμη της μορφής (α):

$$U'(g) \sum_{i=1}^n u'_i(g_i)$$

ΟΠΟΥ, (ζ)

$$\left. \begin{aligned} u'_i(g_i) &= p_i u_i(g_i) \\ u'_i(g_i^*) &= 0 \\ u'_i(g_i^*) &= p_i \end{aligned} \right\}$$

Χρησιμοποιώντας τη συνάρτηση χρησιμότητας της σχέσης (ζ), το πρόβλημα ΓΠ1 έχει μοναδικές άγνωστες τις συναρτήσεις των μερικών χρησιμοτήτων των κριτηρίων αξιολόγησης $u'_i(g_i)$. Όμως οι συναρτήσεις αυτές μπορούν να έχουν οποιαδήποτε συνεχή και αύξουσα μορφή. Συνεπώς, απαιτείται η έκφραση των συναρτήσεων μερικών χρησιμοτήτων βάσει συγκεκριμένων μεταβλητών, η εκτίμηση των οποίων, μέσω της λύσης του γενικού προβλήματος ΓΠ1, θα προσδιορίζει και τη μορφή των συναρτήσεων μερικών χρησιμοτήτων. Αυτό επιτυγχάνεται μοντελοποιώντας τις συναρτήσεις μερικών χρησιμοτήτων ως κατά-τμήματα γραμμικές συναρτήσεις (piece-wise linear functions), ακολουθώντας τη διαδικασία η οποία παρουσιάζεται γραφικά στο Σχήμα 3η.



Σχήμα 3η-[Doumplos, M. and Zopounidis, 2002]

Όπως φαίνεται στο παραπάνω σχήμα, το πεδίο τιμών $[g_i^*, g_i^*]$ του κάθε κριτηρίου αξιολόγησης υποδιαιρείται σε a_i-1 ίσα υποδιαστήματα $[g_i^h, g_i^{h+1}]$, $h=1, 2, \dots, a_i-1$. Ο αριθμός των υποδιαστημάτων αυτών μπορεί να καθοριστεί είτε από τον ίδιο τον αποφασίζοντα, είτε χρησιμοποιώντας τον ακόλουθο απλό ευρεστικό κανόνα:

Το πλήθος των υποδιαστημάτων θα πρέπει να είναι τέτοιο ώστε για κάθε h να υπάρχει τουλάχιστον μια εναλλακτική δραστηριότητα x_j για την οποία $g_{ji} \in [g_i^h, g_i^{h+1}]$

Μέσω αυτής της διαδικασίας υποδιαίρεσης του πεδίου τιμών των κριτηρίων αξιολόγησης, οι αντίστοιχες συναρτήσεις μερικών χρησιμοτήτων μπορούν να εκτιμηθούν υπολογίζοντας τις μερικές χρησιμότητες σε κάθε ένα από τα επίπεδα (σημεία) g_i^2, \dots . Όπως φαίνεται και στο Σχήμα β, ο υπολογισμός των μερικών χρησιμοτήτων στα σημεία αυτά οδηγεί σε μια προσέγγιση της μορφής που μπορεί να έχει η πραγματική συνάρτηση χρησιμότητας. Θα πρέπει βέβαια να τονιστεί ότι ο καθορισμός ενός μεγάλου αριθμού υποδιαστημάτων, δεν οδηγεί απαραίτητα σε καλύτερη αναπαράσταση της πραγματικής συνάρτησης χρησιμότητας που χαρακτηρίζει το σύστημα αξιών και προτιμήσεων του αποφασίζοντος.

Με κάποιες επιπλέον μετατροπές στους περιορισμούς, κάνοντάς τους περιορισμούς μη αρνητικότητας καταφέρνουμε να ελαττώσουμε το φόρτο υπολογισμών για την επίλυση του γραμμικού προγράμματος. Η μετατροπή αυτή επιτυγχάνεται βάσει της διαδικασίας που προτάθηκε από τους Siskos και Yannacopoulos (1985), εισάγοντας μια νέα μεταβλητή, η οποία αναπαριστά τη διαφορά μεταξύ της μερικής χρησιμότητας του σημείου g_i^h και της μερικής χρησιμότητας του σημείου g_i^{h+1} , ως εξής:

$$w_{ih} = u'_i(g_i^{h+1}) - u'_i(g_i^h)$$

Βάσει του μετασχηματισμού αυτού, οι περιορισμοί μετατρέπονται πλέον σε περιορισμούς μη αρνητικότητας $w_{ih} \geq 0$. Εν τέλει το γραμμικό πρόγραμμα παίρνει την εξής μορφή:

$$\text{Min} \left\{ \sum_{k=1}^q \left[\frac{\sum_{\forall x_j \in C_k} (\sigma_j^+ + \sigma_j^-)}{m_k} \right] \right\}$$

Υπό τους περιορισμούς:

$$\sum_{i=1}^n \left(\sum_{p=1}^{r_k-1} w_{ip} + \frac{g_{\beta} - g_i^{r_p}}{g_i^{r_p+1} - g_i^{r_p}} w_{ir_p} \right) - u_1 + \sigma_j^+ \geq \delta_1, \quad \forall x_j \in C_1$$

$$\left. \sum_{i=1}^n \left(\sum_{p=1}^{r_k-1} w_{ip} + \frac{g_{\beta} - g_i^{r_p}}{g_i^{r_p+1} - g_i^{r_p}} w_{ir_p} \right) - u_k + \sigma_j^+ \geq \delta_1 \right\}, \quad \forall x_j \in C_k (k=2, 3, \dots, q-1)$$

$$\sum_{i=1}^n \left(\sum_{p=1}^{r_k-1} w_{ip} + \frac{g_{\beta} - g_i^{r_p}}{g_i^{r_p+1} - g_i^{r_p}} w_{ir_p} \right) - u_{k-1} - \sigma_j^- \leq -\delta_2$$

$$\sum_{i=1}^n \left(\sum_{p=1}^{r_k-1} w_{ip} + \frac{g_{\beta} - g_i^{r_p}}{g_i^{r_p+1} - g_i^{r_p}} w_{ir_p} \right) - u_{q-1} - \sigma_j^- \leq -\delta_2, \quad \forall x_j \in C_q$$

$$\sum_{i=1}^n \sum_{p=1}^{a_i-1} w_{ip} = 1,$$

$$u_k - u_{k+1} \geq s, \quad \forall k=1, 2, \dots, q-2$$

$$\sigma_j^+ \geq 0, \sigma_j^- \geq 0, \quad \forall j=1, 2, \dots, m$$

$$w_{ip} \geq 0, \quad \forall i=1, 2, \dots, n, \forall p=1, 2, \dots, a_i-1$$

Το παραπάνω γραμμικό πρόγραμμα είναι εκείνο το οποίο χρησιμοποιείται για την ανάπτυξη του υποδείγματος ταξινόμησης μέσω της μεθόδου UTADIS και οι εκτιμήσεις που προκύπτουν μέσω αυτού του προγράμματος ως προς τις παραμέτρους του αναπτυσσόμενου υποδείγματος ταξινόμησης, διαθέτουν την ιδιότητα της συνέπειας (consistency), όπως αυτή διατυπώθηκε από τους Charnes et al. (1955). Η εκτίμηση μιας παραμέτρου θεωρείται ότι είναι συνεπής, όταν καθώς αυξάνει το πλήθος των εναλλακτικών δραστηριοτήτων του συνόλου αναφοράς, η εκτιμώμενη τιμή της παραμέτρου προσεγγίζει ασυμπτωτικά την πραγματική της τιμή. Οι συγγραφείς (εκ των θεμελιωτών του χώρου του προγραμματισμού στόχων) θεωρούν ότι η συνέπεια είναι η πλέον σημαντική ιδιότητα που πρέπει να διέπει ένα υπόδειγμα μαθηματικού προγραμματισμού, καθώς υποδεικνύει τα αποτελέσματα τα οποία επιτυγχάνονται όταν χρησιμοποιηθεί όλη η υπάρχουσα πληροφορία.

Στην περίπτωση του τελευταίου μας γραμμικού προγράμματος, καθώς το μέγεθος του συνόλου αναφοράς αυξάνει, και εφόσον οι εναλλακτικές δραστηριότητες που προστίθεται, δεν κυριαρχούνται από τις ήδη υπάρχουσες (άρα προσθέτουν νέα πληροφορία), τότε ουσιαστικά στο πρόγραμμα εντάσσονται νέοι, μη πλεονασματικοί περιορισμοί. Αυτοί περιορίζουν το σύνολο των εφικτών λύσεων του προβλήματος, γεγονός το οποίο ασυμπτωτικά (στην περίπτωση μεγάλων συνόλων αναφοράς) θα οδηγήσει στην εκτίμηση ενός μοναδικού υποδείγματος ταξινόμησης, το οποίο θα είναι εκείνο που πραγματικά χαρακτηρίζει τον αποφασίζοντα.

ΚΕΦΑΛΑΙΟ 4^ο

*ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ
ΔΕΔΟΜΕΝΩΝ-
ΕΦΑΡΜΟΓΗ ΤΩΝ ΕΡΓΑΛΕΙΩΝ
ΚΑΙ ΑΝΑΛΥΣΗ ΤΗΣ ΠΡΟΒΛΕΨΗΣ*

4.1 Ανάλυση δεδομένων

Τα δεδομένα μας, όπως αναφέραμε και προηγουμένως, είναι 553. Όσοι ερωτήθηκαν ήταν άνω των 17 και ανήκαν και στα δύο φύλλα. Συγκεκριμένα ήταν 272 άνδρες και 281 γυναίκες, περίπου ίδια τα δείγματα μας ως προς το φύλο. Επίσης ηλικιακά το μεγαλύτερο ποσοστό των ερωτηθέντων, δηλαδή το 48% περίπου, ήταν ανάμεσα στα 17 και τα 28. Συγκεκριμένα αυτοί ήταν 268, ενώ ηλικίας από 29 ως 45 ήταν 136 ή 24.5%. Τέλος οι μεγάλης ηλικίας από 66 ως 90 ήταν μόλις 15 γιατί άλλωστε δεν αποτελούν και μεγάλη καταναλωτική «δύναμη». Τα άτομα που δεν δήλωσαν καθόλου την ηλικία τους ήταν οκτώ. Όσο για το επίπεδο της μόρφωσής τους το μεγαλύτερο ποσοστό, 60%, ήταν άτομα ανώτερης ή ανώτατης εκπαίδευσης, ενώ μόλις 17 ήταν επιπέδου δημοτικού από αυτούς φυσικά που απάντησαν, διότι 30 δεν απάντησαν καθόλου. Ως προς το μέγεθος της οικογένειας από άποψη ενηλίκων την πρωτιά των ερωτηθέντων κατείχαν με ποσοστό 38.5% οι οικογένειες με δυο ενήλικες στο νοικοκυριό τους, ενώ 154 (27.8%) ήταν αυτές που είχαν μόνο ένα ενήλικο. Τέλος γνωρίζουμε πως 69 άτομα είχαν 2 παιδιά στην οικογένεια, 61 ένα παιδί, 20 τρία παιδιά, 8 τέσσερα παιδιά στην οικογένεια και ένας μόνο ερωτηθείς είχε πάνω από τέσσερα. 394 δεν απάντησαν καθόλου. Άρα παρατηρούμε πως γενικά δεν είναι πάρα πολλές οι πολυμελείς οικογένειες αλλά κινούνται στα κλασικά πλαίσια 4^{ων} ατόμων, οπότε και η κατανάλωση ελαιολάδου θα κινείται στα κανονικά επίπεδα μιας μέσης οικογένειας.

Παρακάτω ακολουθεί μια στατιστική παρουσίαση των προτιμήσεων που εξεδήλωσαν οι ερωτηθέντες για τις περιπτώσεις εναλλακτικών απαντήσεων, δηλαδή παρουσιάζονται τα ποσοστά της κάθε απάντησης, και ύστερα γίνεται μια πιο λεπτομερής ανάλυση ορισμένων από αυτές.

1. What other types of oil do you purchase (or use) ? What do you use it for?

Product	Usage					
	Frying	Marinades	Sauces	Salads	Baking	Other (Define)
Extra virgin olive oil	35%	19%	22%	38%	20%	0.5%
Olive oil	46%	18%	19%	34%	23%	0.1%
Other Vegetable oils (Soya oil, Corn-oil, Sunflower oil etc.)	68%	16%	18%	27%	37%	0.9%

2. Which kinds of oils do you use and how frequently?

Frequency

Product	Frequency		
	3-4 times per week (or more)	Once in two weeks	Once in a month or less
Extra virgin olive oil	29%	27%	38%
Olive oil	12%	24%	26%
Other Vegetable oils (Soya oil,	15%	14%	15%

Corn-oil, , Sunflower oil etc.)

3. From where do you usually buy olive oil?

Supermarket 75%
 Delicatessens 2%
 Healthfood 2%
 Bakery/Butcher/Wine -Cave 1%
 Other 3%

4. Do you buy a particular brand of olive oil?

Yes 29% No 63%

If yes which one? Bertolli 3.4%
 Lorena 2.1%
 Ah organic 1%
 Sitia 0.3%
 Albert Heijn 1%
 other 2.8%

5. Do you know which country the Olive oil you purchase comes from?

Yes 31% No 59%

If yes which one? Italy 6.8%
 Greece 2.3%
 Spain 1.6%
 India 0.1%
 Other 2%

6. How important do you consider the following factors when you choose olive oil ?

	Very Important	Fairly Important	Neutral	Not particularly Important	Not at all Important	I don't know
	1	2	3	4	5	
• Price	31%	37%	16%	5%	4%	3%
• Health	40%	32%	12%	5%	3%	3%
• Flavour	49%	30%	9%	3%	2%	3%
• Color	8%	26%	28%	14%	16%	3%
• Bottle	7%	16%	22%	22%	25%	3%
• Odour	15%	32%	24%	11%	9%	4%
• Label	3%	12%	26%	22%	27%	4%
• Brand Name	6%	15%	20%	20%	30%	3%
• Naturalness	21%	35%	23%	7%	6%	4%
• Country of Origin	4%	11%	17%	24%	22%	5%
• Special offers /	16%	22%	19%	8%	19%	7%

reward points.						
----------------	--	--	--	--	--	--

7. What type of packaging do you prefer when you are considering buying olive oil?

Plastic	14%	Round	22%	1 lit.	38%
Glass	72%	Semi-Round	8%	750 ml	22%
		Square	45%	500 ml	20%
Other (what)	3%	Other (what)	6%	250 ml	4%
				Other (what)	2%

8. What is your opinion about these olive oils according to the factors below?

Products			Lorena 750 ml	AH Organic 500 ml	Sitia 500 ml	Bertolli 500 ml
Color (a)	Very good	1	13%	12%	3%	10%
	Good	2	46%	48%	18%	41%
	Indifferent/Neutral	3	21%	22%	23%	26%
	Bad	4	7,4%	5,4%	20%	10%
	Very Bad	5	0,5%	0,5%	8%	1,5%
	I do not know		7%	7%	21%	5,6%
Package (b)	Very good	1	14%	10%	10%	10%
	Good	2	58%	55%	36%	41%
	Indifferent/Neutral	3	15%	19%	23%	23%
	Bad	4	2%	5%	14%	13%
	Very Bad	5	0,1%	0,7%	2%	2%
	I do not know		4%	5%	8%	5%
Image Reputation (c)	Very good	1	9%	11%	2,5%	25%
	Good	2	24%	32%	7%	36%
	Indifferent/Neutral	3	19%	23%	15%	10%
	Bad	4	16%	8%	25%	7%
	Very Bad	5	9,7%	6%	20%	5%
	I do not know		17%	13%	25%	13%
Quality (d)	Very good	1	15%	10%	5%	11%
	Good	2	32%	28%	9%	35%
	Indifferent/Neutral	3	13%	16%	19%	15%
	Bad	4	2%	1,8%	5%	3%
	Very Bad	5	0,1%	0,5%	2%	1,2%
	I do not know		32%	37%	54%	29%
Price (e)	Very good	1	39%	0,3%	2%	2%
	Good	2	36%	5%	26%	18%
	Indifferent/Neutral	3	10%	14%	37%	36%
	Bad	4	2%	41%	14%	26%
	Very Bad	5	1%	26%	3,6%	5%
	I do not know		6%	6%	9,5%	6%

Product	Price		4.95	9.95	7.95	7.95
---------	-------	--	------	------	------	------

9. Put the following products in numerical order according to your preferences; with 1 being the best?

Order \	Products			
	Lorena 750 ml	AH Organic 500 ml	Sitia 500 ml	Bertolli 500 ml
1	48%	18%	5%	22%
2	24%	25%	17%	26%
3	13%	23%	27%	26%
4	6,6%	25%	41%	16%

10. Do you ever purchase organic produce ?

Yes 72% No 25%

11. Have you ever bought organic olive oil ?

Never 65% Once 15% More than once 18%

(If the respondent has never bought organic olive oil skip the following question)

12. Do you think there would be any advantages in purchasing organic olive oil ? :

Yes No

- Better Quality 33% 66%
- Chemical free . 55% 45%
- Environmentally friendly 49% 51%
- Support Organic Philosophy 38% 62%
- Don't know 17% 83%
- Other (please state) 2% 97%

13. Do you think there would be any dis-advantages in purchasing organic olive oil ?:

Yes No

- High price 71% 29%
- Not trustworthy 8% 91%
- Doesn't interest me 6,7% 93%
- Difficult to find it . 23% 77%
- Don't know 15% 84%

- Other (please state) 2% 98%

14 .If organic olive oil was readily accessible would you ever consider buying it?

Yes 59% No 37%

If no go to question number: 19

15 . Where would you prefer to buy organic olive oil from ?

Supermarket	58%	Open market	11%
Delicatessens	2,5%	Home	6,3%
Healthfood/ Reformhouse	1,4%	Other (please state)	0%
		

16 . Say that the olive oil you currently purchase costs 8.95 for 750ml . Are you prepared to pay a higher price for the same quantity of organic olive oil ?

Yes 32% No 33% other 33%

if the answer is «yes», go to 17 C

if the answer is «no» , go to 17 A

A.Would you pay the same price as for the olive oil you currently purchase

Yes 28% No 6,8%

B.Would you pay a lower price (than for the olive oil) you currently purchase

Yes 14% No 1,6%

C.Would you pay:

9.95	20%
11.45	9,5%
13.95	3%
More than 13.95	1,4%
Other (please state).....	3,6%

17. What would you use organic olive oil for?

Same as usual Olive Oil 58%

Only on special occasions/ for special dishes 8%

18 Gender: Male 49% Female 51%

19 Age: 17-28 48%
29-45 24%
46-65 23%
66-90 3%.....

20 Level of education : Primary School 3% High school 31% college/university 60%

21. Number of adults in household:

1	28%
2	38,5%
3	10%
more than 3	22%

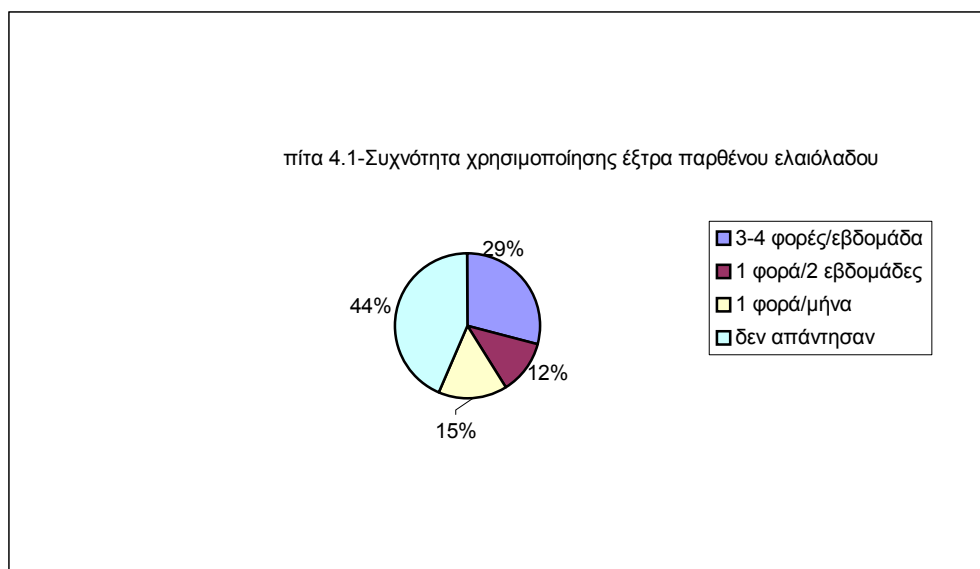
22. Number of children in household:

None	71%
1	11%
2	12%
3	3,6%
4	1,4%
more than 4	0,1%

Ας περάσουμε τώρα σε χαρακτηριστικά που θα μας βοηθήσουν να καταλάβουμε τη σχέση των ερωτηθέντων με το λάδι, είτε μιλάμε για το έξτρα παρθένο ελαιόλαδο, είτε το απλό, είτε μιλάμε για το φυτικό λάδι, και ύστερα θα δούμε κάποιες κρίσεις για τις συγκεκριμένες μάρκες έξτρα παρθένου ελαιολάδου, δηλαδή τα LORENA, AH ORGANIC(οργανικό), SITIA, BERTOLLI.

Ως προς την συχνότητα χρησιμοποίησης του έξτρα παρθένου ελαιολάδου τα στατιστικά μας λένε πως:

- 162 ή 29% το χρησιμοποιούν τρεις με τέσσερις φορές τη βδομάδα,
- 65 ή 11.7% το χρησιμοποιούν μια φορά στις δυο βδομάδες,
- 85 ή 15.3% το χρησιμοποιούν μια φορά το μήνα ή λιγότερο
- τέλος 241 δεν απάντησαν



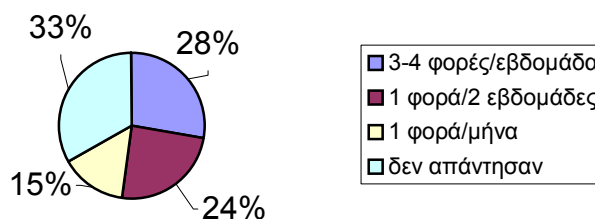
Είναι φανερό πως οι περισσότεροι καταναλωτές χρησιμοποιούν το έξτρα παρθένο ελαιόλαδο τρεις με τέσσερις φορές την εβδομάδα, άρα η χρήση του είναι έντονη, ενώ το αμέσως κατώτερο ποσοστό το χρησιμοποιεί αρκετά αραιά έως και μια φορά το μήνα.

Για τη συχνότητα χρησιμοποίησης του απλού ελαιολάδου απάντησαν:

- 154 ή 27.8% το χρησιμοποιούν τρεις με τέσσερις φορές τη βδομάδα,

- 133 ή 24% το χρησιμοποιούν μια φορά στις δυο εβδομάδες,
- 82 ή 14.8% το χρησιμοποιούν μια φορά το μήνα ή λιγότερο
- τέλος 184 δεν απάντησαν.

πίνα 4.2-Συχνότητα χρησιμοποίησης απλού ελαιολάδου

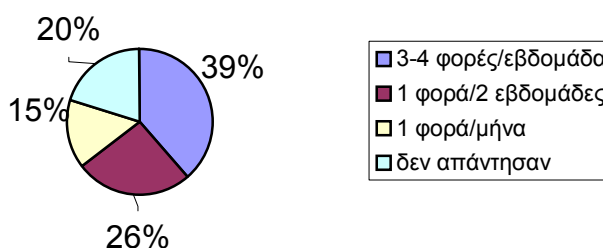


Και για το απλό ελαιόλαδο οι ερωτηθέντες, κατά πλειοψηφία, δήλωσαν ότι το χρησιμοποιούν τρεις με τέσσερις φορές την εβδομάδα, ενώ εδώ το αμέσως χαμηλότερο ποσοστό το χρησιμοποιεί μια φορά στις δυο εβδομάδες. Άρα μπορούμε να πούμε ότι γενικά χρησιμοποιούν περισσότερο το απλό από το έξτρα παρθένο.

Τέλος για τη συχνότητα χρησιμοποίησης του φυτικού λαδιού απάντησαν:

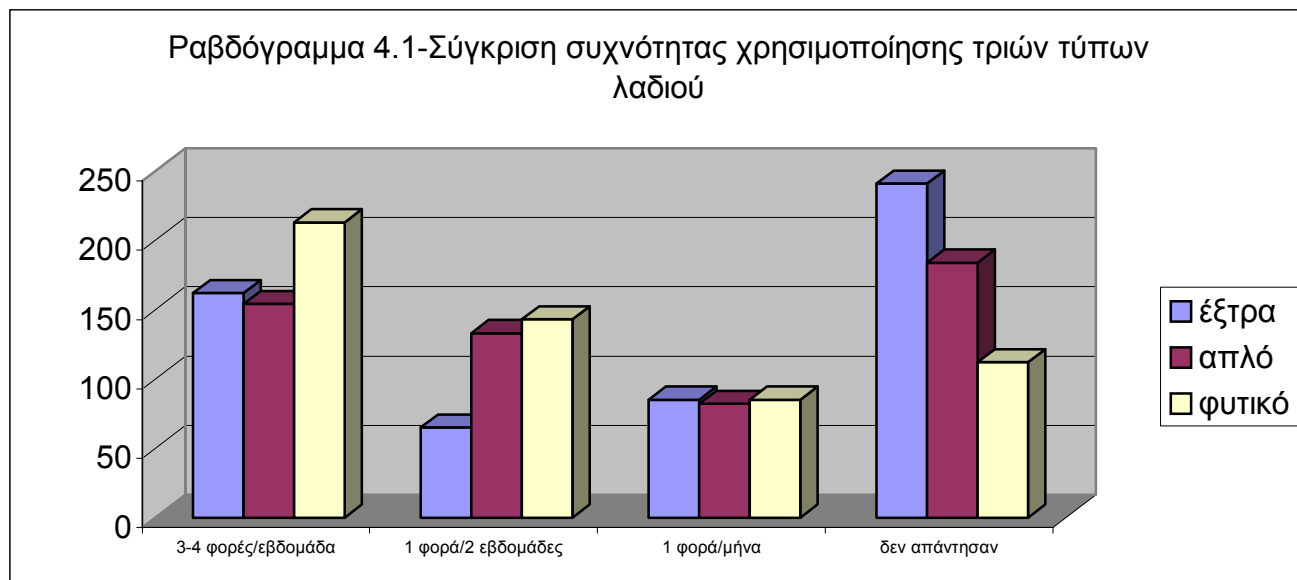
- 213 ή 38,5% το χρησιμοποιούν τρεις με τέσσερις φορές τη εβδομάδα,
- 143 ή 25,8% το χρησιμοποιούν μια φορά στις δυο εβδομάδες,
- 85 ή 15,3% το χρησιμοποιούν μια φορά το μήνα ή λιγότερο
- τέλος 112 δεν απάντησαν.

πίνα 4.3-Συχνότητα χρησιμοποίησης φυτικού λαδιού



Παρόμοια παρουσιάζονται τα ποσοστά και στην περίπτωση του φυτικού λαδιού ως προς τη συχνότητα με την πλειοψηφία να το χρησιμοποιεί τρεις με τέσσερις φορές ανά εβδομάδα.

Συγκεντρωτικά παρουσιάζονται και στο παρακάτω ραβδόγραμμα:



Άρα το λάδι συνήθως χρησιμοποιείται τρεις με τέσσερις φορές την εβδομάδα ανεξαρτήτως τύπου αλλά την πρωτιά έχει το φυτικό λάδι. Άλλωστε αν λάβουμε υπόψη μας και το πλήθος αυτών που δεν απάντησαν καταλήγουμε πως περισσότερο χρησιμοποιείται αυτό το λάδι, ύστερα το απλό και τελευταίο το έξτρα.

Μια άλλη χρήσιμη πληροφορία που μπορούμε να αντλήσουμε από τα δεδομένα μας είναι η χώρα από την οποία προέρχεται το λάδι που αγοράζουν οι ερωτηθέντες. Οι απαντήσεις τους έχουν ως εξής:

- 481 ή 87% δεν απάντησαν καθόλου
- 38 ή 6.8% απάντησαν ότι αγοράζουν λάδι Ιταλίας
- 13 ή 2.3% απάντησαν ότι αγοράζουν ελληνικό λάδι
- 9 ή 1.6% αγοράζει ισπανικό λάδι
- 1 ή 0.18% αγοράζει λάδι Ινδίας
- και τέλος 11 ή 1.9% από διάφορες άλλες χώρες.

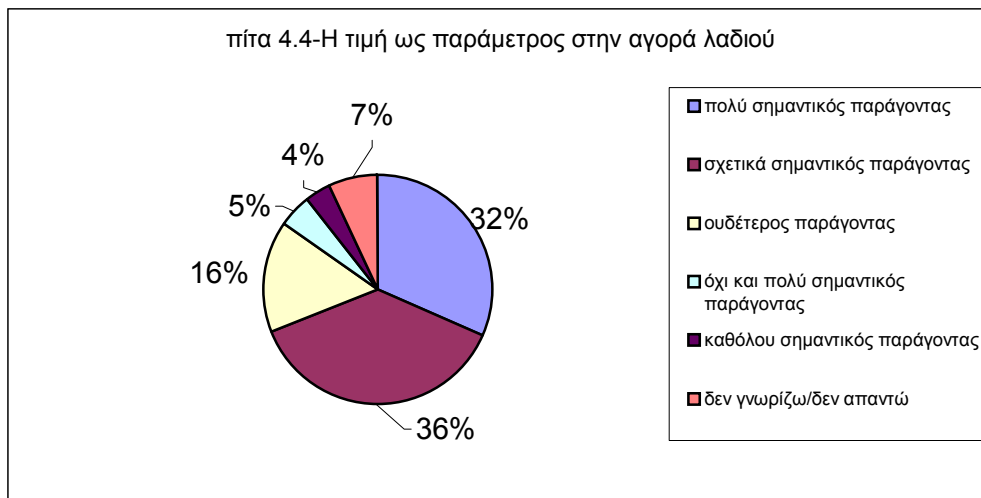
Δηλαδή την πρώτη θέση σε προτιμήσεις από άποψη προέλευσης κατέχει η Ιταλία και ύστερα η Ελλάδα. Δυστυχώς βέβαια ήταν πάρα πολλοί αυτοί που δεν απάντησαν.

Στη συνέχεια θα παρατηρήσουμε πόσο σημαντικό ρόλο στην αγορά λαδιού παίζει η τιμή, ο παράγοντας της υγείας, η γεύση, η χώρα προέλευσης και οι ειδικές προσφορές που γίνονται κατά καιρούς για το λάδι ελιάς. Πιο συγκεκριμένα :

Για το θέμα της τιμής μας δήλωσαν:

- 175 άτομα ή 31.6% ότι είναι πολύ σημαντικός παράγοντας.
- 206 άτομα ή 37.2% ότι είναι σχετικά σημαντικός παράγοντας.
- 88 άτομα ή 16% ότι αποτελεί ουδέτερο παράγοντα.
- 25 άτομα ή 4.5% ότι είναι όχι και πολύ σημαντικός παράγοντας.
- 21 άτομα ή 3.8% ότι δεν είναι καθόλου σημαντικός παράγοντας.

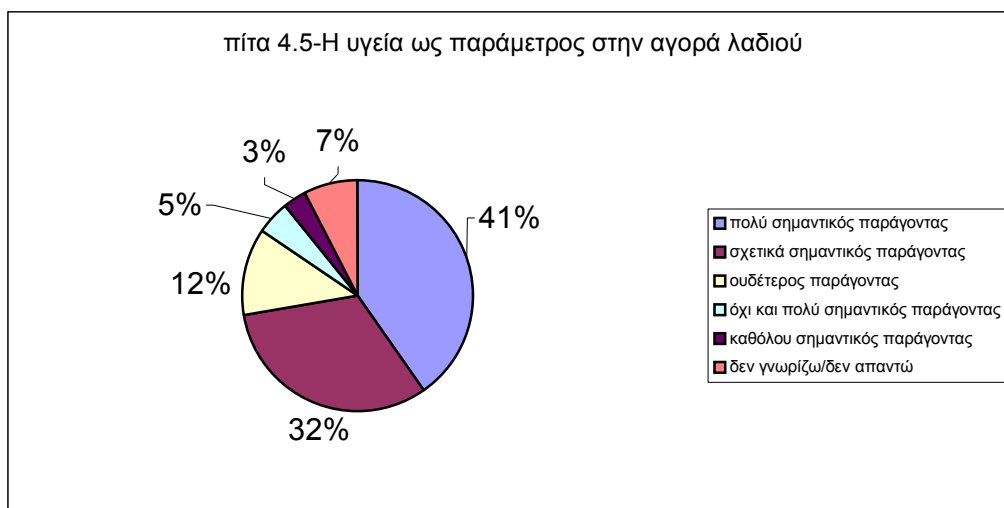
- 18 άτομα ή 3.2% δεν γνωρίζουν πόσο σημαντικός παράγοντας είναι.
- 20 δεν απάντησαν καθόλου



, άρα παίζει αρκετά σημαντικό ρόλο η τιμή του προϊόντος στη αγορά.

Για τον παράγοντα «υγεία» μέσω της χρήσης του ελαιολάδου απάντησαν:

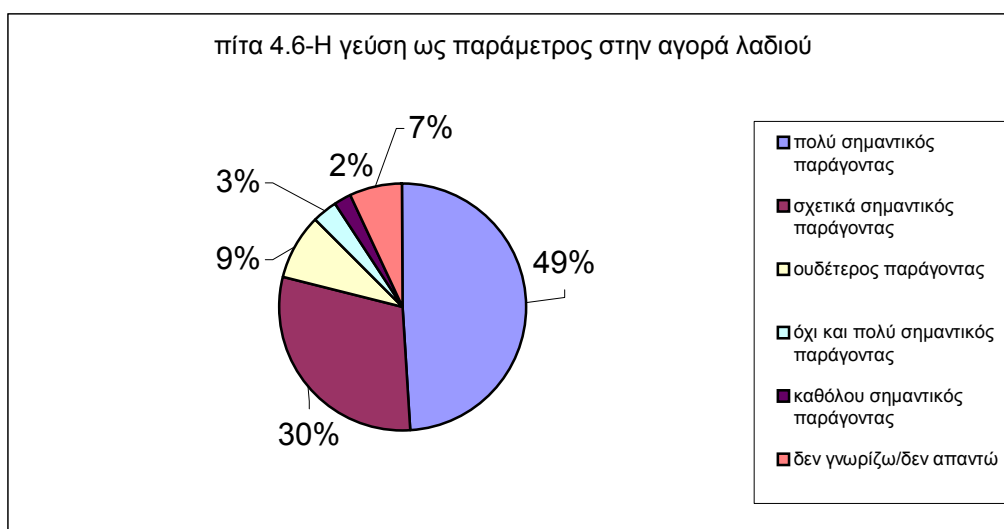
- 223 άτομα ή 40.3% ότι είναι πολύ σημαντικός παράγοντας.
- 177 άτομα ή 32% ότι είναι σχετικά σημαντικός παράγοντας.
- 67 άτομα ή 12.1% ότι αποτελεί ουδέτερο παράγοντα.
- 26 άτομα ή 4.7% ότι είναι όχι και πολύ σημαντικός παράγοντας.
- 19 άτομα ή 3.4% ότι δεν είναι καθόλου σημαντικός παράγοντας.
- 19 άτομα ή 3.4% δεν γνωρίζουν πόσο σημαντικός παράγοντας είναι.
- 22 δεν απάντησαν καθόλου.



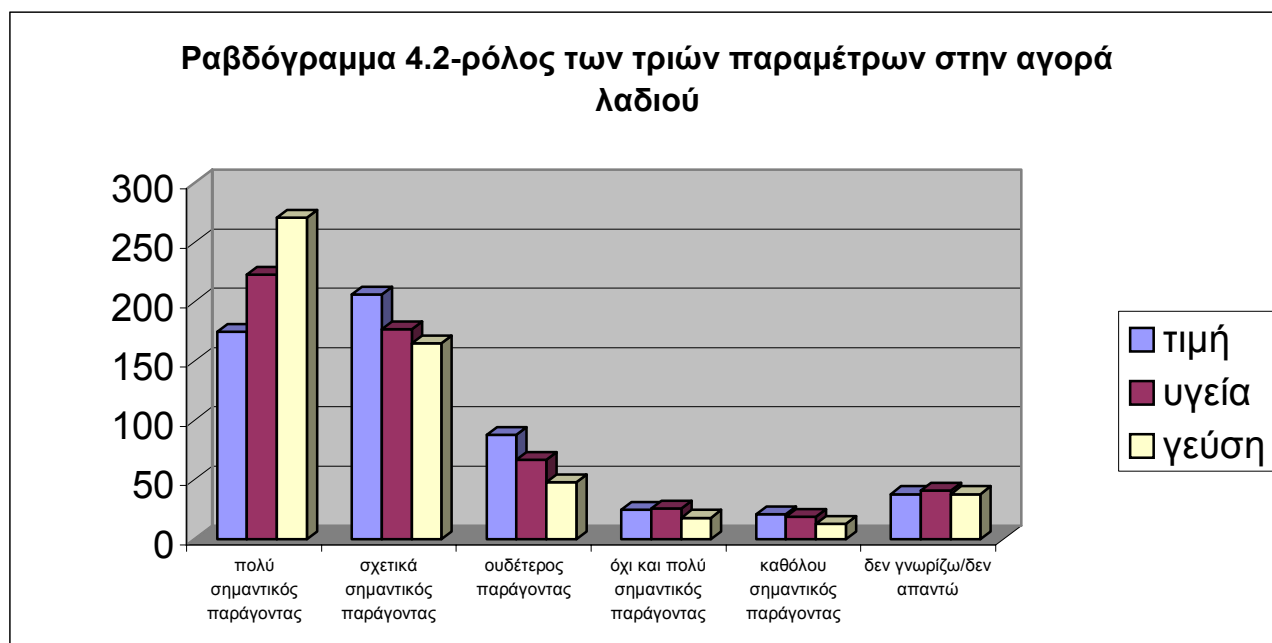
Βλέπουμε πως απασχολεί ιδιαίτερα τον καταναλωτή το πόσο το λάδι συμβάλλει στην υγεία, γιατί ουσιαστικά το 70% πιστεύει πως είναι σημαντικός παράγοντας και μόνο 3% δεν παραδέχεται τη συμβολή του.

Η γεύση είναι ένα χαρακτηριστικό του λαδιού που πολλοί άνθρωποι εμπιστεύονται κύρια για να κρίνουν ένα λάδι. Ας παρατηρήσουμε στατιστικά τι μας δήλωσαν:

- 271 άτομα ή 49% ότι είναι πολύ σημαντικός παράγοντας.
- 165 άτομα ή 29.8% ότι είναι σχετικά σημαντικός παράγοντας.
- 48 άτομα ή 8.6% ότι αποτελεί ουδέτερο παράγοντα.
- 18 άτομα ή 3.2% ότι είναι όχι και πολύ σημαντικός παράγοντας.
- 13 άτομα ή 2.3% ότι δεν είναι καθόλου σημαντικός παράγοντας.
- 19 άτομα ή 3.4% δεν γνωρίζουν πόσο σημαντικός παράγοντας είναι.
- 19 δεν απάντησαν καθόλου.



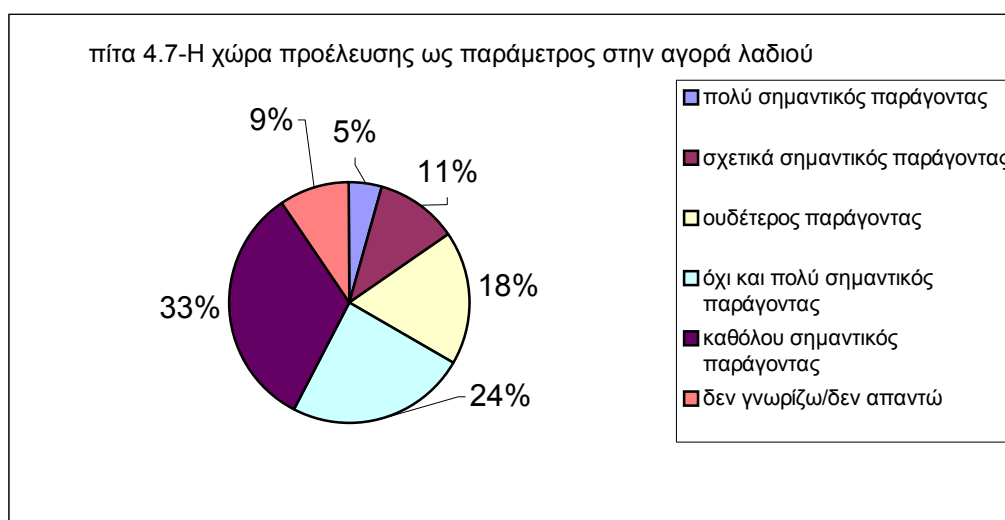
Πραγματικά 436 άτομα συνολικά θεωρούν σημαντικό να έχει ωραία γεύση το ελαιόλαδο και αυτή είναι συνήθως που δοκιμάζει και εμπιστεύεται ο καταναλωτής. Συνολικά παρουσιάζονται και στο παρακάτω ραβδόγραμμα:



Φαίνεται ότι η γεύση και η υγεία είναι πολύ σημαντικοί παράγοντες, ενώ η τιμή και η υγεία είναι σχετικά σημαντικοί παράγοντες όσον αφορά την αγορά ελαιολάδου.

Όσο για τη χώρα προέλευσης του ελαιολάδου μας δήλωσαν επί το πλείστον ότι τελικά δεν παίζει και ιδιαίτερο ρόλο. Πιο συγκεκριμένα:

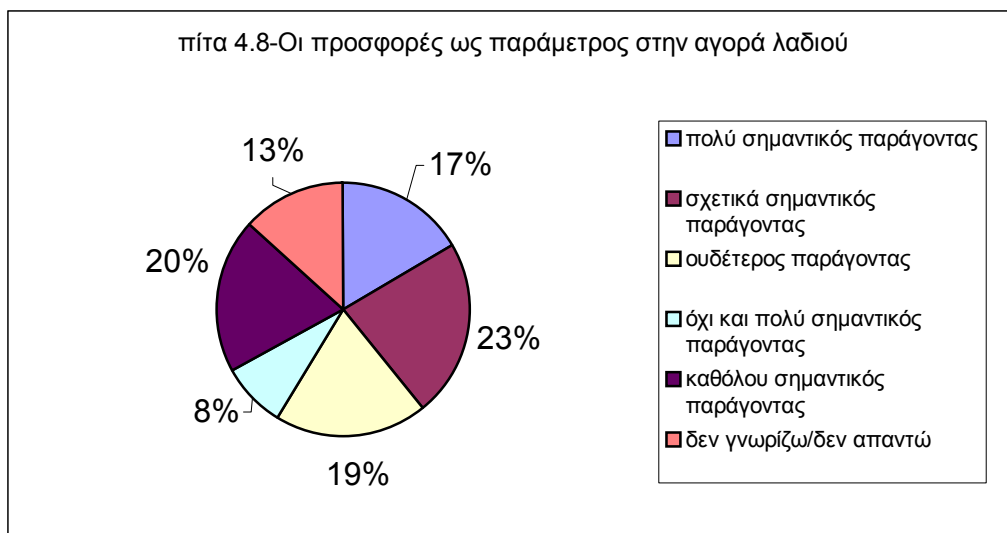
- 25 άτομα ή 4.5% ότι είναι πολύ σημαντικός παράγοντας.
- 61 άτομα ή 11% ότι είναι σχετικά σημαντικός παράγοντας.
- 98 άτομα ή 17.7% ότι αποτελεί ουδέτερο παράγοντα.
- 134 άτομα ή 24.2% ότι είναι όχι και πολύ σημαντικός παράγοντας.
- 183 άτομα ή 33% ότι δεν είναι καθόλου σημαντικός παράγοντας.
- 27 άτομα ή 4.8% δεν γνωρίζουν πόσο σημαντικός παράγοντας είναι.
- 25 δεν απάντησαν καθόλου.



Είναι φανερό ότι η πλειοψηφία δεν ασχολείται με την χώρα προέλευσης και άρα δεν είναι ικανή παράμετρος για την τελική απόφαση αγοράς.

Τέλος τα θέματα προσφορών που κατά καιρούς βγάζουν διάφορες εταιρίες φαίνεται να αγγίζουν τους καταναλωτές αμφίρροπα, δηλαδή:

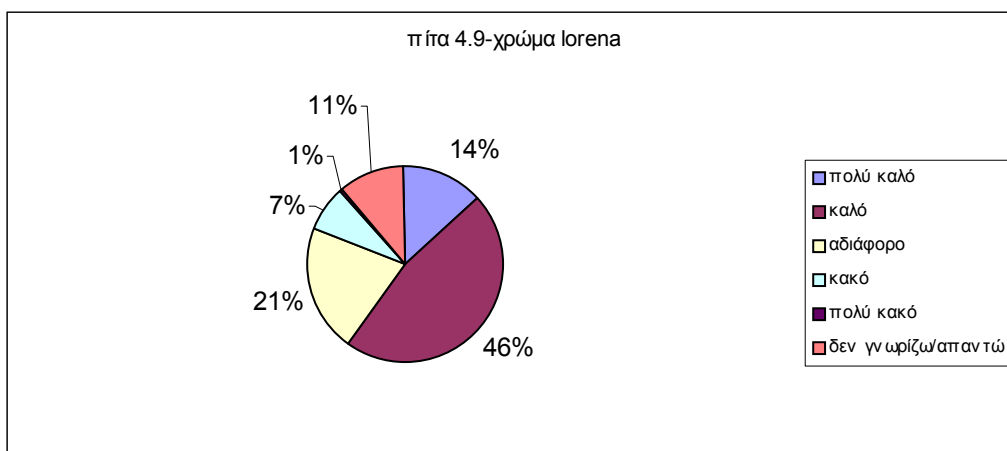
- 92 άτομα ή 16.6% ότι είναι πολύ σημαντικός παράγοντας.
- 125 άτομα ή 22.6% ότι είναι σχετικά σημαντικός παράγοντας.
- 107 άτομα ή 19.3% ότι αποτελεί ουδέτερο παράγοντα.
- 46 άτομα ή 8.3% ότι είναι όχι και πολύ σημαντικός παράγοντας.
- 110 άτομα ή 19.8% ότι δεν είναι καθόλου σημαντικός παράγοντας.
- 41 άτομα ή 7.4% δεν γνωρίζουν πόσο σημαντικός παράγοντας είναι.
- 32 δεν απάντησαν καθόλου.



Ας περάσουμε τώρα σε κάποια χαρακτηριστικά των συγκεκριμένων λαδιών που μας ενδιαφέρουν. Δηλαδή εξετάζουμε τις κρίσεις των καταναλωτών σε καθένα από τα λάδια μας, LORENA, AH ORGANIC, SITIA, BERTOLLI, πάνω σε θέματα χρώματος του προϊόντος, πρεστίζ ή καταξίωσης αυτού, ποιότητας, τιμής και τέλος τους ζητάμε να τα κατατάξουν, με πρώτο, φυσικά, να είναι το καλύτερο. Παρακάτω φαίνονται τα αποτελέσματα και έχουμε:

Για το χρώμα του LORENA:

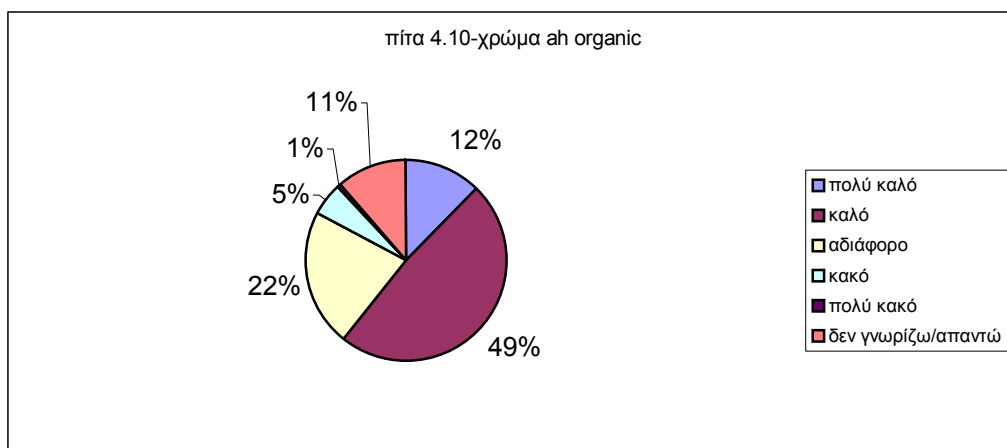
- 75 καταναλωτές ή 13.5% δήλωσαν πως είναι πολύ καλό.
- 254 καταναλωτές ή 46% δήλωσαν πως είναι καλό
- 118 καταναλωτές ή 21.3% δήλωσαν πως είναι αδιάφορο σαν χρώμα
- 41 καταναλωτές ή 7.4% δήλωσαν πως είναι κακό
- 3 καταναλωτές ή 0.5% δήλωσαν πως είναι πολύ κακό
- 62 καταναλωτές ή 11.2% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



Η πλειοψηφία θεωρεί ότι το χρώμα το lorena είναι απλά καλό.

Για το χρώμα του AH ORGANIC:

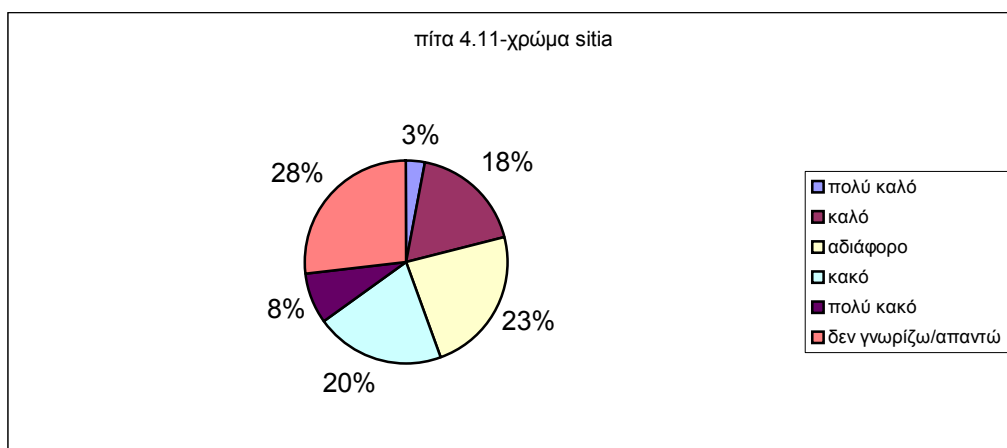
- 69 καταναλωτές ή 12,4% δήλωσαν πως είναι πολύ καλό.
- 266 καταναλωτές ή 48,1% δήλωσαν πως είναι καλό
- 122 καταναλωτές ή 22% δήλωσαν πως είναι αδιάφορο σαν χρώμα
- 30 καταναλωτές ή 5,4% δήλωσαν πως είναι κακό
- 3 καταναλωτές ή 0,5% δήλωσαν πως είναι πολύ κακό
- 63 καταναλωτές ή 11,3% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



Επίσης το μεγαλύτερο ποσοστό θεωρεί καλό το χρώμα του ah organic .

Για το χρώμα του SITIA:

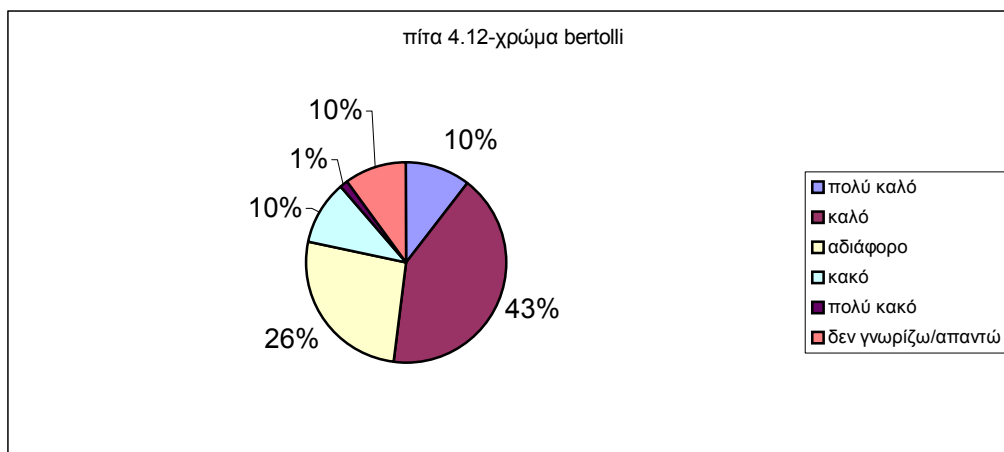
- 17 καταναλωτές ή 3% δήλωσαν πως είναι πολύ καλό.
- 100 καταναλωτές ή 18% δήλωσαν πως είναι καλό
- 129 καταναλωτές ή 23,3% δήλωσαν πως είναι αδιάφορο σαν χρώμα
- 113 καταναλωτές ή 20,4% δήλωσαν πως είναι κακό
- 45 καταναλωτές ή 8,1% δήλωσαν πως είναι πολύ κακό
- 149 καταναλωτές ή 26,9% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



Αδιάφορο βρίσκει η πλειοψηφία το χρώμα του sitia ενώ πάρα πολλοί απάντησαν πως είναι κακό.

Για το χρώμα του BERTOLLI:

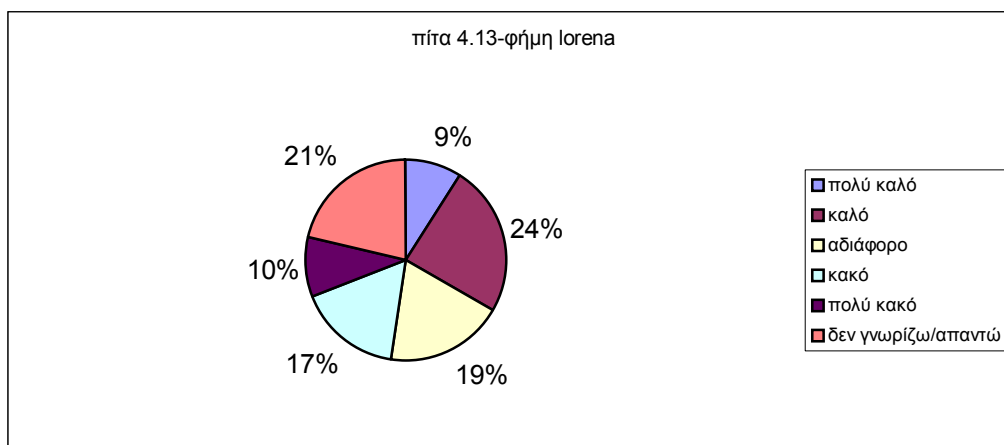
- 58 καταναλωτές ή 10,4% δήλωσαν πως είναι πολύ καλό.
- 230 καταναλωτές ή 41,6% δήλωσαν πως είναι καλό
- 145 καταναλωτές ή 26,2% δήλωσαν πως είναι αδιάφορο σαν χρώμα
- 57 καταναλωτές ή 10,3% δήλωσαν πως είναι κακό
- 8 καταναλωτές ή 1,4% δήλωσαν πως είναι πολύ κακό
- 55 καταναλωτές ή 9,9% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



Καλό ή και αδιάφορο αποδείχθηκε κατά την άποψη των καταναλωτών το χρώμα του bertolli .

Για τη φήμη του LORENA:

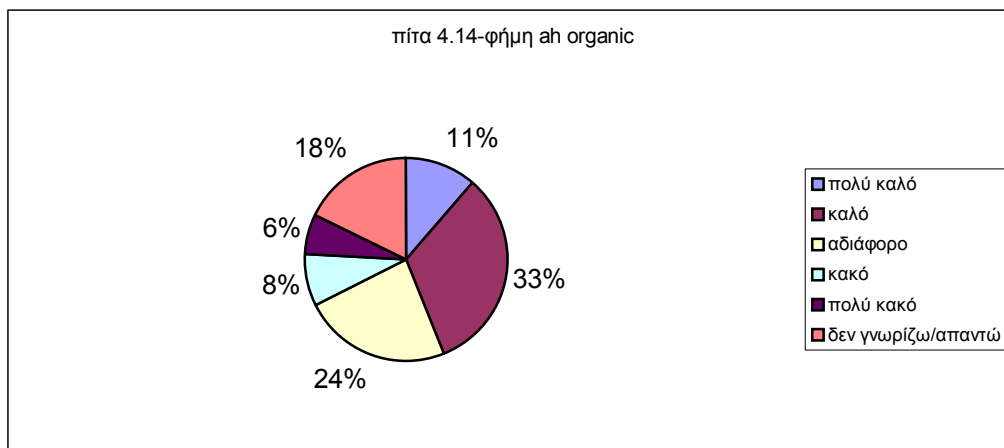
- 51 καταναλωτές ή 9,2% δήλωσαν πως είναι πολύ καλή.
- 133 καταναλωτές ή 24% δήλωσαν πως είναι καλή
- 105 καταναλωτές ή 18,9% δήλωσαν πως είναι αδιάφορο
- 92 καταναλωτές ή 16,6% δήλωσαν πως είναι κακή
- 54 καταναλωτές ή 9,7% δήλωσαν πως είναι πολύ κακή
- 118 καταναλωτές ή 21,3% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



Για τη φήμη του AH ORGANIC:

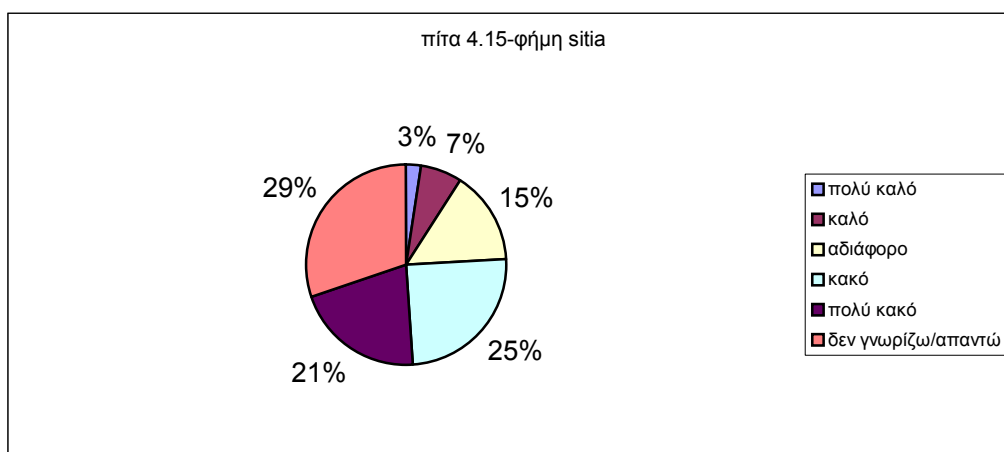
- 63 καταναλωτές ή 11,3% δήλωσαν πως είναι πολύ καλή.
- 180 καταναλωτές ή 32,5% δήλωσαν πως είναι καλή

- 130 καταναλωτές ή 23,5% δήλωσαν πως είναι αδιάφορο
- 47 καταναλωτές ή 8,4% δήλωσαν πως είναι κακή
- 34 καταναλωτές ή 6,1% δήλωσαν πως είναι πολύ κακή
- 99 καταναλωτές ή 17,9% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



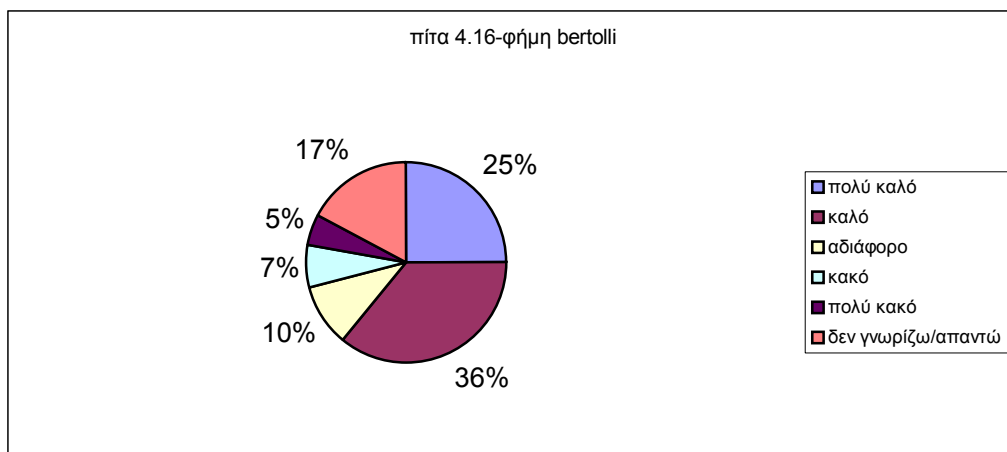
Για τη φήμη του SITIA:

- 14 καταναλωτές ή 2,5% δήλωσαν πως είναι πολύ καλή.
- 37 καταναλωτές ή 6,7% δήλωσαν πως είναι καλή
- 82 καταναλωτές ή 14,8% δήλωσαν πως είναι αδιάφορο
- 138 καταναλωτές ή 25% δήλωσαν πως είναι κακή
- 114 καταναλωτές ή 20,6% δήλωσαν πως είναι πολύ κακή
- 168 καταναλωτές ή 30,3% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.

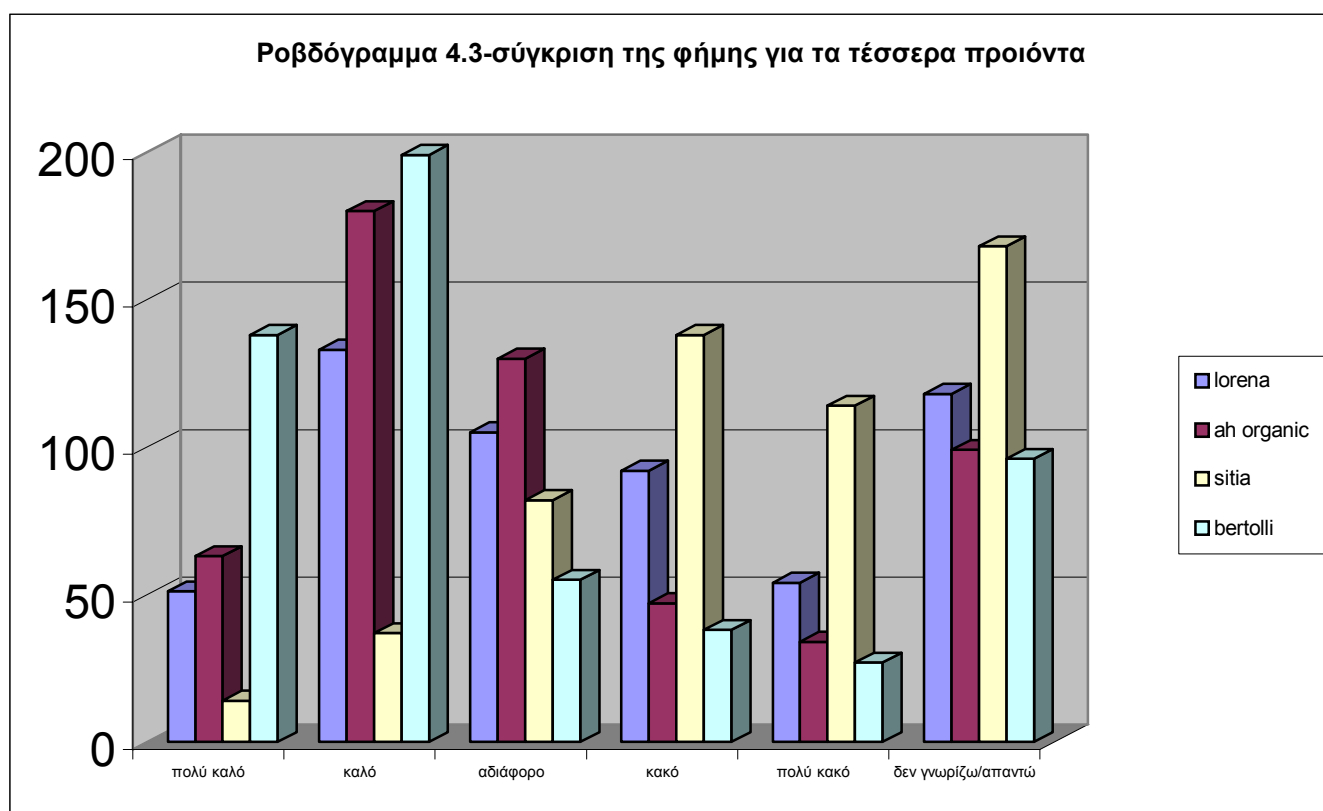


Για τη φήμη του BERTOLLI:

- 138 καταναλωτές ή 25% δήλωσαν πως είναι πολύ καλή.
- 199 καταναλωτές ή 36% δήλωσαν πως είναι καλή
- 55 καταναλωτές ή 9,9% δήλωσαν πως είναι αδιάφορο
- 38 καταναλωτές ή 6,8% δήλωσαν πως είναι κακή
- 27 καταναλωτές ή 4,8% δήλωσαν πως είναι πολύ κακή
- 96 καταναλωτές ή 17,3% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



Συνολικά οι απόψεις των καταναλωτών για τη φήμη των λαδιών φαίνονται στο αντίστοιχο ραβδόγραμμα:

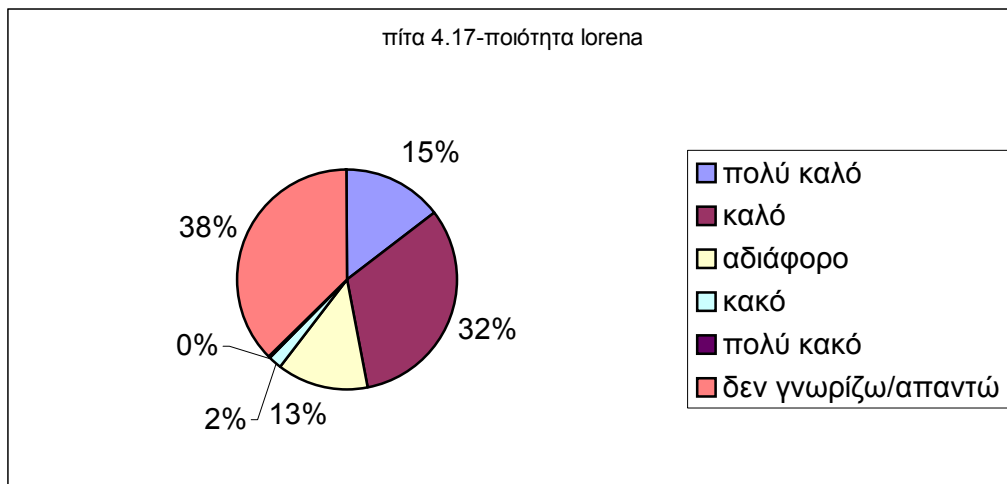


Η φήμη του bertolli είναι σε πολύ υψηλά επίπεδα, όπως και του ah organic, ενώ το lorena ακολουθεί. Όσο για το ελληνικό λάδι η φήμη του δεν είναι και τόσο καλή και πάρα πολλοί δεν το γνωρίζουν γενικότερα, όπως φαίνεται στην τελευταία στήλη.

Για τη ποιότητα του LORENA:

- 81 καταναλωτές ή 14,6% δήλωσαν πως είναι πολύ καλή.
- 178 καταναλωτές ή 32,1% δήλωσαν πως είναι καλή
- 74 καταναλωτές ή 13,3% δήλωσαν πως είναι αδιάφορο
- 11 καταναλωτές ή 2% δήλωσαν πως είναι κακή

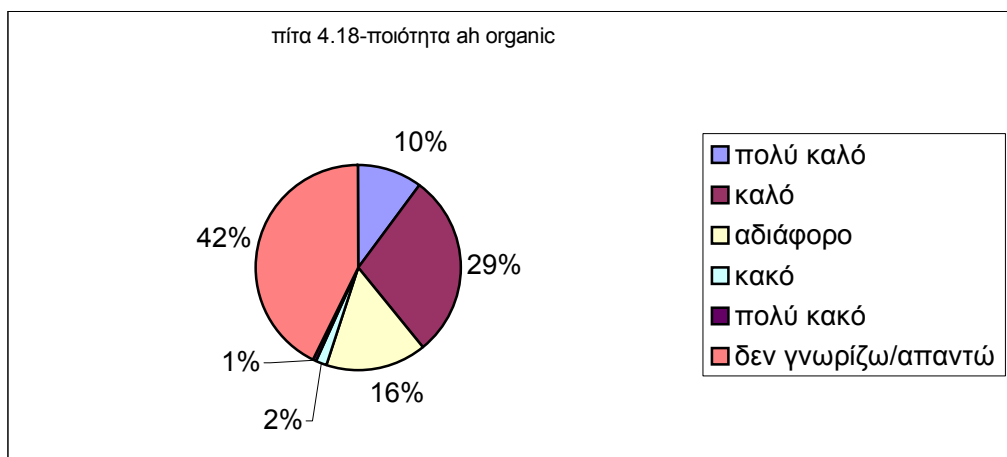
- 1 καταναλωτής ή 0,2% δήλωσαν πως είναι πολύ κακή
- 208 καταναλωτές ή .37,6% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



Στο θέμα της ποιότητας το lorena απέσπασε πολύ καλά σχόλια και θεωρήθηκε καλό.

Για τη ποιότητα του AH ORGANIC:

- 57 καταναλωτές ή 10,3% δήλωσαν πως είναι πολύ καλή.
- 160 καταναλωτές ή 28,9% δήλωσαν πως είναι καλή
- 87 καταναλωτές ή 15,7% δήλωσαν πως είναι αδιάφορο
- 10 καταναλωτές ή 1,8% δήλωσαν πως είναι κακή
- 3 καταναλωτές ή 0,5% δήλωσαν πως είναι πολύ κακή
- 236 καταναλωτές ή .42,6% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.

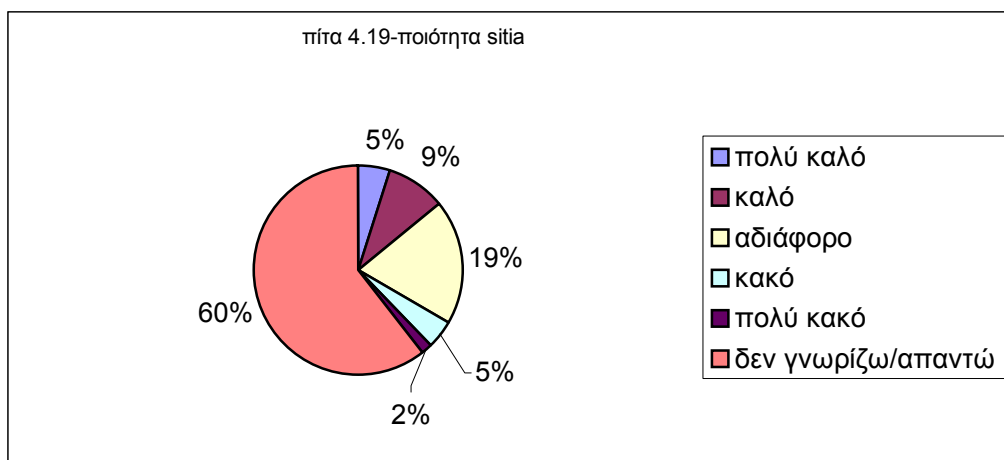


Τα ίδια περίπου σχόλια υπήρξαν και για το ah organic, δηλαδή καλό.

Για τη ποιότητα του SITIA:

- 27 καταναλωτές ή 4,8% δήλωσαν πως είναι πολύ καλή.
- 51 καταναλωτές ή 9,2% δήλωσαν πως είναι καλή
- 106 καταναλωτές ή 19,1% δήλωσαν πως είναι αδιάφορο
- 25 καταναλωτές ή 4,5% δήλωσαν πως είναι κακή

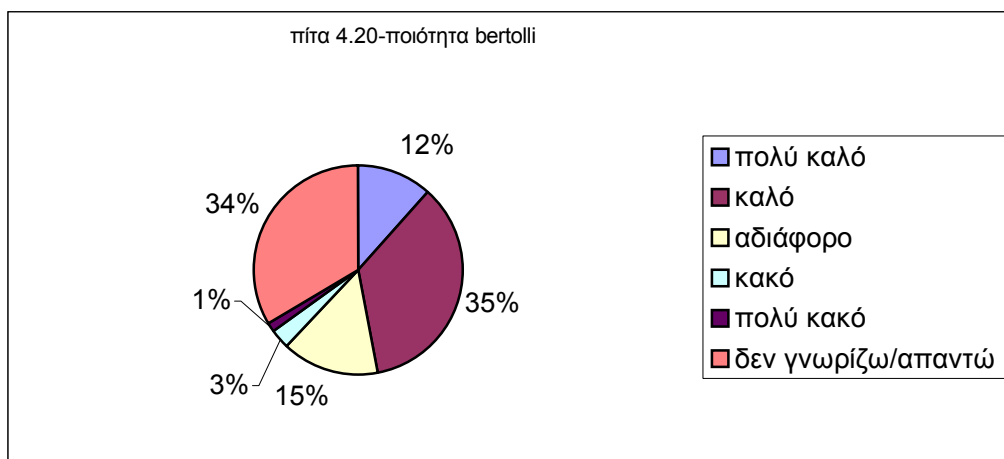
- 9 καταναλωτές ή 1,6% δήλωσαν πως είναι πολύ κακή
- 335 καταναλωτές ή 60,5% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



Αδιάφορο θεωρήθηκε ως επί το πλείστον το επίπεδο της ποιότητας του sitia.

Για τη ποιότητα του BERTOLLI:

- 64 καταναλωτές ή 11,5% δήλωσαν πως είναι πολύ καλή.
- 196 καταναλωτές ή 35,4% δήλωσαν πως είναι καλή
- 83 καταναλωτές ή 15% δήλωσαν πως είναι αδιάφορο
- 17 καταναλωτές ή 3% δήλωσαν πως είναι κακή
- 7 καταναλωτές ή 1,2% δήλωσαν πως είναι πολύ κακή
- 186 καταναλωτές ή 33,6% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.

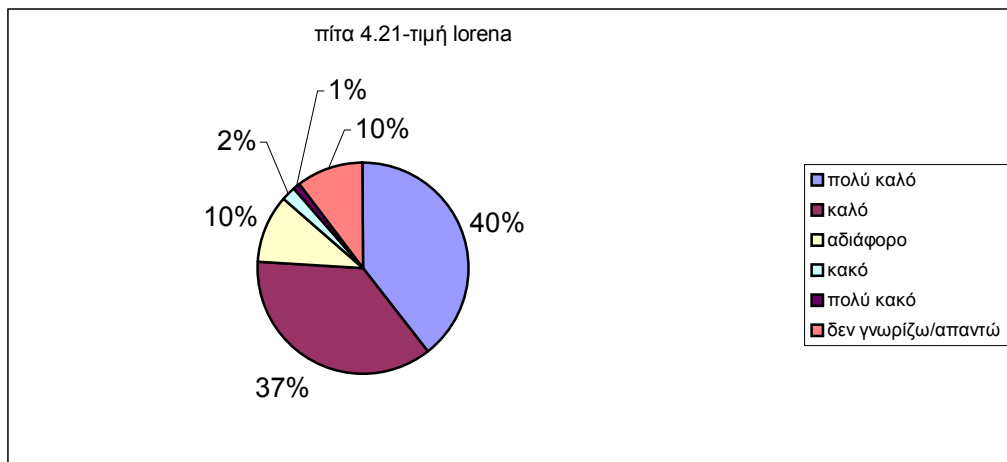


Καλή και η ποιότητα του bertolli σύμφωνα με τις δηλώσεις των καταναλωτών.

Για τη τιμή του LORENA:

- 218 καταναλωτές ή 39,4% δήλωσαν πως είναι πολύ καλή.
- 201 καταναλωτές ή 36,3% δήλωσαν πως είναι καλή
- 58 καταναλωτές ή 10,4% δήλωσαν πως είναι αδιάφορη
- 13 καταναλωτές ή 2,3% δήλωσαν πως είναι κακή

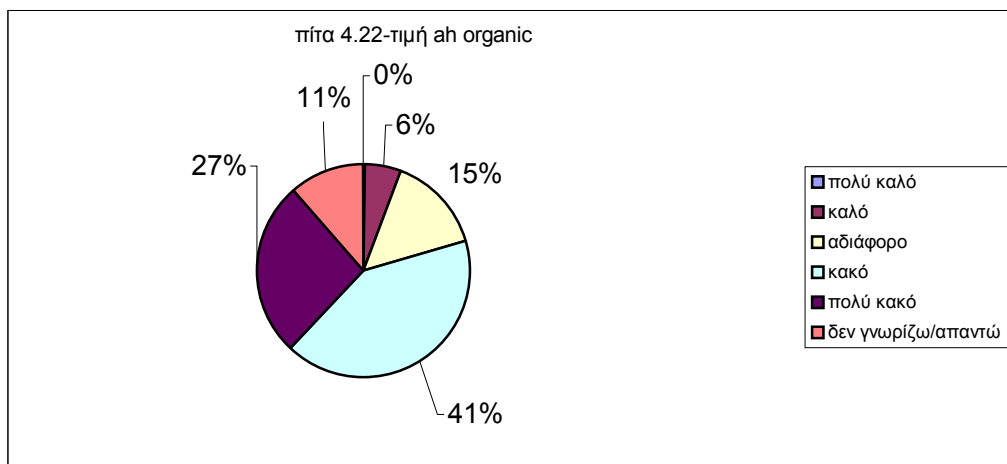
- 6 καταναλωτές ή 1% δήλωσαν πως είναι πολύ κακή
- 57 καταναλωτές ή 10,3% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



Η τιμή του lorena κρίθηκε από τους καταναλωτές πολύ καλή και πολύ λίγοι το βρήκαν ακριβό.

Για τη τιμή του AH ORGANIC:

- 2 καταναλωτές ή 0,3% δήλωσαν πως είναι πολύ καλή.
- 31 καταναλωτές ή 5,6% δήλωσαν πως είναι καλή
- 81 καταναλωτές ή 14,6% δήλωσαν πως είναι αδιάφορη
- 228 καταναλωτές ή 41,2% δήλωσαν πως είναι κακή
- 148 καταναλωτές ή 26,7% δήλωσαν πως είναι πολύ κακή
- 63 καταναλωτές ή 11,4% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.

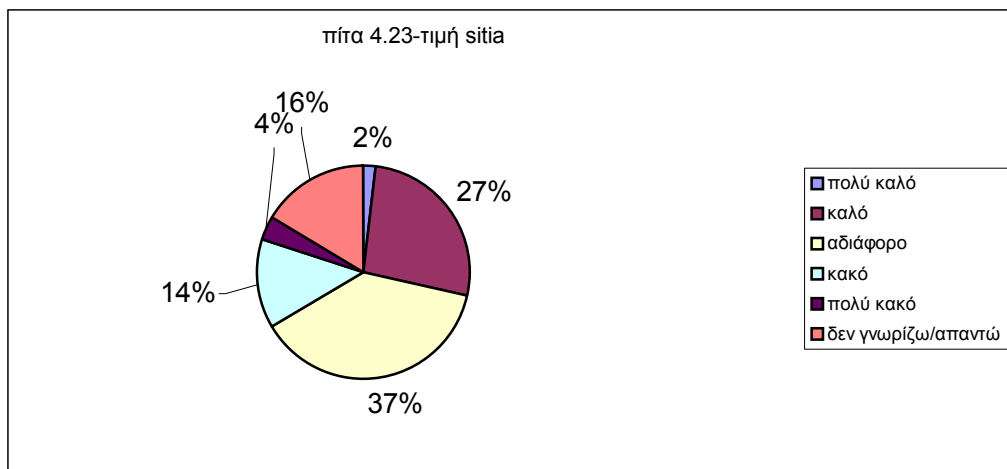


Αντίθετα η τιμή του ah organic μάλλον είναι ακριβή, σύμφωνα με τα λεγόμενα των καταναλωτών.

Για τη τιμή του SITIA:

- 10 καταναλωτές ή 1,8% δήλωσαν πως είναι πολύ καλή.
- 148 καταναλωτές ή 26,7% δήλωσαν πως είναι καλή
- 209 καταναλωτές ή 37,8% δήλωσαν πως είναι αδιάφορη

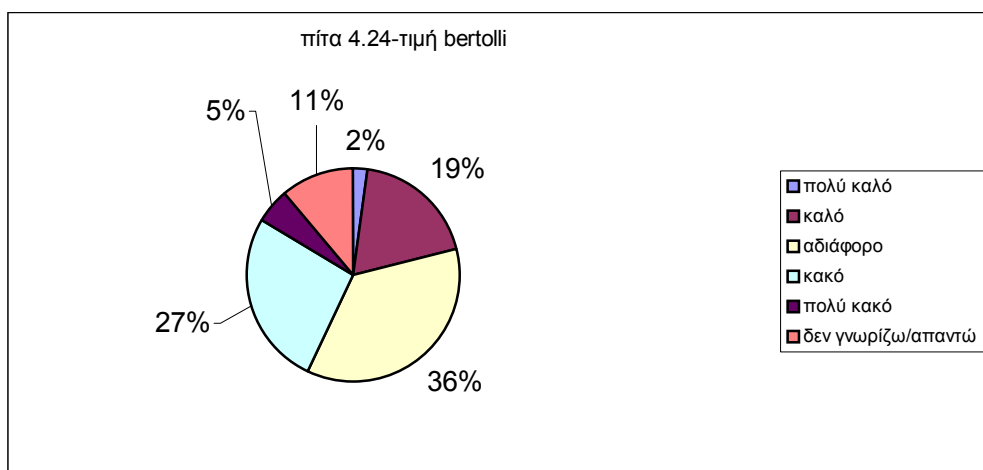
- 76 καταναλωτές ή 13,7% δήλωσαν πως είναι κακή
- 20 καταναλωτές ή 3,6% δήλωσαν πως είναι πολύ κακή
- 90 καταναλωτές ή 16,2% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.



Κανονική βρήκαν την τιμή του sitia οι ερωτηθέντες, δηλαδή ούτε υψηλή, ούτε χαμηλή.

Για τη τιμή του BERTOLLI:

- 12 καταναλωτές ή 2,1% δήλωσαν πως είναι πολύ καλή.
- 104 καταναλωτές ή 18,8% δήλωσαν πως είναι καλή
- 199 καταναλωτές ή 36% δήλωσαν πως είναι αδιάφορη
- 148 καταναλωτές ή 26,7% δήλωσαν πως είναι κακή
- 29 καταναλωτές ή 5,2% δήλωσαν πως είναι πολύ κακή
- 61 καταναλωτές ή 11% δήλωσαν πως δεν γνωρίζουν / δεν απαντούν.

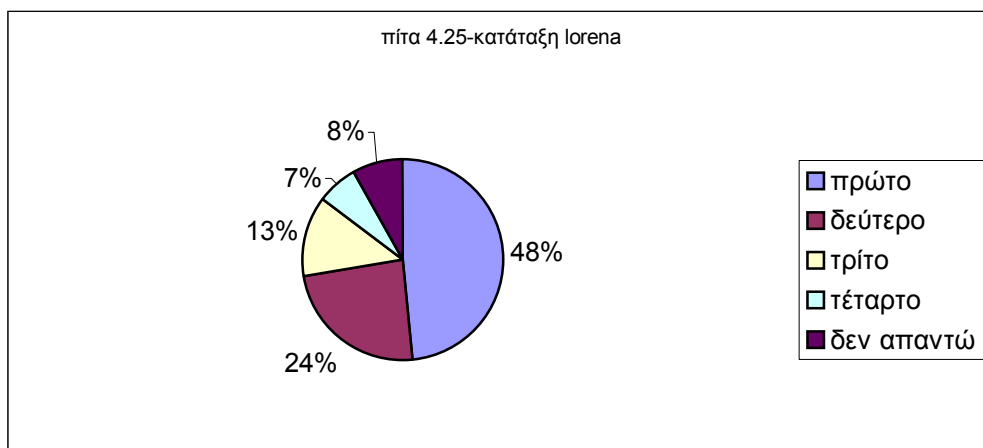


Ομοίως με το sitia η τιμή του bertolli είναι σε αδιάφορο επίπεδο για τους καταναλωτές.

Τέλος για τη θέση που κατατάσσουν το κάθε προϊόν από τα τέσσερα μας απάντησαν ως εξής:

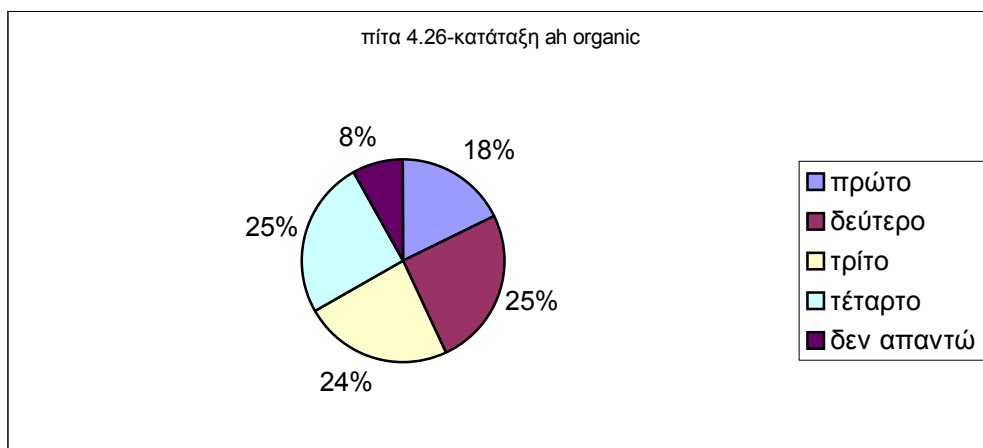
Για το LORENA:

- 267 άτομα το κατάταξαν στην πρώτη θέση, ή το 48,3%.
- 132 άτομα το κατάταξαν στη δεύτερη θέση, ή το 23,8%.
- 73 άτομα το κατάταξαν στην τρίτη θέση, ή το 13,2%.
- 37 άτομα το κατάταξαν στην τέταρτη θέση, ή το 6,7%.
- Τέλος 44 άτομα δεν απάντησαν καθόλου.



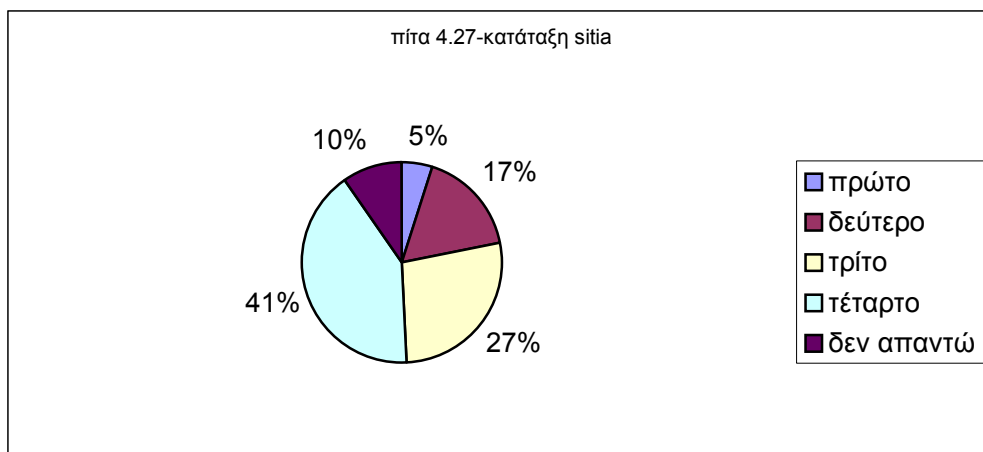
Για το AH ORGANIC:

- 98 άτομα το κατάταξαν στην πρώτη θέση, ή το 17,7%.
- 140 άτομα το κατάταξαν στη δεύτερη θέση, ή το 25,3%.
- 130 άτομα το κατάταξαν στην τρίτη θέση, ή το 23,5%.
- 140 άτομα το κατάταξαν στην τέταρτη θέση, ή το 25,3%.
- Τέλος 45 άτομα δεν απάντησαν καθόλου.



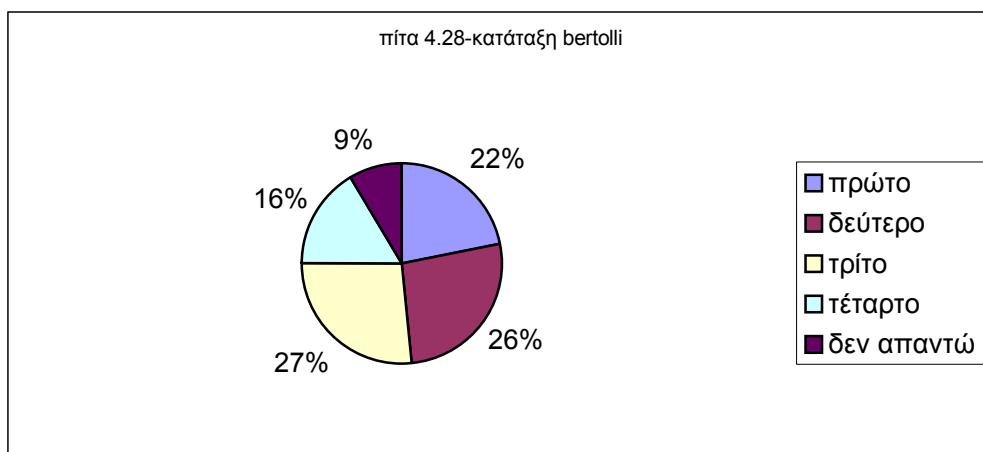
Για το SITIA:

- 28 άτομα το κατάταξαν στην πρώτη θέση, ή το 5%.
- 93 άτομα το κατάταξαν στη δεύτερη θέση, ή το 16,8%.
- 151 άτομα το κατάταξαν στην τρίτη θέση, ή το 27,3%.
- 228 άτομα το κατάταξαν στην τέταρτη θέση, ή το 41,2%.
- Τέλος 53 άτομα δεν απάντησαν καθόλου.



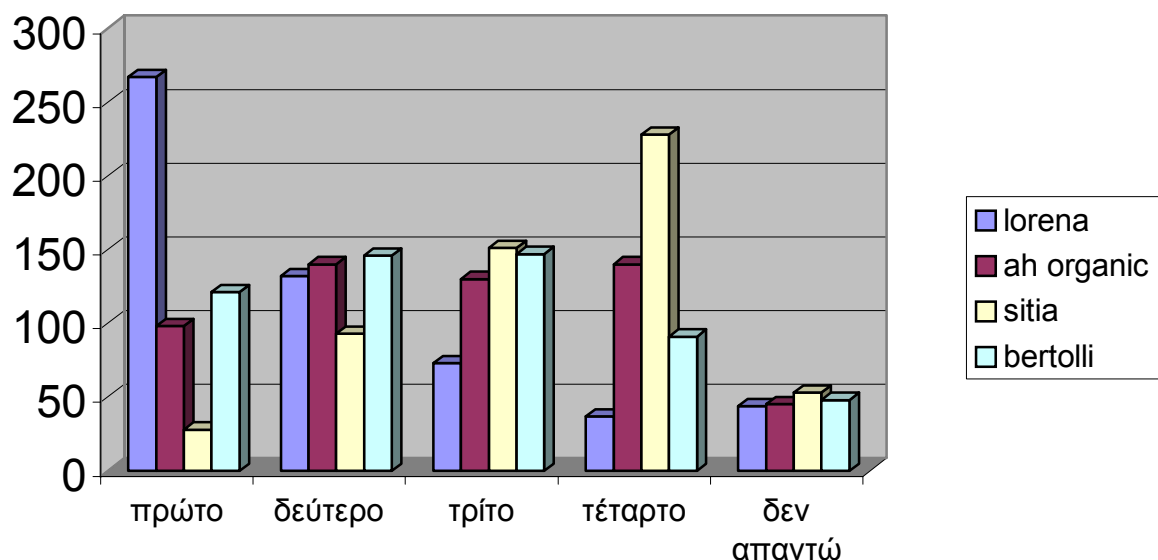
Για το BERTOLLI:

- 121 άτομα το κατάταξαν στην πρώτη θέση, ή το 21,8%.
- 146 άτομα το κατάταξαν στη δεύτερη θέση, ή το 26,4%.
- 147 άτομα το κατάταξαν στην τρίτη θέση, ή το 26,5%.
- 91 άτομα το κατάταξαν στην τέταρτη θέση, ή το 16,4%.
- Τέλος 48 άτομα δεν απάντησαν καθόλου.



Μπορούν να παρουσιαστούν και σε συνολικό ραβδόγραμμα τα αποτελέσματα κατάταξης των προϊόντων λαδιού που εξετάζουμε:

Ραβδόγραμμα 4.4-συγκριτική κατάταξη των τεσσάρων προϊόντων



Είναι φανερό πως την πρώτη θέση στην προτίμηση των καταναλωτών κατέχει το lorena . Παρατηρώντας, βλέπουμε πως πρώτη θέση δεν δίνει η πλειοψηφία σε κανένα από τα άλλα τρία προϊόντα, ωστόσο το ah organic ταλαντεύεται αλλά μάλλον στην δεύτερη θέση κινείται με επόμενο να έρχεται το bertolli , ενώ το sitia καταλήγει στην τελευταία θέση με μεγάλη διαφορά. Το bertolli θα μπορούσε να είναι στη δεύτερη θέση των προτιμήσεων, όπου σ' αυτή τη περίπτωση το ah organic θα κατέληγε στη τέταρτη θέση.

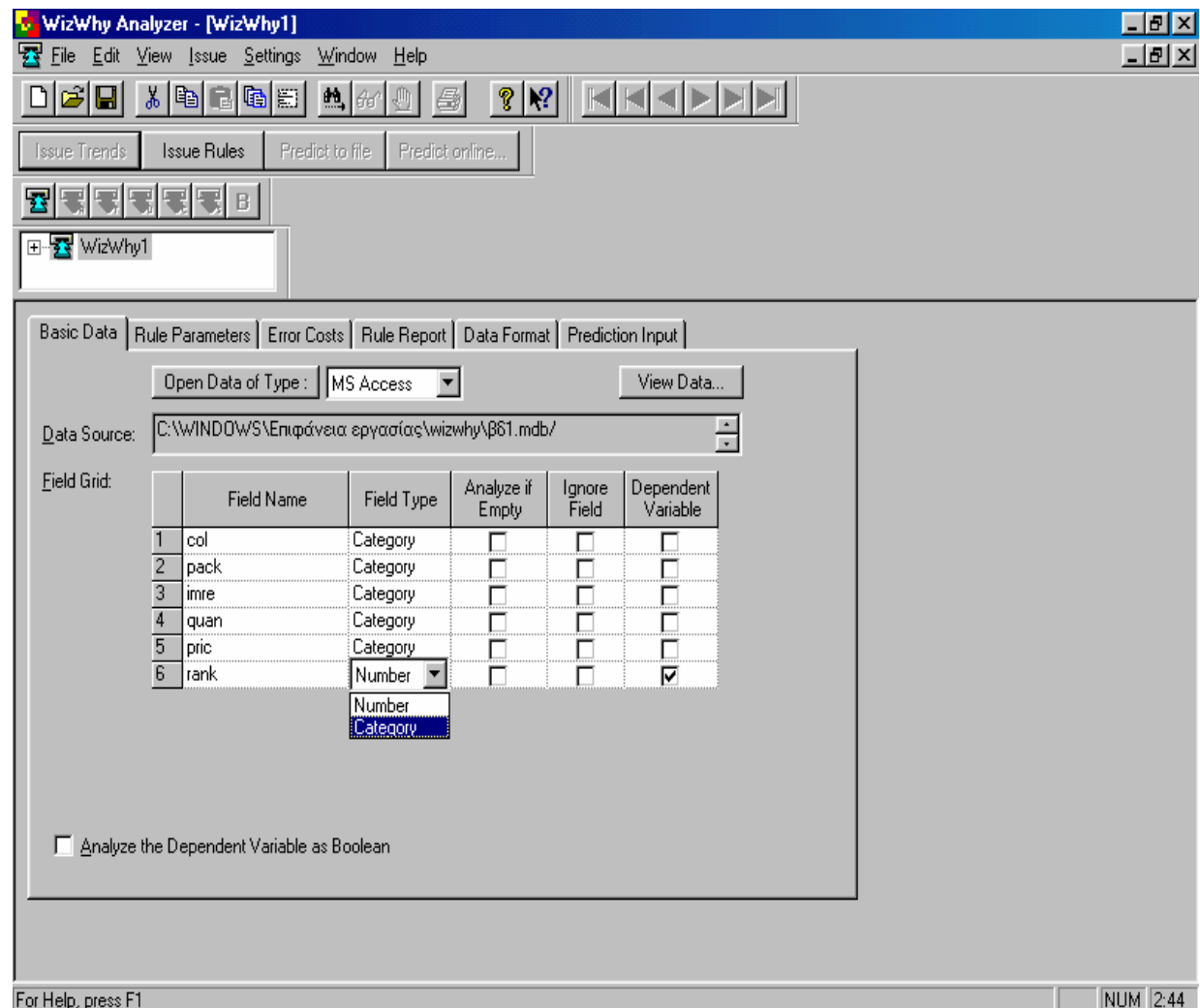
4.2 Εφαρμογή wizwhy

Η έκδοση που χρησιμοποιούμε εμείς είναι η 3.01-demo και η οποία περιορίζεται στα 1000 δεδομένα.

Για να μπορέσουμε να χρησιμοποιήσουμε τα δεδομένα μας θα πρέπει να χειριστούμε μια συμβατή βάση δεδομένων με το πρόγραμμα. Αυτή θα είναι η ACCESS της Microsoft. Άρα αναπροσαρμόζουμε τα δεδομένα μας στη νέα μας βάση κάνοντας έναν διαχωρισμό. Χρησιμοποιούμε δυο βάσεις για κάθε προϊόν, μια για την εκπαίδευση του συστήματος μας(training set) και άλλη μια για τη πρόβλεψη μας. Η πρώτη περιέχει τα πρώτα 500 δεδομένα του αντίστοιχου προϊόντος ενώ η δεύτερη περιέχει τα υπόλοιπα 53.Θα χρησιμοποιηθούν και οι τιμές του prediction field ,δηλαδή του rank, στο αρχείο ,ώστε να φέρουμε σε αντιδιαστολή τις πραγματικές τιμές και αυτές της πρόβλεψης.

Επιλέγουμε στο φάκελο basic data να ανοίγει βάσεις δεδομένων τύπου access και προσέχουμε να δηλώσουμε ότι η εξαρτημένη μεταβλητή μας δεν είναι δυαδική αλλά παίρνει τιμές από μηδέν ως τέσσερα(ποιοτική προδιάταξη). Στον επόμενο φάκελο, τον rule parameters, δηλώνουμε ότι μικρότερος αριθμός περιπτώσεων σε κανόνα είναι είκοσι και ο μέγιστος αριθμός συνθηκών σε κανόνα είναι τρία.Σε

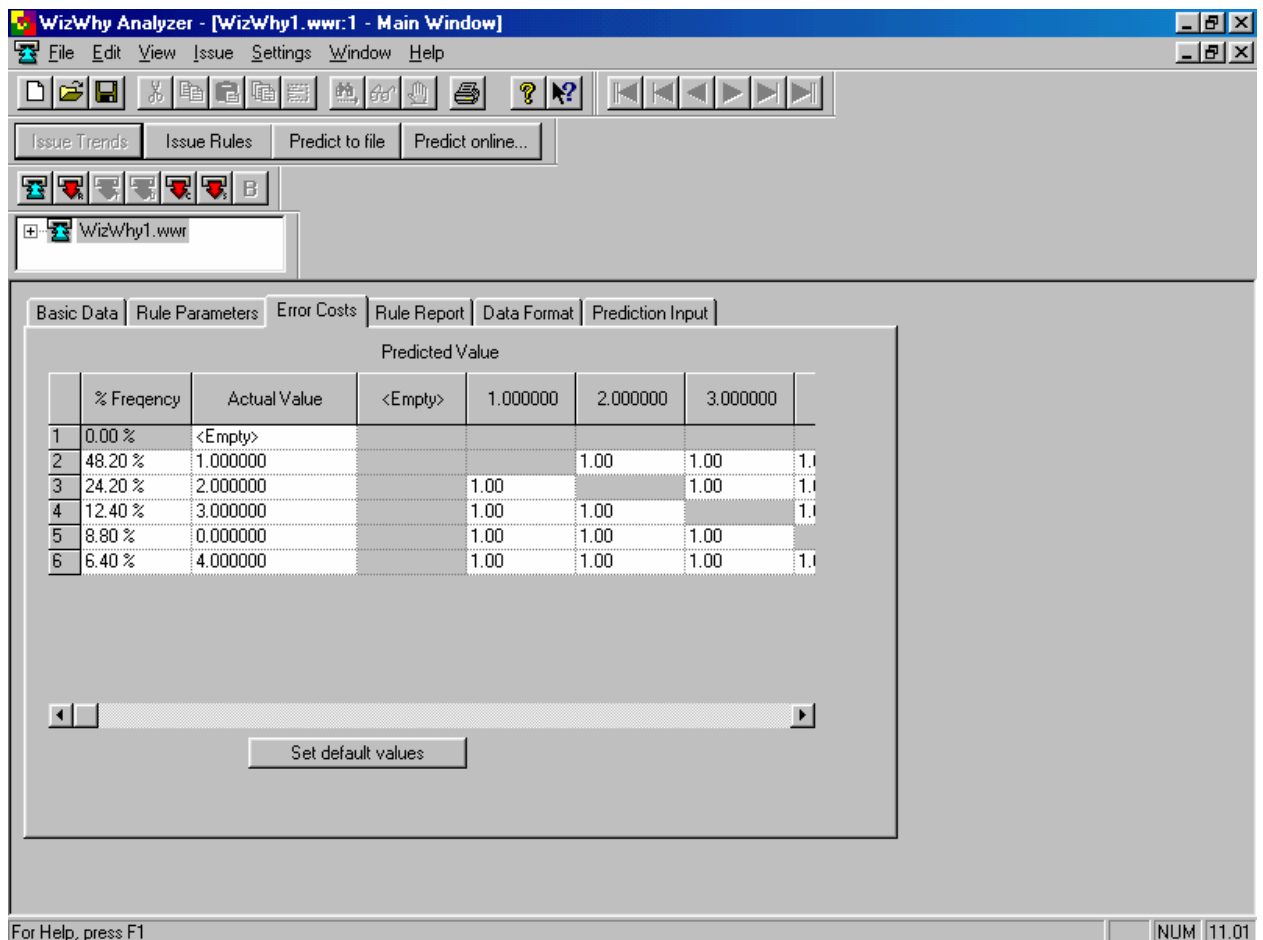
επόμενο φάκελο, rule report, του ζητάμε να μας παρουσιάσει το πολύ 100 κανόνες με maximum παραδείγματα στήριξης ή αποχής από τον κανόνα , τα δέκα.



Σχήμα 4α -Λήλωση μεταβλητών

Επόμενο βήμα είναι να ανοίξουμε τη βάση δεδομένων μας αυτή που περιέχει τα δεδομένα εκπαίδευσης(training-Lorena) και τότε θα μας εμφανίσει τα πεδία μας και θα επιλέξουμε ως εξαρτημένη μεταβλητή το rank. Πρέπει να αλλάξουμε τον τύπο πεδίου από number σε category για να μπορεί το πρόγραμμά μας να καταλάβει ότι τα νούμερα των πεδίων δεν απλά ποσοτικά στοιχεία αλλά ποιοτικά, δηλαδή ότι συμβολίζουν φραστικές εκτιμήσεις.

Το wizwhy θα μας δείξει τότε ότι η εξαρτημένη μεταβλητή είναι δυνατόν να πάρει 5 τιμές:



Σχήμα 4β-Ποσοστά τιμών εξαρτημένης μεταβλητής

με ποσοστό 48,2% παίρνει τη τιμή 1
 με ποσοστό 24,2% παίρνει τη τιμή 2
 με ποσοστό 12,4% παίρνει τη τιμή 3
 με ποσοστό 6,4% παίρνει τη τιμή 4
 με ποσοστό 8,8% παίρνει τη τιμή 0

χρησιμοποιώντας το πλήκτρο issue rules, το wizwhy θα μας παραθέσει τους κανόνες που διέπουν τα δεδομένα μας και κάποιες γενικές πληροφορίες:

Total number of records: **500**

Minimum probability of if-then rules: **0,750**

Minimum number of cases in a rule: **20**

The Demo version is limited to 1000 records.

Dependent Variable: **rank**

Number of rules 153

Number of values : 6

Παρατηρούμε πως ύστερα παρουσιάζονται οι κανόνες if then και ότι η ελάχιστη πιθανότητα τους είναι 75%. Μας δίνει στοιχεία επίσης για το που έχει εφαρμογή ο κάθε κανόνας, δηλαδή σε ποιον απαντάται το κάθε πεδίο

Ύστερα πρέπει να φορτώσουμε και τη βάση δεδομένων που θα κάνουμε τη πρόβλεψη, αυτή που έχει τα 53 δεδομένα(pract-Lorena). Εκεί αφού του δηλώσουμε σε ποιο αρχείο να τυπώσει τα αποτελέσματα της πρόβλεψης, θα μας δώσει τη δυνατότητα να αντιστοιχίσουμε πεδία του training set και του prediction set. Φυσικά

θα κάνουμε ακριβή αντιστοίχιση. Τελικά, δίνοντας μας 153 κανόνες, τα σωστά προβλεπόμενα δεδομένα, που παρατίθενται παρακάτω, είναι 26 στα 53 δηλαδή 49%.

LORENA

"rank","Concl_Prob","Prediction"

"3.000000",0,000,1.000000,
"2.000000",0,482,1.000000,
"1.000000",0,257,1.000000,
"1.000000",0,350,2.000000,
"3.000000",0,347,2.000000,
"2.000000",0,347,2.000000,
"2.000000",0,318,3.000000,
"1.000000",0,674,1.000000,
"1.000000",0,000,1.000000,
"1.000000",0,318,3.000000,
"2.000000",0,000,1.000000,
"1.000000",0,803,1.000000,
"1.000000",0,350,2.000000,
"1.000000",0,350,2.000000,
"1.000000",0,350,2.000000,
"1.000000",0,674,1.000000,
"1.000000",0,035,4.000000,
"1.000000",0,777,1.000000,
"3.000000",0,753,1.000000,
"1.000000",0,699,1.000000,
"1.000000",0,765,1.000000,
"1.000000",0,813,1.000000,
"3.000000",0,766,1.000000,
"4.000000",0,000,2.000000,
"3.000000",0,000,2.000000,
"3.000000",0,000,1.000000,
"2.000000",0,353,2.000000,
"3.000000",0,000,1.000000,
"3.000000",0,000,1.000000,
"3.000000",0,032,4.000000,
"2.000000",0,350,2.000000,
"1.000000",0,783,1.000000,
"1.000000",0,000,1.000000,
"1.000000",0,674,1.000000,
"4.000000",0,061,4.000000,
"4.000000",0,000,2.000000,
"4.000000",0,000,2.000000,
"4.000000",0,061,4.000000,
"3.000000",0,000,2.000000,
"2.000000",0,361,2.000000,
"2.000000",0,049,4.000000,
"1.000000",0,743,1.000000,
"1.000000",0,774,1.000000,
"1.000000",0,695,1.000000,
"2.000000",0,000,2.000000,

"2.000000",0,340,2.000000,
 "1.000000",0,347,2.000000,
 "2.000000",0,347,2.000000,
 "1.000000",0,347,2.000000,
 "1.000000",0,874,1.000000,
 "1.000000",0,350,2.000000,
 "3.000000",0,770,1.000000,
 "1.000000",0,753,1.000000,

Μπορούμε να τα παρουσιάσουμε συνολικά και να φτιάξουμε έναν πίνακα του τύπου:

		predicted			
actual	lorena	1	2	3	4
	1	17	7	1	1
	2	2	7	1	1
	3	7	3	0	1
	4	0	3	0	2

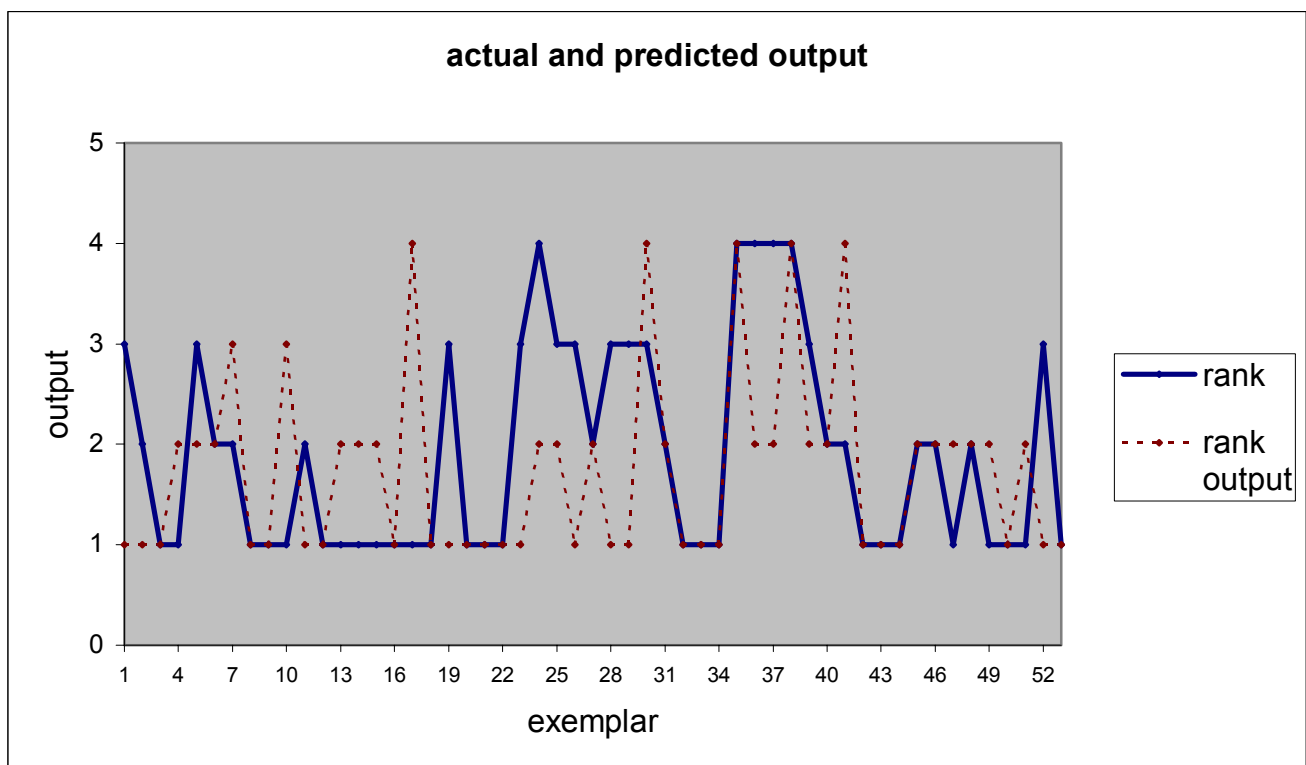
Πίνακας 4.1-Σφάλματα wizwhy για το lorena

Αυτός ο πίνακας έχει, όπως βλέπουμε, τα αποτελέσματα πρόβλεψης για το πρώτο λάδι μας, το lorena , συγκεντρωτικά. Στον οριζόντιο άξονα έχουμε τις πιθανές τιμές της πρόβλεψης ενώ στον κάθετο τις πραγματικές τιμές, οπότε ο αριθμός στο κάθε κελί μέσα στο πίνακα συμβολίζει το πλήθος των προγνώσεων που είχαν πραγματική και υπολογιστική τιμή αυτή που αντιστοιχεί στη γραμμή και τη στήλη του πίνακα αντιστοίχα. Για παράδειγμα στο κελί (1, 2) έχουμε την τιμή 7. Αυτό σημαίνει πως υπήρξαν 7 προβλέψεις που είχαν πραγματική τιμή 1 αλλά το πρόγραμμα υπολόγισε τιμή 2(σφάλμα υποεκτίμησης), δηλαδή το κατέταξε χαμηλότερα από ότι πραγματικά ήταν. Άλλο παράδειγμα είναι το κελί (4, 3), όπου έχει τιμή 0, δηλαδή δεν υπήρξαν περιπτώσεις όπου έπρεπε να υπολογιστεί τιμή 4 αλλά προέκυψε 3(σφάλμα υπερεκτίμησης).

Με το παραπάνω σκεπτικό οι σωστές τιμές της πρόβλεψης είναι μόνο αυτές που βρίσκονται πάνω στη διαγώνιο του πίνακα και στη συγκεκριμένη περίπτωση είναι:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 17 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 7 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 0 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 2 εγγραφές δεδομένων.

Συνολικά προκύπτουν 26 σωστές τιμές προβλέψεων στις 53 που δόθηκαν προς πρόβλεψη δηλαδή είχαμε ποσοστό επιτυχίας 49%. Μπορούμε παρατηρώντας το διάγραμμα πραγματικών και υπολογιστικών τιμών συναρτήσεως των εγγραφών να δούμε και παραστατικά πόσο πολύ πλησιάζουν μεταξύ τους. Είναι :



Διάγραμμα 4α-Διάγραμμα πραγματικών και πειραματικών τιμών(wizwhy-lorena)

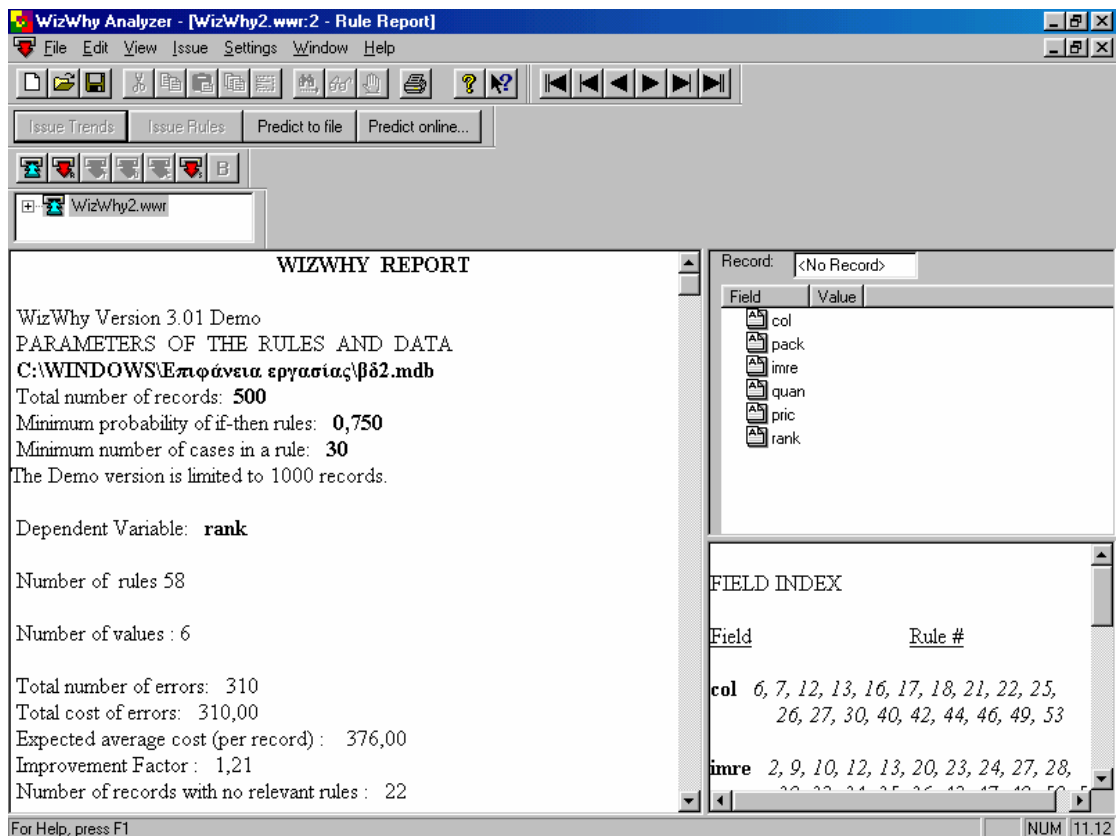
Μπορούμε να δούμε πως σε κάποια σημεία η τιμή της πρόβλεψης ταυτίζεται με την πραγματική αλλά υπάρχουν και πολλές διακυμάνσεις(αποκλίσεις) στις εγγραφές 10 ως 17 και 22 με 30 κυρίως, όπου εκεί δεν μπόρεσε το πρόγραμμα να ανταπεξέλθει με επιτυχία και το συγκεκριμένο λάδι υποβιβάστηκε ή υπερτιμήθηκε.

Ομοίως εφαρμόζουμε και τις υπόλοιπες βάσεις δεδομένων στο wizwhy για να εκπαιδευτεί εκ νέου το σύστημα για τα επόμενα τρία προϊόντα. Συνολικά δηλαδή θα χρησιμοποιήσουμε τις βάσεις της access, δύο για κάθε προϊόν, μια για training και μια για prediction ,όπως προαναφέρθηκε με 500 και 53 εγγραφές αντιστοίχως.

Για το επόμενο προϊόν , που είναι το AH ORGANIC, έχουμε:

Total number of records: **500**
 Minimum probability of if-then rules: **0,750**
 Minimum number of cases in a rule: **30**
 The Demo version is limited to 1000 records.

Dependent Variable: **rank**
 Number of rules 58
 Number of values : 6.



Σχήμα 4γ-Αναφορά γενικών αποτελεσμάτων

Τα αποτελέσματα της πρόβλεψής μας είναι:

AH ORGANIC

"rank","Concl_Prob","Prediction"

"1.000000",0,242,4.000000,
 "1.000000",0,350,2.000000,
 "3.000000",0,427,4.000000,
 "3.000000",0,350,2.000000,
 "1.000000",0,356,3.000000,
 "1.000000",0,356,3.000000,
 "4.000000",0,000,2.000000,
 "2.000000",0,350,2.000000,
 "4.000000",0,000,2.000000,
 "4.000000",0,360,4.000000,
 "3.000000",0,354,4.000000,
 "2.000000",0,348,4.000000,
 "3.000000",0,356,3.000000,
 "4.000000",0,349,2.000000,
 "4.000000",0,349,2.000000,
 "2.000000",0,349,2.000000,
 "3.000000",0,258,4.000000,
 "4.000000",0,348,4.000000,
 "2.000000",0,350,2.000000,
 "4.000000",0,360,4.000000,
 "4.000000",0,349,2.000000,

"3.000000",0,349,2.000000,
 "2.000000",0,348,4.000000,
 "2.000000",0,350,2.000000,
 "2.000000",0,000,2.000000,
 "2.000000",0,000,2.000000,
 "3.000000",0,394,4.000000,
 "4.000000",0,354,4.000000,
 "2.000000",0,347,2.000000,
 "2.000000",0,349,2.000000,
 "4.000000",0,350,2.000000,
 "3.000000",0,349,2.000000,
 "4.000000",0,388,4.000000,
 "3.000000",0,301,4.000000,
 "2.000000",0,350,2.000000,
 "1.000000",0,348,4.000000,
 "1.000000",0,000,2.000000,
 "2.000000",0,350,2.000000,
 "2.000000",0,350,2.000000,
 "3.000000",0,348,4.000000,
 "3.000000",0,348,4.000000,
 "2.000000",0,000,2.000000,
 "3.000000",0,346,3.000000,
 "4.000000",0,427,4.000000,
 "4.000000",0,348,4.000000,
 "1.000000",0,427,4.000000,
 "4.000000",0,356,3.000000,
 "4.000000",0,360,4.000000,
 "2.000000",0,427,4.000000,
 "3.000000",0,000,2.000000,
 "4.000000",0,249,4.000000,
 "4.000000",0,243,4.000000,
 "2.000000",0,294,4.000000,

Τα σωστά προβλεπόμενα αποτελέσματα είναι στην περίπτωση αυτού του λαδιού 24, έχοντας βρει 58 κανόνες, το ποσοστό επιτυχίας της μεθόδου είναι 45,2% και ο πίνακας των συγκεντρωτικών αποτελεσμάτων είναι :

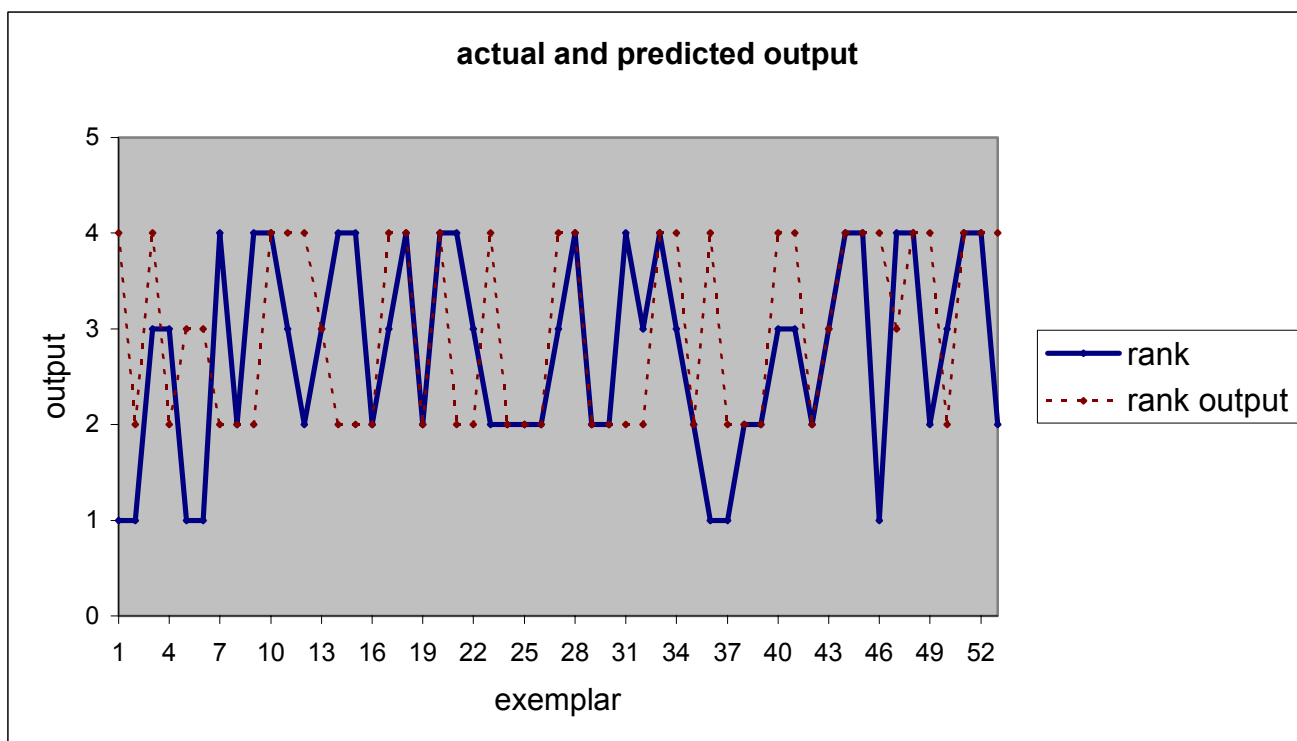
		predicted			
actual	ah-organic	1	2	3	4
	1	0	2	2	3
	2	0	12	0	4
	3	0	4	2	7
	4	0	6	1	10

Πίνακας 4.2-Σφάλματα wizwhy για το ah organic

Ομοίως με πριν έχουμε:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 0 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 12 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 2 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 10 εγγραφές δεδομένων.

Εδώ οι τιμές δυο και τέσσερα προβλέφθηκαν αρκετές φορές σωστά. Πιο ακριβή εικόνα μας δίνει και πάλι το διάγραμμα :



Διάγραμμα 4β-Διάγραμμα πραγματικών και πειραματικών τιμών(wizwhy-ah organic)

Ξεκινάει, όπως φαίνεται, λάθος η πρόβλεψή μας, μετά προσεγγίζει αρκετά καλά τις πραγματικές τιμές, αλλά στη περιοχή των εγγραφών 34 με 40 περίπου οι αποκλίσεις είναι μεγάλες και έχουμε σφάλματα υποεκτίμησης.

Για το επόμενο λάδι, Sitia, βρίσκουμε:

Total number of records: **500**

Minimum probability of if-then rules: **0,750**

Minimum number of cases in a rule: **20**

The Demo version is limited to 1000 records.

Dependent Variable: **rank**

Number of rules 155

Number of values : 6

, και η πρόβλεψη μας δίνει:

SITIA

"rank","Concl_Prob","Prediction"

"4.000000",0,657,4.000000,
"4.000000",0,026,1.000000,
"2.000000",0,039,1.000000,
"2.000000",0,259,3.000000,
"2.000000",0,000,4.000000,
"4.000000",0,593,4.000000,
"3.000000",0,734,4.000000,
"3.000000",0,373,3.000000,
"2.000000",0,038,1.000000,
"2.000000",0,372,3.000000,
"4.000000",0,734,4.000000,
"3.000000",0,373,3.000000,
"4.000000",0,000,4.000000,
"3.000000",0,584,4.000000,
"3.000000",0,372,3.000000,
"3.000000",0,584,4.000000,
"2.000000",0,000,4.000000,
"3.000000",0,657,4.000000,
"4.000000",0,737,4.000000,
"3.000000",0,373,3.000000,
"3.000000",0,000,4.000000,
"4.000000",0,000,4.000000,
"4.000000",0,000,4.000000,
"3.000000",0,657,4.000000,
"4.000000",0,000,4.000000,
"4.000000",0,000,4.000000,
"4.000000",0,318,0.000000,
"2.000000",0,373,3.000000,
"4.000000",0,373,3.000000,
"4.000000",0,372,3.000000,
"3.000000",0,372,3.000000,
"4.000000",0,000,4.000000,
"3.000000",0,000,4.000000,
"4.000000",0,584,4.000000,
"3.000000",0,000,4.000000,
"3.000000",0,028,1.000000,
"3.000000",0,000,4.000000,
"3.000000",0,000,4.000000,
"4.000000",0,657,4.000000,

"4.000000",0,188,3.000000,
 "4.000000",0,657,4.000000,
 "3.000000",0,734,4.000000,
 "2.000000",0,737,4.000000,
 "2.000000",0,375,3.000000,
 "3.000000",0,372,3.000000,
 "3.000000",0,034,1.000000,
 "2.000000",0,675,4.000000,
 "1.000000",0,710,4.000000,
 "4.000000",0,737,4.000000,
 "4.000000",0,000,4.000000,
 "3.000000",0,000,4.000000,
 "2.000000",0,158,4.000000,
 "4.000000",0,373,3.000000,

εδώ έχουμε 21 σωστά αποτελέσματα μέσω πρόβλεψης με 155 κανόνες να έχουν βρεθεί από το wizwhy , και επιτυχία 39,6%. Οι τιμές των σφαλμάτων φαίνονται παρακάτω:

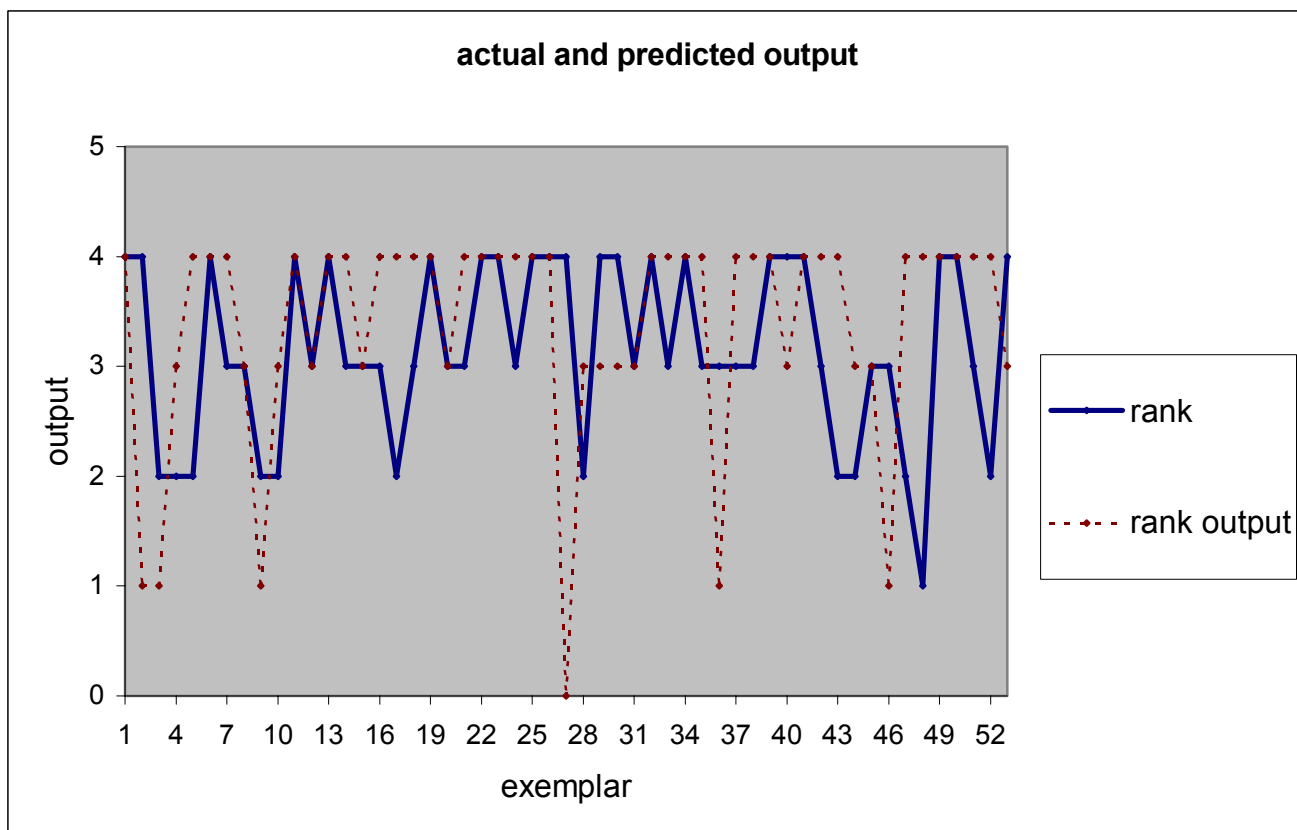
		predicted				
		sitia	1	2	3	4
actual	1	0	0	0	0	1
	2	2	0	4	5	
	3	2	0	6	12	
	4	1	0	4	15	

Πίνακας 4.3-Σφάλματα wizwhy για το sitia

Αναλυτικά είναι:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 0 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 0 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 6 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 15 εγγραφές δεδομένων.

Το ποσοστό επιτυχίας της μεθόδου είναι πιο μικρό από τα προηγούμενα, 39,6% και καθιστά αρκετά αβέβαιη τη μέθοδο αυτή για τα δεδομένα μας. Βλέπουμε και το διάγραμμα παρακάτω:



Διάγραμμα 4γ-Διάγραμμα πραγματικών και πειραματικών τιμών(wizwhy-sitia)

Είναι φανερό πως υφίστανται μεγάλες αποκλίσεις στη περίπτωση αυτού του λαδιού, όπως στις πρώτες τέσσερις εγγραφές, στη 9, στην 27 που αποτελεί και ακραία περίπτωση, γιατί το πρόγραμμα δεν μπόρεσε να λειτουργήσει καθόλου, στην 36 και στην 48 με απόκλιση τριών μονάδων, δίνοντας σφάλμα υπερεκτίμησης.

Τελευταίο προϊόν μας είναι το Bertolli. Οι τελευταίες βάσεις δεδομένων εφαρμοζόμενες στο wizwhy θα μας δώσουν , η πρώτη, τους κανόνες που είναι 170 στη περίπτωση αυτή, και η δεύτερη τις προβλεπόμενες τιμές της εξαρτημένης μεταβλητής, δηλαδή τη κατάταξη μεταξύ των προϊόντων(rank).

Total number of records: **500**
 Minimum probability of if-then rules: **0,750**
 Minimum number of cases in a rule: **20**
 The Demo version is limited to 1000 records.

Dependent Variable: **rank**
 Number of rules 170
 Number of values : 6

BERTOLLI

"rank","Concl_Prob","Prediction"

"2.000000",0,392,3.000000,
"3.000000",0,397,2.000000,
"4.000000",0,377,2.000000,
"4.000000",0,312,4.000000,
"4.000000",0,316,4.000000,
"3.000000",0,395,3.000000,
"1.000000",0,377,2.000000,
"4.000000",0,392,3.000000,
"3.000000",0,316,2.000000,
"3.000000",0,377,2.000000,
"1.000000",0,441,3.000000,
"4.000000",0,371,2.000000,
"2.000000",0,425,1.000000,
"2.000000",0,382,2.000000,
"2.000000",0,265,2.000000,
"4.000000",0,467,3.000000,
"4.000000",0,387,3.000000,
"2.000000",0,389,2.000000,
"1.000000",0,422,1.000000,
"2.000000",0,377,2.000000,
"2.000000",0,467,3.000000,
"2.000000",0,434,3.000000,
"1.000000",0,402,1.000000,
"1.000000",0,302,2.000000,
"1.000000",0,418,2.000000,
"1.000000",0,381,2.000000,
"1.000000",0,385,2.000000,
"1.000000",0,371,1.000000,
"1.000000",0,395,3.000000,
"1.000000",0,418,2.000000,
"1.000000",0,409,1.000000,
"2.000000",0,298,1.000000,
"2.000000",0,467,3.000000,
"2.000000",0,377,2.000000,
"1.000000",0,171,4.000000,
"2.000000",0,414,3.000000,
"2.000000",0,210,4.000000,
"1.000000",0,389,1.000000,
"1.000000",0,402,1.000000,
"1.000000",0,413,1.000000,
"1.000000",0,382,2.000000,
"4.000000",0,387,3.000000,
"4.000000",0,414,3.000000,
"3.000000",0,333,4.000000,
"1.000000",0,385,2.000000,
"4.000000",0,266,3.000000,
"3.000000",0,391,2.000000,

"3.000000",0,395,3.000000,
 "3.000000",0,414,3.000000,
 "2.000000",0,422,1.000000,
 "2.000000",0,385,2.000000,
 "1.000000",0,312,4.000000,
 "3.000000",0,392,3.000000.

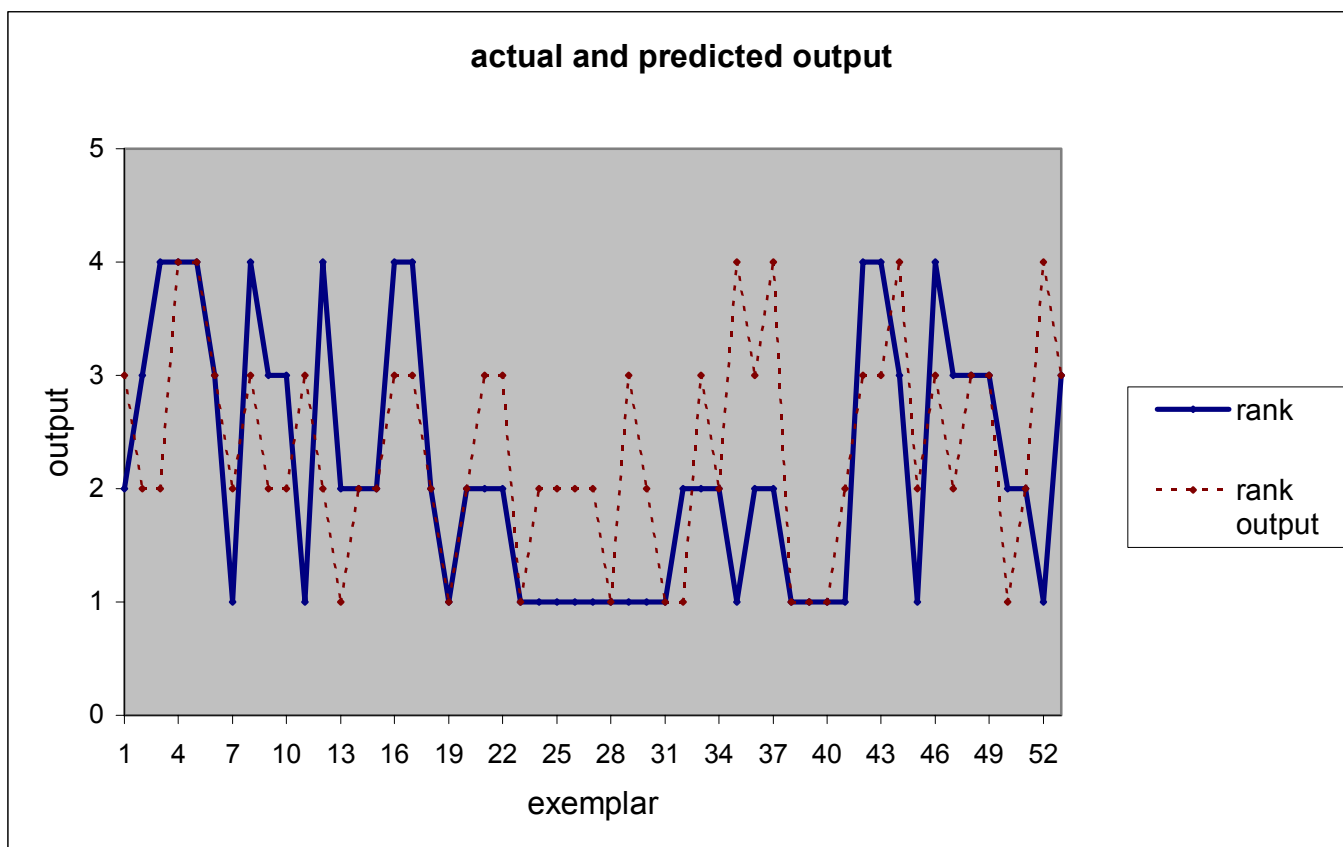
Σε αυτή τη περίπτωση τα σωστά προβλεπόμενα αποτελέσματα είναι 19 στα 53 ή 35,8%. Και εδώ τα αποτελέσματα της πρόβλεψης δεν ήταν καθόλου ικανοποιητικά με πολλά σφάλματα υποεκτίμησης. Ιδού:

		predicted			
actual	bertolli	1	2	3	4
	1	7	8	2	2
	2	3	6	5	1
	3	0	4	4	1
	4	0	2	6	2

Πίνακας 4.4-Σφάλματα wizwhy για το bertolli

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 7 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 6 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 4 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 2 εγγραφές δεδομένων.

Η αξιοπιστία έπεσε στο 35,8% με σωστές προβλέψεις να έχουν γίνει μόνο σε 19 περιπτώσεις από τις 53. Το διάγραμμα εδώ, όπως είναι αναμενόμενο δεν ικανοποιεί καθόλου με τη μορφή του και οι οι προκύπτουσες τιμές με τις πραγματικές δεν πλησιάζουν σε πολλές περιπτώσεις. Έχουμε τελικά την παρακάτω μορφή:



Διάγραμμα 4δ-Διάγραμμα πραγματικών και πειραματικών τιμών(wizwhy-bertolli)

Είναι ολοφάνερο ιδίως στις εγγραφές κεντρικών «περιοχών» πως η μέθοδος έχει αποτύχει, ενώ σε αυτές κοντά στο 37 ερωτηματολόγιο πως υπάρχει η μέγιστη απόκλιση, οπότε η μέθοδος σ' αυτή την περίπτωση δεν μπορεί να μας δώσει αξιόπιστη λύση στο πρόβλημα που του αναθέτουμε.

Παρατηρούμε πως στη περίπτωση του AH ORGANIC χρησιμοποιήσαμε 30 τουλάχιστον εγγραφές για να τεκμηριώσουμε κάθε κανόνα, ενώ στις άλλες περιπτώσεις 20 τουλάχιστον. Αυτό έγινε διότι έτσι είχαμε τα καλύτερα δυνατά αποτελέσματα .Γενικότερα χρησιμοποιώντας διαφορετικές τιμές των παραμέτρων του wizwhy προσπαθώντας να καλυτερεύσουμε τη πρόβλεψη καταλήξαμε στα πιο πάνω συμπεράσματα τα οποία πρακτικά είναι βέλτιστα. Οι παράμετροι του προγράμματος μάς δίνουν τη δυνατότητα να πειραματιστούμε για να βρούμε πότε το πρόγραμμά μας συνεργάζεται καλύτερα με τα δεδομένα μας. Άρα είτε χρησιμοποιήσουμε ως ελάχιστο αριθμό περιπτώσεων σε ένα κανόνα το 20, το 30 κτλ είτε αλλάξουμε το μέγιστο αριθμό συνθηκών σε ένα κανόνα 2, 3, 4 κτλ το ποσοστό της αξιόπιστης πρόβλεψης δεν πρόκειται να αυξηθεί. Αυτά είναι τα ποσοστά αξιοπιστίας που το wizwhy απέδωσε πλησιάζοντας σε δυο περιπτώσεις το 50%.

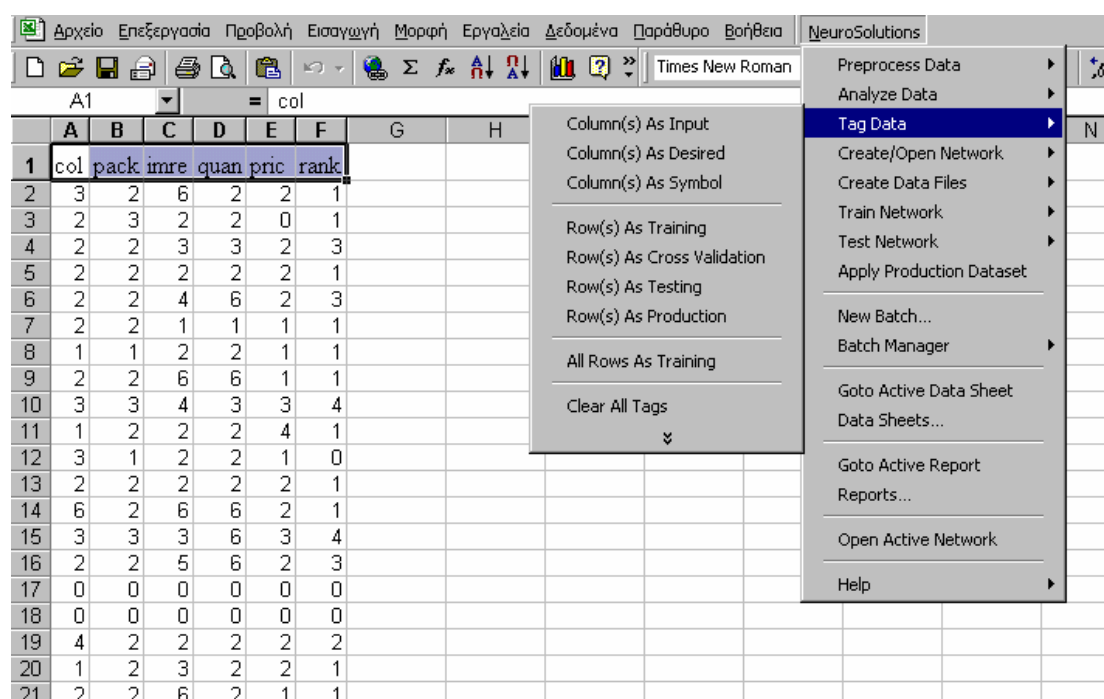
4.3 Εφαρμογή neurosolutions

Ξεκινώντας το neurosolutions ανοίγει ένα παράθυρο το οποίο περιέχει όλα τα επίπεδα τα οποία μπορεί να χρειαστεί κάποιος για να αναπτύξει το νευρωνικό του

δίκτυο. Το neurosolutions δεν χρειάζεται μετατροπή της βάσης δεδομένων μας αλλά απευθείας χρησιμοποιούμε τη βάση μας όπως την έχουμε δημιουργήσει στο excel, γι' αυτό άλλωστε χειριζόμαστε το εργαλείο neurosolutions for excel.

Το πρώτο πράγμα που κάνουμε είναι να ενεργοποιήσουμε λοιπόν το NS EXCEL. Αυτό μας οδηγεί σε ένα νέο παράθυρο που δεν είναι τίποτα άλλο από το γνωστό μας excel, που έχει όμως προσαρμοσμένο το toolbox του neurosolutions. Μόλις ανοίξουμε το αρχείο μας με τα δεδομένα, χρησιμοποιήσουμε ξεχωριστά τη βάση του πρώτου προϊόντος ,μετά του δεύτερου κτλ.

Το πρώτο μας προϊόν, λοιπόν, είναι ως γνωστόν το LORENA. Το neurosolutions λειτουργεί με τον εξής τρόπο: πρέπει να του δηλώσουμε τι αντιπροσωπεύουν αυτά τα δεδομένα που είναι στη βάση μας. Συγκεκριμένα μας δίνει τις εξής επιλογές:



Σχήμα 4δ-Χαρακτηρισμός δεδομένων

- Να δηλώνουμε κάποιες κολώνες ως είσοδο (input) στο σύστημα. Αν κάποιες κολώνες δεν τις χαρακτηρίσουμε input τότε τα δεδομένα τους δεν συμπεριλαμβάνονται στη δημιουργία του αρχείου δεδομένων.
- Να δηλώνουμε κάποιες κολώνες ως επιθυμητές(desired). Αυτές οι κολώνες ουσιαστικά αποτελούν τα δεδομένα της εξαρτημένης μεταβλητής ή αλλιώς της ποσότητας που θέλουμε να προβλέψουμε.
- Να δηλώνουμε κάποιες κολώνες ως συμβολικές(symbol), αν για παράδειγμα δεν είναι ποσοτικές οι τιμές τους αλλά ποιοτικές, με την έννοια ότι είναι κωδικοποιημένα τα νούμερα που εμφανίζονται.(δεν παρέχεται αυτή η δυνατότητα στη shareware έκδοση που διαθέτουμε).
- Να δηλώνουμε κάποιες γραμμές, επίσης, που επιθυμούμε να αποτελούν τις γραμμές εκπαίδευσης(training). Δηλαδή αυτές αποτελούν τα δεδομένα με τα

οποία θα εκπαιδευτεί το νευρωνικό μας δίκτυο και συνήθως δηλώνουμε αρκετά για να δώσουμε όσο μπορούμε μια πληρότητα στη γνώση του συστήματός μας.(εδώ περιοριζόμαστε στα 300 δεδομένα λόγω *shareware version*)

- Να δηλώνουμε γραμμές δοκιμής προκειμένου το νευρωνικό δίκτυο να μας δώσει αποτελέσματα πρόβλεψης (testing) της ποσότητας που έχουμε δηλώσει ως επιθυμητή.

Εφαρμόζουμε στα δικά μας δεδομένα τους παραπάνω χαρακτηρισμούς ως εξής :

Στήλες από A ως E χαρακτηρισμένες ως είσοδος

Στήλη F χαρακτηρισμένη ως επιθυμητή.

Γραμμές από 2 ως 300 χαρακτηρισμένες ως εκπαίδευσης.

Γραμμές από 502 ως 554 χαρακτηρισμένες ως δοκιμής.

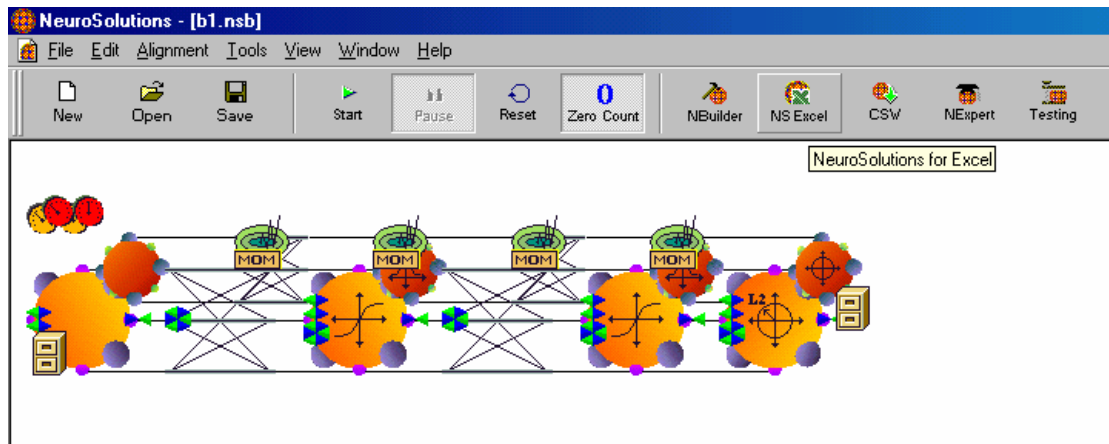
Μόλις έχουμε τελειώσει με αυτά επόμενο βήμα είναι να επιλέξουμε το «κουμπί» δημιουργία/άνοιγμα δικτύου(Create/Open Network) όπου επιλέγουμε «νέο» και ύστερα μπαίνουμε σε διάφορες διαδικασίες επιλογής παραμέτρων.

Αρχικά διαλέγουμε τον τύπο νευρωνικού δικτύου που θα χρησιμοποιηθεί. Εμείς διαλέγουμε τα multilayer perceptron γιατί μπορούν να προσεγγίσουν κάθε πλαίσιο εισόδου, εξόδου και είναι σχετικά εύκολα στο χειρισμό τους. Το κύριο μειονέκτημά τους είναι ότι εκπαιδεύονται αργά και απαιτούν πολλά δεδομένα εκπαίδευσης(τυπικά απαιτούν τρεις φορές περισσότερα δεδομένα από τα βάρη του δικτύου).

Ύστερα επιλέγουμε ως κρυμμένα επίπεδα ένα. Κάθε κρυμμένος νευρώνας(νευρώνας σε κρυμμένο επίπεδο) περιέχει μια μη γραμμική συνάρτηση ενεργοποίησης και αυτό προσφέρει και τις αυξημένες υπολογιστικές δυνατότητες του multilayer perceptron. Αυτοί οι νευρώνες καθιστούν το δίκτυο ικανό να μάθει πολύπλοκα πρότυπα εξάγοντας από αυτά κάποια ιδιαίτερα χαρακτηριστικά τους. Δεν υπάρχει σύνδεση μεταξύ νευρώνων του ίδιου επιπέδου και ως επί το πλείστον υπάρχει πλήρης διασύνδεση μεταξύ των νευρώνων δυο διαδοχικών επιπέδων. Επιπλέον συνήθως δεν επιτρέπονται συνδέσεις μεταξύ νευρώνων που ανήκουν σε επίπεδα που δεν είναι διαδοχικά.

Επόμενο βήμα είναι να επιλέξουμε τον αριθμό των εποχών(epochs). Αυτός ο αριθμός ρυθμίζει το σύστημα σχετικά με το πόσα περάσματα θα κάνει το πρόγραμμα από τα δεδομένα εκπαίδευσης αν κάποιο άλλο κριτήριο δεν επέμβει να σταματήσει τη διαδικασία. Επίσης επιλέγουμε την εκπαίδευση ανά εποχή(off line) και όχι την ανά πρότυπο(on line). Αυτό σημαίνει ότι τα βάρη του δικτύου θα ενημερωθούν ύστερα από την παρουσίαση στο δίκτυο ολόκληρου του συνόλου προτύπων της εποχής και όχι μετά από τη παρουσίαση κάθε προτύπου ξεχωριστά.

Ύστερα επιλέγουμε τον τρόπο παρουσίασης των αποτελεσμάτων και κατασκευάζουμε το δίκτυο.



Σχήμα 4ε-Κατασκευή δικτύου

Δίνουμε την εντολή να κάνει εκπαίδευση στο δίκτυο από το neurosolutions for excel και αφού σώσουμε το δίκτυο που έχουμε δημιουργήσει παίρνουμε τα αποτελέσματα της εκπαίδευσης του νευρωνικού δικτύου στο αρχείο του excel.

Τέλος επιλέγουμε δοκιμή του δικτύου(test) για να προκύψουν τα αποτελέσματα της πρόβλεψης της εξαρτημένης μεταβλητής, επίσης προσαρτημένα στο αρχείο excel.

Τα αποτελέσματά παρουσιάζονται ακολούθως με την ίδια πάντα σειρά προϊόντων:

LORENA

Rank	predicted value(rank)
3	3
2	2
1	3
1	2
3	2
2	2
2	2
1	2
1	2
1	2
2	2
1	2
1	2
1	2
1	2
1	2

1	2
1	2
3	1
1	1
1	1
1	1
3	1
4	3
3	2
3	2
2	2
3	2
3	2
3	2
2	2
1	1
1	2
1	2
4	3
4	3
4	2
4	2
3	1
2	2
2	2
1	1
1	2
1	2
2	3
2	2
1	2
2	2
1	2
1	1
1	2
3	1
1	2

Παρατηρούμε πως στη περίπτωση αυτού του λαδιού οι σωστά προβλεπόμενες τιμές είναι μόλις 17 στα 53 δεδομένα που είχαμε εισάγει στο πρόγραμμα. Αυτό σημαίνει πως η πρόβλεψή μας είχε ποσοστό επιτυχίας 32%.

Ο συγκεντρωτικός πίνακας τώρα θα είναι:

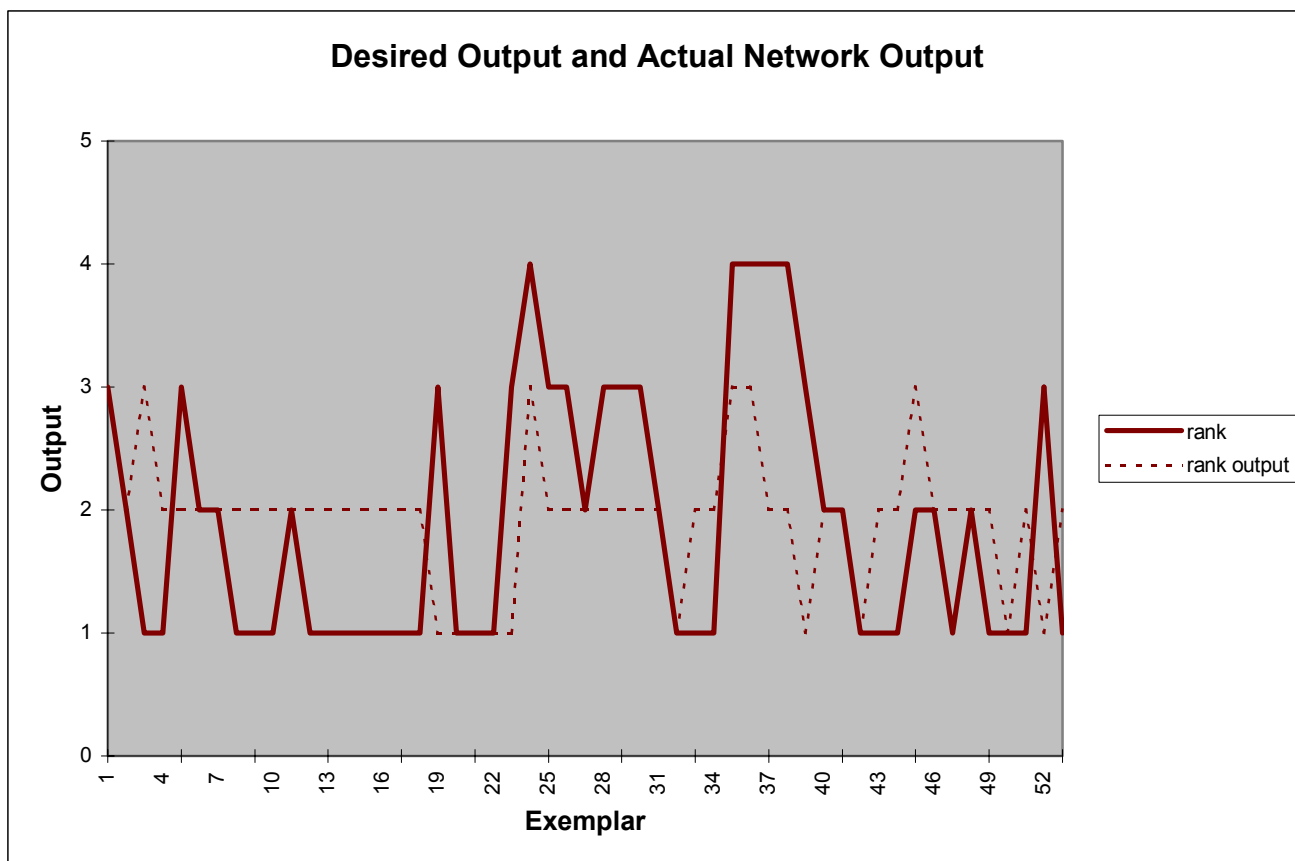
		predicted			
actual	lorena	1	2	3	4
	1	6	19	1	0
	2	0	10	1	0
	3	4	6	1	0
	4	0	2	3	0

Πίνακας 4.5-Σφάλματα neurosolutions για το lorena

Όπως και προηγουμένως τα στοιχεία της κυρίας διαγωνίου απεικονίζουν τις σωστά προβλεπόμενες τιμές. Στη συγκεκριμένη περίπτωση είναι 17 στις 53 περιπτώσεις, δηλαδή 32%, ποσοστό μάλλον απογοητευτικό για τη μελέτη μας. Έχουν γίνει πάρα πολλά λάθη στην περίπτωση (1, 2) δηλαδή πολλές εγγραφές που στην κατάταξη του lorena είχαν την τιμή 1 προβλέφθηκαν στη τιμή 2 λανθασμένα και αποτελούν 19 περιπτώσεις.

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 6 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 10 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 1 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 0 εγγραφές δεδομένων.

Ας έχουμε και μια αναπαράσταση των πραγμάτων:



Διάγραμμα 4ε-Διάγραμμα πραγματικών και πειραματικών τιμών(neurosolutions-lorena)

Η γραφική παράσταση των τιμών της πρόβλεψης ακολουθεί τις πραγματικές τιμές αλλά σε πολλές περιπτώσεις βλέπουμε ότι δεν μπορεί να ελιχθεί σωστά στις μεγάλες διακυμάνσεις των προτιμήσεων των καταναλωτών και να προσεγγίσει τις τιμές αυτές. Στο σημείο, δε, 39(πρότυπο 39) βλέπουμε ότι η παράστασή έχει τη μεγαλύτερη απόκλιση που μπορεί να έχει.

Τα αποτελέσματα για το επόμενο λάδι δίνονται στη παρακάτω μορφή(ah organic):

Microsoft Excel - nsol2.xls

Αρχείο Επεξεργασία Προβολή Εισαγωγή Μορφή Εργαλεία Δεδομένα Παράθυρο Βοήθεια NeuroSolutions

G2 = 3

	A	B	C	D	E	F	G	H	I	J
1	col	pack	imre	quan	ptic	rank	rank output			
2	3	4	2	3	6	1	3			
3	1	2	2	2	2	1	2			
4	2	3	3	2	4	3	3			
5	1	2	3	2	4	3	3			
6	2	2	2	6	3	1	3			
7	2	2	2	6	4	1	3			
8	4	4	4	3	4	4	3			
9	2	2	3	2	4	2	3			
10	4	4	4	4	4	4	3			
11	3	4	6	4	4	4	4			
12	3	4	6	3	5	3	3			
13	1	2	3	3	5	2	3			
14	2	2	2	6	3	3	3			
15	2	2	2	2	5	4	3			
16	2	2	2	2	4	4	2			
17	2	2	2	2	4	2	2			
18	2	3	2	3	6	3	3			
19	2	4	3	2	5	4	3			
20	1	1	1	2	5	2	3			
21	3	1	3	3	4	4	2			

Σχήμα 4στ-Εξαγωγή αποτελεσμάτων(neurosolutions, ah organic)

AH ORGANIC

rank	predicted value(rank)
1	3
1	2
3	3
3	3
1	3
1	3
4	3
2	3
4	3
4	4
3	3
2	3
3	3
4	3
4	2
2	2
3	3
4	3
2	3

4	2
4	2
3	2
2	2
2	2
2	2
2	2
3	3
4	3
2	2
2	2
4	2
3	1
4	3
3	3
2	2
1	3
1	2
2	3
2	2
3	3
3	3
2	2
3	3
4	3
4	3
1	2
4	3
4	3
2	3
3	1
4	3
4	3
2	3

Στη περίπτωση του δεύτερου λαδιού μας έχουμε λίγο καλύτερα αποτελέσματα, δηλαδή 21 σωστές προβλέψεις που αυτό σημαίνει ποσοστό επιτυχίας της μεθόδου μας 39,6%. Ούτε τα αποτελέσματα αυτά είναι ιδιαίτερα ικανοποιητικά

μιας και αυτό σημαίνει ότι στους 10 καταναλωτές εμείς μπορούμε περίπου τους τέσσερις να προβλέψουμε με επιτυχία όσον αφορά την απάντησή τους στο ερώτημα ποιο θεωρούν καλύτερο λάδι μεταξύ των τεσσάρων.

		predicted			
actual	ah-organic	1	2	3	4
	1	0	3	4	0
	2	0	10	6	0
	3	2	1	10	0
	4	0	4	12	1

Πίνακας 4.6-Σφάλματα neurosolutions για το ah organic

Τα σωστά αποτελέσματα (39,6%) είναι λίγο καλύτερα από τα προηγούμενα αλλά δεν παύουν να είναι και πάλι πολύ χαμηλής αξιοπιστίας. Εδώ σύνηθες λάθος είναι η περίπτωση η πραγματική τιμή να είναι 4 αλλά το πρόγραμμα να υπολογίζει 3 και το βρίσκουμε σε 12 περιπτώσεις(σφάλμα υπερεκτίμησης).

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 0 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 10 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 10 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 1 εγγραφές δεδομένων.

Στο αντίστοιχο διάγραμμα μπορούμε να δούμε πως κινείται η παράσταση των προβλέψεων σε σχέση με τις πραγματικές τιμές:

3	3
2	3
2	3
4	3
3	3
4	3
3	3
3	3
3	3
2	2
3	3
4	3
3	3
3	3
4	3
4	1
3	3
4	2
4	3
4	3
2	3
4	3
4	3
3	3
4	3
3	3
4	3
3	3
3	3
3	2
3	3
4	2
4	1
4	3
3	3
2	3
2	3

3	3
3	3
2	3
1	3
4	3
4	3
3	3
2	2
4	3

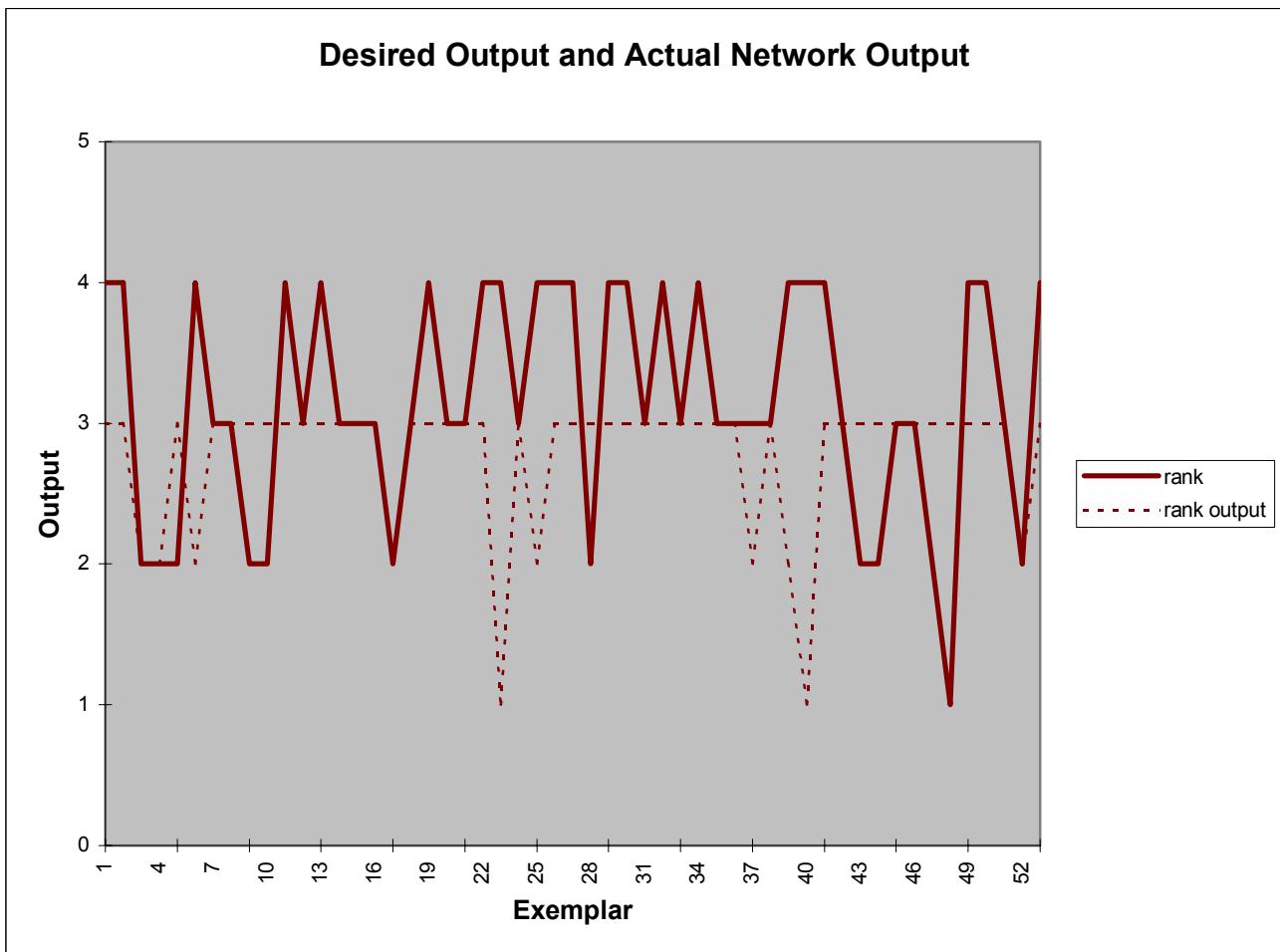
Σε αυτή τη δοκιμή το επίπεδο άρτιας πρόβλεψης έχει ανέβει στο 43,4%, με 23 τιμές της εξαρτημένης μεταβλητής να έχουν προβλεφθεί σωστά. Ας δούμε τον αντίστοιχο πίνακα:

		predicted				
		sitia	1	2	3	4
actual	1	0	0	1	0	
	2	0	4	7	0	
	3	0	1	19	0	
	4	2	3	16	0	

Πίνακας 4.7-Σφάλματα neurosolutions για το sitia

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 0 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 4 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 19 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 0 εγγραφές δεδομένων.

Βλέπουμε αμέσως πως 23 περιπτώσεις από τις 53 είναι αυτές που επαληθεύονται ακριβώς από τα πραγματικά δεδομένα και τα περισσότερα λάθη έχουν γίνει όπου για 4 έχει προβλεφθεί η τιμή 3 και συνολικά είναι 16 περιπτώσεις. Με το ποσοστό επιτυχίας να έχει ανέβει ακόμα πιο πάνω (43,4%) αξίζει να παρατηρήσουμε το γράφημα:



Φαίνεται πως, παρόλο που οι τιμές της εξόδου της συνάρτησης αποκλίνουν μέγιστα σε δυο περιπτώσεις (εγγραφές 23, 40) και παρουσιάζεται σφάλμα υπερεκτίμησης, παρόλα αυτά καταφέρνουν να προσεγγίσουν γενικότερα τη συνάρτηση των πραγματικών τιμών αρκετά ικανοποιητικά.

Προχωρώντας και στο τελευταίο μας προϊόν λαδιού παρατηρούμε πως χαμηλώνει και πάλι λίγο η αξιοπιστία του εργαλείου:

BERTOLLI

rank	predicted value(rank)
2	2
3	3
4	3
4	3
4	3
3	2
1	2
4	3
3	2
3	2
1	3
4	3
2	2
2	2
2	2
4	2
4	2
2	2
1	1
2	2
2	3
2	3
1	2
1	3
1	2
1	3
1	2
1	1
1	2
1	1
1	2
2	2
2	2
2	2
1	3

2	2
2	2
1	1
1	2
1	1
1	2
4	3
4	3
3	3
1	2
4	2
3	2
3	3
3	3
2	1
2	2
1	3
3	2

Σύμφωνα με τα δεδομένα, οι περιπτώσεις σωστών προβλέψεων τώρα είναι 21, άρα 39,6%, ποσοστό ίδιο με του ah organic. Ο μηχανισμός της πρόβλεψης λειτούργησε σχεδόν παρόμοια με το ah organic με λίγο καλύτερη «συμπεριφορά» στα λάθη. Συγκεκριμένα ο συγκεντρωτικός πίνακας έχει ως εξής:

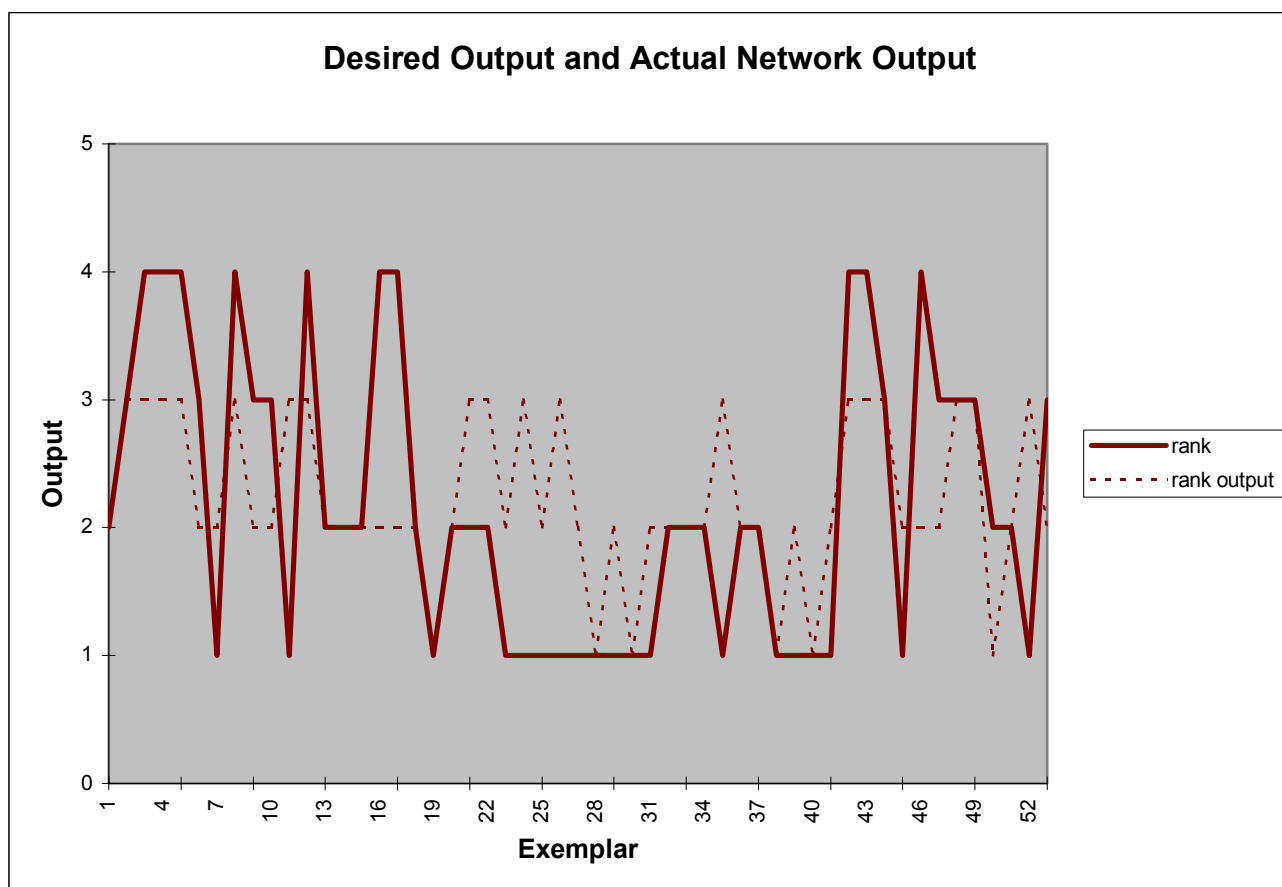
		predicted			
actual	bertolli	1	2	3	4
	1	5	9	5	0
	2	1	12	2	0
	3	0	5	4	0
	4	0	3	7	0

Πίνακας 4.8-Σφάλματα neurosolutions για το bertolli

Φαίνεται ότι:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 5 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 12 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 4 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 0 εγγραφές δεδομένων.

Έχουμε μερικά συνεχή λάθη, ιδίως στους ερωτηθέντες από το νούμερο 20 ως το 37 περίπου, αλλά δεν έχουμε μεγάλες αποκλίσεις. Μπορεί άλλωστε να το δει κανείς αν μελετήσει το τελευταίο διάγραμμα, όπου η παράσταση των προβλέψεων δεν απέχει ούτε συχνά, ούτε πολύ από τις πραγματικές τιμές. Ιδού και το διάγραμμα πραγματικών και υπολογιστικών τιμών:



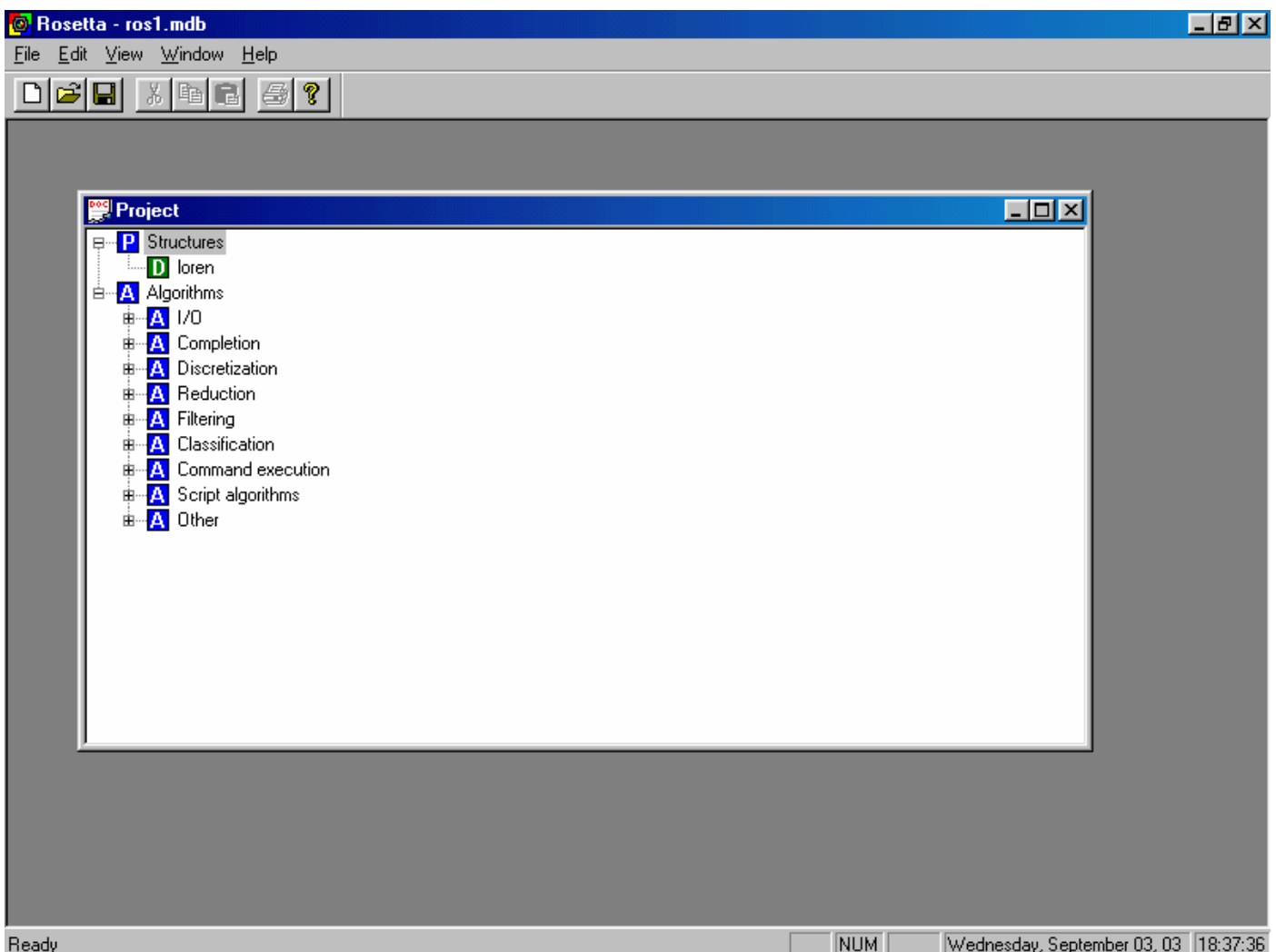
Διάγραμμα 40-Διάγραμμα πραγματικών και πειραματικών τιμών(neurosolutions-bertolli)

Συνολικά το neurosolutions δεν μπόρεσε να ελιχθεί τόσο καλά όσο το wizwhy και μας έδωσε, συνολικά, χειρότερα αποτελέσματα, ξεπερνώντας το 40% αξιοπιστίας μόνο στο ένα από τα τέσσερα προϊόντα (το sitia).

4.4 Εφαρμογή Rosetta

Το πρόγραμμα αυτό ξεκινάει με ένα παράθυρο που δίνει ως μόνες επιλογές το άνοιγμα κάποιας βάσης δεδομένων ή τη δημιουργία κάποιου καινούριου project. Όταν επιλεγεί το άνοιγμα ενός αρχείου βάσης δεδομένων, οποιασδήποτε κατάληξης, το πρόγραμμα ζητά τον τρόπο αναγνώρισής της σύμφωνα με κάποιες ρουτίνες εισαγωγής που διαθέτει. Αυτό που ταιριάζει στην προκειμένη περίπτωση είναι ως Decision Table Importer(ODBC).

Συγκεκριμένα τα δεδομένα μας για τον παραπάνω σκοπό είναι τροποποιημένα σε M.Access. Όλες οι εγγραφές(553) περιέχονται λοιπόν στο ros1.mdb για το πρώτο λάδι, μιλώντας. Αφού γίνει η εισαγωγή των στοιχείων του πίνακα μας ανοίγει το παράθυρο του project μας. Αυτό περιέχει σε μορφή λίστας τη βάση δεδομένων και από κάτω τις δυνατότητες του προγράμματος, δηλαδή ποιους αλγορίθμους προσφέρει και πως μπορούμε να χειριστούμε τα δεδομένα μας.



Σχήμα 4η-Βασικό παράθυρο του project(rosetta –lorena)

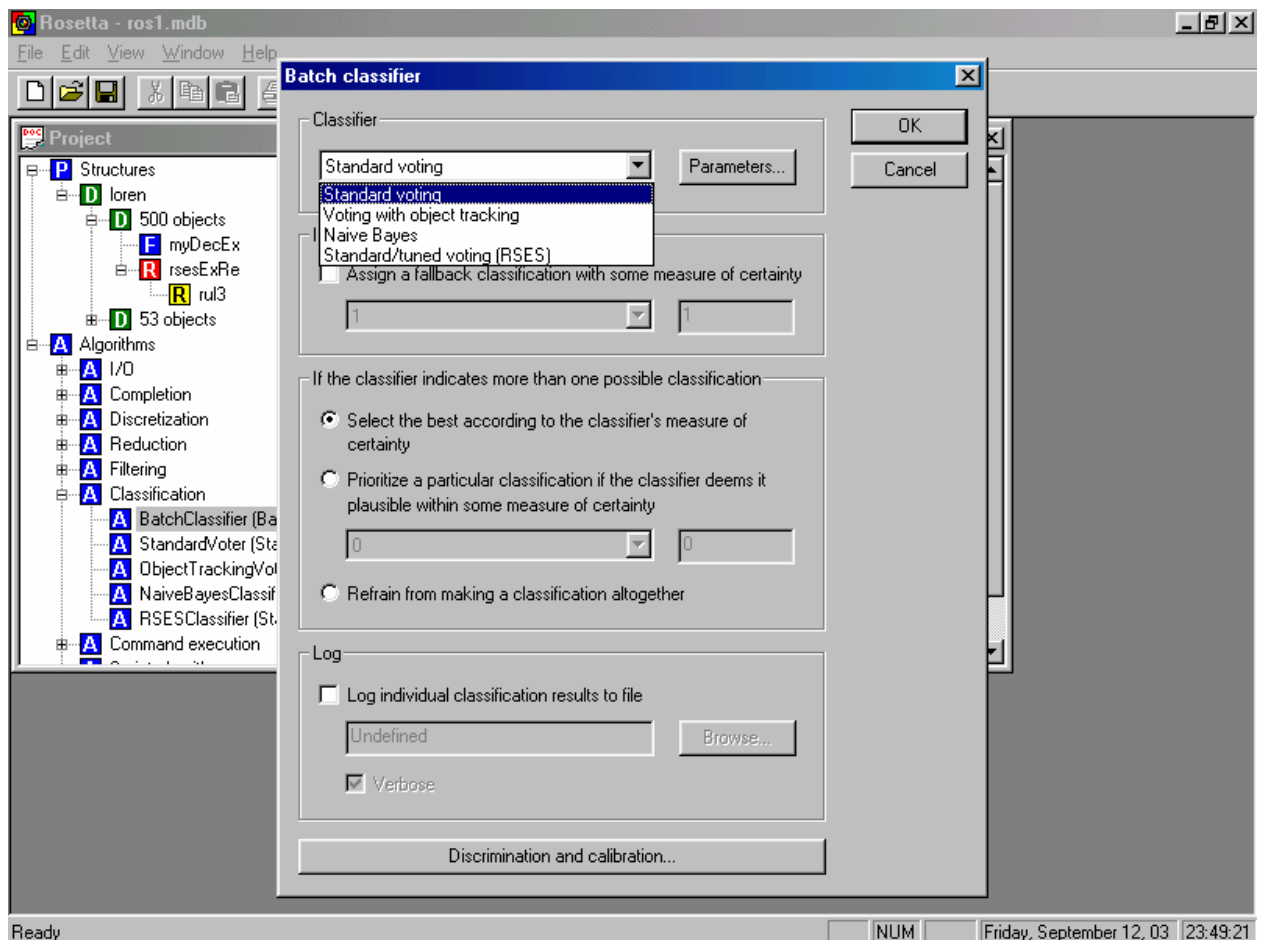
Επειδή σε αυτό το αρχείο(ros1) υπάρχουν όλα τα δεδομένα, θα πρέπει να χωριστούν σε training set και σε prediction set, με σκοπό να παρέχουμε κατά τα γνωστά τη γνώση μέσω της εκπαίδευσης των 500 δεδομένων και τα υπόλοιπα 53 να

τα δώσουμε ως είσοδο στο πρόγραμμα να παράγει τη πρόβλεψη. Το χωρισμό των δεδομένων θα τον κάνει μια εντολή του προγράμματος που λέγεται binary splitter και είναι στο μενού other. Εκεί δηλώνουμε ως συντελεστή 0.9041 που είναι το ποσοστό του 500 στα 553. Έτσι λοιπόν χωρίζεται η βάση σε δύο άλλες 500 και 53 εγγραφών.

Ύστερα επιλέγουμε από το μενού I/O την εντολή MyDecisionTableExporter (plain format) και αυτό γίνεται apply στη βάση δεδομένων μας(των 500^{ων} δεδομένων) και δημιουργεί μια βάση δεδομένων δική του. Απλά εκεί δηλώνεται ότι τα νούμερα μας είναι ακέραιοι και ότι υπάρχουν 6 πεδία και η χρησιμότητά της είναι μόνο βοηθητική για το πρόγραμμα.

Εν συνεχεία από το μενού Reduction επιλέγουμε να κάνουμε εφαρμογή του αλγορίθμου RSESExhaustiveReducer (Exhaustive Calculation(RSES)) πάνω στα 500 δεδομένα μας ούτως ώστε να γίνει η επεξεργασία και η παραγωγή κανόνων, ουσιαστικά, δηλαδή, η εκπαίδευση του συστήματος. Στην ερώτηση της διακριτότητας επιλέγουμε object related ώστε να δημιουργηθούν οι κανόνες και οι σχέσεις μεταξύ όλων των πεδίων. Έτσι δημιουργείται ο πίνακας με τους κανόνες τον οποίο εμείς θα ονομάσουμε rules. Συγκεκριμένα στη βάση αυτή που δουλεύουμε(Iorena) οι κανόνες που προκύπτουν είναι 438.

Τελευταίο βήμα είναι να πάμε στο μενού Classification, όπου επιλέγουμε την εντολή Batch Classifier για να μας δημιουργήσει τα αποτελέσματα της πρόβλεψης πάνω φυσικά στα 53 δεδομένα. Διαλέγουμε ως classifier το Standard Voting για το πρώτο και τέταρτο προϊόν λαδιού και Standard/Tuned Voting(RSES) για το δεύτερο και τρίτο και επιλέγουμε στις παραμέτρους να διαβάσει τους κανόνες που είχαμε δημιουργήσει νωρίτερα και είχαμε ονομάσει rules. Διαλέξαμε διαφορετικό αλγόριθμο classifier μεταξύ των τεσσάρων προϊόντων μας γιατί έτσι λειτούργησε βέλτιστα το πρόγραμμα και είχαμε καλύτερα αποτελέσματα.



Σχήμα 40-Επιλογή classifier

Τελικά αυτό παράγει τον πίνακα με τις τιμές που προβλέφθηκαν αλλά μόνο ως προς το πλήθος και όχι αναλυτικά όπως τα άλλα προγράμματα. Επειδή στην περίπτωση χρήσης αυτού του προγράμματος τα δεδομένα μας δεν συνεργάστηκαν απόλυτα με τον αλγόριθμό του, πήραμε πολλές τιμές είτε απροσδιόριστες είτε μηδενικές. Αυτές, φυσικά δεν μπορούν να μας παρέχουν πληροφόρηση για την πρόβλεψη. Μας ενδιαφέρουν οι τιμές ranking από ένα ως τέσσερα οπότε και αυτές θεωρούμε εξαγόμενες τιμές πρόβλεψης. Έτσι καταλήγουμε να έχουμε πλήθος δεδομένων πρόβλεψης στην περίπτωση του lorena 44, για το ah organic 41, για το sitia 40 και για το bertolli τελικά 39. Έτσι τα ποσοστά θα υπολογιστούν στην προκειμένη περίπτωση αναγόμενα στις παραπάνω τιμές.

Τα σωστά αποτελέσματα στη περίπτωση του πρώτου λαδιού(loreana) είναι 22 στα 44 που σημαίνει αξιοπιστία 50%, με αντίστοιχο πίνακα συγκεντρωτικών αποτελεσμάτων αυτόν στη συνέχεια:

		predicted			
actual	lorena	1	2	3	4
	1	21	4	2	0
	2	8	1	1	0
	3	3	1	0	0
	4	2	1	0	0

Πίνακας 4.9-Σφάλματα rosetta για το lorena

Αναλυτικά είναι:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 21 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 1 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 0 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 0 εγγραφές δεδομένων.

Άρα συνολικά η αξιοπιστία μας εδώ είναι 50%. Βλέπουμε όμως πως έχει γίνει συχνό λάθος η περίπτωση να είναι 2 η πραγματική τιμή του rank και να προβλεφθεί 1 (σφάλμα υπερεκτίμησης), όπως και ότι η πρόβλεψη για την τιμή 4 δεν έχει λειτουργήσει καθόλου. Τα υπόλοιπα λάθη είναι λίγα και με μικρή απόκλιση λάθους.

Ομοίως αν εφαρμόσουμε και τις υπόλοιπες βάσεις δεδομένων στο rosetta ακολουθώντας ακριβώς τα ίδια βήματα τότε:

Στην περίπτωση του ah organic θα προκύψουν 19 σωστά προβλεπόμενες τιμές στις 41 εγγραφές δηλαδή επιτυχία 46,3%, μέσω 501 κανόνων. Βλέπουμε τον αντίστοιχο πίνακα:

		predicted			
actual	ah-organic	1	2	3	4
	1	1	3	1	0
	2	1	6	5	1
	3	0	2	5	3
	4	1	1	4	7

Πίνακας 4.10-Σφάλματα rosetta για το ah organic

Αναλυτικά θα είναι:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 1 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 6 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 5 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 7 εγγραφές δεδομένων.

Εδώ η αξιοπιστία της πρόβλεψης είναι χαμηλότερη από πριν αλλά δεν απέχει πολύ(46,3%). Λάθη έχουν γίνει 4 και 5 φορές στην περίπτωση (4, 3) και (2, 3) αντίστοιχα που κι αυτές είναι μικρές αποκλίσεις. Δηλαδή είχαμε καλή συμπεριφορά του rosetta στα μέχρι εδώ δεδομένα μας.

Για το τρίτο λάδι, το sitia προκύπτουν 16 επιτυχείς προβλέψεις δηλαδή 40% μέσω 576 κανόνων. Ο επόμενος πίνακας θα μας δια φωτίσει:

		predicted			
actual	sitia	1	2	3	4
	1	1	1	0	1
	2	0	2	2	6
	3	0	0	3	7
	4	0	2	5	10

Πίνακας 4.11-Σφάλματα rosetta για το sitia

Λεπτομερειακά είναι:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 1 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 2 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 3 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 10 εγγραφές δεδομένων.

Οι σωστά προβλεπόμενες τιμές αυτή τη φορά είναι 16 στις 40 εγγραφές τώρα οπότε το ποσοστό μας κατά τα γνωστά βρίσκεται στο 40%. Συνήθη λάθη βρίσκουμε στις περιπτώσεις, όπου έπρεπε να προβλεφθεί τιμή 2 ή 3 ενώ προβλέφθηκε 4 με 6 και 7 περιπτώσεις αντίστοιχα.

Και τέλος για το bertolli έχουμε 13 σωστές προβλέψεις στα 39 δεδομένα άρα το ποσοστό αξιόπιστης πρόβλεψης είναι στα πιο χαμηλά του επίπεδα στο 33%,μέσω 547 κανόνων, και είναι σαφώς χαμηλότερη από τις άλλες. Ας δούμε όμως:

		predicted				
		bertolli	1	2	3	4
actual	1	5	2	2	1	
	2	5	3	0	2	
	3	3	6	4	1	
	4	0	2	2	1	

Πίνακας 5.12-Σφάλματα rosetta για το bertolli

Το παραπάνω μας λέει:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 5 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 3 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 4 εγγραφές δεδομένων.

- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 1 εγγραφές δεδομένων.

Τα σωστά αποτελέσματα, όπως βλέπουμε, είναι στην τελευταία μας περίπτωση 13 σε πλήθος 39 τιμών πρόβλεψης, άρα εδώ το ποσοστό επιτυχίας της μεθόδου είναι 33%. Γενικά σε αυτή τη πρόβλεψη έχουν γίνει πολλά λάθη εκ των οποίων τα σημαντικότερα είναι στις περιπτώσεις (2, 1) με 5 περιπτώσεις και (3, 2) με 6 περιπτώσεις.

Συνολικά το rosetta μπόρεσε να μεταχειριστεί τα δεδομένα αρκετά καλά και να επεξεργαστεί τη πληροφορία που αυτά προσφέρουν για να κάνει τη ζητούμενη πρόβλεψη. Τα αποτελέσματα του ξεπέρασαν το neurosolutions αλλά όχι και του wizwhy .

4.5 Εφαρμογή Utadis

Επειδή, όπως αποδείχτηκε στη πράξη, υπήρχαν κάποια δεδομένα, τα οποία ενδεχομένως έρχονταν σε αντίθεση ως προς την πληροφορία που παρείχαν και εμπόδιζαν τον αλγόριθμο του Utadis να λειτουργήσει σωστά, αναγκαστήκαμε να απομακρύνουμε αυτές τις προβληματικές εγγραφές για να προκύψει η ζητούμενη πρόβλεψη. Χρησιμοποιήθηκε, δηλαδή, μικρότερος όγκος πληροφόρησης ενώ το δείγμα του prediction set μεγάλωσε. Έτσι το Utadis έκανε testing σε περισσότερα δεδομένα ώστε να μας δώσει τις τιμές της πρόβλεψης του ranking. Συγκεκριμένα το test set για τα τέσσερα προϊόντα είχε μέγεθος 102 δεδομένων για το lorena και το ah organic και 100, 101 για το sitia και το bertolli αντίστοιχα. Ας ξεκινήσουμε να δούμε τι αποτελέσματα πήραμε από τη μέθοδο αυτή.

Όσον αφορά το lorena συνολικά χρησιμοποιήθηκαν 509 δεδομένα εκ των οποίων 407 αποτέλεσαν το πακέτο εκπαίδευσης του προγράμματος, 102 δόθηκαν ως input στο σύστημα για επεξεργασία πρόβλεψης και τελικά πήραμε 40 σωστά αποτελέσματα. Συνολικά η ακρίβεια που πήραμε είναι 39,54% και τελικά προέκυψε:

		predicted			
actual	lorena	1	2	3	4
	1	10	24	17	0
	2	1	11	9	1
	3	1	2	12	2
	4	1	3	7	4

Πίνακας 4.13-Σφάλματα Utadis για το lorena

Δείχνει αναλυτικότερα:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 10 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 11 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 12 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 4 εγγραφές δεδομένων.

Ενώ η δεύτερη και η τρίτη γραμμή έχουν καλό «σκορ», δηλαδή υψηλά ποσοστά επιτυχίας, οι αποκλίσεις στην περίπτωση (1, 2), (1, 3) είναι 24, 17 αντίστοιχα(σφάλματα υποεκτίμησης), τιμές μεγάλες για την έρευνά μας.

Επόμενο λάδι είναι το ah organic , όπου χρησιμοποιήθηκαν 508 δεδομένα συνολικά με τα 102 να είναι το prediction set. Συνολικά, το πρόγραμμά μας, έδωσε επιτυχία σ' αυτή τη περίπτωση 37,38%. Ας δούμε επιμέρους:

		predicted			
actual	ah-organic	1	2	3	4
	1	8	9	3	0
	2	2	16	7	1
	3	3	10	12	0
	4	3	8	20	0

Πίνακας 4.14-Σφάλματα Utadis για το ah organic

Φαίνεται ότι:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 8 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 16 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 12 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 0 εγγραφές δεδομένων.

Στη περίπτωση αυτή τα κελιά (2, 2) και (3, 3) έχουν ισχυρή θέση, με 16 και 12 επιτυχίες αντίστοιχα, ενώ σχετικά καλά κινείται και το (1, 1), όμως στη περίπτωση που έπρεπε να προβλεφθεί η τιμή 4 το σύστημα απέτυχε τελείως και επέλεξε τη τιμή 3 σε 20 εγγραφές. Βέβαια είναι μικρή η απόκλιση της τιμής αλλά δεν παύει να είναι λανθασμένη εκτίμηση που δεν μας δίνει καμιά πληροφόρηση.

Τρίτο λάδι αποτελεί το sitia. Εδώ έχει χρησιμοποιηθεί πακέτο 500 εγγραφών. Το ελληνικό λάδι είχε τελικά αξιοπιστία στη πρόβλεψή του 35,23%, ποσοστό που σημαίνει πως από τα 100 δεδομένα που δόθηκαν για να προκύψει υπολογιστική τιμή βρέθηκαν σωστά τα 35. Παρακάτω είναι ο αντίστοιχος πίνακας:

		predicted			
actual	sitia	1	2	3	4
	1	6	0	3	1
	2	5	0	8	3
	3	6	0	14	6
	4	6	0	29	13

Πίνακας 4.15-Σφάλματα Utadis για το sitia

Πιο αναλυτικά μας δείχνει:

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 6 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 0 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 14 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 13 εγγραφές δεδομένων.

Τα αποτελέσματα κινούνται όπως και στην προηγούμενη περίπτωση με υψηλές τιμές στις περιπτώσεις (1, 1), (3, 3) αλλά πάλι το σύστημα δε μπόρεσε να λειτουργήσει για να προβλέψει τη τιμή 2 και το πρόβλημα ήταν γενικότερο είτε αφορούσε σωστή είτε εσφαλμένη τιμή για τη τιμή 2, όλη η στήλη 2 έχει μηδενικά.

Τελευταίο και από άποψη κατάταξης και από άποψη σωστής πρόβλεψης είναι το bertolli. Με συνολικά 505 δεδομένα να αποτελούν και το prediction και το training set, η εφαρμογή μας έδωσε 25 δεδομένα να συμβαδίζουν με τις πραγματικές τιμές σε πλήθος 101 δεδομένων. Άρα επιτύχαμε 24,22% ακρίβεια. Ο πίνακας συγκεντρώνει τα αποτελέσματα:

		predicted			
actual	bertolli	1	2	3	4
	1	0	0	23	0
	2	1	0	27	0
	3	1	0	25	0
	4	1	0	17	0

Πίνακας 4.16-Σφάλματα Utadis για το bertolli

- Η πραγματική και η υπολογιστική τιμή να είναι 1, προέκυψε για 0 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 2, προέκυψε για 0 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 3, προέκυψε για 25 εγγραφές δεδομένων.
- Η πραγματική και η υπολογιστική τιμή να είναι 4, προέκυψε για 0 εγγραφές δεδομένων.

Παρόλο που σε 25 περιπτώσεις ορθώς προβλέφθηκε 3, εντούτοις αποτελεί μια πολύ κακή πρόβλεψη συνολικά αφού στις υπόλοιπες περιπτώσεις έχουμε μηδενική αξιοπιστία. Συγκεκριμένα και το συνολικό ποσοστό 24,22% είναι τελείως εικονικό γιατί προέρχεται από μόνο μία περίπτωση πρόβλεψης και σημαίνει ότι το πρόγραμμα αδυνατεί να λειτουργήσει με τα συγκεκριμένα δεδομένα. Αυτό γίνεται φανερό από όλη τη λειτουργία του Utadis, που δεν μπορούσε να δώσει ομοιόμορφα αποτελέσματα σε όλες τις πιθανές τιμές παρά μόνο αποσπασματικά σε δυο τρεις περιπτώσεις. Αυτό δείχνει ότι δεν μπορούμε πολύ περισσότερο από τα άλλα να το εμπιστευθούμε στη πρόβλεψή του.

ΚΕΦΑΛΑΙΟ 5^ο

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο παρόν κεφάλαιο θα σχολιάσουμε τα αποτελέσματα της πρόβλεψης, ως προς την αξιοπιστία τους, και θα γίνουν συγκρίσεις μεταξύ των συστημάτων που χρησιμοποιήσαμε για να καταλήξουμε ποιο από όλα συμπεριφέρθηκε καλύτερα πάνω

5.1 Συμπεράσματα

Τα τέσσερα προγράμματα, που χρησιμοποιήσαμε, σκοπό είχαν, όπως έχουμε προαναφέρει, να λειτουργήσουν με είσοδο τα δεδομένα μας και να παράγουν τη ζητούμενη πρόβλεψη ώστε να αποφανθούμε στο θέμα της αξιοπιστίας της πρόβλεψής τους, πάντα με γνώμονα τη δική μας βάση δεδομένων.

Ήταν φανερό από την πρώτη κιόλας επαφή με τα δεδομένα πως δεν αποτελούν άριστη πηγή πληροφοριών. Δυστυχώς είχαν πολλές ατέλειες, πολλά κενά και ενδεχομένως αρκετές φορές αλληλοσυγκρουόμενα «γεγονότα» τύπου: κατάταξη ενός λαδιού στη πρώτη θέση αλλά χαμηλή αξιολόγηση των παραμέτρων του λαδιού αυτού. Αυτό αποτελεί αντιφατικό γεγονός στη ανθρώπινη λογική και κατά προέκταση και στην υπολογιστική λογική. Δηλαδή ενώ το εκάστοτε εργαλείο λογισμικού εκπαιδεύεται από μια αλληλουχία παραμέτρων και συσχετίσεων, παρουσιάζονται κάποιες τέτοιες μεμονωμένες περιπτώσεις και αποσυντονίζουν την ανοικοδόμηση της εμπειρίας του.

Παρόμοιο με την παραπάνω κατάσταση είναι, σε πολλές περιπτώσεις προγραμμάτων, αυτά να μπερδεύονται την ώρα του training, λόγω παραπλήσιων κανόνων. Δηλαδή πάνω στη δημιουργία κανόνων το πρόγραμμα εξετάζει κάθε εγγραφή ξεχωριστά προσπαθώντας να μοντελοποιήσει τη διαδικασία επιλογής της υπο εξέτασης ποσότητας. Όταν λοιπόν κάποιες εγγραφές έχουν ακριβώς τα ίδια δεδομένα και αλλάζει μόνο η τιμή του prediction field τότε το πρόγραμμα ενδεχομένως να καταχωρήσει τη μία από τις δύο εγγραφές στο πλαίσιο δημιουργίας κανόνων και όταν εμφανιστεί η άλλη να δημιουργηθεί σύγχυση στο μοντέλο μας.

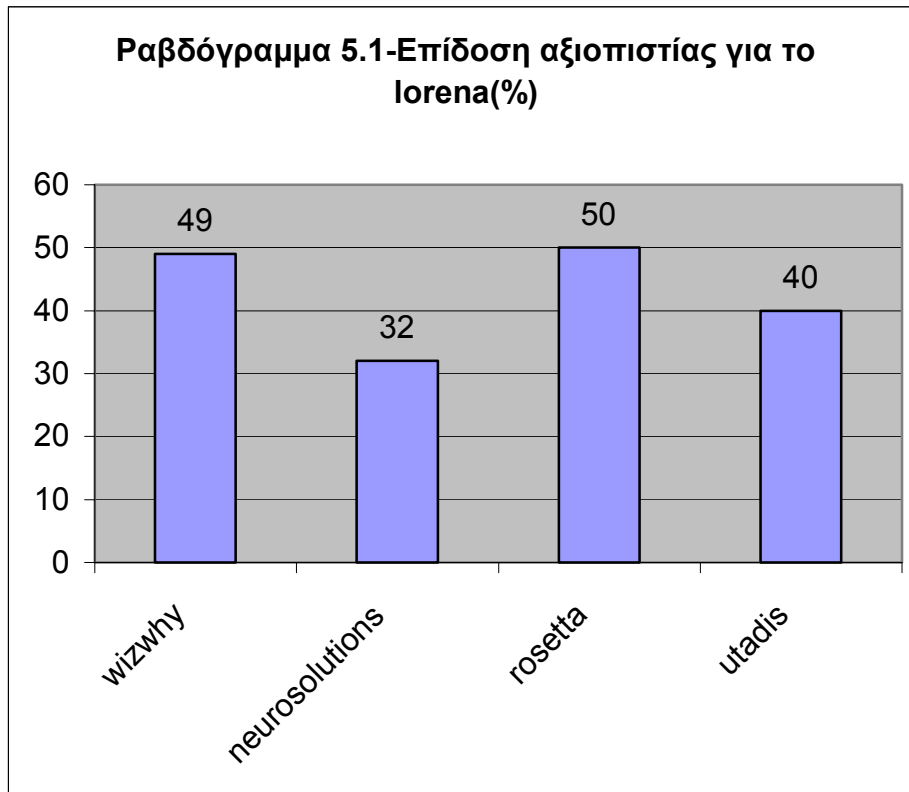
Σημαντικό ρόλο πάντα στην εκπαίδευση ενός μοντέλου τεχνητής νοημοσύνης είναι να δοθεί πλήρης πληροφορία μέσω των δεδομένων εκπαίδευσης. Μια σημαντική κίνηση προς αυτή την κατεύθυνση είναι να ανακατεύονται οι εγγραφές των δεδομένων μας. Πρέπει δηλαδή να χρησιμοποιούνται όλα τα δεδομένα που έχουμε στη διάθεσή μας, είτε αυτά ανήκουν στο prediction είτε στο training set, να ανακατεύονται με τυχαίο τρόπο και το πρόγραμμα να τα σαρώνει αρκετές φορές (πραγματοποιεί training) ώστε να αντλήσει όση περισσότερη ποσότητα πληροφορίας μπορεί.

Λόγω της ιδιαίτερης ιδιομορφίας κάθε προγράμματος λογισμικού που χρησιμοποιήσαμε είναι λογικό να μην είναι απόλυτα όμοιες οι συνθήκες εφαρμογής τους. Μερικά είχαν περιορισμούς λόγω περιορισμένης έκδοσης, άλλα είχαν επιπρόσθετες δυνατότητες που βοηθούσαν σημαντικά τον χρήστη, ή πολλές φορές διέφεραν στον τρόπο παρουσίασης των αποτελεσμάτων τους και της ανάλυσής τους. Αυτό δημιουργούσε εξ ορισμού μια ιδιαιτερότητα στη μελέτη τους, αφού έπρεπε όλα να αναχθούν σε κοινή βάση.

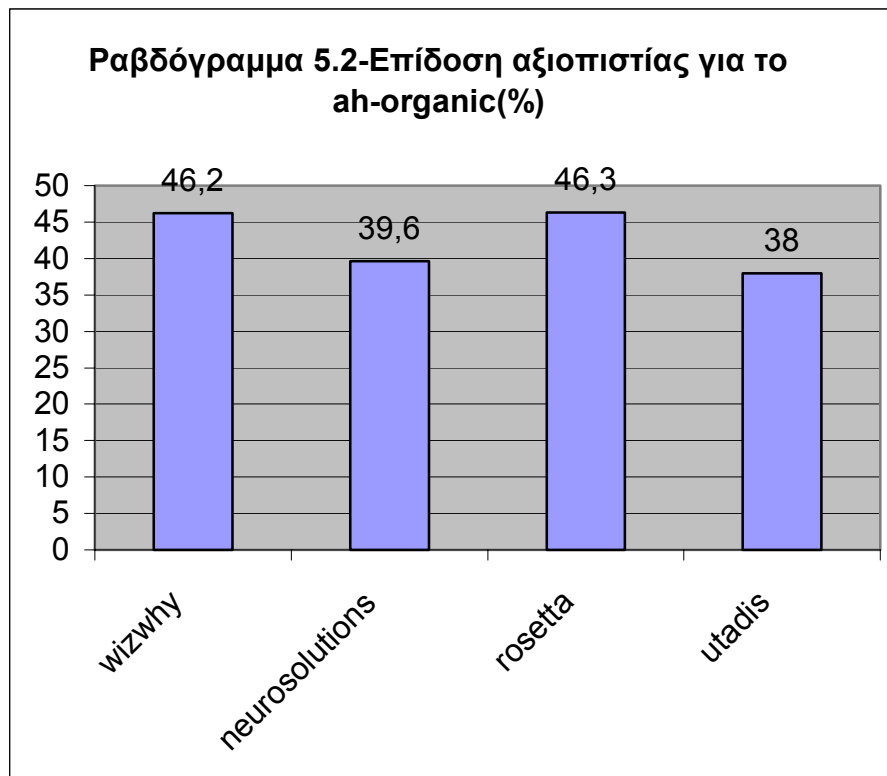
Υπό το πρίσμα όλων των παραπάνω παρατηρήσεων μπορούμε πλέον να δούμε συνολικά τη λειτουργία των προγραμμάτων που χρησιμοποιήσαμε, μέσω των δεδομένων μας. Καταρχήν θα εξετάσουμε μεμονωμένα κάθε μάρκα λαδιού για να δούμε πως συμπεριφέρθηκε το κάθε εργαλείο επί ίσης βάσης και ύστερα θα δούμε το θέμα καθολικά.

1. Όσον αφορά το lorena μπορούμε να πούμε πως την καλύτερη αξιοποίηση της βάσης δεδομένων του την έκανε το rosetta με επιτυχία 50%. Ίσως η φιλοσοφία των rough sets λειτούργησε καλύτερα από κάθε άλλη μέθοδο και

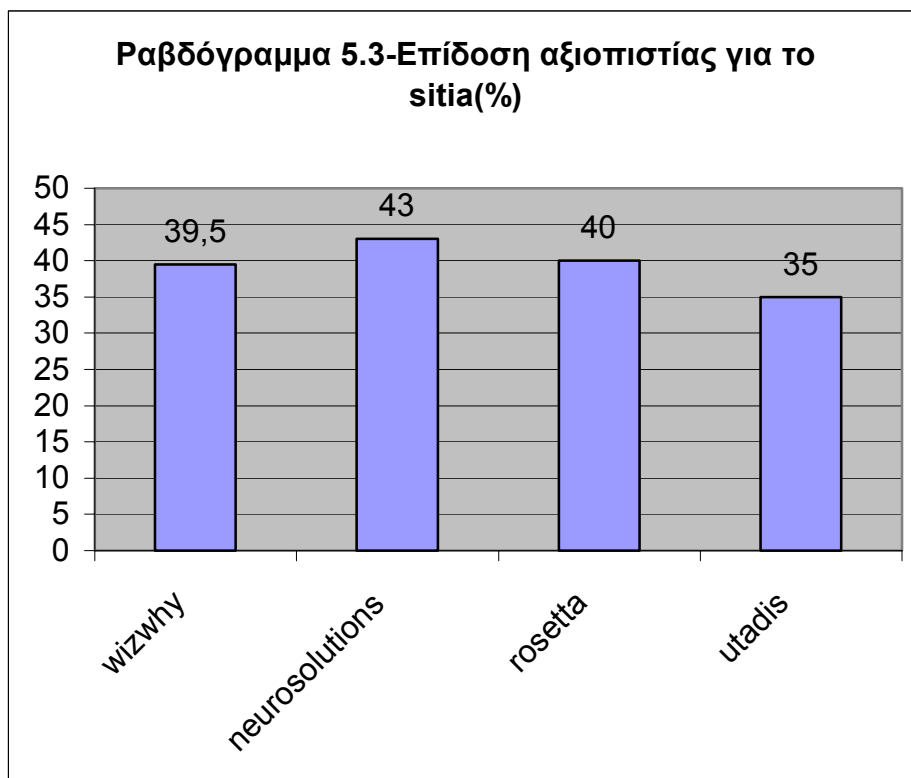
μας έδωσε τη δυνατότητα ανα δύο εγγραφές να έχουμε μια σωστή πρόβλεψη. Πολύ κοντά κινήθηκε και το wizwhy με «σκορ» 49% και ακολούθησαν τα Utadis και neurosolutions με 40% και 32% αντιστοίχως. Πάντως γενικά τα δεδομένα του lorena ήταν αξιόπιστα σε πληροφόρηση και σχετικά ποιοτικά για εκπαίδευση γιατί έδωσαν καλά ποσοστά επιτυχίας, συγκριτικά.



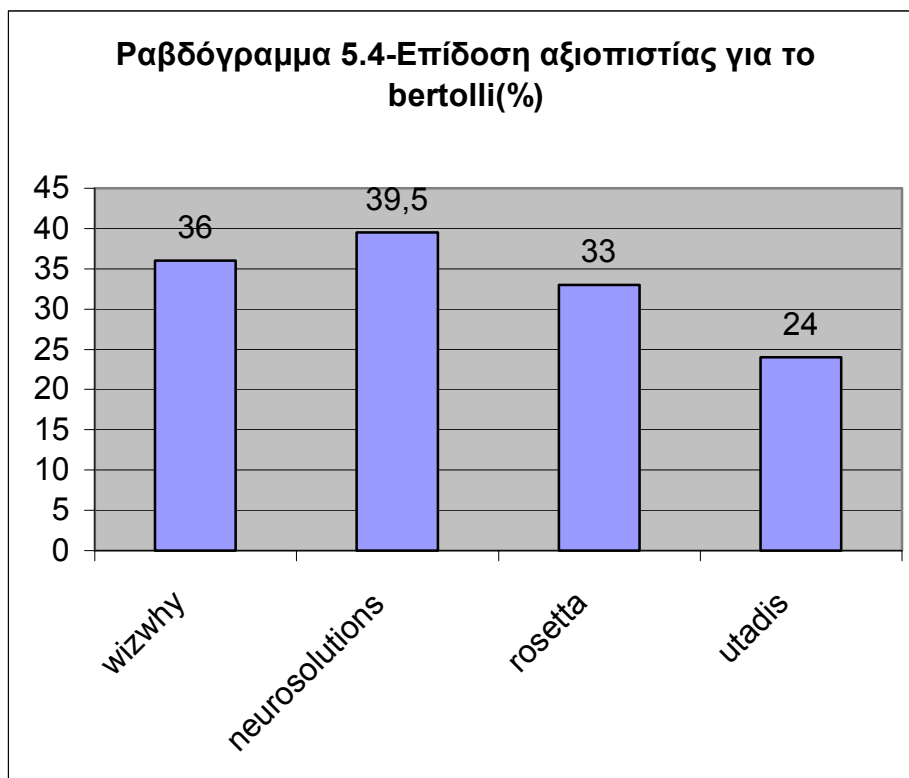
2. Δεύτερη βάση ήταν τα δεδομένα που αφορούσαν το ah organic. Σε αυτό τη καλύτερη πρόβλεψη έδωσε το wizwhy και το rosetta με σχεδόν ίδια ποσοστά επιτυχίας 46%. Η λογική του data mining εδώ λειτούργησε με τον καλύτερο δυνατό τρόπο και μπόρεσε να πλησιάσει το 50% αξιόπιστης πρόβλεψης. Ακολούθησαν με τη σειρά τα neurosolutions και Utadis με ακρίβεια 39,6% και 38% αντίστοιχα.



3. Το τρίτο πακέτο εγγραφών είναι αυτό που αφορά το ελληνικό λάδι, το sitia. Εδώ καλύτερα αποφάνθηκε το neurosolutions, θέτοντας ως καλύτερη τη μοντελοποίηση μέσω νευρωνικών δικτύων. Επέτυχε αξιοπιστία 43%, έχοντας τα επίπεδα πρόβλεψης του rosetta και του wizwhy σε απόσταση αναπνοής να ακολουθούν με 40% και 39,5% . Τελευταίο είναι το Utadis όπου επέτυχε ακρίβεια 35% περίπου.



4. Τέλος για το bertolli δυστυχώς γενικά δεν μπορέσαμε να επιτύχουμε πολύ υψηλά επίπεδα πρόβλεψης τουλάχιστον αντίστοιχης με τα προηγούμενα με κανένα από τα εργαλεία μας. Το καλύτερο που μπορέσαμε να πάρουμε είναι 39,5% ακρίβεια μέσω του neurosolutions , όπου και πάλι τα νευρωνικά δίκτυα επικράτησαν. Έπειτα ακολούθησαν το wizwhy, το rosetta και το Utadis με ποσοστά επιτυχίας στη πρόβλεψη 36%, 33% και 24% αντίστοιχα. Περισσότερο να τονίσουμε πως το επίπεδο που έφτασε το Utadis είναι πολύ χαμηλό και δεν αποτελεί αξιόπιστη πληροφόρηση. Ίσως αυτή η τελευταία βάση που περιέγραφε το bertolli να είχε περισσότερα «προβληματικά» δεδομένα γιατί γενικά δεν μπόρεσε να την επεξεργασθούν τόσο καλά τα πακέτα λογισμικού μας ώστε να παρέχουν πολύ ικανοποιητικά αποτελέσματα.



Μια επιπλέον αναφορά μπορεί να γίνει για τη «συμπεριφορά» γενικά των προγραμμάτων που χρησιμοποιήσαμε. Μπορούμε να πούμε πως σύμφωνα με τα παραπάνω την καλύτερη προσαρμοστικότητα στα δεδομένα μας έδειξε το WIZWHY με συνολική ακρίβεια και στα τέσσερα προϊόντα λαδιού 42,65%. Το επόμενο πρόγραμμα που ήταν σχεδόν εξίσου ακριβές στην πρόβλεψή του είναι το ROSETTA με συνολικό «δείκτη επιτυχίας» 42,32%, ενώ την τρίτη θέση κατέχει το λογισμικό με χρήση νευρωνικών δικτύων, δηλαδή το NEUROSOLUTIONS με ακρίβεια 38,65% και τελευταίο έρχεται το UTADIS που δεν μπόρεσε να λειτουργήσει πολύ καλά και παρουσίασε ποσοστό αξιοπιστίας μόλις 34,1%.

Όλα τα παραπάνω μπορούν να συγκεντρωθούν σε έναν πίνακα διπλής εισόδου όπου φαίνεται καθαρά η επίδοση του κάθε προγράμματος που χρησιμοποιήσαμε αναφορικά με τις τέσσερις επωνυμίες ελαιολάδου.

Ποσοστά επιτυχίας	WIZWHY	NEUROSOLUTIONS	ROSETTA	UTADIS
LORENA	49,00%	32,00%	50,00%	39,50%
AH-ORGANIC	46,20%	39,60%	46,30%	37,40%
SITIA	39,60%	43,40%	40,00%	35,20%
BERTOLLI	35,80%	39,60%	33,00%	24,20%
συνολικά:	42,65%	38,65%	42,32%	34,10%

Πίνακας 5.1-Συγκεντρωτικά αποτελέσματα προγράμματος ανά προϊόν

Είναι προφανές γενικά πως τα ποσοστά που έχουμε στη διάθεσή μας δεν ικανοποιούν ιδιαίτερα τον σκοπό που μας ενδιαφέρει, δηλαδή την πρόβλεψη της συμπεριφοράς καταναλωτών. Σκοπός είναι να εξάγουμε κάποια συμπεράσματα για το πώς θα αντιδράσουν οι καταναλωτές(τι απόφαση θα πάρουν) αν κληθούν να επιλέξουν ελαιόλαδο. Σύμφωνα με την έρευνα αγοράς που εν δυνάμει μπορούμε να κάνουμε, κάποια γενικά χαρακτηριστικά τους ή κάποιες απαντήσεις τους σε δεδομένες ερωτήσεις μπορούν να μας βοηθήσουν να αποφανθούμε για την επιλογή τους, η οποία θα αποτελέσει και τη βάση για τη διαμόρφωση των μεριδίων αγοράς. «Πόσα άτομα θα επιλέξουν το αντίστοιχο λάδι; Γιατί; Σε τι ποσοστό της καταναλωτικής δύναμης αυτοί αντιστοιχούν; ». Αυτά απαντώνται αξιοποιώντας τις προβλέψεις που μας δίνουν τα προγράμματα αν είμαστε σε θέση να κρίνουμε κατά πόσο τις εμπιστευόμαστε.

Ανάλογα με την έρευνα και την εταιρία, που χρηματοδοτεί το όλο πρόγραμμα, επιλέγονται κάποια κατώφλια εμπιστοσύνης της πρόβλεψης. Αυτά υπάρχουν για να υπολογίζουμε την πιθανότητα στη χειρότερη περίπτωση αυτό που προσδωκούμε να μην γίνει. Στην συγκεκριμένη περίπτωση λόγω κακής ποιότητας δεδομένων τα προϊόντα λογισμικού μας έδωσαν τα αποτελέσματα που προαναφέρθηκαν και τα οποία συγκεντρωτικά παρουσιάστηκαν στον πίνακα 5.1. Είναι σχετικά χαμηλά το ποσοστό επιτυχίας όλων των μεθόδων και το κατώφλι εμπιστοσύνης θα πρέπει να επιλεγεί πιο κάτω και από 50% για να μπορεί κάποια εταιρία να κάνει προγραμματισμό παραγωγής ή διάθεσης προϊόντων ελαιολάδου. Το ρίσκο, δυστυχώς, βρίσκεται σε υψηλά επίπεδα, κάνοντας ριψοκίνδυνη την οποιαδήποτε απόφαση από πλευράς ενδιαφερομένου(ενδεχομένως εταιρία).

Στην περίπτωση που επιλέξουμε να εμπιστευτούμε τα αποτελέσματα της πρόβλεψης οποιουδήποτε εργαλείου, είναι φανερό πως πρώτη επιλογή για τους καταναλωτές αποτελεί το lorena. Σε αυτή την περίπτωση, η ενδιαφερόμενη εταιρία διάθεσης ελαιολάδου, μπορεί να αποφασίσει την εισαγωγή, ενδεχομένως, μεγαλύτερης ποσότητας lorena από τα υπόλοιπα τρία. Αυτό θα είναι το προϊόν που οι καταναλωτές γνωρίζουν, θα αναζητήσουν και τελικά θα αγοράσουν περισσότερο. Άρα αυτό θα αποτελεί και το μεγαλύτερο αγοραστικό μερίδιο στη συγκεκριμένη αγορά και θα αποφέρει το μεγαλύτερο κέρδος στην εταιρία.

Μια καλή κίνηση που θα μπορούσε να κάνει η παραπάνω εταιρία είναι να πλησιάσει τους καταναλωτές που προτιμούν το lorena, αφού τα στοιχεία τους είναι στη διάθεσή της, να φροντίσει για τη μεταπωλητική ικανοποίησή τους και να στραφεί στη προώθηση του προϊόντος και σε άλλα άτομα του περιβάλλοντός τους,

επεκτείνοντας το καταναλωτικό κοινό της. Άλλωστε η πρώτη μορφή διαφήμισης, ενδεχομένως, θα έχει γίνει από τους ικανοποιημένους πελάτες, μέσω της μεθόδου «από στόμα σε στόμα».

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Γεώργιος Ι.Δουκίδης, Μάριος Κ.Αγγελίδης , «Έμπειρα συστήματα, Τεχνητή Νοημοσύνη και LISP»,Ι.Σιδέρης, 1998
2. Θ. Μοδης, «Προβλέψεις, προσεγγίζοντας επιστημονικά τα προμηνύματα του αύριο», Πανεπιστημιακές εκδόσεις Κρήτης, 1996.
3. “Wizwhy, User’s guide”, Wizsoft, 1996
4. Α. Λύκας, «Υπολογιστική νοημοσύνη», Σημειώσεις , Πανεπιστήμιο Ιωαννίνων, 1999
5. Aleksander Ohrn, “ROSETTA Technical reference manual”, 2001
6. NeuroDimension, “Manual of neurosolutions”, 2003
7. Νικ.Φ.Ματσατσίνης, «Συστήματα Υποστήριξης Αποφάσεων», σημειώσεις μαθήματος.
8. Doumpos, M. and Zopounidis, C., «Multicriteria Decision Aid Classification Methods», Kluwer Academic Publishers, Dordrecht), (2002)
9. <http://www.wizsoft.com>
10. <http://www.neurosolutions.com>
11. <http://www.idi.ntnu.no/-aleks/rosetta/>

12. http://www.idss.cs.put.poznan.pl/research/rough_sets/main.html
13. <http://sound.eti.pg.gda.pl/SRS/rough.html>
14. Taisuke Sato, Yoshitaka Kameya, “Parameter learning of logic programs for symbolic-statistical modelling”, *Journal of AI Research* 15, 391-454, 2001
15. Sander M. Bohte, Enrico Gerding, Han La Poutré , “Market-based Recommendation: Agents that Compete for Consumer Attention”, *ACM Transactions on Internet Technology*, Vol. V, No. N, Month 20YY.
16. Charu Chandra, Armen Tumanyan “Supply Chain Reconfiguration: Designing Information Support With System Taxonomy Principles”, 2002.
17. Del.I.Hawkins, Roger G.Best, Kenneth A.Coney, “Consumer Behavior-Building Marketing Strategy”, McGraw-Hill Companies, eight edition, 2001.
18. Carl McDaniel and Roger Gates, “Marketing Research”, John Wiley and Sons Inc., 2002.
19. Philip Cotler, “Marketing Management”, Pearson Education Inc., 2003.
20. Γρηγορούδης Β. και Γ. Σίσκος, «Ποιότητα υπηρεσιών και μέτρηση ικανοποίησης του πελάτη», Εκδόσεις Νέων Τεχνολογιών, Αθήνα, 2000.

ΠΑΡΑΡΤΗΜΑ

A. ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

QUESTIONNAIRE OLIVE OIL MARKET RESEARCH IN THE NETHERLANDS

City Date /.....
/.....
Interviewer :

1. What other types of oil do you purchase (or use) ? What do you use it for?

Product	Usage					
	Frying	Marinades	Sauces	Salads	Baking	Other (Define)
Extra virgin olive oil						
Olive oil						
Other Vegetable oils (Soya oil, Corn-oil, Sunflower oil etc.)						

2. Which kinds of oils do you use and how frequently?

Frequency

Product	3-4 times per week (or more)	Once in two weeks	Once in a month or less
Extra virgin olive oil			
Olive oil			
Other Vegetable oils (Soya oil, Corn-oil, , Sunflower oil etc.)			

3. From where do you usually buy olive oil?

Supermarket

Delicatessens

Healthfood

Bakery/Butcher/Wine -Cave

Other

4. Do you buy a particular brand of olive oil?

Yes No

If yes which one?

5. Do you know which country the Olive oil you purchase comes from?

Yes No

If yes which one?

6. How important do you consider the following factors when you choose olive oil ?

	Very Important	Fairly Important	Neutral	Not particularly Important	Not at all Important	I don't know
	1	2	3	4	5	
• Price						
• Health						
• Flavour						
• Color						
• Bottle						
• Odour						
• Label						
• Brand Name						
• Naturalness						
• Country of Origin						
• Special offers / reward points.						

7. What type of packaging do you prefer when you are considering buying olive oil?

Plastic

Round

1 lit.

Glass

Semi-Round
Square

750 ml

500 ml

Other
(what)

Other
(what)

250 ml

Other
(what)

8. What is your opinion about these olive oils according to the factors below?

Products			Lorena 750 ml	AH Organic 500 ml	Sitia 500 ml	Bertolli 500 ml
Color (a)	Very good	1				
	Good	2				
	Indifferent/Neutral	3				
	Bad	4				
	Very Bad	5				
	I do not know					
Package (b)	Very good	1				
	Good	2				
	Indifferent/Neutral	3				
	Bad	4				
	Very Bad	5				
	I do not know					
Image Reputation (c)	Very good	1				
	Good	2				
	Indifferent/Neutral	3				
	Bad	4				
	Very Bad	5				
	I do not know					
Quality (d)	Very good	1				
	Good	2				
	Indifferent/Neutral	3				
	Bad	4				
	Very Bad	5				
	I do not know					
Price (e)	Very good	1				
	Good	2				
	Indifferent/Neutral	3				
	Bad	4				
	Very Bad	5				
	I do not know					
Product	Price		4.95	9.95	7.95	7.95

- a) Which is your opinion about the color ofolive oil ?
- b) Which is your opinion about the packaging of olive oil ?
- c) Which is your opinion about the brand name olive oil ?
- d) Which is your opinion about the quality of olive oil ?

e) How would you describe the prices of olive oil ?

9. Put the following products in numerical order according to your preferences; with 1 being the best?

<div>Order</div>	Products			
	Lorena 750 ml	AH Organic 500 ml	Sitia 500 ml	Bertolli 500 ml

10. Do you ever purchase organic produce ?

Yes

No

11. Have you ever bought organic olive oil ?

Never

Once

More than once

(If the respondent has never bought organic olive oil skip the following question)

12. Do you think there would be any advantages in purchasing organic olive oil ? :

- Better Quality
- Chemical free .
- Environmentally friendly
- Support Organic Philosophy
- Don't know
- Other (please state)

13. Do you think there would be any dis-advantages in purchasing organic olive oil ?

:

- High price
- Not trustworthy
- Doesn't interest me
- Difficult to find it .
- Don't know
- Other (please state)

14 .If organic olive oil was readily accessible would you ever consider buying it?

Yes No
If no go to question number: 19

15 . Where would you prefer to buy organic olive oil from ?

Supermarket	Open market
Delicatessens	Home
Healthfood/ Reformhouse	Other (please state)

16 . Say that the olive oil you currently purchase costs 8.95 for 750ml . Are you prepared to pay a higher price for the same quantity of organic olive oil ? Yes
No

if the answer is «yes», go
to 17 C

if the answer is «no» , go
to 17 A

A. Would you pay the same price as for the olive oil you currently purchase
Yes No (if the answer is «no» , then :)

B. Would you pay a lower price (than for the olive oil) you currently purchase
Yes No

C. Would you pay:
9.95
11.45
13.95
More than 13.95
Other (please state).....

17. What would you use organic olive oil for?

Same as usual Olive Oil
Only on special occasions/ for special dishes

18 Gender: Male Female

20 Age:

20 Level of education : Primary School High school college/university

21. Number of adults in household: 1 2 3 more than 3

22. Number of children in household: None 1 2 3 4 more than 4