

Α.Τ.Ε.Ι. ΠΑΤΡΩΝ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ & ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Η ΕΠΙΧΕΙΡΗΜΑΤΙΚΗ ΕΥΦΥΪΑ ΚΑΙ Η ΕΦΑΡΜΟΓΗ

ΤΗΣ ΣΤΗΝ ΛΟΓΙΣΤΙΚΗ

ΦΟΙΤΗΤΕΣ:

ΔΗΜΟΠΟΥΛΟΥ ΕΥΓΕΝΙΑ  
ΦΙΛΕΡΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ  
ΦΡΑΓΚΟΥΛΗΣ ΕΥΑΓΓΕΛΟΣ

ΕΙΣΗΓΗΤΗΣ:

ΤΑΜΠΑΚΑΣ ΒΑΣΙΛΕΙΟΣ



## ΠΕΡΙΕΧΟΜΕΝΑ

ΕΞΩΦΥΛΛΟ-ΘΕΜΑ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ	1
ΠΕΡΙΕΧΟΜΕΝΑ	2
ΠΡΟΛΟΓΟΣ	5
ΚΕΦΑΛΑΙΟ 1 - Τι είναι η Επιχειρηματική Νοημοσύνη (BI). Τα συστήματα υποστήριξης αποφάσεων στις επιχειρήσεις. Η φιλοσοφία της επιχειρηματικής νοημοσύνης.	7
1.1 Τι είναι η επιχειρηματική νοημοσύνη;	8
1.2 Τα συστήματα υποστήριξης αποφάσεων στις επιχειρήσεις	9
1.3 Τα συστήματα υποστήριξης αποφάσεων – Η φιλοσοφία της επιχειρηματικής νοημοσύνης	11
1.4 Τι μπορεί να κάνει η Επιχειρηματική Νοημοσύνη για μια επιχείρηση;	13
1.5 Σε ποιους απευθύνεται η Επιχειρηματική Νοημοσύνη;	14
1.6 Πώς λειτουργεί;	14
1.7 Τι χρειάζεται μια εταιρία;	14
1.8 Εργαλεία Επιχειρηματικής Νοημοσύνης	15
ΚΕΦΑΛΑΙΟ 2 - Οι νέες τεχνολογίες στον χώρο των Βάσεων Δεδομένων στην υπηρεσία της Επιχειρηματικής Νοημοσύνης. Οι Αποθήκες Δεδομένων – Data Warehouse	18
2.1 Εισαγωγή στις Αποθήκες Δεδομένων (Data Warehouse), το OLAP (On-Line Analytical Processing) και την Εξόρυξη Γνώσης (Data Mining)	19
2.1.1 Αποθήκη Δεδομένων (Data Warehouse)	22
2.1.2 OLAP	22
2.1.3 Εξόρυξη Γνώσης	23
2.2 Το μέσο: οι Αποθήκες Δεδομένων	25
2.2.1 Σχεδίαση Αποθηκών Δεδομένων	30
2.2.2 Λειτουργία της αποθήκης δεδομένων	33
2.2.3 Ροή δεδομένων	34
2.2.4 Η αποθήκη δεδομένων όπως την γνωρίζουν οι αναλυτές	34
2.2.5 OLAP Πράξεις στις αποθήκες δεδομένων	36
2.3 Το εργαλείο: η Εξόρυξη Γνώσης	39
2.3.1 Τα αποτελέσματα της εξόρυξης γνώσης	39

2.3.2 Οι στόχοι της εξόρυξης γνώσης	40
2.3.3 Οι κυριότερες τεχνικές	41
2.3.4 Πως λειτουργεί η εξόρυξη γνώσης	42
2.3.5 Λογισμικό Εξόρυξης Γνώσης	44
2.3.6 Η διαδικασία εξόρυξης γνώσης από μια αποθήκη δεδομένων	45

ΚΕΦΑΛΑΙΟ 3 - Περιγραφή των πιο χρησιμοποιούμενων μεθόδων Επιχειρηματικής νοημοσύνης: Κανόνες Συσχετισμού (Association Rules), Ταξινόμηση (Classification), Ομαδοποίηση (Clustering) και Πρόβλεψη (Forecasting -Prediction)	49
--	----

3.1 Βασικές εργασίες εξόρυξης γνώσης	50
3.2 Ταξινόμηση (Classification)	50
3.3 Τμηματοποίηση (Clustering)	55
3.4 Εξαγωγή κανόνων συσχέτισης (Association rules)	59
3.5 Πρόβλεψη (Prediction)	62

ΚΕΦΑΛΑΙΟ 4 - ΛΟΓΙΣΤΙΚΗ ΕΦΑΡΜΟΓΗ	65
---------------------------------	----

ΠΕΡΙΛΗΨΗ	66
----------	----

ΕΙΣΑΓΩΓΗ	67
----------	----

4.1 Εξελικτικοί Αλγόριθμοι	69
4.2 Περιγραφή λειτουργίας των Ε.Α	69
4.3 Μοντέλα Εξελικτικών Αλγορίθμων	71
4.3.1 Εξελικτικός Προγραμματισμός (Evolutionary Programming)	71
4.3.2 Εξελικτικές Στρατηγικές	72
4.3.3 Γενετικός Προγραμματισμός	73
4.3.4 Γενετικοί Αλγόριθμοι	74
4.4 Τεχνικές και Τελεστές	74
4.5 Γιατί οι Ε.Α. είναι αποτελεσματικοί	78
4.6 Πλεονεκτήματα – Μειονεκτήματα	80
4.7 Εφαρμογές	81
4.8 Εξελικτικοί Αλγόριθμοι και Χρηματιστήριο	82
4.9 Πράξεις (θέσεις) στο Χρηματιστήριο Αξιών Αθηνών και στο Χρηματιστήριο Παραγώγων	83
4.10 Επενδυτικά μοντέλα	84
4.10.1 Είδη επενδυτικών μοντέλων	86

4.11 Τεχνική Ανάλυση	86
4.11.1 Απλός Κινητός Μέσος (Simple Moving Average)	88
4.11.2 Εκθετικός Κινητός Μέσος (Exponential Moving Average)	89
4.11.3 Γραμμική Παλινδρόμηση (Linear Regression) και $r^2$	90
4.11.4 Μέση Πραγματική Διακύμανση (Average True Range)	91
4.11.5 Δείκτης Σχετικής Ισχύος (Relative Strength Index) – Stochastic	92
4.11.6 Συμμετρικά Τρίγωνα	94
4.12 Το πρόβλημα στη δημιουργία επενδυτικών μοντέλων	94
4.13 Προηγούμενες Μελέτες για τη δημιουργία επενδυτικών μοντέλων	96
4.14 Σχεδιασμός της εφαρμογής	103
4.15 Οι παράμετροι του επενδυτικού μοντέλου	104
4.15.1 Κανόνες	104
4.15.2 Σήματα Αγορών / Πωλήσεων	104
4.16 Προϋποθέσεις – Παραδοχές	106
4.17 Οι κανόνες	106
4.18 Υλοποίηση - Οι παράμετροι του Εξελικτικού Αλγορίθμου	112
4.18.1 Αναπαράσταση	112
4.18.2 Γενετικοί τελεστές	113
4.18.3 Μηχανισμός επιλογής	114
4.18.4 Τοπικά βέλτιστα	114
4.18.5 Αρχεία ιστορικών τιμών	115
4.18.6 Αποθήκευση καλύτερων λύσεων	115
4.18.7 Υπολογισμός της τιμής της Συνάρτησης Ποιότητας (Fitness Function)	115
4.19 Αποτελέσματα – Συμπεράσματα	116
4.20 Οι Δοκιμές	117
4.21 Συμπεράσματα από τις εκτελέσεις του προγράμματος	119
4.22 Προτάσεις βελτίωσης	120
ΕΠΙΛΟΓΟΣ	122
ΒΙΒΛΙΟΓΡΑΦΙΑ	123

## ΠΡΟΛΟΓΟΣ

Η ραγδαία ανάπτυξη στους τομείς της πληροφορικής και τεχνολογίας που έλαβε χώρα τις τελευταίες δεκαετίες του εικοστού αιώνα οδήγησε στη σύσταση πολυεθνικών επιχειρήσεων, έχουσών ως αποδέκτη ένα κοινό ευρύ, του οποίου οι ανάγκες ποικίλουν ανάλογα με το χώρο και το χρόνο. Η αύξηση της αναλογίας προσφορά –ζήτηση έχει ως αποτέλεσμα τη στελέχωση της επιχείρησης με πολυάριθμο υπαλληλικό δυναμικό. Η σύγχρονη αυτή τάση της επιχειρηματικής επέκτασης ενέχει διάφορες εκφάνσεις στη λειτουργία της επιχείρησης. Πλέον σημαντική αποτύπωση της, ενυπάρχει στη δυσκολία ελέγχου αφενός των δραστηριοτήτων του συνόλου των εργαζομένων(π.χ. λόγω της δαιδαλώδους διάρθρωσης της επιχείρησης τα ανώτερα στελέχη της δεν δύνανται να γνωρίζουν την ταυτότητα και δράση του κάθε υπαλλήλου) και αφετέρου στην επιχειρηματική συμπεριφορά και συνέπεια των πελατών(καθώς η ανάπτυξη μεταφορών και συγκοινωνιών και το άνοιγμα των συνόρων έχει οδηγήσει σε ευρεία συνεργασία ανάμεσα σε επιχειρήσεις, που εδρεύουν ακόμα και σε διαφορετικές ηπείρους). Το δυσχερές έργο του ελέγχου εργαζομένων και πελατών με σκοπό την αναγωγή γενικών –πλην ασφαλών- συμπερασμάτων που συμβάλλουν στην ορθή λειτουργία της επιχείρησης καλείται να διεκπεραιώσει η επιστήμη της επιχειρηματικής ευφυΐας(Business Intelligence).

Η επιχειρηματική ευφυΐα περιλαμβάνει διάφορες μεθόδους για την επεξεργασία των δεδομένων μιας επιχείρησης προκειμένου να μοιράζεται η κάθε πληροφορία έτσι ώστε να λαμβάνονται όσο το δυνατόν καλύτερες αποφάσεις. Σκοπός της παρούσας εργασίας είναι η ανάλυση των μεθόδων που χρησιμοποιεί η επιστήμη της επιχειρηματικής ευφυΐας προκειμένου να αντλήσει συμπεράσματα ως προς τις δραστηριότητες υπαλλήλων και πελατών.

Η εισαγωγή αναφέρεται στα συστήματα υποστήριξης αποφάσεων τα οποία κατέχουν συμβουλευτικό ρόλο στην διαδικασία λήψης αποφάσεων με σκοπό την καταχώριση αφηρημένων πληροφοριών προς εξαγωγή συγκεκριμένων συμπερασμάτων.

Στο κεφάλαιο δύο επιχειρείται μια επισκόπηση των εργαλείων των συστημάτων υποστήριξης αποφάσεων. Τα συστήματα υποστήριξης αποφάσεων περιλαμβάνουν διάφορα εργαλεία και μέσα τα οποία αποσκοπούν στην εκπλήρωση του στόχου της επιχειρηματικής ευφυΐας ο οποίος είναι ο συγκερασμός διάφορων πληροφοριών με σκοπό την απάντηση σε σύνθετα ερωτήματα σε θέματα σχετιζόμενα με την επιχείρηση. Υπάρχει λοιπόν μια εκτενής αναφορά στα διάφορα εργαλεία (π.χ. αποθήκες δεδομένων, εφαρμογή OLAP πράξεων) και στον επικουρικό ρόλο καθενός από αυτά στο σύνολο της διεργασίας. Επιπλέον επιχειρείται μια σύντομη ιστορική αναφορά σε ότι αφορά την τεχνολογική εξέλιξη και τις υπάρχουσες φιλοσοφίες και επιρροές που αυτές ασκούν στη λειτουργία των επιχειρήσεων.

Το κεφάλαιο τρία καλύπτει το ευρύτερο και σημαντικότερο ίσως φάσμα της επιχειρηματικής νοημοσύνης καθώς αναφέρεται στην περιγραφή των μεθόδων που συμπεριλαμβάνει η επιχειρηματική ευφυΐα, προκειμένου να κατορθώσει την επίτευξη των στόχων της. Η ανάλυση των μεθόδων αποτελεί όπως προαναφέρθηκε το σημαντικότερο κομμάτι της επιστήμης της επιχειρηματικής νοημοσύνης καθώς ένα λάθος στην εφαρμογή κάποιου μέσου μπορεί να κοστίζει την ασφάλεια του αποτελέσματος και συνεπώς να οδηγήσει την επιχείρηση σε μια ασύμφορη και πιθανώς κοστοβόρα επιλογή. Έτσι, αναλύονται οι βασικότερες μέθοδοι και επεξηγούνται με παραδείγματα.

Η εφαρμογή (Κεφάλαιο 4) ασχολείται με τους εξελικτικούς αλγόριθμους ένα μέσο το οποίο δύναται να δώσει ικανοποιητικές λύσεις. Συμπεριλαμβάνεται μια επισκόπηση στον τρόπο λειτουργίας του, στα είδη του, στα πλεονεκτήματα και μειονεκτήματα του. Η αναφορά αυτή προσδοκά μέσω της διεξοδικής ανάλυσης του μέσου αυτού και της μελέτης διάφορων παραμέτρων να τον καταστήσει κατά το δυνατόν αποτελεσματικότερο. Επιπλέον καταγράφονται οι προηγούμενες μελέτες με σκοπό να κατανοηθούν πληρέστερα οι επενδυτικές δυνατότητες και οι δυσκολίες στην ανάπτυξη τους. Η εφαρμογή κλείνει με τον τρόπο σχεδιασμού της, δηλαδή την προσέγγιση του προβλήματος, τα εργαλεία που χρησιμοποιήθηκαν και τα αποτελέσματα της εκτέλεσης του προγράμματος καθώς και προτάσεις βελτίωσης των εργαλείων των εξελικτικών αλγόριθμων.

## ΚΕΦΑΛΑΙΟ 1<sup>ο</sup>

Τι είναι η Επιχειρηματική Νοημοσύνη (BI). Τα συστήματα υποστήριξης αποφάσεων στις επιχειρήσεις. Η φιλοσοφία της επιχειρηματικής νοημοσύνης.

## 1.1 Τι είναι η επιχειρηματική νοημοσύνη;

Οι ανάγκες της σύγχρονης εποχής απαιτούν γρήγορη, αποδοτική και συνεπή διαχείριση τεράστιου όγκου πληροφοριών και κάθε είδους δεδομένων από την πλευρά των επιχειρήσεων βάσει των οποίων προσδοκούν να πετύχουν βραχυπρόθεσμους και μακροπρόθεσμους στόχους τους. Στον πυρήνα μιας λύσης για αυτό το ζήτημα βρίσκεται η επιχειρηματική ευφυΐα. Η επιχειρηματική ευφυΐα επιτρέπει να μοιράζεται η πληροφορία σε συνεργάτες, πελάτες και επιχειρησιακούς συμμάχους, έτσι ώστε όλα τα συμμετέχοντα μέρη να μπορούν άμεσα να λαμβάνουν καλύτερες και εξυπνότερες επιχειρηματικές αποφάσεις.

Αναλυτικότερα θα μπορούσαμε να πούμε ότι το πλήθος των δεδομένων οδήγησε μερικές επιχειρήσεις στην δημιουργία αυτόνομων επιχειρησιακών μονάδων που ως σκοπό έχουν την τροφοδότηση της διοίκησης ή άλλων τομέων σχετικούς με την λήψη αποφάσεων, με πληροφορίες σχετικές με πελάτες, ανταγωνιστές κλπ. Η λειτουργία αυτή ονομάζεται Επιχειρηματική Νοημοσύνη (Business Intelligence).

Η επιχειρηματική νοημοσύνη (BI) δεν είναι καινούρια στον χώρο της πληροφορικής. Πρώιμες εκδόχές της συναντάμε ακόμα και στα τέλη της δεκαετίας του '70, στη δεκαετία του '90 όμως συνέβησαν τα μεγαλύτερα άλματα. Όσο και αν φαίνεται παράξενο σήμερα τα αποτελέσματα του BI τα συναντάμε συχνά σε πολλές και καθημερινές μας δραστηριότητες.

Στα μεγάλα super markets με μεθόδους επιχειρηματικής ευφυΐας καθορίζεται πλέον η τοποθέτηση των προϊόντων στα ράφια. Δεν είναι τυχαίο ότι σχεδόν σε όλα τα super markets συγκεκριμένα προϊόντα θα τα δούμε να γειτνιάζουν (π.χ. τα snacks δίπλα στα ποτά, τις τομάτες δίπλα στα ζυμαρικά, κ.α.)

Τέτοιες επιλογές δεν γίνονται τυχαία, παρά μόνο μετά από ανάλυση των καταναλωτικών συνηθειών των πελατών με μεθόδους BI. Και αν πάμε λίγο πιο μακριά, η BI είναι αυτή που δίνει ώθηση σε μύριες άλλες περιοχές, από την ώρα προβολής των διαφημιστικών μηνυμάτων μέχρι τις παροχές των ασφαλιστήριων συμβολαίων.

Πριν μερικά χρόνια οι επιχειρηματικές ανάγκες περιορίζονταν στην οργάνωση και επεξεργασία της πληροφορίας με την δομή που γνωρίζουμε σήμερα ως "διαχείριση



και προγραμματισμός επιχειρηματικών πόρων” ή αλλιώς ERP (Enterprise Resource Planning). Τέτοιες δομές είχαν ως αποτέλεσμα τη συσσώρευση ενός μεγάλου όγκου από πληροφορίες.

Με το πέρασμα του χρόνου τα μεγέθη αυξήθηκαν και ο συσσωρευμένος όγκος πληροφοριών οργανώθηκε σε δομές που τις αποκαλούμε πλέον Αρχιτεκτόνημα Δεδομένων – Αποθήκες Δεδομένων (Data Warehouse – DW). Ένα DW πλαισιώνεται με την χρήση ειδικών εφαρμογών διαχείρισης (Clients) αποθηκών δεδομένων οι οποίες μπορούν και αναλύουν τις πληροφορίες πολυδιάστατα, παρέχοντας την δυνατότητα για εξελιγμένο επιτελικό reporting.

Ο τεράστιος αυτός όγκος πληροφοριών κρύβει μέσα του “γνώση” η οποία είναι ανεξερεύνητη. Συσχετίσεις μεταξύ δεδομένων, αλληλουχίες γεγονότων, προβλέψεις για στόχους και πολλά άλλα είναι θαμμένα μέσα στα δεδομένα και περιμένουν με κάποιο τρόπο να έρθουν στην επιφάνεια.

Η γνώση αυτή είναι κρυμμένη διότι δεδομένα εξαρχής είναι οργανωμένα σε δομές οι οποίες εξυπηρετούν συγκεκριμένους σκοπούς. Ο κρυμμένος θησαυρός λοιπόν μπορεί να έρθει στην επιφάνεια με ειδικές μεθόδους το σύνολο των οποίων ονομάζεται BI.

Συμπερασματικά λοιπόν μπορούμε να πούμε ότι ως BI θεωρείται η χρησιμοποίηση μεθόδων οι οποίες αναλύουν τα δεδομένα σε ένα DW ή σε μία πολύ μεγάλη βάση δεδομένων και είτε προτείνουν είτε βοηθούν στην επιχειρηματική απόφαση. Πρόκειται δηλαδή για στατιστικούς κανόνες ή/και αναλυτικές μεθόδους που βοηθούν στην εξαγωγή συμπερασμάτων.

## **1.2 Τα συστήματα υποστήριξης αποφάσεων στις επιχειρήσεις**

Η επιχειρηματική νοημοσύνη χρησιμοποιεί λογισμικό και τα δεδομένα της εταιρίας ώστε να υποστηρίξει την διαδικασία λήψης αποφάσεων. Αποτέλεσμα αυτού ήταν η δημιουργία των συστημάτων υποστήριξης αποφάσεων στις επιχειρήσεις.

Τα Συστήματα Υποστήριξης Αποφάσεων (ΣΥΑ) αναπτύχθηκαν από δυο κύριες ερευνητικές δραστηριότητες – τις θεωρητικές μελέτες πάνω στην οργανωτική λήψη αποφάσεων που έγιναν στο Carnegie Institute of Technology (Simon, Cyert, March και

άλλοι) στα τέλη του 1950 και στις αρχές του 1960 και στην τεχνική εργασία του MIT (Gerrity, Ness και άλλοι) το 1960.

Τα συστήματα υποστήριξης αποφάσεων σχεδιάζονται για να βοηθούν αυτούς που λαμβάνουν αποφάσεις να τις πάρουν με βραχυπρόθεσμο ή μακροπρόθεσμο ορίζοντα. Θα πρέπει εξ αρχής να τονιστεί ότι δεν αναλαμβάνουν να πάρουν αποφάσεις αφού αξιολογήσουν κάποια στοιχεία. Ο ρόλος τους είναι καθαρά συμβουλευτικός και υποστηρικτικός.

Κάποιος θα μπορούσε να αναλογιστεί τι γινόταν τα περασμένα χρόνια. Πως δηλαδή λαμβάνονταν τότε οι αποφάσεις χωρίς τη χρήση τέτοιων συστημάτων. Η απάντηση είναι ότι ασφαλώς λαμβάνονταν αποφάσεις απλά αυτό απαιτούσε πολύ χρόνο και πάρα πολύ κόπο γιατί απασχολούσε πολύ προσωπικό. Την εποχή που τα Πληροφοριακά Συστήματα δεν είχαν την σημερινή διάδοση η λήψη αποφάσεων απαιτούσε τη χειροκίνητη συλλογή πληροφοριών από τα τμήματα της εταιρίας και τη δημιουργία των λεγόμενων εκθέσεων (reports). Αυτά αξιολογούσαν οι μάνατζερ για να μπορέσουν να αποφασίσουν ποια πορεία θα ακολουθήσουν.

Γίνεται επομένως σαφές ότι τέτοιες διαδικασίες δεν θα μπορούσαν να επιζήσουν στο σημερινό επιχειρηματικό περιβάλλον που απαιτούνται γρήγορες ταχύτητες για να ανταπεξέλθει μια εταιρία στον ανταγωνισμό. Σήμερα οι πληροφορίες που έχουν στη διάθεση τους οι επιχειρήσεις κρατούνται σε βάσεις δεδομένων και είναι διαθέσιμες σε όλα τα τμήματα ανά πάσα στιγμή. Οι εφαρμογές υποστήριξης αποφάσεων έχουν πρόσβαση σε όλο το εύρος αυτών και χρησιμοποιούν κατάλληλα εργαλεία και σύγχρονες τεχνολογίες ώστε να παρουσιάζουν τα δεδομένα έγκαιρα, περιληπτικά και με γραφικό τρόπο στους αποφασίζοντες ώστε να μπορούν αυτοί να πάρουν κατευθυνόμενες από τη γνώση αποφάσεις.

Αυτή η παρουσίαση των δεδομένων δεν γίνεται αυτόματα. Προκύπτει μετά από μια έντονη διαλεκτική σχέση μεταξύ χρήστη και συστήματος. Ο χρήστης έχει στη διάθεση του διάφορες δυνατότητες τις οποίες εφαρμόζει πάνω στα δεδομένα και παίρνει τα ανάλογα αποτελέσματα. Έτσι, ανάλογα με τις δυνατότητες που προσφέρουν, έχουν προκύψει τρεις τύποι συστημάτων υποστήριξης αποφάσεων.

- **Κατευθυνόμενα από μμοντέλο (model-driven)**

Πρόκειται για ολοκληρωμένα συστήματα που έχουν τη δυνατότητα εκτέλεσης what-if σεναρίων καθώς και άλλων τύπων ανάλυσης...

- **Κατευθυνόμενα από τα δεδομένα (data-driven)**

Επιτρέπουν στο χρήστη να εξάγει και να αναλύει χρήσιμη πληροφορία από μεγάλες βάσεις δεδομένων.

- **Εξόρυξης Γνώσης (data mining)**

Εύρεση κρυμμένων τυποποιημένων μορφών (patterns) και κάποιων σχέσεων (relationships) σε μεγάλες βάσεις δεδομένων. Το αποτέλεσμα είναι η εξαγωγή κάποιων κανόνων ώστε να είναι δυνατή η πρόβλεψη μελλοντικών συμπεριφορών.

Στις δυο πρώτες κατηγορίες ανήκουν κυρίως τα σημερινά ΣΥΑ. Η εργασία θα εξερευνήσει την τρίτη κατηγορία και θα προσπαθήσει να δείξει τη δυναμική που μπορούν να προσδώσουν οι τεχνολογίες που σχετίζονται με την εξόρυξη γνώσης στην υποστήριξη αποφάσεων.

Η υποστήριξη αποφάσεων οπωσδήποτε ταυτίζεται με την πληροφορία. Αυτή είναι που βοηθά κάποιον να καταλήξει σε μια πολιτική, σε μια απόφαση, σε μια πορεία. Η πληροφορία προέρχεται από το πρόσφατο παρελθόν. Η μελέτη αυτού κρύβει μυστικά που αν αποκαλυφθούν θα αποτελέσουν σημαντικό οδηγό για το πώς πρέπει να κινηθεί μελλοντικά μια εταιρία ή ένας οργανισμός.

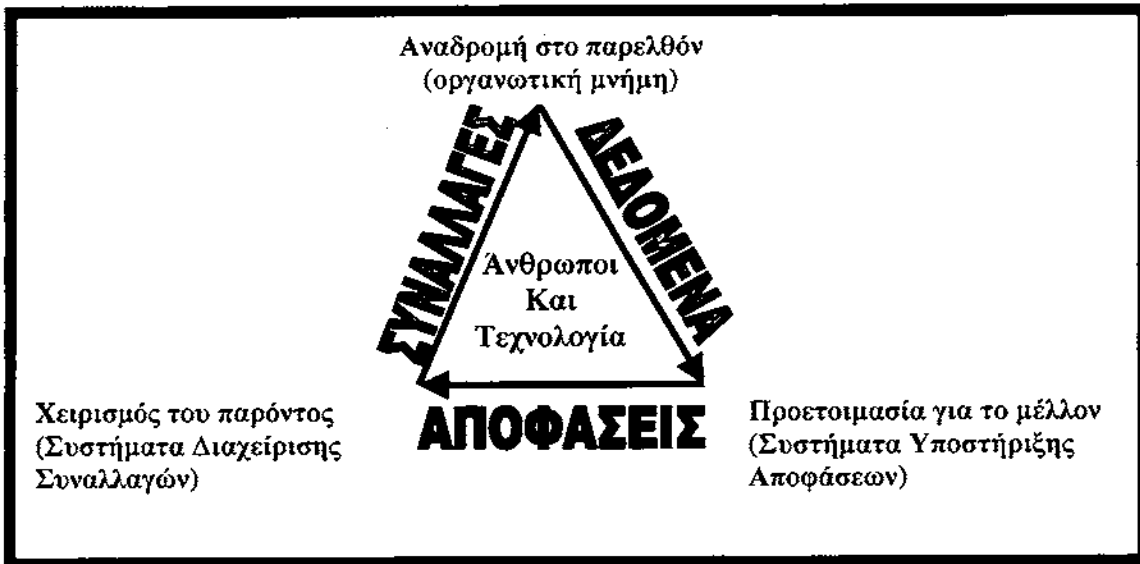
### **1.3 Τα συστήματα υποστήριξης αποφάσεων – Η φιλοσοφία της επιχειρηματικής νοημοσύνης**

Τα συστήματα υποστήριξης αποφάσεων αποτελούν το σημαντικότερο κομμάτι στην υποδομή ενός επιχειρησιακού πληροφοριακού συστήματος (σχήμα 1) γιατί παρέχουν τη δυνατότητα στις εταιρίες να μετατρέψουν μεγάλες ποσότητες επιχειρηματικών πληροφοριών σε χειροπιαστά και επίκερδη αποτελέσματα. Παρόλα

αυτά η συλλογή, συντήρηση και ανάλυση τεράστιων ποσοτήτων δεδομένων αποτελούν δύσκολες εργασίες και απαιτούν τεχνικές δεξιότητες, κόστος και οργανωτικές δεσμεύσεις.

### Σχήμα 1

Η εικόνα μιας οποιαδήποτε επιχείρησης – ο ρόλος των OLTP, ΣΥΑ



Ας αναλύσουμε λίγο το σχήμα για να εξηγήσουμε την κυκλική πορεία που παρουσιάζεται.

Τα OLTP (On Line Transaction Processing Systems) συστήματα επιτρέπουν στους οργανισμούς να συγκεντρώνουν πληροφορίες για τις καθημερινές συναλλαγές τους με τους πελάτες (για παράδειγμα πληροφορίες από τις πωλήσεις). Οι OLTP εφαρμογές τυπικά αυτοματοποιούν δομημένες και επαναληπτικές διαδικασίες επεξεργασίας δεδομένων όπως είναι η εισαγωγή στοιχείων και οι τραπεζικές συναλλαγές.

Λεπτομερέστερα, «φρέσκα» (up-to-date) δεδομένα από διάφορα, ανεξάρτητα μεταξύ τους σημεία θα πρέπει να συγκεντρωθούν σε μια τοποθεσία πριν μπορέσουν οι αναλυτές να εξάγουν χρήσιμα συμπεράσματα. Οι μάνατζερ χρησιμοποιούν καθημερινά αυτό το σύνολο δεδομένων για να παίρνουν αποφάσεις για οτιδήποτε.

Για παράδειγμα ένας αγοραστής βιομηχανικού εξοπλισμού επιθυμεί να έχει μια λίστα με τους - μέχρι αυτή τη στιγμή - παγκόσμιους προμηθευτές καθώς και τις κάθε λεπτό ανανεώσιμες τιμές που δίνει κάθε πωλητής. Η άμεση εμφάνιση αυτών των δεδομένων στην επιφάνεια εργασίας του μαζί με όλες τις σχετικές πληροφορίες θα βοηθήσουν στην έγκαιρη και σωστή απόφαση αγοράς.

Φαίνεται λοιπόν και στο σχήμα 1 ότι οι απλές καθημερινές συναλλαγές δημιουργούν δεδομένα στα οποία στηρίζονται τα ΣΥΑ για να μπορέσουν να παρέχουν υπηρεσίες υποστήριξης αποφάσεων. Οι αποφάσεις αυτές θα οδηγήσουν ίσως την εταιρία σε μια νέα πορεία. Ο κύκλος επαναλαμβάνεται αφού και μετά τις όποιες αλλαγές θα συνεχιστούν οι συναλλαγές με τους πελάτες κ.ο.κ.

Το κλειδί της επιτυχίας αυτών των συστημάτων είναι το κατά πόσο η παρεχόμενη πληροφορία καλύπτει τις ανάγκες ενός λήπτη αποφάσεων. Είναι επομένως κατανοητό ότι ένα ΣΥΑ για έναν αντιπρόεδρο μιας τράπεζας ο οποίος ενδιαφέρεται για παράδειγμα για τους ρυθμούς χορήγησης δανείων είναι διαφορετικό από ένα σύστημα που θα είναι χρήσιμο σε ένα ταξιδιωτικό γραφείο το οποίο αναζητά τις χαμηλότερες τιμές σε ξενοδοχεία μιας περιοχής.

Στις επόμενες παραγράφους του κεφαλαίου αυτού περιγράφεται η νοητή διαδικασία η οποία θα ωθούσε πιθανόν μια εταιρία να εκμεταλλευτεί τα δεδομένα της για να λάβει αποφάσεις για το μέλλον.

#### **1.4 Τι μπορεί να κάνει η Επιχειρηματική Νοημοσύνη για μια επιχείρηση;**

Οι υπέρ-ανταγωνιστικές αγορές, οι αυξανόμενες απαιτήσεις των πελατών, οι ραγδαίες τεχνολογικές αλλαγές και η αλματώδης ανάπτυξη των επιχειρήσεων προκαλούν το ενδιαφέρον πολλών βιομηχανικών τομέων. Για να είναι δυνατή η βελτίωση της απόδοσης της εταιρίας και η ικανοποίηση των απαιτήσεων των πελατών θα πρέπει να αξιοποιηθούν καλύτερα οι κρίσιμες επιχειρηματικές πληροφορίες και μάλιστα ταχύτερα από ότι οι ανταγωνιστές. Αυτό εξηγεί γιατί η Επιχειρηματική Νοημοσύνη αποτελεί το

στρατηγικό πλεονέκτημα για κάθε οργανισμό. Αυτό σημαίνει ότι εξοπλίζει το προσωπικό της εταιρίας με τα πλέον ενήμερα και ακριβή στοιχεία.

### **1.5 Σε ποιους απευθύνεται η Επιχειρηματική Νοημοσύνη;**

Η Επιχειρηματική Νοημοσύνη (ΕΝ) είναι κατάλληλη για τον οικονομικό διευθυντή μιας νομικής εταιρίας όπως επίσης για τον διευθυντή μιας κλινικής όπως και για τον διευθυντή του τμήματος μάρκετινγκ προϊόντων σε μια εταιρία υψηλής τεχνολογίας. Απευθύνεται, με άλλα λόγια, σε ανθρώπους που χρειάζονται στοιχεία και πληροφορίες για να λάβουν αποφάσεις.

### **1.6 Πώς λειτουργεί;**

Η Επιχειρηματική Νοημοσύνη ενοποιεί τις παραδοσιακά διακριτές λειτουργίες της πρόσβασης σε δεδομένα, της εξερεύνησης και της ανάλυσης. Αυτή η ενοποίηση παρέχει τα μέσα για την μετατροπή απομονωμένων νησίδων πληροφορίας σε περιεκτική γνώση που αποτελεί τη βάση των δυναμικών επιχειρηματικών αποφάσεων. Αυτή η γνώση επιτρέπει γρήγορη, μη χρονοβόρα δράση επειδή τα διαθέσιμα δεδομένα είναι εύκολο να εντοπιστούν.

### **1.7 Τι χρειάζεται μια εταιρία;**

Οι περισσότερες εταιρίες έχουν ήδη στήσει τα θεμέλια της Επιχειρηματικής Νοημοσύνης – υπάρχουν βάσεις δεδομένων που περιέχουν στοιχεία σχετικά με την εταιρία, τους πελάτες της, τα προϊόντα της, το προσωπικό και τα τμήματα της. Το πρόβλημα είναι ότι οι πληροφορίες αυτές είναι συνήθως «θαμμένες» σε διαφορετικές πηγές και η διαδικασία ανάκτησης αυτών είναι χρονοβόρα και επίπονη.

Οι εφαρμογές EN προσφέρουν τον μηχανισμό της απλής και γρήγορης ενοποίησης επιχειρηματικών στοιχείων από ανόμοιες πηγές ώστε να είναι δυνατός ο διαμοιρασμός πληροφοριών στο εσωτερικό της εταιρίας μεταξύ υπαλλήλων αλλά και στο εξωτερικό με πελάτες, προμηθευτές. Με αυτό τον τρόπο μπορούμε να χρησιμοποιήσουμε αποτελεσματικά τα δεδομένα για ανάλυση, αναφορές ή για υποστήριξη αποφάσεων ώστε να βελτιώσουμε την λειτουργικότητα και για να χτίσουμε πιο επικερδής πελατειακές σχέσεις.

## **1.8 Εργαλεία Επιχειρηματικής Νοημοσύνης**

Υπάρχει κατάλληλο λογισμικό που επιτρέπει την παρακολούθηση και χρήση μεγάλων ποσοτήτων δεδομένων. Οι τρεις παρακάτω τύποι εργαλείων είναι γνωστά ως Εργαλεία Επιχειρηματικής Νοημοσύνης:

- **Λογισμικό Πολυδιάστατης Ανάλυσης**

Γνωστό και ως OLAP (On- Line Analytical Processing) – Δίνει την δυνατότητα στον χρήστη να κοιτάξει τα δεδομένα από πολλές διαφορετικές διαστάσεις.

- **Εργαλεία ερωτήσεων (queries)**

Λογισμικό που επιτρέπει στον χρήστη να κάνει ερωτήσεις σχετικά με τυποποιημένες μορφές των δεδομένων ή ακόμα και για λεπτομέρειες.

- **Εργαλεία Εξόρυξης Γνώσης**

Λογισμικό που αυτομάτως αναζητά σημαντικές τυποποιημένες μορφές ή συσχετίσεις στα δεδομένα.

Ο Lesca1 το 1994 περιέγραψε την Επιχειρηματική Νοημοσύνη ως μια κυκλική διαδικασία αποτελούμενη από πέντε φάσεις (σχήμα 2).

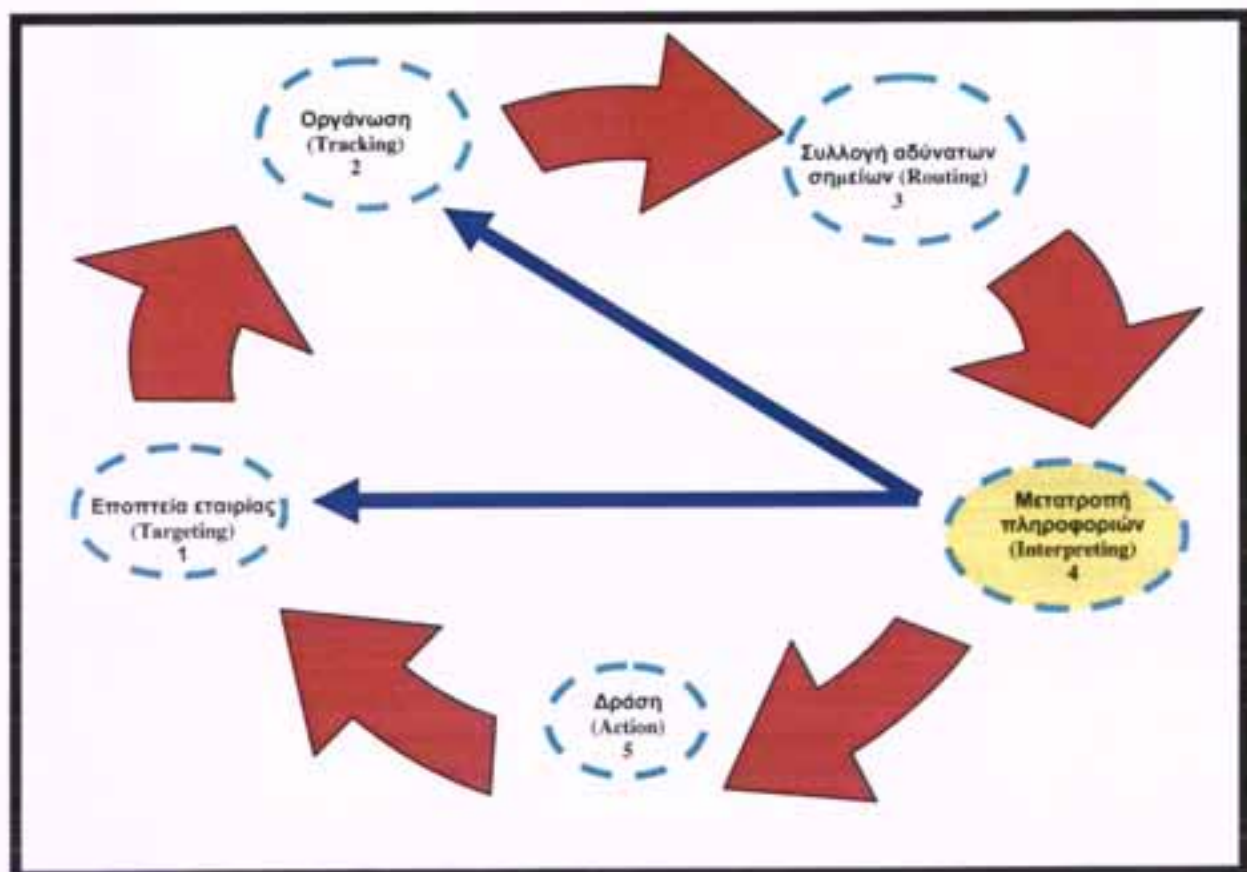
Στην πρώτη φάση “targeting”, γίνεται η εποπτεία της εταιρίας με σκοπό να γίνει κάποια κατανομή των προτεραιοτήτων της. Η δεύτερη φάση “tracking”, είναι

περισσότερο οργανωτική αλλά περιλαμβάνει και τον εντοπισμό των πιο κρίσιμων αδύνατων σημείων. Στην τρίτη φάση "routing", γίνεται με επαναλαμβανόμενο τρόπο η συλλογή των αδύνατων σημείων της εταιρίας από έξω προς τα μέσα. Η τέταρτη φάση "interpreting" αποτελείται από την μετατροπή των πληροφοριών που συλλέχθηκαν σε χρήσιμη πληροφορία. Αν αυτή η φάση αποδώσει η διαδικασία προχωράει στη φάση 5. Διαφορετικά θα πρέπει η πληροφορία να επεξεργαστεί ξανά με πιο ιδιαίτερο τρόπο(επιστροφή στη φάση 2) αν αυτή είναι ανακριβής ή θα πρέπει να επανεξεταστεί ξανά ο στόχος(επιστροφή στη φάση 1) αν η πληροφορία είναι πολύ μεγάλη.

Από αυτές τις φάσεις, η πιο σημαντική και πιο δύσκολη είναι η τέταρτη (interpreting). Ανεπαρκείς μετατροπές θα οδηγήσουν σε λανθασμένη χρήση της Επιχειρηματικής Νοημοσύνης.

## Σχήμα 2

Οι φάσεις της επιχειρηματικής νοημοσύνης





Η εργασία στο εξής θα περιγράψει το πώς μπορεί να επιτευχθεί η φιλοσοφία της Επιχειρηματικής Νοημοσύνης η οποία είναι και η καρδιά των συστημάτων υποστήριξης αποφάσεων. Πιο συγκεκριμένα θα αναλυθούν σε αρκετό βάθος οι τεχνολογίες εκείνες οι οποίες αποτελούν τα συστατικά στοιχεία των συστημάτων υποστήριξης αποφάσεων που ανήκουν στην Τρίτη από τις κατηγορίες που παρουσιάστηκαν στην παράγραφο 1.2.

## ΚΕΦΑΛΑΙΟ 2<sup>ο</sup>

Οι νέες τεχνολογίες στον χώρο των Βάσεων Δεδομένων στην υπηρεσία της Επιχειρηματικής Νοημοσύνης. Οι Αποθήκες Δεδομένων – Data Warehouse.

## 2.1 Εισαγωγή στις Αποθήκες Δεδομένων (Data Warehouse), το OLAP (On-Line Analytical Processing) και την Εξόρυξη Γνώσης (Data Mining)

Στις αρχές της δεκαετίας του 1990 τέσσερα πανίσχυρα εργαλεία εμφανίστηκαν στην περιοχή της ανάπτυξης συστημάτων υποστήριξης αποφάσεων. Το πρώτο νέο εργαλείο ήταν οι αποθήκες δεδομένων. Τα δυο επόμενα που ακολούθησαν ήταν η επεξεργασία δεδομένων σε πραγματικό χρόνο (OLAP) και η εξόρυξη γνώσης. Το τέταρτο εργαλείο είναι η τεχνολογία που σχετίζεται με τον παγκόσμιο ιστό. Ο πίνακας στην επόμενη σελίδα αναπαριστά τα χαρακτηριστικά τεχνολογικά βήματα της κάθε εποχής. Διακρίνονται οι επικρατούσες φιλοσοφίες στο χώρο των βάσεων δεδομένων και πως αυτές επηρεάζουν τις δυνατότητες των επιχειρήσεων και των συστημάτων υποστήριξης αποφάσεων

Θα εστιάσουμε την προσοχή μας στις δυο τελευταίες κατηγορίες και θα δείξουμε πως μπορούν να βασιστούν τα συστήματα υποστήριξης αποφάσεων πάνω στις τεχνολογίες αυτές.

Στον παρακάτω πίνακα φαίνεται ότι ο τομέας της εξόρυξης γνώσης είναι ανερχόμενος. Σε αυτό το αποτέλεσμα όμως δεν φτάσαμε τυχαία. Οι επιχειρήσεις θεώρησαν πολύ σημαντικό να διαθέτουν απαντήσεις σε ερωτήσεις όπως αυτή που φαίνεται στον πίνακα. Αυτή ακριβώς η ανάγκη επέβαλε την ανάπτυξη συστημάτων υποστήριξης αποφάσεων βασισμένα σε εξόρυξη γνώσης.

<u>Εξελικτικό βήμα</u>	<u>Επιχειρηματική ερώτηση</u>	<u>Βοηθητικές τεχνολογίες</u>	<u>Κατασκευαστές προϊόντων</u>	<u>Χαρακτηριστικά</u>
Συλλογή Δεδομένων (1960)	"Ποια ήταν τα συνολικά μου έσοδα τα τελευταία 5 χρόνια;"	Υπολογιστές, ταινίες, δισκέτες	IBM, CDC	Αναδρομική, στατική ανάκτηση δεδομένων
Πρόσβαση σε δεδομένα (1980)	"Ποιες ήταν οι πωλήσεις μου στην Πάτρα τον τελευταίο Μάρτιο;"	Σχεσιακές βάσεις δεδομένων (RDBMS), γλώσσα SQL, ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Αναδρομική, δυναμική ανάκτηση δεδομένων σε επίπεδο εγγραφής
Αποθήκες Δεδομένων & Υποστήριξη Αποφάσεων (1990)	"Ποιες ήταν οι πωλήσεις μου στην Πάτρα τον τελευταίο Μάρτιο; Βάση αυτών παίρνω αποφάσεις για την Αθήνα;"	Επεξεργασία σε πραγματικό χρόνο (OLAP), πολυδιάστατες βάσεις δεδομένων, αποθήκες δεδομένων	Pilot, Comshare, Arbor, Cognos, Microstrategy	Αναδρομική, δυναμική ανάκτηση δεδομένων σε πολλαπλά επίπεδα
Εξόρυξη Γνώσης (Ανερχόμενος τομέας σήμερα)	"Ποιες είναι οι πιθανές πωλήσεις του επόμενου μήνα στην Αθήνα;"	Αλγόριθμοι, πολύεπεξεργαστικά Υπολογ. συστήματα, μεγάλες βάσεις δεδομένων	Pilot, Lockheed, IBM, SGI	Προφητική ανάκτηση πληροφορίας

Ας θεωρήσουμε μια εταιρία που κατασκευάζει υποδήματα και αναλύσουμε τις δικές της ανάγκες ώστε να δούμε αν αυτή χρειάζεται αυτές τις τεχνολογίες για να βρει απαντήσεις στα ερωτήματα της. Η εταιρία αυτή πουλάει τα προϊόντα της με δυο τρόπους. Είτε κατευθείαν στους πελάτες, είτε μέσω μεταπωλητών. Οι ειδικοί του τμήματος μάρκετινγκ της εταιρίας χρειάζεται να εξάγουν τις παρακάτω πληροφορίες από το «βουνό» πληροφοριών της εταιρίας:

- τις πέντε μεγαλύτερες αυξήσεις σε πωλήσεις στην κατηγορία νέων προϊόντων για τα περασμένα χρόνια,
- τις συνολικές πωλήσεις σε υποδήματα στη Νέα Υόρκη τον τελευταίο μήνα ανά προϊόν παραγωγής,
- τις πενήντα πόλεις με τον μεγαλύτερο αριθμό «καλών» πελατών,
- ένα εκατομμύριο πελάτες που αποτελούν τους πιο πιθανούς αγοραστές του νέου τύπου Walk-On-Air.

Για να βρεθούν οι απαντήσεις σε αυτά είναι σαφές ότι δεν αρκεί μια απλή ανάγνωση των δεδομένων που διαθέτει η εταιρία. Χρειάζεται μια διαφορετική προσέγγιση διαχείρισης των δεδομένων ώστε να προκύψουν πληροφορίες που ουσιαστικά είναι κρυμμένες.

Έτσι λοιπόν η ανάγκη θα την ωθήσει στην επέκταση ή ακόμα και αλλαγή του πιθανού υπάρχοντος συστήματος υποστήριξης αποφάσεων. Ο στόχος είναι να βρεθούν απαντήσεις σε σύνθετα ερωτήματα. Την λύση μπορεί να την παρέχει η παρακάτω διαδικασία.

- Δημιουργία αποθήκης δεδομένων
- Εφαρμογή OLAP πράξεων
- Εφαρμογή αλγορίθμων εξόρυξης γνώσης πάνω στην αποθήκη δεδομένων

Αυτή τη διαδικασία ακολουθεί μια εταιρία που επιθυμεί να προσδώσει στο σύστημα υποστήριξης αποφάσεων της δυνατότητες εξόρυξης γνώσης. Φαίνεται λοιπόν ότι η πολυκόστητη εξόρυξη γνώσης δεν μπορεί να εφαρμοστεί αμέσως στα δεδομένα της εταιρίας. Οι λόγοι θα αναλυθούν παρακάτω. Στο σχήμα 3 φαίνεται ποια είναι η αρχιτεκτονική ενός τέτοιου συστήματος δηλαδή πως θα χρησιμοποιηθούν οι προαναφερόμενες τεχνολογίες για να πετύχουμε το στόχο της υποστήριξης αποφάσεων.

Για να μπορέσει να κατανοήσει ο αναγνώστης καλύτερα το σχήμα παραθέτουμε τις βασικές παραμέτρους των τεχνολογιών αυτών.

### 2.1.1 Αποθήκη Δεδομένων (Data Warehouse) :

Περιλαμβάνει δεδομένα που συσσωρεύονται εκεί από τις βάσεις δεδομένων της επιχείρησης και συχνά το μέγεθος τους φτάνει τα gigabytes ή ακόμα και terabytes. Τυπικά η αποθήκη δεδομένων συντηρείται ξεχωριστά από τις βάσεις δεδομένων του οργανισμού γιατί οι απαιτήσεις των εφαρμογών ανάλυσης δεν συμπίπτουν με τις δυνατότητες των βάσεων δεδομένων. Οι αποθήκες δεδομένων εξυπηρετούν τα συστήματα υποστήριξης αποφάσεων γιατί παρέχουν ιστορικά, ομαδοποιημένα και συγκεντρωτικά δεδομένα αντί για λεπτομερείς εγγραφές.

Υπάρχουν εμφανείς διαφορές μεταξύ των κλασικών βάσεων δεδομένων και των αποθηκών δεδομένων. Στην αποθήκη δεδομένων καταλήγουν κατάλληλα επεξεργασμένα δεδομένα των επιμέρους βάσεων δεδομένων, διαφοροποιώντας την έτσι ως προς το περιεχόμενο των πληροφοριών, πολλές φορές μάλιστα αυτά τα δεδομένα αποτελούν δομημένες πληροφορίες και όχι απλά μια καταγραφή απλών στοιχείων –πράγματα που εμφανίζονται στις απλές βάσεις δεδομένων.

Επειδή όμως η κατασκευή μιας αποθήκης δεδομένων μπορεί να διαρκέσει πολλά χρόνια, μερικοί οργανισμοί αντί αυτών κτίζουν τα λεγόμενα data marts που περιλαμβάνουν πληροφορίες για κάποια συγκεκριμένα τμήματα. Έτσι μπορεί ένα data mart μπορεί να ανήκει στο τμήμα Μάρκετινγκ, ένα άλλο στο Λογιστήριο. Όλα αυτά μαζί αποτελούν την κεντρική αποθήκη δεδομένων.

Μια ακόμα σημαντική παράμετρος είναι και ο τρόπος υλοποίησης της αποθήκης δεδομένων. Αν δηλαδή θα βασίζεται στο σχεσιακό ή το πολυδιάστατο μοντέλο (ROLAP εναντίον MOLAP). Η επιλογή παίζει ρόλο στην απόδοση της αποθήκης δεδομένων όχι όμως και στις δυνατότητες που αυτή μπορεί να προσφέρει.

### 2.1.2 OLAP :

Οι αποθήκες δεδομένων παρέχουν τη δυνατότητα για Συνεχή Αναλυτική Επεξεργασία (On-Line Analytical Processing – OLAP) των δεδομένων περιέχοντας ιστορικά και συγκεντρωτικά δεδομένα χρήσιμα για υποστήριξη αποφάσεων.

Η ανάπτυξη και η εξέλιξη της συνεχούς αναλυτικής διαδικασίας (OLAP) οφείλεται κυρίως σε δύο λόγους: Στη ραγδαία αύξηση των ποσοτήτων των δεδομένων και την ταυτόχρονη ανάγκη για ταχεία ανάλυση τους.

Με τα εργαλεία OLAP παρέχονται περισσότερες δυνατότητες από αυτές που οι απλές ερωτήσεις και οι αναφορές (reports) μπορούν να δώσουν. Βοηθούν τους αναλυτές, τους μάνατζερ και τα υψηλόβαθμα στελέχη των επιχειρήσεων στη ταχεία πρόσβαση και πολυδιάστατη επεξεργασία των δεδομένων τους με σκοπό τη παρουσίαση και τη λύση των προβλημάτων της επιχείρησης στις πραγματικές τους διαστάσεις.

Βοηθούν τον χρήστη να δημιουργεί αναλύσεις μέσα από πολλαπλές ερωτήσεις του τύπου “what-if” και έτσι να μοντελοποιεί το σενάριο του.

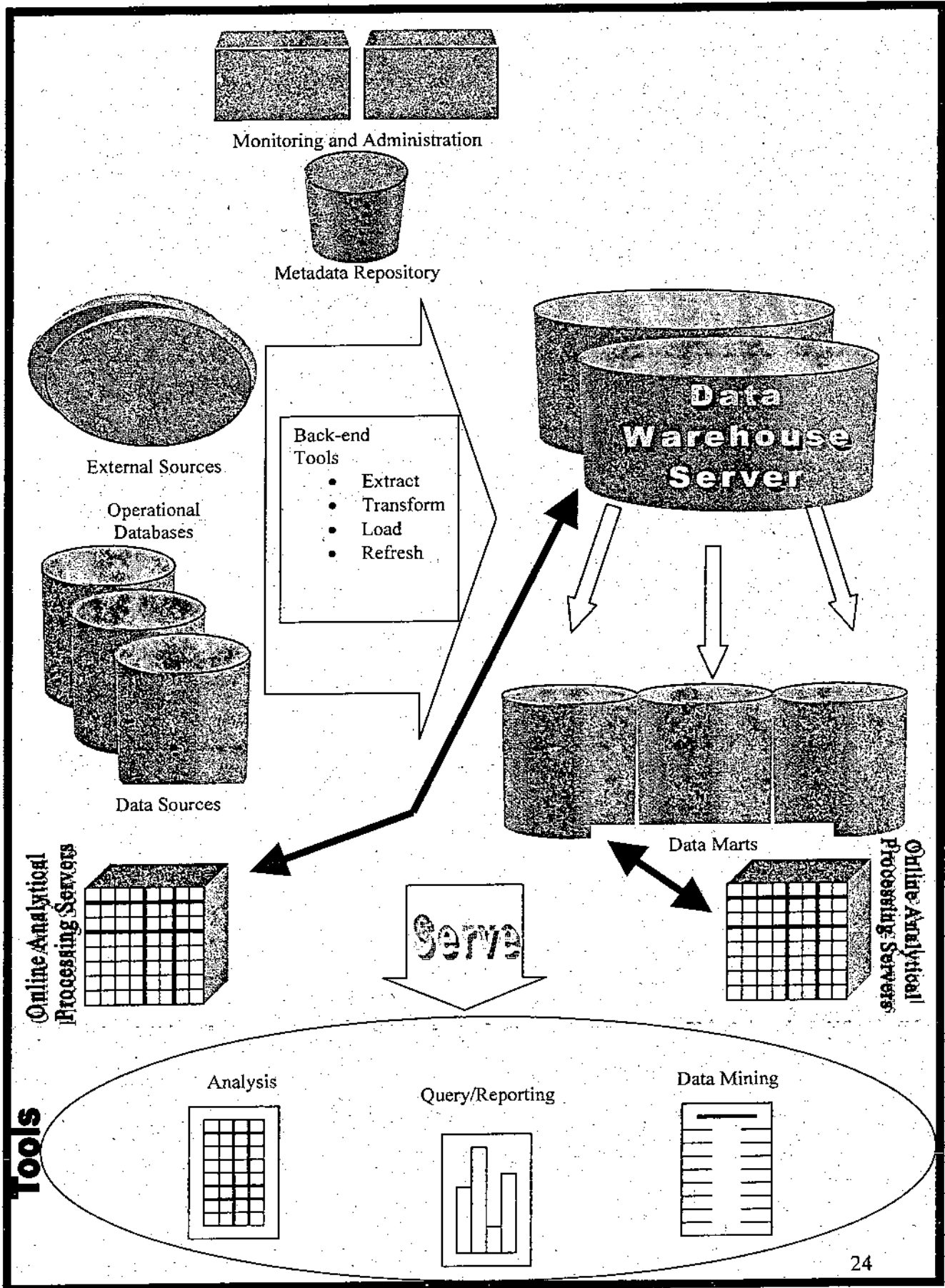
Οι εφαρμογές OLAP έχουν γίνει συνώνυμα με τη πολυδιάστατη παρουσίαση των δεδομένων. Αυτή η πολυδιάστατη παρουσίαση ενισχύεται και υποστηρίζεται από τις πολυδιάστατες βάσεις δεδομένων παρέχοντας έτσι στις OLAP εφαρμογές τη βάση για τον υπολογισμό και την ανάλυση των δεδομένων. Η ανάγκη για πολυδιάστατη ανάλυση αναδεικνύει τις αποθήκες δεδομένων ως την κύρια πηγή άντλησης πληροφοριών.

### 2.1.3 Εξόρυξη Γνώσης :

Αφού λοιπόν έχουμε δημιουργήσει την αποθήκη δεδομένων και έχουμε εκμεταλλευτεί τις δυνατότητες που προσφέρει η τεχνολογία OLAP, μπορούμε να προχωρήσουμε ακόμα ένα βήμα και να ψάξουμε με εξελιγμένους αλγόριθμους για κρυμμένη πληροφορία που βρίσκεται στην αποθήκη δεδομένων. Η εξόρυξη γνώσης από αποθήκη δεδομένων είναι ότι πιο σύγχρονο χρησιμοποιούν οι αναλυτές σήμερα.

### Σχήμα 3

Η αρχιτεκτονική ενός Συστήματος Υποστήριξης Αποφάσεων που αποτελείται από τρία κύρια μέρη: έναν data Warehouse server, εργαλεία ανάλυσης και εξόρυξης γνώσης καθώς και backend εργαλεία για την αποθήκη δεδομένων.





Τώρα λοιπόν που υπάρχει μια σχετική εξοικείωση με τους μέχρι τώρα άγνωστους όρους μπορούμε να δούμε καλύτερα από το σχήμα το πώς μπορεί ένα ΣΥΑ να στηρίζεται σε μια αποθήκη δεδομένων που μπορεί να αποτελείται από πολλά data marts και η οποία γεμίζει με στοιχεία που προέρχονται μετά από επεξεργασία των βάσεων δεδομένων της εταιρίας ή από άλλες εξωτερικές πηγές(από το internet για παράδειγμα). Θέματα που έχουν να κάνουν με φυσική διαχείριση της αποθήκης και των μεταδεδομένων της δεν θα μας απασχολήσουν. Παρατηρούμε όμως ότι η αποθήκη δεδομένων μπορεί να εξυπηρετήσει τόσο την εφαρμογή OLAP πράξεων καθώς επίσης και την εξόρυξη γνώσης.

Η παραπάνω συνοπτική παρουσίαση ασφαλώς δεν έχει αγγίξει βαθύτερα θέματα των τεχνολογιών αυτών. Έγινε μια πρώτη εισαγωγή για να μπορεί ο αναγνώστης να παρακολουθήσει την ακόλουθη ιεραρχική(από χαμηλό σε υψηλότερο επίπεδο) υλοποίηση ενός ΣΥΑ.

Για να γίνουμε πιο συγκεκριμένοι, θα προχωρήσουμε με **θέματα σχεδίασης και εισαγωγής στοιχείων** σε μια αποθήκη δεδομένων. Θα αναφερθούμε σε θέματα εφαρμογής **OLAP πράξεων** πάνω στην αποθήκη αφού αυτές εξυπηρετούν την υποστήριξη αποφάσεων και θα φθάσουμε να δούμε **αλγόριθμους εξόρυξης γνώσης** όπου έχουν την ικανότητα να φανερώνουν σχέσεις και εξαρτήσεις που δεν είναι ορατές.

## 2.2 Το μέσο: οι Αποθήκες Δεδομένων

**Η αποθήκευση δεδομένων (Data Warehousing) παρέχει και εξασφαλίζει ένα είδος μνήμης στην επιχείρηση.**

Μια μικρή επιχείρηση χτίζει τις σχέσεις με τους πελάτες της με το να κατανοεί τις ανάγκες τους, να θυμάται τις προτιμήσεις τους και να μαθαίνει από τις προηγούμενες αλληλεπιδράσεις πώς να τους εξυπηρετήσει καλύτερα στο μέλλον. Πώς μπορεί μια μεγάλη επιχείρηση να ολοκληρώσει κάτι παρόμοιο όταν οι περισσότεροι πελάτες δεν μπορούν να αλληλεπιδράσουν προσωπικά με τους υπάλληλους επιχείρησης; Ακόμη και εκεί όπου υπάρχει αλληλεπίδραση πελατών, είναι πιθανό να είναι με διαφορετικό υπάλληλο πωλήσεων ή ανώνυμο υπάλληλο τηλεφωνικού κέντρου κάθε φορά. Οπότε πώς μπορεί η επιχείρηση να παρατηρήσει, να θυμάται και να μαθαίνει από αυτές τις

αλληλεπιδράσεις; Τι είναι αυτό που μπορεί να αντικαταστήσει η δημιουργική διαίσθηση του “μοναδικού” ιδιοκτήτη ο οποίος αναγνωρίζει τους πελάτες από το όνομα, το πρόσωπο και τη φωνή, και θυμάται τις συνήθειες και τις προτιμήσεις τους;

Με μια λέξη, τίποτα. Αλλά μέσω της έξυπνης εφαρμογής της τεχνολογίας πληροφοριών, ακόμη και η μεγαλύτερη επιχείρηση μπορεί να έρθει εκπληκτικά κοντά στο να το κάνει πραγματικότητα. Στις μεγάλες εμπορικές επιχειρήσεις, το πρώτο βήμα- παρατηρώντας το τι κάνει ο πελάτης- έχει ήδη κατά ένα μεγάλο μέρος αυτοματοποιηθεί. Τα συστήματα επεξεργασίας δεδομένων σε πραγματικό χρόνο (OLTP) βρίσκονται παντού, συλλέγοντας στοιχεία που αφορούν σχεδόν τα πάντα.

Αυτές τις μέρες, όλοι έχουμε σε μια “ρουτίνα” ζωής που παράγει ένα σταθερό ρεύμα αρχείων συναλλαγής. Όταν παίρνετε το τηλέφωνο για να παραγγείλετε ένα ανταλλακτικό για το αυτοκίνητό σας από την ανάλογη αντιπροσωπεία ή ένα γεύμα από το αγαπημένο σας εστιατόριο, ένα αρχείο συναλλαγής δημιουργείται στην τοπική τηλεφωνική επιχείρηση (πχ. ΟΤΕ, TELEDOME, FORTHNET) που παρουσιάζει το χρόνο της κλήσης σας, τον αριθμό που σχηματίσατε, και η μεγάλης απόστασης επιχείρηση με την οποία έχετε συνδεθεί (πχ. αντιπροσωπεία BMW, εστιατόρια PIZZA HAT). Στη μεγάλης απόστασης επιχείρηση, παράγονται επιπλέον αρχεία καταγράφοντας τη διάρκεια της κλήσης σας και της ακριβούς δρομολόγησης της μέσω του συστήματος μεταγωγής-διακλάδωσης των τηλεφωνικών γραμμών. Αυτά τα στοιχεία θα συνδυαστούν με άλλα αρχεία που συγκροτούν την αναλυτική κατάσταση τηλεφωνικών κλήσεων, του χρόνου χρέωσης, το όνομα, και τη διεύθυνση τιμολόγησής σας προκειμένου να δημιουργηθεί και να εκδοθεί ο πληρωτέος λογαριασμός σας. Στην επιχείρηση καταλόγων, η κλήση σας καταγράφεται πάλι μαζί με τις πληροφορίες για τον συγκεκριμένο κατάλογο από τον οποίο παραγγείλατε και οποιεσδήποτε ειδικές διαφημίσεις για τις οποίες ενδιαφερθήκατε. Όταν ο αντιπρόσωπος εξυπηρέτησης πελατών που απάντησε στην κλήση σας, ζητά τον αριθμό της πιστωτικής σας κάρτας και την ημερομηνία λήξης αυτής, οι πληροφορίες αναμεταδίδονται αμέσως σε ένα σύστημα επαλήθευσης πιστωτικών καρτών για να εγκρίνουν τη συναλλαγή. Σε σύντομο χρονικό διάστημα, η συναλλαγή φθάνει στην τράπεζα που εξέδωσε την πιστωτική κάρτα σας όπου θα εμφανιστεί στην επόμενη μηνιαία ανάλυση κίνησης του εκάστοτε λογαριασμού σας. Όταν η παραγγελία σας, με τον κωδικό προϊόντος, το μέγεθος, και το χρώμα,

πηγαίνει στο σύστημα εισόδων παραγγελιών του καταλόγου, θα παράγει ακόμα περισσότερα αρχεία στο σύστημα τιμολόγησης και το σύστημα ελέγχου καταλόγων απογραφής. Εάν η παραγγελία σας έχει γίνει μέσω του διαδικτύου, εντός ολίγων ωρών, παράγονται επιπλέον αρχεία συναλλαγής σε ένα συγκρότημα ηλεκτρονικών υπολογιστών (ανάλογα την εταιρεία από την οποία παραγγείλατε το προϊόν που σας ενδιέφερε) όπου μπορεί να ανιχνευθεί πολλές φορές μεταξύ της αποθήκης εμπορευμάτων και του σπιτιού σας. Αυτά τα αρχεία επιτρέπουν σε σας να καλέσετε έναν τηλεφωνικό αριθμό εξυπηρέτησης πελατών ή να ελέγξετε την ιστοσελίδα του ναυλωτή για να ακολουθήσετε την πρόοδο της παραγγελίας σας.

Η εστιάζουσα στον πελάτη επιχείρηση θεωρεί κάθε αρχείο μιας αλληλεπίδρασης με έναν πελάτη ή πιθανό αγοραστή - κάθε κλήσης στην υποστήριξη πελατών, κάθε συναλλαγής σημείου πώλησης, κάθε σειρά καταλόγων, κάθε επίσκεψη σε ιστοσελίδα της επιχείρησης -ως ευκαιρία εκμάθησης. Αλλά, η εκμάθηση απαιτεί περισσότερο από απλή συλλογή στοιχείων. Στην πραγματικότητα, πολλές επιχειρήσεις συλλέγουν εκατοντάδες gigabytes των στοιχείων από και για τους πελάτες τους χωρίς να αποκτούν γνώση για τίποτα! Τα δεδομένα συγκεντρώνονται επειδή απαιτούνται για κάποιο λειτουργικό σκοπό, όπως ο έλεγχος καταλόγων ή η τιμολόγηση. Και, μόλις εξυπηρετήσει εκείνο τον σκοπό, "μαραζώνουν" σε κάποιο φάκελο αρχειοθέτησης, σε δισκέτες ή πετάγονται στα απορρίμματα.

Για πραγματοποιηθεί η εκμάθηση, δεδομένα από πολλές πηγές - αρχεία τιμολόγησης, στοιχεία ανιχνευτών, αιτήσεις εγγραφής, εφαρμογές, αρχεία κλήσης, εξαγορές δελτίων, έρευνες - πρέπει πρώτα να μαζευτούν μαζί και να οργανωθούν με έναν συνεπή και χρήσιμο τρόπο. Αυτό καλείται Αποθήκευση Στοιχείων - Data Warehousing. Η αποθήκευση στοιχείων επιτρέπει στην επιχείρηση να θυμάται το τι έχει παρατηρήσει για τους πελάτες της. Έπειτα, τα στοιχεία πρέπει να αναλυθούν, να γίνουν κατανοητά, και να μετατραπούν σε ενεργές - επεξεργάσιμες πληροφορίες. Από το σημείο αυτό αρχίζει η εφαρμογή της Εξόρυξης Γνώσης (Data Mining) στην οποία και θα αναφερθούμε αναλυτικότερα σε στις προσεχείς παραγράφους.

Από τα μέσα της δεκαετίας του '70, η αλματώδης παραγωγή πολύ ισχυρών συστημάτων διαχείρισης βάσεων δεδομένων βοήθησε στην ανάπτυξη πληροφοριακών συστημάτων που καλύπτουν τις λειτουργικές ανάγκες οργανισμών και επιχειρήσεων. Τα

μεγαλύτερα και ισχυρότερα συστήματα αναπτύχθηκαν με στόχο τον αυτοματισμό βασικών αναγκών των οργανισμών όπως η διεκπεραίωση των τραπεζικών εργασιών και τα λογιστικά συστήματα. Η λειτουργία αυτών των πληροφοριακών συστημάτων είναι πλέον κρίσιμη και πολύτιμη για τη ζωή των οργανισμών στους οποίους έχουν εγκατασταθεί, η δε βάση δεδομένων ενός τέτοιου συστήματος αποτελεί τον πυρήνα τους. Η ορθή σχεδίαση, ανάπτυξη και λειτουργία της βάσης είναι ο σημαντικότερος παράγοντας για την επιτυχία ενός πληροφοριακού συστήματος. Τα συστήματα αυτά παρέχουν τη δυνατότητα επεξεργασίας μεγάλου αριθμού δοσοληψιών που διαχειρίζονται τα δεδομένα του οργανισμού (OLTP). Ένα άλλο είδος πληροφοριακών συστημάτων που αναπτύσσονται στους οργανισμούς είναι τα συστήματα υποστήριξης αποφάσεων που σκοπό έχουν να βοηθήσουν τα στελέχη των οργανισμών να σχεδιάσουν τις δραστηριότητές του. Η επιτυχία των συστημάτων αυτών είναι επίσης βασικός παράγοντας επιτυχίας του οργανισμού. Μία βασική απαίτηση των συστημάτων υποστήριξης αποφάσεων είναι η αποδοτική πρόσβαση στα δεδομένα των συστημάτων αυτοματισμού. Το πρόβλημα που προκύπτει, όμως, είναι ότι τα συστήματα αυτοματισμού έχουν ήδη πολύ σοβαρό υπολογιστικό φορτίο από μόνα τους και επιπλέον, είναι σχεδιασμένα για την εκτέλεση διαφορετικών λειτουργιών.

Ένας τηλεπικοινωνιακός οργανισμός, για παράδειγμα, συνήθως διαθέτει ένα μεγάλο πληροφοριακό σύστημα ελέγχου του τηλεφωνικού δικτύου του. Αυτό το σύστημα ελέγχει την ομαλή λειτουργία του δικτύου και παράλληλα των παροχή υπηρεσιών και την χρέωση των συνδρομητών του. Η βάση δεδομένων του συστήματος περιέχει όλα τα δεδομένα των παραπάνω εργασιών. Είναι σαφές ότι αυτό το σύστημα λειτουργεί συνεχώς (24 ώρες ημερησίως) με μεγάλο όγκο δοσοληψιών (transactions) να εξυπηρετούνται στη βάση δεδομένων. Από αυτή τη βάση θα πρέπει να αντλήσει και ένα σύστημα υποστήριξης αποφάσεων τα απαραίτητα δεδομένα, για να μπορέσει να βοηθήσει στο σχεδιασμό της λειτουργίας του οργανισμού. Μελετώντας λίγο πιο προσεκτικά την περίπτωση αυτή, θα δούμε ότι είναι πρακτικά αδύνατο το σύστημα ελέγχου του δικτύου και το σύστημα υποστήριξης αποφάσεων να λειτουργούν, χρησιμοποιώντας την ίδια βάση δεδομένων. Διάφορα προβλήματα κάνουν αδύνατη την εφαρμογή αυτού του σεναρίου. Τα κυριότερα από αυτά τα προβλήματα είναι τα παρακάτω:

I. Τα δύο συστήματα αναπτύχθηκαν πιθανότατα από διαφορετικούς ανθρώπους και κυρίως με τη χρήση διαφορετικών τεχνολογιών. Είναι πιθανό η τεχνολογία του συστήματος αποφάσεων να αδυνατεί να επιτρέψει άμεση πρόσβαση (on-line) στη βάση δεδομένων του συστήματος ελέγχου του δικτύου. Πολύ συχνά, σε μεγάλα συστήματα, όπως στην προκειμένη περίπτωση, το σύστημα ελέγχου του δικτύου έχει αναπτυχθεί με τη χρήση παρωχημένης τεχνολογίας, όπως, για παράδειγμα, αρχαία COBOL. Εφαρμογές που χρησιμοποιούν μοντέρνα τεχνολογία αντιμετωπίζουν προβλήματα στο να διαχειριστούν πληροφορία που προέρχεται από μια βάση δεδομένων παλαιάς τεχνολογίας.

II. Η βάση δεδομένων του συστήματος ελέγχου του δικτύου σχεδιάστηκε με βάση αποκλειστικά τις απαιτήσεις αυτής της εφαρμογής. Βασικό χαρακτηριστικό σε εφαρμογές αυτού του είδους είναι η όσο το δυνατό αποδοτικότερη ικανοποίηση μικρών δοσοληψιών που εισάγουν ή τροποποιούν πολύ μικρό αριθμό εγγραφών της βάσης. Μία τυπική δοσοληψία που θα αφορούσε τη χρέωση μίας υπεραστικής συνδιάλεξης θα εισήγαγε μία εγγραφή με τον κωδικό του συνδρομητή και τη διάρκεια της συνδιάλεξης. Στη σχεδίαση μιας τέτοιας βάσης δεδομένων, με την εφαρμογή των κανόνων κανονικοποίησης, καταλήγουμε σε μεγάλο αριθμό από πίνακες που ο κάθε ένας έχει περιορισμένο αριθμό πεδίων. Σε αντίθεση με τα παραπάνω, μία εφαρμογή που αντλεί στοιχεία λειτουργίας του δικτύου για λόγους ανάλυσης και λήψης αποφάσεων, δεν κάνει καμία αλλαγή στη βάση του δικτύου αλλά απαιτεί αποδοτική απόκριση από το σύστημα στις ερωτήσεις που θέτει. Αυτές οι ερωτήσεις συνήθως απαιτούν πρόσβαση σε μεγάλο αριθμό δεδομένων, θέτοντας διαφορετικούς κανόνες σχεδίασης της βάσης δεδομένων του συστήματος. Για μία ερώτηση σχετική με τη στρατηγική του οργανισμού, που θα είχε πρόσβαση σε μεγάλο αριθμό δεδομένων, το κόστος σε μία βάση με πολλούς πίνακες θα ήταν σημαντικό, καθώς θα έπρεπε να εκτελεστεί μεγάλο αριθμός από πράξεις join μεταξύ των πινάκων αυτών.

III. Κάθε σύστημα υποστήριξης αποφάσεων εκτελεί μεγάλο αριθμό ερωτήσεων, θα δεσμεύσει μεγάλο αριθμό πόρων του συστήματος διαχείρισης της βάσης δεδομένων με αποτέλεσμα να μειώσει την απόδοση του συστήματος ελέγχου του δικτύου. Για παράδειγμα, μία ερώτηση σχετική με τις χρεώσεις των πελατών του δικτύου, που απαιτεί κάποιο χρονικό διάστημα για να εκτελεστεί, θα κλείδωνε τον πίνακα με τις χρεώσεις των πελατών εμποδίζοντας οποιαδήποτε μεταβολή από το σύστημα ελέγχου (π.χ. μια νέα χρέωση).

Από τα παραπάνω γίνεται σαφές ότι είναι εξαιρετικά δυσχερής η χρήση των βάσεων δεδομένων των πληροφοριακών συστημάτων των οργανισμών από τα συστήματα υποστήριξης αποφάσεων. Όμως, η αποδοτική χρήση των συστημάτων υποστήριξης αποφάσεων απαιτεί όπως προαναφέρθηκε, πρόσβαση σε αυτά τα δεδομένα. Η εισαγωγή των αποθηκών δεδομένων είναι η λύση στο κρίσιμο αυτό πρόβλημα.

### 2.2.1 Σχεδίαση Αποθηκών Δεδομένων

Το μεγαλύτερο ποσοστό των επιχειρήσεων επιλέγουν τον σχεδιασμό και την εφαρμογή κεντρικών και ανεξάρτητων αποθηκών δεδομένων. Οι σχεδιαστές των συστημάτων αποσκοπούν σε:

- Μία απλή και ανεξάρτητη εφαρμογή που να διαχειρίζεται εύκολα μεγάλο όγκο δεδομένων
- Συγκεντρωτική παρουσίαση των δεδομένων της επιχείρησης στους αποφασίζοντες, συλλέγοντας τα διάσπαρτα δεδομένα από όλες τις πτυχές της.
- Διαχωρισμό ανά κατηγορία και θέμα των δεδομένων (πολλές φορές σε διαφορετικές αποθήκες δεδομένων); αφού πολλές φορές η συγκέντρωση σε μία και μόνη αποθήκη δυσχεραίνει την πρόσβαση.

Είναι αυτονόητο ότι μία εφαρμογή αποθήκης δεδομένων -η ακόμα περισσότερο μία έτοιμη εμπορική εφαρμογή- δεν ταιριάζει σε μία επιχείρηση. Για αυτό τον λόγο και

τις περισσότερες φορές ο σχεδιασμός από την αρχή είναι επιβεβλημένος. Πολλές εταιρίες παρέχουν την δυνατότητα προσαρμογής ήδη υπαρχόντων εφαρμογών, ανάλογα με τις απαιτήσεις της κάθε επιχείρησης.

Σε κάθε περίπτωση όμως, οι παράγοντες που θα πρέπει να λαμβάνονται υπόψη είναι:

1. Πολλές επιχειρήσεις γνωρίζουν την αναγκαιότητα μιας αποθήκης δεδομένων, άλλοι δεν είναι σε θέση να θέσουν τις προτεραιότητες και τις προδιαγραφές, πράγματα που μόνο οι εταιρίες εφαρμογών μπορούν να ξέρουν και να θέσουν. Σαν τέτοιες χαρακτηρίζονται το μέγεθος, η εγκατάσταση, η συχνότητα χρήσης και η συντήρηση. Για πολλές εταιρίες, όμως, μία ολοκληρωμένη μελέτη και εφαρμογή κρίνεται αντιοικονομική.
2. Η κατανόηση του ήδη υπάρχοντος διαχειριστικού συστήματος είναι ένας σημαντικός παράγοντας. Πάνω σε αυτό το σύστημα θα στηριχτεί η όποια εφαρμογή αποθήκης δεδομένων. Έτσι, πρώτος στόχος είναι να διασαφηνιστεί η ακριβής προέλευση των δεδομένων. Έτσι, πρώτος στόχος είναι να διασαφηνιστεί η ακριβής προέλευση των δεδομένων και της δομής τους, ώστε να επιλεγθούν τα κατάλληλα εργαλεία για την εφαρμογή.
3. Η ανάγκη για μετακίνηση των δεδομένων είναι ένας άλλος παράγοντας που θα πρέπει να εξεταστεί. Πολλές φορές κρίνεται συμφέρουσα και οικονομική η παραμονή των δεδομένων στις αρχικές βάσεις δεδομένων. Τα κριτήρια για την μετακίνηση ή όχι των δεδομένων είναι:
  - Η ποιότητα των δεδομένων
  - Το μέγεθος και η χρηστικότητα τους
  - Η δομή τους
  - Η δυνατότητα να εφαρμοστούν στη νέα εφαρμογή
  - Η ευκολία πρόσβασης
  - Η αναγκαιότητα τους στο ήδη υπάρχον σύστημα

4. Ένα πολύ σημαντικό σημείο στη μελέτη και τον σχεδιασμό είναι και ο τελικός αποδέκτης των μετακινούμενων δεδομένων. Πριν καν μελετηθεί η μετακίνηση των δεδομένων, πρέπει να γίνει ουσιαστικά ο σχεδιασμός της δομής των δεδομένων που θα συμπεριληφθούν στην αποθήκη δεδομένων.
5. Για να γίνει περισσότερο κατανοητή η χρήση της αποθήκης δεδομένων πρέπει να εντάσσεται σε ένα πληροφοριακό σύστημα διοίκησης ή σε ένα σύστημα υποστήριξης αποφάσεων. Τα ήδη υπάρχοντα συστήματα είναι ένας παράγοντας που πρέπει να λαμβάνεται υπόψη κατά τον σχεδιασμό μιας αποθήκης δεδομένων.
6. Ο βαθμός μετατροπής των δεδομένων, καθώς και τα εργαλεία που απαιτούνται για αυτό.
7. Η επιλογή των κατάλληλων εργαλείων (στην περίπτωση όπου δεν υπάρχει εφαρμογή ΣΥΑ) για την επεξεργασία των δεδομένων της αποθήκης δεδομένων. Τις περισσότερες των περιπτώσεων απαιτούνται διαφορετικά εργαλεία, ανάλογα σε τι είδους χρήση απευθύνεται (για παράδειγμα σε έμπειρους χρήστες αποθηκών δεδομένων, βελτιωτές εφαρμογών, υψηλόβαθμα στελέχη της επιχείρησης κ.α.).
8. Καθορισμός των κανόνων χρήσης και λειτουργίας της αποθήκης δεδομένων. Καθορισμός των χρηστών με πρόσβαση μερική ή ολική σε αυτή.

Οι παραπάνω παράγοντες δεν είναι οι μοναδικοί για τον σχεδιασμό μιας αποθήκης δεδομένων. Η ευελιξία που δίνεται στον σχεδιαστή μέσω των διάφορων τεχνολογικών εφαρμογών που ποικίλουν πλέον στην αγορά, καθώς και η πληθώρα των λογισμικών που είναι διαθέσιμα για τον σχεδιασμό βάσεων δεδομένων, έχουν ως αποτέλεσμα την εμφάνιση πολλών και διαφορετικών αποθηκών δεδομένων.

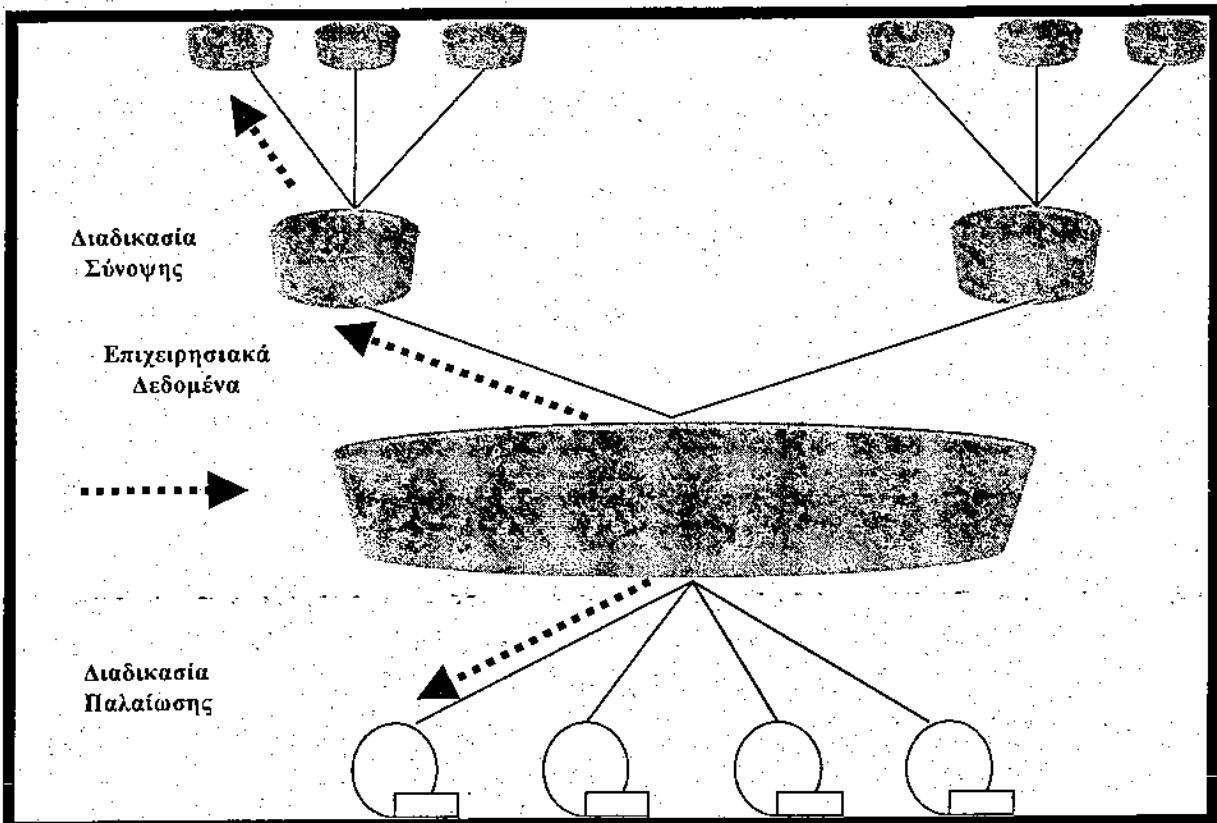


### 2.2.2 Λειτουργία της αποθήκης δεδομένων

Στο σχήμα 4 παρουσιάζεται η ροή των δεδομένων από την αρχική πηγή τους μέχρι τον χρήστη, συμπεριλαμβανομένων και των εφαρμογών που παρεμβάλλονται. Λόγω των ετερογενών πηγών των δεδομένων (διαφορετικές διαχειριστικές βάσεις δεδομένων), και λοιπών προβλημάτων που προκύπτουν σε τέτοια ανομοιογενή συστήματα, παρεμβάλλονται μηχανισμοί που τα αποκαθιστούν, μετατρέποντας τα και παραδίδοντας τα στις αποθηκευτικές βάσεις δεδομένων. Αυτός ο μηχανισμός βασίζεται σε προεπιλεγμένο μοντέλο (σε αυτό βασίζεται ουσιαστικά και ο σχεδιασμός των αποθηκών δεδομένων). Κύριο συστατικό – περιγράφει το μοντέλο ενώ παράλληλα διευκρινίζει τα στοιχεία των δεδομένων-είναι το metadata.

Σχήμα 4

Ροή δεδομένων μέσα σε μια αποθήκη δεδομένων



### 2.2.3 Ροή δεδομένων

Τα περισσότερα δεδομένα εισέρχονται μέσα στην αποθήκη δεδομένων απ' ευθείας από τις διαχειριστικές βάσεις δεδομένων. Στο σχήμα παρουσιάζεται η ροή των δεδομένων μέσα σε μια αποθήκη δεδομένων, ανάλογα με τη χρήση που υπόκεινται.

Έτσι, τα δεδομένα διακρίνονται σε:

- Διαγραφέντα ( τα δεδομένα προς διαγραφή, μικρή χρηστικότητα )
- Συνοπτικά ( είναι δεδομένα που χρησιμοποιούνται σε μεγαλύτερη συχνότητα)
- Δεδομένα-αρχεία.

Η διαδικασία “παλαιώσης” (aging process) σε μία βάση δεδομένων μετατρέπει τα δεδομένα ανάλογα με την ηλικία τους από σύγχρονα δεδομένα (current data) σε παλαιά (older data) και για αυτό τον λόγο, τις περισσότερες φορές τα αποθέτει σε αποθηκευτικούς δίσκους. Η διαδικασία σύνοψης (summarized process) χρησιμοποιεί προσεκτικά επιλεγμένα δεδομένα με σκοπό να τα μετατρέψει σε συνοπτικά δεδομένα (lightly summarized data), είτε πολύ συνοπτικά δεδομένα (hightly summarized data).

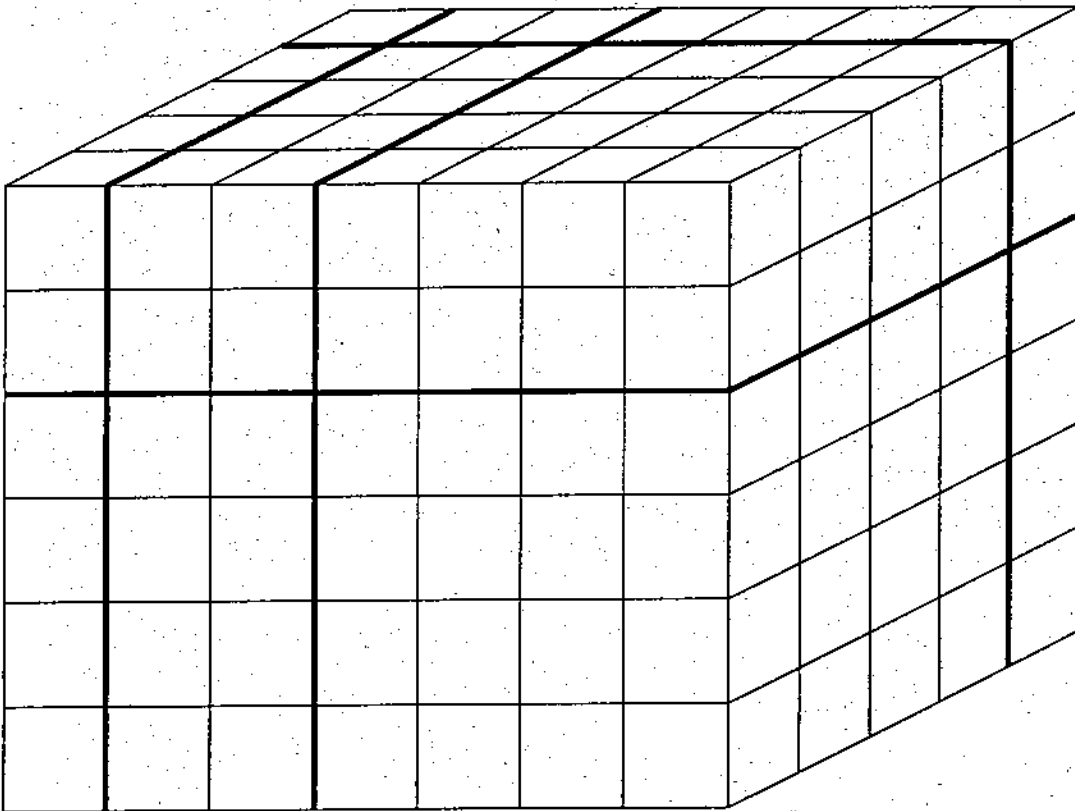
Όσο πιο συγκεντρωτικά και συνοπτικά είναι τα δεδομένα σε μία αποθήκη δεδομένων, τόσο ταχύτερη και αποτελεσματικότερη είναι η πρόσβαση για τη χρήση τους.

### 2.2.4 Η αποθήκη δεδομένων όπως την γνωρίζουν οι αναλυτές

Όταν κάποιος ασχολείται με την αποθήκη δεδομένων μιας εταιρίας δεν βλέπει μπροστά του πίνακες συνδεδεμένους μεταξύ τους, όπως θα έκανε κάποιος ειδικός επί των βάσεων δεδομένων. Ανεξάρτητα με το πώς έχει γίνει η φυσική αποθήκευση των δεδομένων οι αναλυτές γνωρίζουν ότι έχουν μπροστά τους ένα κύβο με πολλές ίσως διαστάσεις, δηλαδή όπως λένε ένα υπερκύβο.

### Σχήμα 5

Ένας κύβος. Τα κελιά του αντιστοιχούν σε μια τιμή από κάθε διάσταση του.



Κάθε διάσταση ουσιαστικά αντιπροσωπεύει και μια παράμετρο που λαμβάνεται υπόψη κατά την ανάλυση. Έτσι για παράδειγμα ο παραπάνω κύβος θα μπορούσε να έχει τις εξής τρεις διαστάσεις: προϊόν, χρόνος και τοποθεσία αν ασχολούμασταν με την αποθήκη δεδομένων ενός super-market. Στις διαστάσεις αυτές κυριαρχεί η ιεραρχία. Δηλαδή στη διάσταση χρόνος μπορούμε να συναντήσουμε έτος – τρίμηνο – μήνας – ημέρομηνία.

Τα κελιά του υπερκύβου περιέχουν συγκεντρωτικά δεδομένα όπως έχει ήδη αναφερθεί. Αυτό σημαίνει πως εκεί υπάρχουν μέσοι όροι, σύνολα, μέγιστα και ελάχιστα κτλ. Το κάθε κελί αντιστοιχεί σε ένα συγκεκριμένο τόπο, για ένα συγκεκριμένο χρονικό διάστημα και για κάποιο από τα προϊόντα του super-market. Τα κελιά μπορούν να

γεμίσουν με την εκτέλεση κατάλληλων SQL ερωτήσεων, συγκεκριμένα SQL aggregated ερωτήσεων (SELECT AVG, COUNT, MIN, MAX), προς τη βάση δεδομένων.

Οι δυνατότητες υποστήριξης αποφάσεων που προσφέρουν οι αποθήκες δεδομένων είναι μεγάλες. Πολλά στελέχη επιχειρήσεων όταν βρίσκονται μακριά από την έδρα της εταιρίας τους, άρα και μακριά από τις βάσεις δεδομένων, φροντίζουν να έχουν στο φορητό υπολογιστή τους συγκεντρωτικά, ιστορικά στοιχεία για την τελευταία εβδομάδα για παράδειγμα ώστε να έχουν εικόνα της επιχείρησης. Με άλλα λόγια μια αποθήκη δεδομένων είναι εύκολα μεταφέρσιμη αφού το μέγεθος της μπορεί να καθοριστεί μιας και ο χρήστης επιλέγει σε τι βάθος των ενδιαφερόντων τα στοιχεία.

Θα μπορούσε κάποιος να αναρωτηθεί; Τώρα που έχουμε την αποθήκη δεδομένων πως θα μπορέσουμε να βγάλουμε συμπεράσματα. Υπάρχουν απλές πράξεις που μπορούν να γίνουν όπως να αναζητηθούν ο συνολικός αριθμός προϊόντων που πωλήθηκαν το μήνα Μάρτιο, στην κατηγορία ειδών διατροφής στην περιοχή της Πάτρας. Μπορούμε όμως να εφαρμόσουμε ποιο σύνθετες πράξεις όπως αυτές που περιγράφονται στην επόμενη παράγραφο με βαθύτερο σκοπό πάντα να πάρουμε μια καλύτερη εικόνα πριν προχωρήσουμε στην λήψη αποφάσεων.

### 2.2.5 OLAP Πράξεις στις αποθήκες δεδομένων

Η αποθήκη δεδομένων έχει ως κύριο χαρακτηριστικό την εύκολη πλοήγηση μεταξύ των σημαντικών δεδομένων που επιλέξαμε να φυλάξουμε εκεί. Τα σύγχρονα OLAP εργαλεία επιτρέπουν την εκτέλεση διάφορων πράξεων πάνω στον υπερκύβο και μάλιστα με πλήρως γραφικό τρόπο. Ο σκοπός βέβαια δεν είναι άλλος από την καλύτερη ανάλυση των δεδομένων και την εξυπηρέτηση ακόμα και των πιο απαιτητικών χρηστών.

- **Roll-up:** Κατάλληλη για εμφάνιση των δεδομένων σε διαφορετικό (ανώτερο) επίπεδο ιεραρχίας. Πιο συγκεκριμένα κάνοντας roll-up πάμε από λεπτομερή σε πιο συγκεντρωτικά δεδομένα. Για παράδειγμα αν ο χρήστης βλέπει τα καλύτερα, σε πωλήσεις, υποκαταστήματα supermarket ανά πόλη μπορεί κάνοντας roll-up να δει τα καλύτερα υποκαταστήματα και πάλι αλλά ανά χώρα.

- **Drill-down:** Ουσιαστικά πρόκειται για το αντίθετο του roll-up. Απλά σε αυτή την περίπτωση αντί να ανεβαίνει ο χρήστης σε πιο ομαδοποιημένο επίπεδο, αναζητά πιο λεπτομερή δεδομένα. Για παράδειγμα βλέποντας τους ετήσιους μέσους όρους πωλήσεων ενός προϊόν ανά χώρα να αναζητήσει να δει τι συμβαίνει ανά πόλη.

Κάποιος μπορεί πολύ εύκολα και χωρίς ιδιαίτερες κινήσεις να πραγματοποιήσει την πράξη drill-down. Συνήθως τα εργαλεία OLAP έχουν δίπλα από τα δεδομένα σε ανώτερο επίπεδο ιεραρχίας (όπως είναι για παράδειγμα οι ήπειροι όταν μιλάμε για τον υπερκύβο μιας πολυεθνικής εταιρίας) το σύμβολο +. Έτσι πατώντας το σύμβολο + δίπλα από τη λέξη «Ευρώπη»

[+] Ευρώπη

[+] Αμερική

[+] Αφρική

.....

.....

θα προκύψει το εξής οπτικό αποτέλεσμα...

[-] Ευρώπη

[+] Ελλάδα

[+] Γερμανία

[+] Μεγάλη Βρετανία

.....

.....

[+] Αμερική

[+] Αφρική

.....

Βέβαια όσο υπάρχουν και πιο χαμηλά επίπεδα ιεραρχίας θα εμφανίζεται το σύμβολο +. Πατώντας το σύμβολο - που εμφανίστηκε δίπλα από την λέξη «Ευρώπη» εκτελούμε την αντίθετη πράξη, δηλαδή κάνουμε roll-up.

- **Slicing:** Με αυτή την πράξη δίνουμε μια συγκεκριμένη τιμή σε μια από τις διαστάσεις του κύβου και ασχολούμαστε με τις υπόλοιπες. Για παράδειγμα αν σε ένα κύβο τριών διαστάσεων με διαστάσεις: χρόνος, τοποθεσία και προϊόν εφαρμοστεί η πράξη θεωρώντας ο χρήστης ότι τον ενδιαφέρει αποκλειστικά το έτος 2000 τότε ουσιαστικά μένει μια «φέτα» από όλο τον κύβο - ένας δυδιάστατος πίνακας.
- **Dicing:** Κατάλληλη για ερωτήσεις διαστήματος. Μπορεί κάποιος να ενδιαφέρεται για διάστημα ενός μηνός (παράδειγμα από 1/1 μέχρι 1/2). Τότε η διάσταση του χρόνου τεμαχίζεται σε 30 τμήματα που αναπαριστούν τις μέρες που ενδιαφέρουν τον χρήστη.
- **Pivoting:** Με αυτή την πράξη είναι δυνατή η αλλαγή της διάταξης των διαστάσεων. Ο σκοπός είναι να βοηθηθεί ο χρήστης στην ανάλυση. Δεν συντελείται καμία αλλαγή (ούτε αύξηση, ούτε μείωση) των δεδομένων του υπερκύβου.

Τα οπτικά αποτελέσματα που σε κάθε περίπτωση βλέπει ο χρήστης είναι εξαιρετικά φιλικά και καλαίσθητα ώστε να διευκολύνεται στο έργο της λήψης μιας απόφασης. Αποδεικνύεται λοιπόν έτσι ότι και η πολυδιάστατη ανάλυση (OLAP) αποτελεί σημαντικό κομμάτι της σύγχρονης υποστήριξης αποφάσεων.

Δεν αγγίξαμε όμως ακόμα την κορυφή. Δεν φτάσαμε στο σημείο που οι εταιρίες προσπαθούν να κατακτήσουν. Μιλήσαμε για δυνατότητες που προσφέρουν οι αποθήκες δεδομένων αλλά τα αποτελέσματα δεν φανερώνουν κρυμμένη γνώση που βρίσκεται μεταξύ των εκατομμυρίων εγγραφών της βάσης δεδομένων.

## 2.3 Το εργαλείο: η Εξόρυξη Γνώσης

Είδαμε λοιπόν με πιο τρόπο μπορούν οι σχετικές με τις αποθήκες δεδομένων τεχνολογίες να παίξουν πρωταγωνιστικό ρόλο στο χώρο της υποστήριξης αποφάσεων. Υπάρχουν όμως κατηγορίες αποφάσεων που για να ληφθούν σωστά απαιτούν να έχει ο αποφασίζων στη διάθεση του απαντήσεις και πληροφορίες που δεν είναι εύκολο να προκύψουν με τις τεχνολογίες που μέχρι τώρα αναλύθηκαν. Αυτές τις κατηγορίες αποφάσεων καλείται να υποστηρίξει η εξόρυξη γνώσης.

Η εξόρυξη γνώσης, δηλαδή η διαδικασία εφαρμογής μεθόδων ανάλυσης σε μεγάλο όγκο δεδομένων, είναι μια πολύ ισχυρή νέα τεχνολογία που μπορεί να βοηθήσει τις εταιρίες να εστιάσουν μόνο στα πιο σημαντικά δεδομένα των αποθηκών δεδομένων τους. Τα εργαλεία εξόρυξης γνώσης δίνουν τη δυνατότητα στο χρήστη να προφητεύει μελλοντικές συμπεριφορές και ροπές επιτρέποντας έτσι στις επιχειρήσεις να παίρνουν κατευθυνόμενες από τη γνώση αποφάσεις. Το αποτέλεσμα είναι να μπορούν τα εργαλεία αυτά να απαντήσουν σε επιχειρηματικές ερωτήσεις που παραδοσιακά απαιτούσαν πολύ χρόνο ανάλυσης.

Οι περισσότερες εταιρίες ήδη συλλέγουν και επεξεργάζονται τεράστιες ποσότητες δεδομένων. Οι τεχνικές εξόρυξης γνώσης μπορούν να αναπτυχθούν γρήγορα χωρίς να χρειάζονται αλλαγές στην υλικοτεχνική υποδομή και ως σκοπό έχουν την αξιοποίηση των πηγών πληροφοριών.

### 2.3.1 Τα αποτελέσματα της εξόρυξης γνώσης

Η εφαρμογή των μεθόδων εξόρυξης γνώσης αποσκοπεί στην ανακάλυψη πληροφοριών που είναι πολύ χρήσιμες για τις επιχειρήσεις. Πληροφορίες για **συσχετίσεις** όπως «όταν ένας πελάτης αγοράζει βίντεο τότε αγοράζει επίσης κάποια άλλη ηλεκτρονική συσκευή» ή για **τυποποιημένες μορφές** όπως «ο πελάτης που θα ψωνίσει περισσότερο από δύο φορές σε περίοδο εκπτώσεων είναι πιθανό να αγοράσει τουλάχιστο μία φορά κατά τη διάρκεια των Χριστουγέννων» αποτελούν πραγματικό θησαυρό για τους διοικούντες που μπορούν έτσι να αποφασίσουν για διάφορα θέματα λειτουργίας της επιχείρησής τους, όπως είναι το ωράριο, το ύψος και η διάρκεια των

εκπτώσεων και η τοποθέτηση των πραγμάτων μέσα στα καταστήματα αν βέβαια μιλάμε για εμπορικού τύπου επιχειρήσεις.

Τέτοιες πληροφορίες μπορούν επίσης να χρησιμοποιηθούν για τον προγραμματισμό χρήσης πρόσθετων αποθηκευτικών χώρων, για τον συνδυασμό διαφόρων πραγμάτων (βίντεο-ηλεκτρική σκούπα για παράδειγμα) στις διαφημίσεις ή για τη σχεδίαση ανάλογα την εποχή διαφορετικών στρατηγικών μάρκετινγκ.

### 2.3.2 Οι στόχοι της εξόρυξης γνώσης

Παρακάτω αναλύονται αυτά που μπορεί να προσφέρει η εξόρυξη. Τις δυνατότητες αυτές καλείται να εκμεταλλευτεί το μανάτζμεντ της εταιρίας ή ενός οργανισμού και να προχωρήσει σε αποφάσεις που θα μετατρέψουν τη γνώση σε χειροπιαστά αποτελέσματα. Αν το πετύχει τότε οι αρχές της επιχειρηματικής νοημοσύνης, που αποτελούν και την κεντρική ιδέα των συστημάτων υποστήριξης αποφάσεων, εφαρμόζονται και είναι σίγουρο ότι τα οφέλη θα είναι μεγάλα.

- Πρόβλεψη τάσεων και συμπεριφορών.

Δηλαδή η προσπάθεια ανακάλυψης κάποιων μελλοντικών συμπεριφορών ώστε να παρθούν οι κατάλληλες αποφάσεις με σκοπό τη μεγιστοποίηση του κέρδους ή την πρόληψη δυσμενών καταστάσεων. Παράδειγμα εφαρμογής της προφητικής εξόρυξης αποτελεί το λεγόμενο επιλεκτικό μάρκετινγκ που χρησιμοποιείται στο web mining και αναλύεται παρακάτω. Τα αποτελέσματα αυτού του είδους εξόρυξης μπορεί να είναι η πρόβλεψη για το που θα φτάσουν οι πωλήσεις ενός καταστήματος για μια συγκεκριμένη χρονική περίοδο και το αν το κλείσιμο μιας γραμμής παραγωγής προϊόντων θα ενεργούσε θετικά σε ότι αφορά αυτές (τις πωλήσεις). Σε επιστημονικό επίπεδο, η μελέτη παλαιότερων σεισμικών ακολουθιών μπορεί ίσως να οδηγήσει στην πρόβλεψη, με υψηλά ποσοστά επιτυχίας, σεισμικής δραστηριότητας.

- Αναγνώριση.

Οι τυποποιημένες μορφές ανάμεσα στα δεδομένα μπορούν να χρησιμοποιηθούν για να αποκαλύψουν την ύπαρξη ενός γεγονότος, μια δραστηριότητας. Για



παράδειγμα οι εισβολείς στη προσπάθεια να σπάσουν ένα σύστημα ασφαλείας μπορούν να αναγνωριστούν από τα προγράμματα που εκτέλεσαν, τα αρχεία που προσπέλασαν και τον χρόνο που απασχόλησαν την CPU.

- **Ταξινόμηση.**

Η εξόρυξη γνώσης μπορεί να διαχωρίσει έτσι τα δεδομένα ώστε να προκύψουν διαφορετικές κλάσεις ή κατηγορίες βάση κάποιων παραμέτρων. Για παράδειγμα οι πελάτες ενός super-market μπορούν να χωριστούν σε κατηγορίες, όπως φίλων-εκπτώσεων, παρορμητικοί, πιστοί-κανονικοί, και σπάνιοι πελάτες. Αυτή η κατηγοριοποίηση μπορεί να χρησιμοποιηθεί στην ανάλυση των πωλήσεων ώστε να μπορεί για παράδειγμα ο μάνατζερ να λάβει αποφάσεις για να προσελκύσει σε μεγαλύτερο βαθμό κάποια από τις παραπάνω κατηγορίες.

- **Βελτιστοποίηση.**

Ένας τελικός στόχος της εξόρυξης γνώσης μπορεί να είναι και η βέλτιστη χρήση περιορισμένων πόρων όπως είναι ο χρόνος, ο χώρος, το χρήμα ή τα υλικά και η μεγιστοποίηση, κάτω από ορισμένους περιορισμούς, κάποιων «ποσοτήτων» όπως είναι οι πωλήσεις ή τα κέρδη. Έτσι σε ότι αφορά τουλάχιστον αυτό το στόχο, η εξόρυξη γνώσης έχει κοινά στοιχεία με την επιχειρησιακή έρευνα που επίσης ασχολείται με θέματα βελτιστοποίησης κάτω από περιορισμούς.

### **2.3.3 Οι κυριότερες τεχνικές**

Τα εργαλεία εξόρυξης γνώσης συνήθως δίνουν τη δυνατότητα στον χρήστη να επιλέξει ποια τεχνική-αλγόριθμο θέλουν να εφαρμόσουν. Παρακάτω γίνεται μια σύντομη γνώριμία με αυτούς τους αλγόριθμους.

- **Τα νευρωνικά δίκτυα:**

Μη γραμμικά, προφητικά και μπορούν να εκπαιδευτούν. Σε ότι αφορά τη δομή μοιάζουν στα βιολογικά νευρωνικά δίκτυα.

- **Δένδρα απόφασης:**

Δενδρικές δομές που αναπαριστούν σύνολα απόφασης. Αυτές οι αποφάσεις γεννούν κανόνες για τη ταξινόμηση ενός συνόλου δεδομένων.

- **Γενετικοί αλγόριθμοι:**

Τεχνικές βελτιστοποίησης που χρησιμοποιούν διαδικασίες όπως γενετικοί συνδυασμοί, μετάλλαξη.

- **Επαγωγή κανόνα:**

Η εξαγωγή χρήσιμων, και με στατιστική σημασία, if-then κανόνων από τα δεδομένα.

### 2.3.4 Πως λειτουργεί η εξόρυξη γνώσης

Πως ακριβώς μπορεί να μας πει η εξόρυξη γνώσης πράγματα που δεν ξέρουμε ή που θα συμβούν στο μέλλον; Η τεχνική που χρησιμοποιείται για να επιτευχθούν αυτά λέγεται μοντελοποίηση. Με άλλα λόγια η σκέψη του κτισίματος ενός μοντέλου για μια κατάσταση όπου γνωρίζουμε την απάντηση και στη συνέχεια η εφαρμογή του σε μια άλλη που δεν τη ξέρουμε. Για παράδειγμα, αν αναζητούσαμε μια βυθισμένη ισπανική γαλέρα στην ανοικτή θάλασσα το πρώτο πράγμα που ίσως σκεφτόμασταν θα ήταν να ερευνήσουμε όλες τις περασμένες περιπτώσεις εύρεσης ισπανικών θησαυρών από άλλους. Ίσως λοιπόν να παρατηρούσαμε ότι αυτά τα πλοία στην πλειονότητα τους βρέθηκαν στις ακτές Βερμούδα και ότι υπήρχαν κάποιες βέβαιες πορείες που ακολουθούσαν οι καπετάνιοι των πλοίων αυτών κείνη την εποχή. Αυτές οι ομοιότητες σημειώνονται και κτίζεται ένα μοντέλο που περιλαμβάνει τα χαρακτηριστικά που είναι κοινά στις τοποθεσίες αυτών των βυθισμένων θησαυρών. Με αυτό το μοντέλο αρχίζει το ψάξιμο σε περιοχές που δείχνει αυτό ότι είναι πιθανό να υπήρξε μια παρόμοια κατάσταση στο παρελθόν. Αν το μοντέλο είναι καλό ο θησαυρός θα βρεθεί.

Επομένως τη σκέψη του κτισίματος μοντέλων την είχαν οι άνθρωποι εδώ και πολύ καιρό και σίγουρα πριν την έλευση των υπολογιστών και της τεχνολογίας της εξόρυξης γνώσης. Πάντως αυτό που συμβαίνει στους υπολογιστές δεν διαφέρει πολύ από τον τρόπο με τον οποίο οι άνθρωποι κτίζουν μοντέλα. Οι υπολογιστές φορτώνονται με πληροφορίες για μια ποικιλία καταστάσεων ενώ μια απάντηση είναι γνωστή. Τότε το λογισμικό εξόρυξης γνώσης τρέχει πάνω σε αυτά τα δεδομένα και ξεχωρίζει εκείνα τα χαρακτηριστικά που πρέπει να συμπεριληφθούν στο μοντέλο. Όταν τελειώσει η

διαδικασία κτισίματος μπορεί το μοντέλο να χρησιμοποιηθεί σε παρόμοιες καταστάσεις που όμως η απάντηση δεν είναι γνωστή.

### Σχήμα 6

Η φιλοσοφία της Εξόρυξης γνώσης

## **Εξόρυξη Γνώσης = Μοντελοποίηση**

Είμαστε γνώστες μιας  
κατάστασης

-1-

Το εφαρμόζουμε σε μια  
άλλη κατάσταση που δεν  
γνωρίζουμε

-3-

Φτιάχνουμε πάνω σε αυτή ένα  
μοντέλο

-2-

Για παράδειγμα ας υποθέσουμε ότι βρισκόμαστε στη θέση του διευθυντή μάρκετινγκ μιας εταιρίας τηλεπικοινωνιών και θέλουμε να αποκτήσουμε μερικούς πελάτες που κάνουν τηλεφωνήματα μεγάλων αποστάσεων. Βρισκόμαστε δηλαδή αντιμέτωποι με ένα πρόβλημα απόφασης, σε ποιους να απευθυνθούμε. Θα μπορούσαμε να ταχυδρομήσουμε με τυχαίο τρόπο κουπόνια στο γενικό πληθυσμό όπως θα μπορούσαμε να ταξιδεύουμε στις θάλασσες ψάχνοντας για βυθισμένους θησαυρούς. Πάντως σε καμιά από τις δυο περιπτώσεις δεν θα είχαμε τα επιθυμητά αποτελέσματα. Αντί αυτού θα μπορούσαμε να χρησιμοποιήσουμε την εμπειρία της εταιρίας που βρίσκεται αποθηκευμένη στις βάσεις δεδομένων και να κτίσουμε ένα μοντέλο.

Ο διευθυντής μάρκετινγκ έχει πρόσβαση σε πολλές πληροφορίες σχετικές με τους πελάτες μας: την ηλικία τους, το φύλο τους, το αν είναι καλοί πληρωτές, το πόσα τηλεφωνήματα μεγάλων αποστάσεων κάνουν. Το καλό είναι ότι υπάρχουν πληροφορίες και για τους πιθανούς πελάτες της εταιρίας:

την ηλικία τους, το φύλο τους, το πόσο γρήγορα θα πληρώνουν κτλ. Το πρόβλημα είναι ότι δεν γνωρίζουμε πόσο πολύ θα κάνουν χρήση τηλεφωνημάτων σε απομακρυσμένες περιοχές. Επειδή θέλουμε αυτούς που κάνουν πολλά τέτοια τηλεφωνήματα μπορούμε να το πετύχουμε αυτό κτίζοντας ένα μοντέλο.

Ένα απλό μοντέλο που θα ταίριαζε σε μια τηλεπικοινωνιακή εταιρία είναι το παρακάτω:

*98% των πελατών που έχουν λογαριασμό μεγαλύτερο από 60.000\$ το χρόνο δαπανούν περισσότερα από 80\$ το μήνα για τηλεφωνήματα σε μακρινές περιοχές*

Αυτό το μοντέλο θα μπορούσε να εφαρμοστεί στα δεδομένα των πιθανών πελατών και να δοθεί απάντηση στο πρόβλημα απόφασης. Αφού γίνει αυτό θα ξέρει σε ποιους να απευθυνθεί η εταιρία.

### 2.3.5 Λογισμικό Εξόρυξης Γνώσης

Η εξόρυξη γνώσης είναι κατά κάποιο τρόπο μια επέκταση της στατιστικής με κάποια στοιχεία τεχνητής νοημοσύνης και μηχανική μάθηση(machine learning). Όπως και η στατιστική, η εξόρυξη γνώσης δεν αποτελεί επιχειρηματική λύση. Είναι απλά μια τεχνολογία. Για παράδειγμα φανταστείτε ότι πρέπει από ένα κατάλογο εμπορών λιανικής να αποφασιστεί σε ποιους θα σταλούν πληροφορίες για κάποιο νέο προϊόν. Η πληροφορία που αναζητείται από την διαδικασία εξόρυξης γνώσης περιλαμβάνεται σε βάσεις ιστορικών δεδομένων προηγούμενων συναλλαγών με τους πελάτες και στα χαρακτηριστικά των πελατών όπως η ηλικία, ο ταχυδρομικός τους κώδικας, το αν αποκρίθηκαν στο παρελθόν. Το λογισμικό εξόρυξης γνώσης θα χρησιμοποιήσει αυτές τις πληροφορίες από το παρελθόν για να χτίσει ένα μοντέλο συμπεριφοράς πελάτη που θα μπορεί να χρησιμοποιηθεί για να προβλέψουμε ποιοι πελάτες θα ήταν πιθανό να ανταποκριθούν στο νέο προϊόν. Ένας διευθυντής μάρκετινγκ κάνοντας χρήση αυτής της πληροφορίας μπορεί να επιλέξει μόνο τους πελάτες που είναι πιο πιθανό να ανταποκριθούν. Το λογισμικό της επιχείρησης μπορεί τότε να τροφοδοτήσει με τα αποτελέσματα της απόφασης τα κατάλληλα «σημεία επαφής» (τηλεφωνικά κέντρα, web

servers, e-mails κτλ) ώστε οι κατάλληλοι πελάτες να λαμβάνουν τις κατάλληλες πληροφορίες.

### 2.3.6 Η διαδικασία εξόρυξης γνώσης από μια αποθήκη δεδομένων

Σε αυτή την παράγραφο θα ασχοληθούμε με πιο τεχνικά θέματα των συστημάτων υποστήριξης αποφάσεων ώστε ο αναγνώστης να έχει συνολική εικόνα του θεμάτων με τα οποία ασχολείται η εργασία. Πιο συγκεκριμένα θα δούμε αναλυτικά τα στάδια που μεσολαβούν μέχρι να είναι δυνατή η ανάλυση και η ερμηνεία των αποτελεσμάτων.

Η ανακάλυψη γνώσης – η διαδικασία καθορισμού και επίτευξης ενός σκοπού μέσω επαναληπτικής εξόρυξης γνώσης – τυπικά αποτελείται από τρεις φάσεις:

- Προετοιμασία των δεδομένων,
- Υλοποίηση και αποτίμηση του μοντέλου και
- Ανάπτυξη του μοντέλου

Πιο αναλυτικά έχουμε:

- **Η φάση της προετοιμασίας των δεδομένων**

Στη φάση της προετοιμασίας των δεδομένων, ο αναλυτής προετοιμάζει ένα σύνολο δεδομένων που περιλαμβάνει αρκετές πληροφορίες για να κτιστεί ένα σωστό μοντέλο σε ακόλουθες φάσεις. Προσδιορίζοντας αυτές τις απαραίτητες πληροφορίες για μια εταιρία, ένα αποτελεσματικό μοντέλο θα μπορούσε να προβλέψει αν υπάρχει πιθανότητα να αγοράσει κάποιος πελάτης προϊόντα που διαφημίζονται σε ένα νέο κατάλογο. Επειδή οι προβλέψεις βασίζονται σε παράγοντες που πιθανότατα επηρεάζουν τις αγορές των πελατών, ένα μοντέλο συνόλου δεδομένων θα μπορούσε να περιλάμβανε όλους τους πελάτες που ανταποκρίθηκαν σε καταλόγους μέσω e-mails, ταχυδρομείων κτλ τα τελευταία τρία χρόνια, τις δημογραφικές πληροφορίες τους, τα δέκα πιο ακριβά προϊόντα που αγόρασε κάθε πελάτης και πληροφορίες για τους καταλόγους από τους οποίους έγιναν οι αγορές.

Η προετοιμασία των δεδομένων μπορεί να περιλαμβάνει πολύπλοκες ερωτήσεις με τεράστια αποτελέσματα-απαντήσεις. Για παράδειγμα στην υποθετική εταιρία που αναφέρθηκε και στα προηγούμενα παραδείγματα, η προετοιμασία του μοντέλου περιλαμβάνει joins μεταξύ του πίνακα των πελατών και του πίνακα των πωλήσεων καθώς επίσης και τον προσδιορισμό των δέκα κορυφαίων προϊόντων για κάθε πελάτη. Όλα τα θέματα που έχουν να κάνουν με την αποτελεσματική επεξεργασία ερωτήσεων υποστήριξης αποφάσεων σχετίζονται με το περιβάλλον της εξόρυξης γνώσης.

Η εξόρυξη γνώσης τυπικά περιλαμβάνει επαναληπτικό κτίσιμο μοντέλων πάνω σε ένα ήδη προετοιμασμένο σύνολο δεδομένων και στη συνέχεια την ανάπτυξη ενός ή περισσότερων μοντέλων. Επειδή το κτίσιμο των μοντέλων σε μεγάλα σύνολα δεδομένων μπορεί να είναι δαπανηρό, οι αναλυτές συχνά εργάζονται επαναληπτικά με δείγματα συνόλων δεδομένων.

- **Η φάση της υλοποίησης και υπολογισμός του μοντέλου**

Μόνο όταν έχει αποφασιστεί ποιο μοντέλο θα αναπτυχθεί, κτίζει ο αναλυτής το μοντέλο πάνω στο συνολικά προετοιμασμένο σύνολο δεδομένων. Ο σκοπός της φάσης της υλοποίησης είναι ο εντοπισμός των τυποποιημένων μορφών που καθορίζουν ένα χαρακτηριστικό-στόχο(target attribute). Ένα παράδειγμα τέτοιου χαρακτηριστικού-στόχου σε ένα σύνολο δεδομένων θα μπορούσε να ήταν το αν αγόρασε ένας πελάτης τουλάχιστον ένα προϊόν από ένα περασμένο κατάλογο.

Μερικές κλάσεις μοντέλων εξόρυξης γνώσης βοηθούν την πρόβλεψη τόσο ρητά καθορισμένων όσο και κρυφών χαρακτηριστικών. Δυο σημαντικά θέματα που επηρεάζουν την επιλογή του μοντέλου είναι η ακρίβεια του και η αποτελεσματικότητα του αλγορίθμου κατασκευής του μοντέλου πάνω σε μεγάλα σύνολα δεδομένων. Από στατιστικής πλευράς η ακρίβεια των περισσότερων μοντέλων βελτιώνεται με το πλήθος των δεδομένων που χρησιμοποιούνται, οπότε οι αλγόριθμοι που επηρεάζουν τα μοντέλα εξόρυξης πρέπει να κάνουν αποτελεσματική και κλιμακωτή επεξεργασία μεγάλων συνόλων δεδομένων σε ένα λογικό χρονικό διάστημα.

- **Τύποι Μοντέλων**

Τα μοντέλα ταξινόμησης είναι προφητικά. Μπορούν να προβλέψουν αν μία νέα πλειάδα ανήκει σε ένα από τα σύνολα των κλάσεων-στόχων. Κάνοντας εφαρμογή πάνω στον κατάλογο μιας εταιρίας για παράδειγμα, ένα μοντέλο ταξινόμησης θα μπορούσε να προσδιορίσει, βασιζόμενο σε παλιότερες συμπεριφορές, αν υπάρχει πιθανότητα να αγοράσει κάποιος πελάτης από τον κατάλογο αυτό. Τα δέντρα αποφάσεων και τα μοντέλα του Bayes αποτελούν δυο δημοφιλείς τύπους μοντέλων ταξινόμησης.

Τα δέντρα παλινδρόμησης και η λογιστική παλινδρόμηση είναι δυο δημοφιλείς τύποι μοντέλων παλινδρόμησης που μπορούν να προβλέψουν αριθμητικά χαρακτηριστικά όπως τον μισθό ή την ηλικία του πελάτη.

Σε κάποιες εφαρμογές ο αναλυτής δεν γνωρίζει ρητά το σύνολο των κλάσεων-στόχων και τις θεωρεί κρυφές. Ο αναλυτής χρησιμοποιεί μοντέλα-ομαδοποίησης όπως ο K-Means και ο Birch για να προσδιορίσει το κατάλληλο σύνολο κλάσεων και για να ταξινομήσει μια νέα πλειάδα σε κάποια από αυτές τις κρυφές κλάσεις.

Οι αναλυτές χρησιμοποιούν τα λεγόμενα βασισμένα στους κανόνες (rule-based) μοντέλα για να εξερευνήσουν αν για παράδειγμα η αγορά ενός καθορισμένου συνόλου προϊόντων υποδημάτων είναι ενδεικτική, με κάποιο βαθμό εμπιστοσύνης, της αγοράς κάποιου άλλου προϊόντος.

- **Η φάση της ανάπτυξης του μοντέλου**

Στη τελευταία φάση, ο αναλυτής εφαρμόζει το μοντέλο που επιλέχτηκε για το σύνολο δεδομένων για να προβλέψει ένα χαρακτηριστικό-στόχο με άγνωστη τιμή. Για κάθε τρέχον σύνολο πελατών σε μια εταιρία, η πρόβλεψη μπορεί να είναι για παράδειγμα αν θα αγοράσουν κάτι από το νέο κατάλογο. Αναπτύσσοντας ένα μοντέλο πάνω σε ένα σύνολο δεδομένων εισόδου – σε ένα υποσύστημα ή σε ένα αυτόνομο κομμάτι του συνόλου δεδομένων εισόδου – μπορεί να οδηγήσει σε ένα άλλο σύνολο δεδομένων.

Στη συνέχεια θα υπάρξει μια ανακεφαλαίωση όλων όσων αναφέρθηκαν με σκοπό την παρουσίαση της αρχιτεκτονικής που έχει προκύψει και που ουσιαστικά αποτελεί τον σκελετό ενός συστήματος υποστήριξης αποφάσεων.



### ΚΕΦΑΛΑΙΟ 3<sup>ο</sup>

Περιγραφή των πιο χρησιμοποιούμενων μεθόδων  
Επιχειρηματικής νοημοσύνης: Κανόνες Συσχετισμού  
(Association Rules), Ταξινόμηση (Classification),  
Ομαδοποίηση (Clustering) και Πρόβλεψη (Forecasting -  
Prediction).

### 3.1 Βασικές εργασίες εξόρυξης γνώσης

Οι μέθοδοι που χρησιμοποιούνται για την επίτευξη των στόχων του data mining εκτελούν κατά την εφαρμογή τους ένα σύνολο από εργασίες, οι βασικότερες εκ των οποίων είναι οι ακόλουθες:

- Ταξινόμηση (Classification)
- Τμηματοποίηση (Clustering)
- Εξαγωγή κανόνων συσχέτισης (Association rules extraction)
- Πρόβλεψη (Prediction - Forecasting)

Στη συνέχεια αναλύουμε κάθε επιμέρους εργασία και παραθέτουμε ενδεικτικά παραδείγματα για καλύτερη κατανόηση.

### 3.2 Ταξινόμηση (Classification)

Η ταξινόμηση, ο πιο κοινός στόχος εξόρυξης δεδομένων, φαίνεται να είναι μια ανθρώπινη αναγκαστική πράξη. Προκειμένου να καταλάβουμε και να επικοινωνήσουμε μέσα στον κόσμο, ταξινομούμε συνεχώς, κατηγοριοποιούμε, και βαθμολογούμε. Διαιρούμε τα πράγματα διαβίωσης σε φύλα, είδη, και γένη, θέματα σε στοιχεία, σκυλιά σε ράτσες και ανθρώπους σε φυλές.

Η ταξινόμηση αποτελείται από την εξέταση των χαρακτηριστικών γνωρισμάτων ενός πρόσφατα παρουσιασμένου αντικειμένου και την κατάταξη του σε ένα προκαθορισμένο σύνολο κατηγοριών. Για τους σκοπούς μας, τα αντικείμενα που ταξινομούνται αντιπροσωπεύονται γενικά από τα αρχεία σε μια βάση δεδομένων και η

πράξη της ταξινόμησης αποτελείται από την ενημέρωση κάθε αρχείου με το να συμπληρώσει έναν τομέα με έναν κώδικα κατηγορίας κάποιου είδους.

Δοθέντων:

- ενός προβλήματος με  $N$  κλάσεις:  $C_1, C_2, \dots, C_N$  όπου κάθε στιγμιότυπο του προβλήματος έχει  $m$  ιδιότητες (γνωρίσματα):  $A_1, A_2, \dots, A_m$
- και ενός συνόλου στιγμιότυπων του προβλήματος για τα οποία γνωρίζουμε εκ των προτέρων σε ποια κλάση ανήκουν - το σύνολο αυτό είναι γνωστό ως **σύνολο εκπαιδευτικών στιγμιότυπων** (*training set*),

το ζητούμενο είναι

- η δημιουργία ενός μοντέλου για την ταξινόμηση νέων άγνωστων στιγμιότυπων του προβλήματος. Με τον όρο ταξινόμηση εννοούμε την τοποθέτηση ενός στιγμιότυπου σε μία από τις προκαθορισμένες κλάσεις του προβλήματος.

Η επιτυχής έκβαση της ταξινόμησης εξαρτάται από δύο βασικούς παράγοντες:

- το σαφή καθορισμό των κλάσεων του προβλήματος - οι κλάσεις είναι προκαθορισμένες και δεν μεταβάλλονται κατά τη διάρκεια της ταξινόμησης.
- την «ποιότητα» του συνόλου των στιγμιότυπων εκπαίδευσης – τα στιγμιότυπα αυτά θα πρέπει να είναι αντιπροσωπευτικά του προβλήματος.

Όπως έχουμε ήδη αναφέρει το σύνολο των εκπαιδευτικών στιγμιότυπων χρησιμοποιείται για την κατασκευή του ταξινομητή. Υπάρχει ωστόσο ένα ακόμη σύνολο στιγμιότυπων, το **σύνολο των στιγμιότυπων ελέγχου** (*test set*) βάσει του οποίου ελέγχεται η απόδοση του ταξινομητή. Με τον όρο απόδοση εννοούμε την ακρίβεια με την οποία ο ταξινομητής απαντά στο πρόβλημα της ταξινόμησης νέων άγνωστων στιγμιότυπων του προβλήματος. Η απόδοση ισούται με τον αριθμό των στιγμιότυπων

του συνόλου ελέγχου για τα οποία ο ταξινομητής προέβλεψε σωστά την κλάση προς το συνολικό αριθμό των στιγμιότυπων του συνόλου ελέγχου.

Στο ακόλουθο σχήμα (Σχήμα 1.3) διαφαίνεται ο ρόλος των δύο επιμέρους συνόλων.

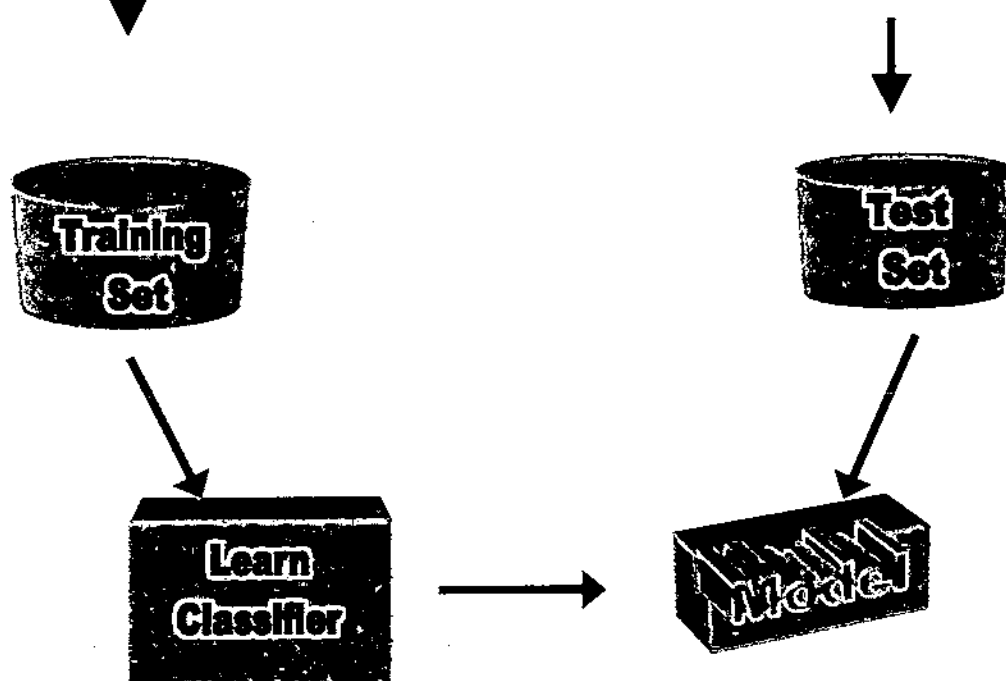
### Σχήμα 7

Η λειτουργία της ταξινόμησης

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



Refund	Marital Status	Taxable Income	Cheat
No	Single	75K	?
Yes	Married	50K	?
No	Married	150K	?
Yes	Divorced	90K	?
No	Single	40K	?
No	Married	80K	?



Στη συνέχεια παραθέτουμε κάποια ενδεικτικά παραδείγματα ταξινόμησης:

➤ Παράδειγμα πολιτικής πίστωσης μιας τράπεζας

Στην περίπτωση της πολιτικής πίστωσης, η τράπεζα θα ήθελε να γνωρίζει πότε μπορεί να δίνει δάνειο σε κάποιο πελάτη χωρίς να λαμβάνει μεγάλο ρίσκο.

Εφαρμόζοντας τη διαδικασία της ταξινόμησης το πρόβλημα ορίζεται ως εξής:

- Οι πελάτες χωρίζονται στις κλάσεις: "άριστος", "καλός", "μέτριος" και "κακός" ανάλογα με την πιθανότητα κάθε πελάτη να εξοφλήσει το δάνειο.
- Κάθε πελάτης χαρακτηρίζεται από την ηλικία του, την εκπαίδευσή του, το ετήσιο εισόδημά του, κ.α.
- Έχουμε στη διάθεση μας δεδομένα πελατών που έχουν δανειστεί από την τράπεζα στο παρελθόν.

Ένα ενδεχόμενο αποτέλεσμα της ταξινόμησης για τις κλάσεις "άριστος" και "καλός" θα μπορούσε να είναι το ακόλουθο:

- Για κάθε πελάτη  $P$ , με  $P$ .πτυχίο = μεταπτυχιακό και  $P$ .εισόδημα  $> 75,000$   $P$ .κλάση = άριστος
- Για κάθε πελάτη  $P$ , με  $P$ .πτυχίο = πτυχίο πανεπιστημίου ή ( $P$ .εισόδημα  $\geq 25,000$  και  $P$ .εισόδημα  $\leq 75,000$ )  $P$ .κλάση = καλός

Αξιοποιώντας το αποτέλεσμα μιας τέτοιας ταξινόμησης η τράπεζα αναμένεται να μειώσει το ρίσκο ο πελάτης στον οποίο χορήγησε κάποιο δάνειο να είναι ασυνεπής ως προς την εξόφληση του δανείου.

#### ➤ Παράδειγμα οργάνωσης διαφημιστικής καμπάνιας

Στην περίπτωση της οργάνωσης μιας διαφημιστικής καμπάνιας, η εταιρία θα ήθελε να γνωρίζει ποιοι πελάτες είναι πιο πιθανό να απαντήσουν θετικά στην καμπάνια.

Στόχος της εταιρίας είναι να προωθήσει την καμπάνια μόνο σε (πιθανά) ενδιαφερόμενα άτομα μειώνοντας έτσι το συνολικό κόστος.

Εφαρμόζοντας τη διαδικασία της ταξινόμησης το πρόβλημα ορίζεται ως εξής:

- Οι πελάτες χωρίζονται στις κλάσεις: θετικοί και αρνητικοί αποδέκτες διαφημιστικών φυλλαδίων.
- Κάθε πελάτης χαρακτηρίζεται από το όνομά του, την ηλικία του, το επάγγελμά του κ.α.
- Έχουμε στη διάθεση μας δεδομένα πελατών που είχαν απαντήσει σε παλαιότερες διαφημιστικές καμπάνιες της εταιρίας.

Ένα ενδεχόμενο αποτέλεσμα της ταξινόμησης για την κλάση των θετικών αποδεκτών διαφημιστικών φυλλαδίων θα μπορούσε να είναι το ακόλουθο:

- Για κάθε πελάτη  $P$ , με  $(P.ηλικία > 25$  και  $P.ηλικία < 55)$  και  $Περιοχή = Ν.Προάστεια$   
 $P.κλάση =$  θετικός αποδέκτης

Η εταιρία θα μπορούσε να αξιοποιήσει το αποτέλεσμα αποστέλλοντας το νέο διαφημιστικό υλικό μόνο στους θετικούς αποδέκτες μειώνοντας έτσι το συνολικό κόστος της διαφημιστικής καμπάνιας.

### 3.3 Τμηματοποίηση (Clustering)

Η τμηματοποίηση είναι η εργασία του διαχωρισμού ενός ετερογενή πληθυσμού σε διάφορες πιο ομοιογενείς υποομάδες ή τμήματα (clusters).

Στη τμηματοποίηση, δεν υπάρχει καμία προκαθορισμένη κατηγορία και κανένα παράδειγμα. Τα αρχεία συγκεντρώνονται μαζί βάσει της ομοιότητας τους. Εξαρτάται από εσάς να καθοριστεί ποιο νόημα, εάν υπάρχει, θα συνδεθεί με τα συμπερασματικά

τμήματα. Παραδείγματος χάριν τα τμήματα των συμπτωμάτων υγείας μπορεί να δείξουν τις διαφορετικές ασθένειες. Τα τμήματα των ιδιοτήτων φύλλων και πυρήνων μπορεί να δείξουν το διαφορετικό γένος - είδος του καλαμποκιού.

Δοθέντων

- ενός προβλήματος όπου κάθε στιγμιότυπο του προβλήματος έχει  $m$  ιδιότητες (γνωρίσματα):  $A_1, A_2, \dots, A_m$
- και ενός συνόλου στιγμιότυπων του προβλήματος

το ζητούμενο είναι:

- ο διαχωρισμός των στιγμιότυπων του προβλήματος σε τμήματα (*clusters*), έτσι ώστε στιγμιότυπα με παρόμοια χαρακτηριστικά να ανήκουν στο ίδιο τμήμα και
- η εύρεση του προφίλ κάθε τμήματος.

Η τμηματοποίηση είναι κατάλληλη για την εύρεση τμημάτων αντικειμένων με παρόμοια χαρακτηριστικά. Έτσι όταν θέλουμε να εξάγουμε κανόνες σχετικά με τη συμπεριφορά των αντικειμένων ενός συγκεκριμένου τμήματος δε χρειάζεται να εξετάσουμε τις ανεξάρτητες εγγραφές του συνόλου των δεδομένων, αρκεί να εξετάσουμε τα χαρακτηριστικά του συγκεκριμένου τμήματος. Η ιδέα / διαίσθηση είναι πως τα στοιχεία που ανήκουν στο ίδιο τμήμα θα συμπεριφέρονται με ενιαίο τρόπο, καθώς έχουν παρόμοια χαρακτηριστικά. Συνεπώς, ένας κανόνας που είναι έγκυρος για κάποιο από τα στοιχεία ενός τμήματος αναμένεται να είναι έγκυρος και για τα υπόλοιπα στοιχεία του τμήματος.

Η βασική διαφορά μεταξύ της ταξινόμησης (*classification*) και της τμηματοποίησης (*clustering*) έγκειται στο γεγονός ότι στην ταξινόμηση οι κλάσεις είναι προκαθορισμένες, ενώ στην τμηματοποίηση δεν υπάρχουν προκαθορισμένες κλάσεις, τα



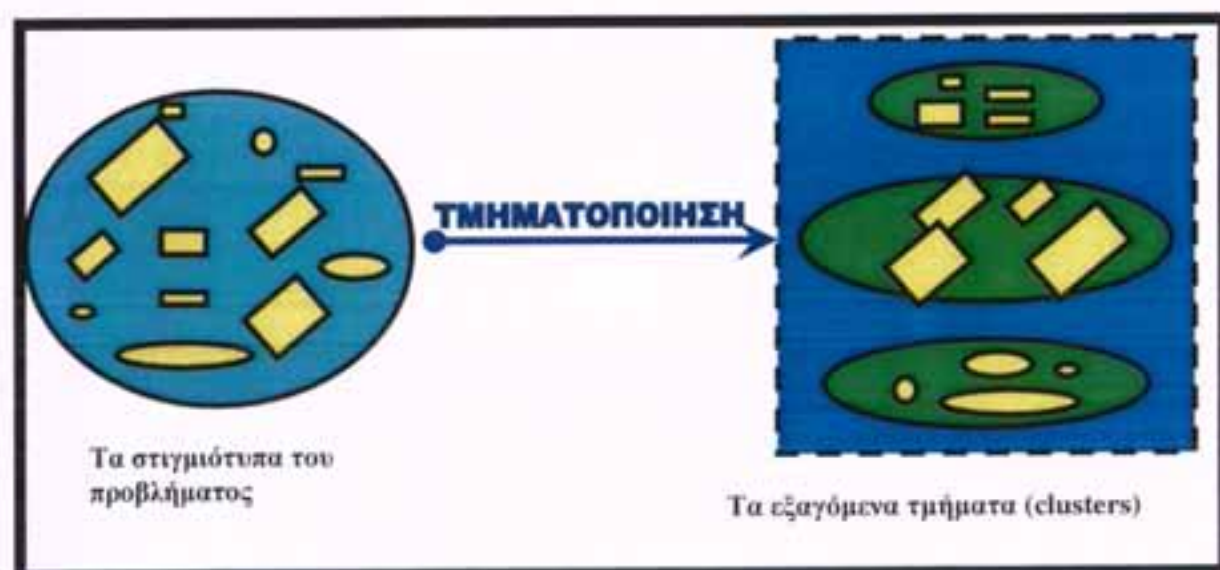
στιγμιότυπα διασπώνται σε τμήματα βάσει της ομοιότητας που παρουσιάζουν μεταξύ τους ως προς τα γνωρίσματα της τμηματοποίησης. Συνεπώς, αν εφαρμόσουμε τμηματοποίηση σε ένα σύνολο δεδομένων, δεν υπάρχει κάποιο συγκεκριμένο σύνολο παραδειγμάτων το οποίο θα μπορούσε να μας υποδείξει ποιες είναι οι επιθυμητές σχέσεις που θα πρέπει να ισχύουν μεταξύ των δεδομένων.

Αρκετά συχνά η τμηματοποίηση χρησιμοποιείται και σαν πρώτο βήμα σε κάποια άλλη μορφή εξόρυξης δεδομένων (data mining) εργασίας. Για παράδειγμα, μπορεί να χρησιμοποιηθεί σαν πρώτο βήμα στην προσπάθεια μερισμού της αγοράς. Αντί δηλαδή να προσπαθήσουμε να προσδιορίσουμε τι είδους διαφήμιση ταιριάζει καλύτερα σε κάθε πελάτη, μπορούμε να διασπάσουμε τους πελάτες σε τμήματα με βάση τις συνήθειές τους κατά την αγορά προϊόντων, να φτιάξουμε το προφίλ κάθε τμήματος και στη συνέχεια να προσδιορίσουμε το είδος της διαφήμισης που ταιριάζει καλύτερα στο κάθε τμήμα.

Στο ακόλουθο σχήμα (Σχήμα 8) παρατίθεται ένα γενικό παράδειγμα τμηματοποίησης των ετερογενών στιγμιότυπων ενός προβλήματος σε τμήματα με κοινά χαρακτηριστικά.

### Σχήμα 8

Η λειτουργία της τμηματοποίησης



Στη συνέχεια παραθέτουμε κάποια ενδεικτικά παραδείγματα τμηματοποίησης:

➤ Παράδειγμα τμηματοποίησης του πληθυσμού μιας χώρας σε διάφορα μορφωτικά επίπεδα

Στην περίπτωση αυτή κάθε στιγμιότυπο του προβλήματος έχει ιδιότητες όπως ηλικία, τόπος κατοικίας, οικονομική κατάσταση, μόρφωση, κ.α., οπότε εφαρμόζοντας τμηματοποίηση μπορούμε να διασπάσουμε τον πληθυσμό με βάση τα γνωρίσματα αυτά και να βρούμε το προφίλ του κάθε τμήματος.

➤ Παράδειγμα διαχωρισμού των χρηστών ενός δικτυακού τόπου με βάση τα κινηματογραφικά τους ενδιαφέροντα

Στην περίπτωση αυτή κάθε στιγμιότυπο του προβλήματος έχει ιδιότητες όπως ηλικία, προηγούμενες προτιμήσεις σε ταινίες, επάγγελμα, μόρφωση, κ.α., οπότε μέσω της τμηματοποίησης μπορούμε να διασπάσουμε τους χρήστες σε τμήματα και να βρούμε το προφίλ του κάθε τμήματος.

### Σχήμα 9

Τμηματοποίηση των χρηστών με βάση τα κινηματογραφικά τους ενδιαφέροντα



Έτσι όταν εμφανίζεται ένας νέος χρήστης, μπορούμε να βρούμε το πιο κοντινό στο χρήστη τμήμα χρηστών (ανάλογα με τις προηγούμενες προτιμήσεις του) και να του προτείνουμε ταινίες με βάση τις προτιμήσεις του τμήματος στο οποίο ανήκει. Η διαίσθησή μας είναι πως θα τον ενδιαφέρουν ταινίες που ενδιαφέρουν και τα υπόλοιπα μέλη του τμήματος.

### 3.4 Εξαγωγή κανόνων συσχέτισης (Association rules)

Οι κανόνες συσχέτισης είναι κατάλληλοι για την εύρεση συσχετίσεων μεταξύ διαφορετικών αντικειμένων. Ένας κανόνας συσχέτισης μεταξύ δύο αντικειμένων  $A$  και  $B$  δηλώνει πως η παρουσία του  $A$  σε κάποιο στιγμιότυπο του προβλήματος συνεπάγεται και την παρουσία του  $B$  στο ίδιο στιγμιότυπο του προβλήματος και συμβολίζεται με  $A \Rightarrow B$ .

Η εξαγωγή των κανόνων συσχέτισης γίνεται με τη βοήθεια κάποιων αλγορίθμων, οι οποίοι αποδεικνύονται αρκετά αποδοτικοί. Μετά την ανάλυση και την εύρεση των κανόνων θα πρέπει να διαπιστωθεί κατά πόσο είναι έγκυροι και σημαντικοί για την εκάστοτε εφαρμογή. Για το σκοπό αυτό υπάρχουν δύο συντελεστές: η υποστήριξη (*support*) και η σιγουριά (*confidence*).

- Η **υποστήριξη** (*support*) ισούται με το ποσοστό του συνόλου των στιγμιότυπων, έστω  $N$  το σύνολο των στιγμιότυπων, που ικανοποιεί το συνδυασμό  $A$  και  $B$ .

$$support = [AB]/N.$$

Έστω για παράδειγμα ο κανόνας συσχέτισης *γάλα ! κατσαβίδια*, αν υποθέσουμε πως μόνο το 0.001 όλων των αγορών περιλαμβάνει γάλα και κατσαβίδια, τότε η υποστήριξη του κανόνα συσχέτισης είναι χαμηλή. Συνήθως, οι επιχειρήσεις δεν ενδιαφέρονται για κανόνες με χαμηλή υποστήριξη, δεδομένου ότι αφορούν ένα πολύ μικρό ποσοστό των πελατών τους.

Από την άλλη αν το 50% των αγορών περιλαμβάνει γάλα και ψωμί, τότε η υποστήριξη για τον κανόνα συσχέτισης *γάλα  $\Rightarrow$  ψωμί* είναι μεγάλη. Τέτοιοι κανόνες



παρουσιάζουν ενδιαφέρον για τις επιχειρήσεις καθώς αφορούν ένα μεγάλο ποσοστό των πελατών.

- Η **σιγουριά** (*confidence*) ισούται με το ποσοστό του συνόλου των στιγμιότυπων για τα οποία όταν ισχύει το A ισχύει και το B.

$$Confidence = [AB] / [A]$$

Για παράδειγμα, ο κανόνας συσχέτισης *ψωμί* ⇒ *γάλα* έχει μια σιγουριά 80% αν το 80% των αγορών που περιλαμβάνουν ψωμί περιλαμβάνει επίσης και γάλα. Για τις επιχειρήσεις ένας κανόνας με χαμηλή σιγουριά δεν παρουσιάζει ενδιαφέρον.

Να σημειώσουμε πως η σιγουριά του κανόνα *ψωμί* ⇒ *γάλα* μπορεί να διαφέρει από τη σιγουριά του κανόνα *γάλα* ⇒ *ψωμί*, παρόλο που και οι δύο κανόνες έχουν την ίδια υποστήριξη.

Στη συνέχεια παραθέτουμε κάποια ενδεικτικά παραδείγματα κανόνων συσχέτισης:

- **Παράδειγμα: σχεδιασμός καταλόγου σε καταστήματα**

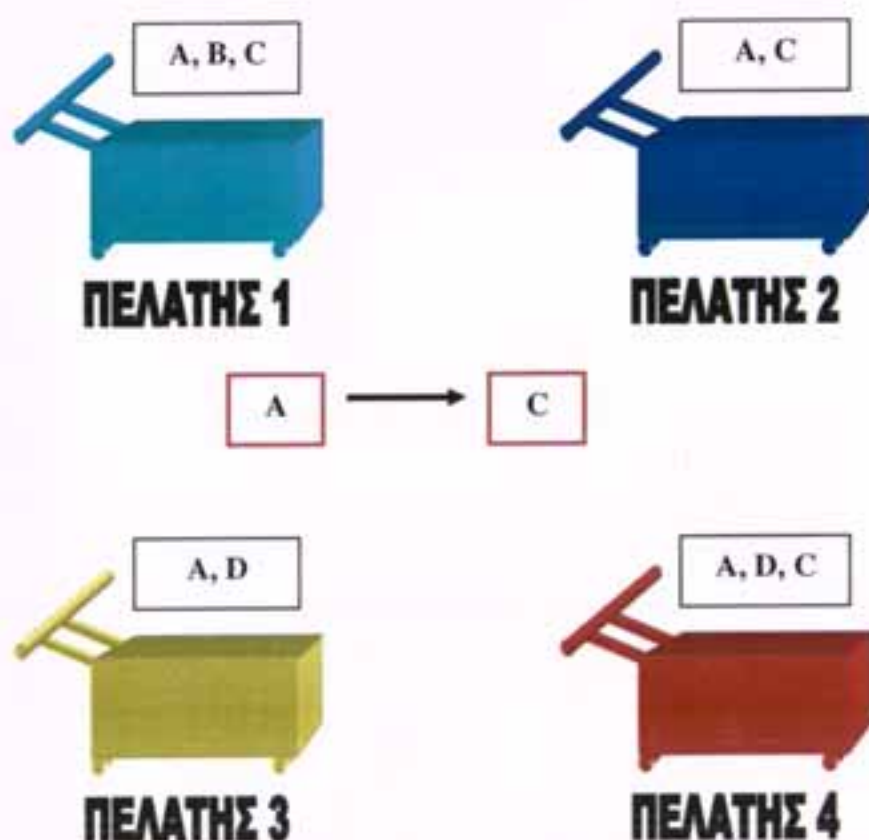
Τα μαγαζιά λιανικής πώλησης ενδιαφέρονται να βρουν συσχετίσεις μεταξύ των διαφορετικών προϊόντων που αγοράζουν οι πελάτες του. Παραδείγματα τέτοιων συσχετίσεων θα μπορούσαν να είναι:

- Κάποιος που αγοράζει ψωμί είναι πολύ πιθανό να αγοράσει και γάλα (Σχήμα 9). Γνωρίζοντας ένα σουπερμάρκετ αυτόν τον κανόνα θα μπορούσε να τοποθετήσει σε διπλανά ράφια το ψωμί και το γάλα, δεδομένου ότι τα δύο αυτά προϊόντα αγοράζονται συχνά μαζί.
- Κάποιος που αγοράζει το βιβλίο "Database System Concepts" είναι πολύ πιθανό να αγοράσει και το βιβλίο "Operating System Concepts". Γνωρίζοντας ένα online

βιβλιοπωλείο αυτόν τον κανόνα θα μπορούσε να προτείνει το βιβλίο "Operating System Concepts" στους πελάτες του που αγοράζουν το βιβλίο "Database System Concepts", δεδομένου ότι τα δύο αυτά βιβλία αγοράζονται συχνά μαζί.

### Σχήμα 9

Παράδειγμα του κανόνα συσχέτισης  $A \Rightarrow C$  που μπορεί να προκύψει από τις αγορές των πελατών ενός σουπερμάρκετ.



Ένας κανόνας συσχέτισης πρέπει να σχετίζεται με κάποιο πληθυσμό: ο πληθυσμός αυτός αποτελείται από ένα σύνολο στιγμιότυπων. Στην περίπτωση του σουπερμάρκετ για παράδειγμα, ο πληθυσμός προκύπτει από το ιστορικό των αγορών των

πελατιών του και κάθε στιγμιότυπο του προβλήματος αποτελείται από τα προϊόντα που αγόρασε ο πελάτης κατά της διάρκεια μιας αγοράς. Στην περίπτωση του online καταστήματος από την άλλη, ο πληθυσμός αποτελείται από όλους τους πελάτες του καταστήματος και κάθε στιγμιότυπο του προβλήματος περιλαμβάνει τις προτιμήσεις και τα προϊόντα που αγόρασε ο πελάτης καθ' όλη τη διάρκεια λειτουργίας του καταστήματος.

Παρατηρούμε λοιπόν ότι η έννοια του πληθυσμού καθορίζεται κάθε φορά από το πρόβλημα που καλούμαστε να αντιμετωπίσουμε. Έτσι στο πρώτο παράδειγμα επικεντρωθήκαμε στις επιμέρους αγορές ενός πελάτη ενώ στο δεύτερο επικεντρωθήκαμε στη συνολική αγοραστική εικόνα του πελάτη χωρίς να λαμβάνουμε υπόψη τα προϊόντα που αγοράστηκαν στις επιμέρους αγορές

### **3.5 Πρόβλεψη (Prediction)**

Η πρόβλεψη είναι η ίδια με την ταξινόμηση (classification) εκτός από το ότι τα αρχεία είναι ταξινομημένα σύμφωνα με κάποια προβλεπόμενη μελλοντική συμπεριφορά ή εκτιμώμενη μελλοντική αξία. Σε έναν στόχο πρόβλεψης, ο μόνος τρόπος να ελεγχθεί η ακρίβεια της ταξινόμησης είναι να περιμένεις και να δεις.

Οποιοσδήποτε από τις τεχνικές που χρησιμοποιούνται για την ταξινόμηση και την εκτίμηση μπορούν να προσαρμοσθούν για τη χρήση στην πρόβλεψη με τη χρησιμοποίηση των παραδειγμάτων κατάρτισης όπου η αξία της μεταβλητής που προβλέπεται είναι ήδη γνωστή, μαζί με τα ιστορικά στοιχεία για εκείνα τα παραδείγματα. Τα ιστορικά στοιχεία χρησιμοποιούνται για να χτιστεί ένα μοντέλο που εξηγεί την τρέχουσα παρατηρηθείσα συμπεριφορά. Όταν αυτό το μοντέλο εφαρμόζεται στις τρέχουσες εισαγωγές, το αποτέλεσμα είναι μια πρόβλεψη της μελλοντικής συμπεριφοράς.

Η τεχνική ανάλυσης καλαθιών αγοράς, που χρησιμοποιείται για να ανακαλύψει ποια στοιχεία είναι πιθανό να αγοραστούν μαζί σε ένα κατάστημα παντοπωλείων, μπορεί επίσης να προσαρμοστεί στο μοντέλο που ασχολείται με το ποιες μελλοντικές αγορές ή ενέργειες τείνουν να υπονοηθούν από τα τρέχοντα στοιχεία.



Η Ανάλυση καλαθιών αγοράς, ο βασισμένος στην μνήμη συλλογισμός, τα δέντρα απόφασης και τα τεχνητά νευρωνικά δίκτυα είναι όλα διάφορες τεχνικές κατάλληλες για τη χρήση στην πρόβλεψη. Η επιλογή της τεχνικής εξαρτάται από τη φύση των δεδομένων εισόδου, τον τύπο αξίας που προβλέπεται, και τη σημασία που αποδίδεται στην ευκολία εξήγησης της πρόβλεψης.

Δοθέντος ενός μοντέλου πρόβλεψης και ενός νέου στιγμιότυπου του προβλήματος, το ζητούμενο είναι η πρόβλεψη της τιμής ενός συγκεκριμένου γνωρίσματος του στιγμιότυπου αυτού (Σχήμα 11).

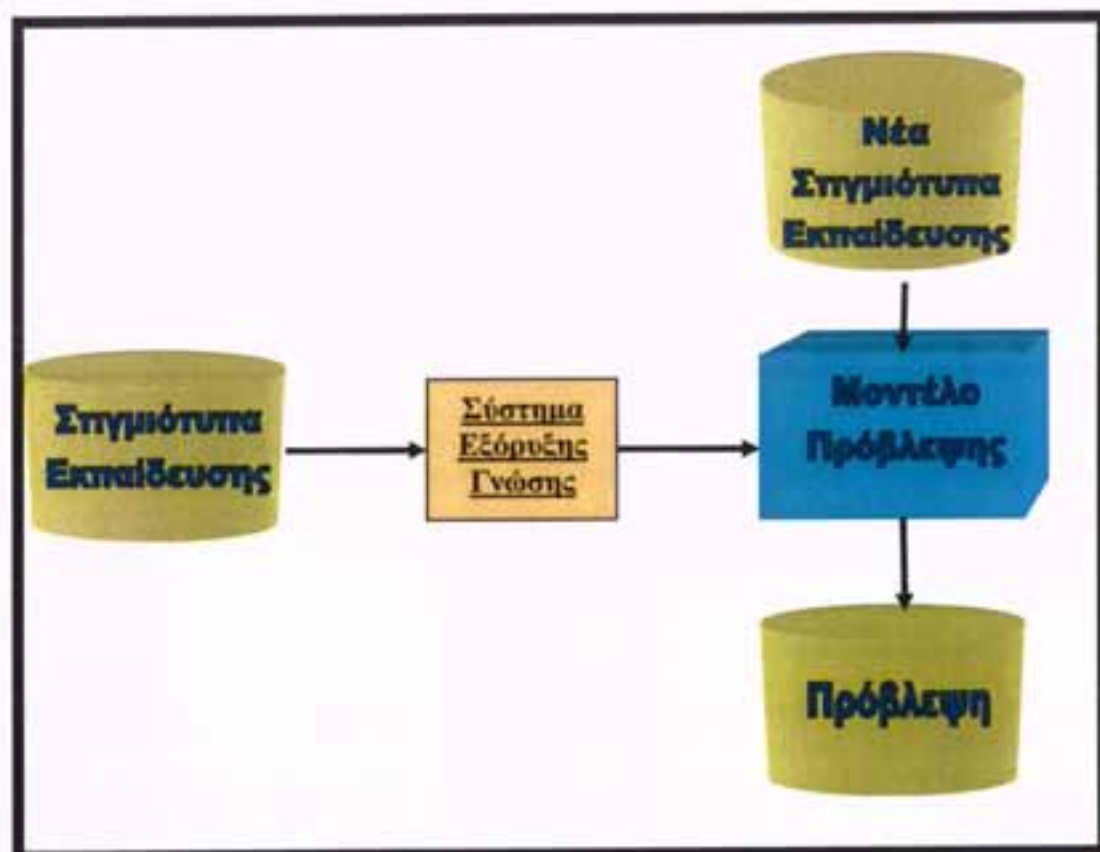
Το μοντέλο πρόβλεψης «χτίζεται» μέσω των παραδειγμάτων του συνόλου εκπαίδευσης (πρόκειται για παραδείγματα στα οποία η τιμή του προς πρόβλεψη γνωρίσματος είναι γνωστή). Πέραν του συνόλου εκπαίδευσης υπάρχει και το σύνολο ελέγχου, το οποίο αποτελείται από παραδείγματα στα οποία η τιμή του προς πρόβλεψη γνωρίσματος είναι γνωστή – το σύνολο αυτό συνήθως ισούται αριθμητικά με το 1/3 των παραδειγμάτων του συνόλου εκπαίδευσης και χρησιμοποιείται για την αξιολόγηση του μοντέλου πρόβλεψης. Το σύνολο ελέγχου ελέγχει κατά κάποιο τρόπο την απόδοση του μοντέλου πρόβλεψης, δηλαδή την ακρίβεια με την οποία το μοντέλο πρόβλεψης προβλέπει την τιμή ενός άγνωστου γνωρίσματος στα νέα στιγμιότυπα του προβλήματος. Για να βρούμε την άγνωστη τιμή ενός γνωρίσματος σε κάποιο νέο στιγμιότυπο του προβλήματος, θα πρέπει να περάσουμε το στιγμιότυπο αυτό από το μοντέλο πρόβλεψης.

Συγκρίνοντας την πρόβλεψη με την ταξινόμηση που είδαμε στην αμέσως προηγούμενη ενότητα μπορούμε να πούμε πως η ταξινόμηση αποτελεί μια ειδική περίπτωση πρόβλεψης, καθώς αναφέρεται στην πρόβλεψη της κλάσης των στιγμιότυπων του προβλήματος.

Μερικά ενδεικτικά παραδείγματα πρόβλεψης είναι: πρόβλεψε την πιθανότητα ένας ασθενής να πάσχει από μία συγκεκριμένη ασθένεια, πρόβλεψε το πλήθος των αγορών που θα κάνει ένας νέος πελάτης στον πρώτο χρόνο κ.α.

Σχήμα 11

Λειτουργία της πρόβλεψης





## ΚΕΦΑΛΑΙΟ 4<sup>ο</sup>

### ΛΟΓΙΣΤΙΚΗ ΕΦΑΡΜΟΓΗ

## ΠΕΡΙΛΗΨΗ

Τα Επενδυτικά Μοντέλα είναι αυτοματοποιημένα συστήματα προτάσεων για Αγορές και Πωλήσεις μετοχών, βασισμένα στην εκτίμηση που κάνουν για την μελλοντική πορεία των τιμών. Οι δυσκολίες για την δημιουργία τέτοιων μοντέλων είναι πολύ μεγάλες λόγω της χαώδους συμπεριφοράς που χαρακτηρίζει τις χρηματαγορές και ως εκ τούτου της μικρής δυνατότητας πρόβλεψης που έχουν. Παρόλα αυτά, στην πράξη κάποιοι κανόνες φαίνεται να δίνουν ικανοποιητικά αποτελέσματα.

Στην εργασία αυτή γίνεται μια προσπάθεια δημιουργίας Επενδυτικών Μοντέλων με χρήση Εξελικτικών Αλγορίθμων. Ειδικότερα, έχουν δημιουργηθεί κάποιοι επενδυτικοί κανόνες, οι οποίοι χρησιμοποιούν ορισμένους δείκτες Τεχνικής Ανάλυσης και με τη βοήθεια του Εξελικτικού Αλγορίθμου αναζητούνται οι κανόνες και οι παράμετροι που δίνουν τα καλύτερα αποτελέσματα.

Ύστερα από πολλές εκτελέσεις του προγράμματος που συνοδεύει την παρούσα εργασία, δημιουργείται η εκτίμηση ότι τα αποτελέσματα είναι αρκετά ενθαρρυντικά., οι αποδόσεις των μοντέλων που προτάθηκαν από το πρόγραμμα, μερικές φορές είναι εντυπωσιακές. Πριν όμως εφαρμοσθούν στην πράξη, θα πρέπει να γίνει στατιστική αξιολόγηση της απόδοσής τους στη διάρκεια του χρόνου και να δοκιμαστούν σε νέες περιοχές δεδομένων. Οι δύο ανωτέρω προτάσεις θα μπορούσαν να αποτελέσουν τη βάση για περαιτέρω έρευνα και βελτίωση των αποτελεσμάτων της παρούσας εργασίας.

## ΕΙΣΑΓΩΓΗ

Για την επίλυση δύσκολων προβλημάτων, ο χώρος της πληροφορικής οδηγήθηκε σε νέες τεχνικές αναζήτησης λύσεων και βελτιστοποίησης. Ένα πρόβλημα εξαιρετικά δύσκολο στη λύση του είναι και αυτό της πρόβλεψης της μελλοντικής πορείας των τιμών των μετοχών. Για την επίλυσή του χρησιμοποιήθηκαν πολλές διαφορετικές προσεγγίσεις. Μια από αυτές, που χρησιμοποιήθηκε σχετικά πρόσφατα με ικανοποιητικά αποτελέσματα, είναι η χρήση Εξελικτικών Αλγορίθμων. Λόγω του τρόπου λειτουργίας τους, είναι σε θέση να βρουν ικανοποιητικές λύσεις σε ένα πραγματικά πολύ μεγάλο χώρο λύσεων.

Στην εργασία αυτή χρησιμοποιήθηκαν μερικοί κλασικοί δείκτες Τεχνικής Ανάλυσης για την δημιουργία κάποιων επενδυτικών κανόνων. Οι δείκτες αυτοί χρησιμοποιούνται από τους αναλυτές εδώ και πολλά χρόνια ενώ η δημιουργία των κανόνων βασίστηκε περισσότερο στην εμπειρία από τη χρήση των δεικτών αυτών. Το πρόβλημα εντοπίζεται στην επιλογή των καλύτερων κανόνων και των παραμέτρων των δεικτών που συνθέτουν τους κανόνες αυτούς. Αυτό επιτυγχάνεται με τη βοήθεια του Εξελικτικού Αλγορίθμου. Με το πρόγραμμα που αναπτύχθηκε στο πλαίσιο της εργασίας αυτής, έγινε προσπάθεια εύρεσης των καλύτερων παραμέτρων του αλγορίθμου, ώστε αυτός να είναι πιο αποτελεσματικός. Οι μελέτες γίνονται χρησιμοποιώντας ιστορικές τιμές μετοχών και δεικτών του Χρηματιστηρίου Αξιών Αθηνών.

Στις παραγράφους 4.1 έως 4.7 που ακολουθούν, γίνεται μια σύντομη αναφορά στον τρόπο λειτουργίας τους, στα είδη των Εξελικτικών Αλγορίθμων, παρουσιάζονται τα μέρη που απαρτίζουν έναν αλγόριθμο, αναφέρονται τα πλεονεκτήματα και τα μειονεκτήματά τους, εξηγείται γιατί οι Ε.Α. είναι αποτελεσματικοί και τέλος παρουσιάζονται κάποιοι ενδεικτικοί τομείς στους οποίους βρίσκουν εφαρμογή.

Στις παραγράφους 4.8 έως 4.13, γίνεται μια παρουσίαση των επενδυτικών δυνατοτήτων στο Χρηματιστήριο, αναφέρονται μερικά είδη επενδυτικών μοντέλων και οι δυσκολίες που υπάρχουν στην ανάπτυξή τους, οι δείκτες της Τεχνικής Ανάλυσης που χρησιμοποιούνται στην εργασία αυτή καθώς και προηγούμενες μελέτες και ανακοινώσεις με διαφορετικές προσεγγίσεις στο πρόβλημα.

Οι παράγραφοι 4.14 έως 4.17 αναφέρεται στο σχεδιασμό της εφαρμογής. Παρουσιάζεται ο τρόπος που προσεγγίστηκε το πρόβλημα, πώς έγινε η σχεδίασή του, περιγράφονται αναλυτικά οι αποδεκτές και μη καταστάσεις για την εξαγωγή έγκυρων σημάτων Αγοράς και Πώλησης, αναφέρονται μερικοί συμβιβασμοί που έγιναν για πρακτικούς λόγους και τέλος γίνεται αναλυτική περιγραφή των επενδυτικών κανόνων.

Στην παράγραφο 4.18 περιγράφεται πώς αναπτύχθηκε η εφαρμογή που υλοποιεί τη σχεδίαση που προηγήθηκε, αναφέρονται τα εργαλεία που χρησιμοποιήθηκαν για την ανάπτυξή της και εξηγείται τι και γιατί έγινε σε κάθε στάδιο της εφαρμογής.

Στις παραγράφους 4.19 έως 4.22 παρουσιάζονται συνοπτικά τα αποτελέσματα των εκτελέσεων του προγράμματος, αναφέρονται τα συμπεράσματα που προήλθαν από τις εκτελέσεις και τέλος, γίνονται διάφορες προτάσεις βελτίωσης τόσο της παρούσας σχεδίασης όσο και γενικότερων προσεγγίσεων στο πρόβλημα της δημιουργίας επενδυτικών μοντέλων.

## 4.1 Εξελικτικοί Αλγόριθμοι

Η αυξημένη διαθεσιμότητα υπολογιστικής ισχύος τις τελευταίες δεκαετίες επέτρεψε την ανάπτυξη μιας νέας γενιάς «έξυπνων» τεχνικών με τη χρήση Η/Υ, όπως Έμπειρα Συστήματα (Expert Systems), Νευρωνικά Δίκτυα (Neural Networks) Ασαφή Συστήματα (Fuzzy Systems), Εξελικτικοί Αλγόριθμοι (Evolutionary Algorithms) κλπ. Οι «έξυπνες» αυτές τεχνικές είναι ευέλικτες και μπορούν να προσαρμόζονται σε νέες καταστάσεις.

Οι Εξελικτικοί Αλγόριθμοι είναι τεχνικές επίλυσης προβλημάτων αναζήτησης-βελτιστοποίησης εμπνευσμένοι από την βιολογική εξέλιξη. Βασίζονται σε μια τεχνητή προσομοίωση της διαδικασίας της φυσικής εξέλιξης ή επιβίωσης του καλύτερου, γνωστής από τη Θεωρία της Εξέλιξης των Ειδών του Δαρβίνου. Χρησιμοποιούνται για την επίλυση δύσκολων προβλημάτων που με παραδοσιακές μεθόδους θα ήταν πολύ δύσκολο, αν όχι αδύνατο, να λυθούν.

## 4.2 Περιγραφή λειτουργίας των Ε.Α.

Οι Ε.Α. βασίζονται σε μια επαναλαμβανόμενη διαδικασία εξέλιξης ενός πληθυσμού. Πληθυσμός είναι ένα σύνολο από πιθανές λύσεις του προβλήματος.

Η εξέλιξη του πληθυσμού οφείλεται στην οργανωμένη ανταλλαγή πληροφοριών ανάμεσα στα διάφορα μέλη του πληθυσμού, που έχουν επιλεγεί κατάλληλα με βάση κάποιους κανόνες επιλογής (selection rules), ώστε ο νέος πληθυσμός που θα δημιουργηθεί να είναι πιθανόν καλύτερος από προηγούμενους.

Η διαδικασία της ανταλλαγής πληροφοριών μεταξύ των μελών ενός πληθυσμού είναι δανεισμένη από τη βιολογία. Στη φύση, δομικά στοιχεία κάθε οργανισμού είναι τα γονίδια (genes), καθένα από τα οποία κωδικοποιεί και ένα χαρακτηριστικό του οργανισμού (π.χ. χρώμα ματιών). Οι τιμές ενός γονιδίου λέγονται alleles (π.χ. γαλανά μάτια). Τα γονίδια συνθέτουν τα χρωμοσώματα (chromosomes) που είναι αλυσίδες DNA. Σε κάθε οργανισμό, στο στάδιο της αναπαραγωγής, γίνεται τυχαία ανταλλαγή γονιδίων μεταξύ των γονέων για τη δημιουργία των χρωμοσωμάτων του νέου

οργανισμού. Έτσι ο νέος οργανισμός κληρονομεί χαρακτηριστικά και από τους δύο γονείς. Συχνά παρατηρούνται και μεταλλάξεις, που είναι τυχαίες μεταβολές των τιμών τους, με αποτέλεσμα ο οργανισμός να παρουσιάζεται με νέα χαρακτηριστικά τα οποία δεν υπήρχαν στους γονείς.

Ουσιώδες χαρακτηριστικό κάθε οργανισμού είναι η ποιότητά του. Πολλές φορές αυτό είναι καθοριστικό για την επιβίωσή του και κατά συνέπεια για την ικανότητά να αναπαράγει απογόνους.

Κατά παρόμοιο τρόπο οι Ε.Α. δημιουργούν (συνήθως με τυχαίο τρόπο) έναν αρχικό πληθυσμό. Στη συνέχεια γίνεται αξιολόγηση του πληθυσμού (evaluation). Κάθε στοιχείο του πληθυσμού αξιολογείται με βάση κάποια μέθοδο (συνάρτηση αξιολόγησης - fitness function) και του αποδίδεται μία τιμή. Η συνάρτηση αξιολόγησης μετρά την ποιότητα μιας λύσης που αντιστοιχεί σε μια συγκεκριμένη γενετική δομή. Είναι ειδική για κάθε πρόβλημα και καθορίζει το «περιβάλλον» της εξέλιξης.

Νέες λύσεις δημιουργούνται από κατάλληλα επιλεγμένα μέλη, με κάποιους κανόνες επιλογής (selection rules). Μπορεί να γίνουν ανασυνδυασμοί γενετικού υλικού (recombinations) μεταξύ των γονέων και μεταλλάξεις (mutations) που μεταβάλουν ακόμα περισσότερο τους απογόνους. Η μεταβολή στη δομή τους μας επιτρέπει να διερευνούμε νέες περιοχές στο χώρο λύσεων του προβλήματος. Στη διαδικασία επιλογής μπορεί να περιλαμβάνονται και οι απόγονοι, ώστε στο νέο πληθυσμό να μεταφέρονται τα καλύτερα μέλη.

Ο νέος πληθυσμός αξιολογείται και η διαδικασία αυτή επαναλαμβάνεται (εξελικτικός κύκλος) έως ότου ικανοποιηθεί κάποια προκαθορισμένη συνθήκη τερματισμού. Αυτή μπορεί να είναι η επίτευξη της βέλτιστης λύσης, μια λύση που δεν βελτιώνεται περαιτέρω, κάποιο κριτήριο πλήρωσης χρόνου, ένας καθορισμένος αριθμός γενεών κ.ά. Ο τελικός πληθυσμός περιέχει μια συλλογή από λύσεις, από τις οποίες μία ή περισσότερες μπορεί να εφαρμοστεί στο συγκεκριμένο πρόβλημα.

Η αλγοριθμική δομή ενός τυπικού Ε.Α. φαίνεται στο Σχήμα 12.

## Σχήμα 12

Ένας τυπικός E.A.

```
Procedure EA (); {  
  T=0;  
  Initialize population P (t);  
  Evaluate P (t);  
  Until (done) {  
    T=t+1;  
    Parent selection P (t);  
    Recombine P (t);  
    Mutate P (t);  
    Evaluate P (t);  
    Survive P (t);  
  }  
}
```

### 4.3 Μοντέλα Εξελικτικών Αλγορίθμων

Τα βασικότερα μοντέλα E.A. είναι:

- Εξελικτικός Προγραμματισμός (Evolutionary Programming)
- Εξελικτικές Στρατηγικές (Evolution Strategies)
- Γενετικός Προγραμματισμός (Genetic Programming)
- Γενετικοί Αλγόριθμοι (Genetic Algorithms)

Αν και οι βασικές αρχές τους είναι όμοιες κάθε ένα υλοποιεί τον εξελικτικό αλγόριθμο με διαφορετικό τρόπο και οι διαφορές αυτές αφορούν σχεδόν όλες τις πλευρές, όπως την αναπαράσταση της δομής τους, τους μηχανισμούς επιλογής, τους γενετικούς τελεστές και την μέτρηση της απόδοσης.

#### 4.3.1 Εξελικτικός Προγραμματισμός (Evolutionary Programming)

Με την αρχικοποίηση του πληθυσμού ( $N$ ), όλα τα στοιχεία επιλέγονται σαν γονείς και ύστερα μεταλλάσσονται δημιουργώντας  $N$  απογόνους. Οι απόγονοι αξιολογούνται και επιλέγονται τα  $N$  καλύτερα στοιχεία από τα συνολικά  $2N$ , χρησιμοποιώντας μια συνάρτηση πιθανοτήτων βασισμένη στην ποιότητα (fitness).

Δηλαδή, τα στοιχεία με την καλύτερη ποιότητα έχουν περισσότερες πιθανότητες να επιβιώσουν. Ο τρόπος μετάλλαξης εξαρτάται από την αναπαράσταση των στοιχείων και είναι συχνά προσαρμόσιμος (adaptive). Ανασυνδυασμοί γενικά δεν χρησιμοποιούνται καθώς οι τύποι των μεταλλάξεων είναι αρκετά ευέλικτοι και παρουσιάζουν αποτελέσματα όμοια με τον ανασυνδυασμό.

### Σχήμα 13

#### Ένας αλγόριθμος Εξελικτικού Προγραμματισμού

```
Procedure EA () {  
  T=0;  
  Initialize population P (t);  
  Evaluate P (t);  
  Until (done) {  
    T=t+1;  
    Parent selection P (t);  
    Recombine P (t);  
    Mutate P (t);  
    Evaluate P (t);  
    Survive P (t);  
  }  
}
```

#### 4.3.2 Εξελικτικές Στρατηγικές

Αναπτύχθηκαν από τον Rechenberg (1973) με μέγεθος πληθυσμού ένα και χρησιμοποιώντας μόνο μετάλλαξη. Στη συνέχεια προστέθηκε και ο τελεστής ανασυνδυασμού καθώς και πληθυσμοί με περισσότερα μέλη. Χρησιμοποιούνται κυρίως σε προβλήματα βελτιστοποίησης και χρησιμοποιούν πραγματικούς αριθμούς για την αναπαράσταση των μελών τους. Η μετάλλαξη είναι ο κύριος γενετικός τελεστής.

Η επιβίωση των μελών του πραγματοποιείται με δύο τρόπους. Με τον πρώτο όλα τα μέλη του νέου πληθυσμού επιλέγονται μόνο από τους απογόνους και αυτά αντικαθιστούν τους γονείς τους ενώ με τον δεύτερο ο νέος πληθυσμός δημιουργείται από τους καλύτερους γονείς και απογόνους. Καθοριστικό ρόλο κι εδώ παίζει η αυτοπροσαρμογή της μετάλλαξης.

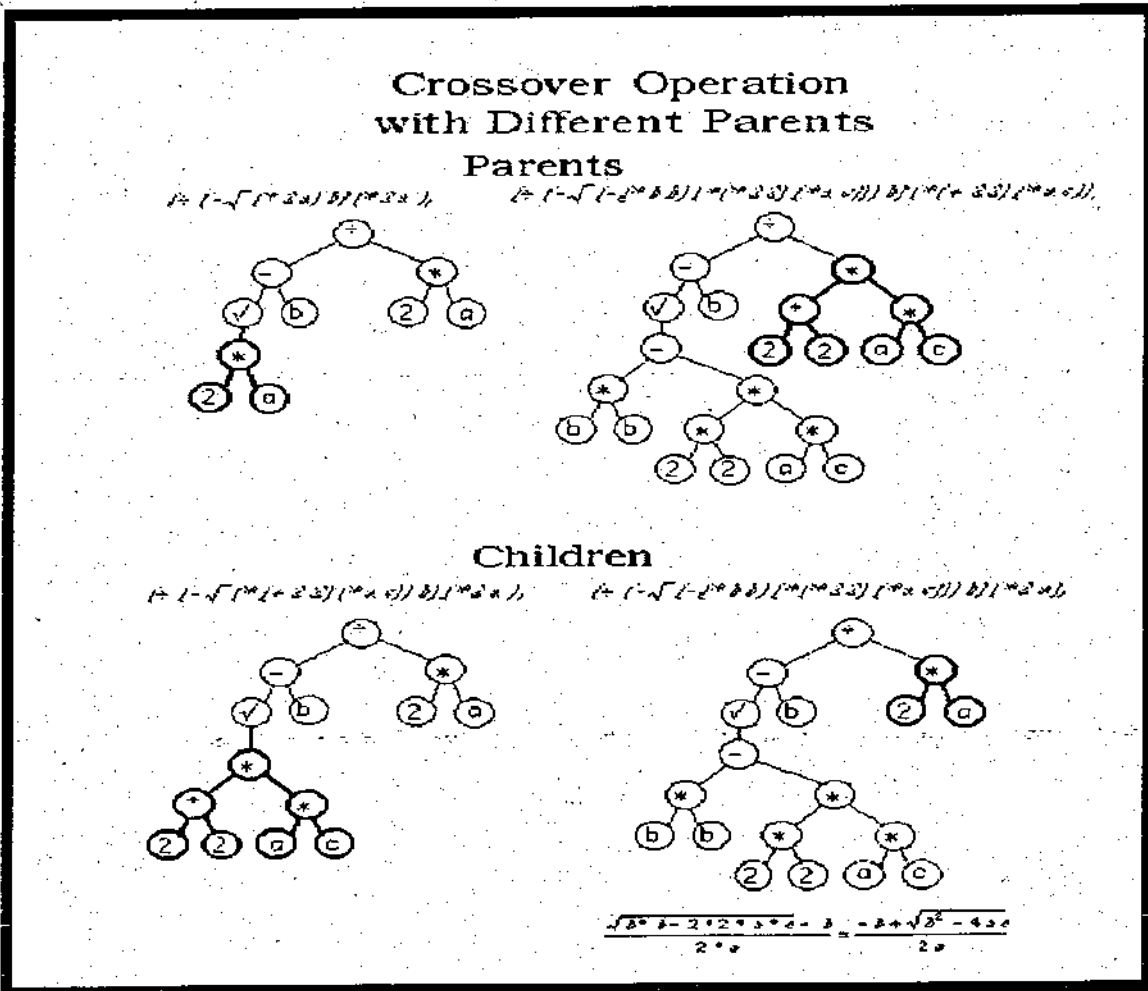


### 4.3.3 Γενετικός Προγραμματισμός

Εδώ οι λύσεις αναπαριστάνονται σαν ιεραρχικές συνθέσεις από συναρτήσεις εμφανίζοντας μια δομή δένδρου (tree-like structure). Οι ακολουθίες κάθε κόμβου παρέχουν τις παραμέτρους της συνάρτησης που ορίζει ο κόμβος. Οι τελικοί κόμβοι, που δεν έχουν ακολουθίες, αντιστοιχούν στα δεδομένα εισόδου. Όλο το δένδρο επίσης αναπαριστά μια συνάρτηση. Κύριο χαρακτηριστικό τους είναι ότι δεν έχουν προκαθορισμένη δομή. Οι ανασυνδυασμοί γίνονται αντικαθιστώντας ένα τυχαία επιλεγμένο τμήμα του δένδρου του ενός γονέα με ένα τμήμα από τον άλλο γονέα. Οι μεταλλάξεις γίνονται με μικρή πιθανότητα και υλοποιούνται με τη χρήση ενός τυχαία δημιουργημένου δένδρου στη θέση του δεύτερου γονέα

**Σχήμα 14.**

Η διαδικασία του ανασυνδυασμού



#### 4.3.4 Γενετικοί Αλγόριθμοι

Αναπτύχθηκαν τη δεκαετία του '60 από τον John Holland και αρχικά χρησιμοποιούσαν δυαδική κωδικοποίηση σταθερού μήκους. Εφαρμόζονται κυρίως για προβλήματα βελτιστοποίησης. Η επιλογή των γονέων γίνεται σύμφωνα με μια συνάρτηση πιθανότητας, βασισμένη στη σχετική ποιότητα των ατόμων του πληθυσμού. Έτσι αυτά με την καλύτερη ποιότητα έχουν μεγαλύτερη πιθανότητα να επιλεγούν σαν γονείς.

Τα μέλη του νέου πληθυσμού δημιουργούνται από τμήματα και των δύο γονέων. Η διαδικασία του ανασυνδυασμού εδώ παίζει καθοριστικό ρόλο και λέγεται διασταύρωση (crossover). Η μετάλλαξη χρησιμοποιείται με μικρή πιθανότητα και έχει δευτερεύοντα ρόλο.

#### 4.4 Τεχνικές και Τελεστές

##### *Αναπαράσταση*

Η δυαδική αναπαράσταση έχει το πλεονέκτημα ότι χρησιμοποιείται ανεξάρτητα από το πρόβλημα. Μπορούν όμως να χρησιμοποιηθούν και άλλες μορφές, ανάλογα με τη φύση του προβλήματος, όπως πραγματικοί αριθμοί, ακέραιοι, αλφαριθμητικά κ.λπ. Δυστυχώς δεν υπάρχει ένας τρόπος αναπαράστασης που να είναι ο καλύτερος για όλα τα προβλήματα.

##### *Αρχικοποίηση*

Το πρώτο βήμα στην υλοποίηση ενός Γενετικού Αλγόριθμου είναι η δημιουργία ενός αρχικού πληθυσμού από πιθανές (και σίγουρα όχι τις καλύτερες) λύσεις του προβλήματος (χρωμοσώματα). Η δημιουργία τους γίνεται συνήθως με τυχαίο τρόπο. Δεν υπάρχουν αυστηροί κανόνες για το μέγεθος του πληθυσμού. Γενικά επιλέγονται τιμές από 100 έως 200. Μεγάλοι πληθυσμοί πιθανόν να οδηγήσουν σε καλύτερες λύσεις, όμως χρησιμοποιούν περισσότερους πόρους του συστήματος.

Ο πληθυσμός στην πορεία της εξέλιξης θα συγκλίνει. Αυτό σημαίνει ότι οι απόγονοι που θα υφίστανται γενετικούς μετασχηματισμούς με τη χρήση των γενετικών τελεστών (ανασυνδυασμού και/ή μετάλλαξης) θα γίνονται όλο και περισσότερο όμοιοι

με τη πάροδο των γενεών. Αν υπάρχει σαφής γνώση για το πρόβλημα που καλείται να λύσει ο αλγόριθμος μπορεί να χρησιμοποιηθεί στη δημιουργία του αρχικού πληθυσμού π.χ. να περιοριστεί η αρχικοποίηση σε έναν τομέα που μας ενδιαφέρει.

### *Συνάρτηση ποιότητας (Fitness function)*

Η συνάρτηση ποιότητας (fitness function) ή συνάρτηση αξιολόγησης (evaluation function) αξιολογεί τα μέλη του πληθυσμού. Η τιμή της αποτελεί την ποιότητα του ατόμου.

Στόχος του Εξελικτικού Αλγόριθμου είναι να βρει το άτομο που αναπαριστά μια συγκεκριμένη λύση του προβλήματος, που ανάλογα με το πρόβλημα, μεγιστοποιεί ή ελαχιστοποιεί την συνάρτηση ποιότητας.

### *Επιλογή (Selection)*

Η επιλογή καθορίζει ποια μέλη του πληθυσμού θα μεταφέρουν τις πληροφορίες που περικλείονται στη δομή τους, στα μέλη της επόμενης γενιάς. Σκοπός της μεθόδου επιλογής είναι να δώσει στα μέλη με καλύτερο δείκτη ποιότητας περισσότερες πιθανότητες επιλογής για αναπαραγωγή.

Υπάρχουν πολλές εναλλακτικές διαδικασίες επιλογής. Η πιο συνηθισμένη είναι με χρήση τροχού ρουλέτας (roulette wheel) που προτάθηκε από τον Holland.

#### *Roulette wheel selection*

Σύμφωνα με αυτήν, τμήματα του τροχού της ρουλέτας αντιστοιχούν σε μέλη του πληθυσμού αναλογικά με την τιμή της ποιότητάς τους.

Η επιλογή γίνεται με το «γύρισμα» του τροχού. Η διαδικασία είναι τυχαία αλλά τα άτομα με υψηλό βαθμό ποιότητας έχουν μεγαλύτερες πιθανότητες να επιλεγούν. Σε σχετικά μικρό πληθυσμό, που συνήθως χρησιμοποιείται σε τυπικές εφαρμογές, μια σειρά από μη επιθυμητές επιλογές μπορεί να δημιουργήσει ένα δυσανάλογα μεγάλο αριθμό από απογόνους με χαμηλή συγκριτικά ποιότητα.

### *Επιλογή με τουρνουά (Tournament selection)*

Είναι μια ανταγωνιστική διαδικασία κατά την οποία ένας αριθμός ατόμων προεπιλέγονται τυχαία και από αυτά επιλέγεται τελικά αυτό με την καλύτερη τιμή ποιότητας για να συμμετάσχει στην εξέλιξη του πληθυσμού. Με τον τρόπο αυτό τα χειρότερα άτομα δεν θα επιλεγούν ποτέ.

Άλλες μέθοδοι είναι η επιλογή αυστηρά μόνο των καλύτερων και η χρήση ρουλέτας με βαθμολόγηση.

Ο τρόπος επιλογής επηρεάζει σημαντικά την ισορροπία ανάμεσα στην «αξιοποίηση» (exploitation) και την «διερεύνηση» (exploration). Με την έννοια «αξιοποίηση» εννοούμε την εκμετάλλευση των πληροφοριών που αποκτήθηκαν έως τώρα και περιέχονται στη δομή των ατόμων ενώ με την έννοια «διερεύνηση» εννοούμε τη διερεύνηση για λύσεις σε νέες περιοχές του χώρου αναζήτησης (search space). Για παράδειγμα, αν κάποιος ανακαλύψει μια καλή στρατηγική στο σκάκι μπορεί απλά να επικεντρωθεί στην αξιοποίησή της. Αυτό όμως θα τον εμποδίσει να ανακαλύψει νέες και ίσως πολύ καλύτερες στρατηγικές. Βελτιώσεις ανακαλύπτονται μόνο δοκιμάζοντας νέες στρατηγικές. Ένα από τα σημαντικότερα χαρακτηριστικά των Ε.Α. είναι η ισορροπία ανάμεσα στις δύο παραπάνω έννοιες. Η αναζήτηση επιτυγχάνεται με την χρήση κυρίως της μετάλλαξης ως γενετικού τελεστή.

### *Γενετικοί τελεστές*

#### *Ανασυνδυασμός (recombination)*

Είναι ένας τρόπος αναζήτησης λύσεων, βασισμένος σε πληροφορίες που πάρθηκαν από υπάρχουσες λύσεις. Αν δύο άτομα με υψηλό βαθμό ποιότητας ανταλλάξουν μεταξύ τους μερικά γονίδια, μπορεί να δημιουργηθούν απόγονοι με ακόμη καλύτερη ποιότητα. Μπορεί να υλοποιηθεί με πολλούς τρόπους, όπως ανασυνδυασμός ενός σημείου, δύο σημείων, n-σημείων, ενιαίος (uniform) κ.ά. όπως φαίνεται στο σχήμα 15.

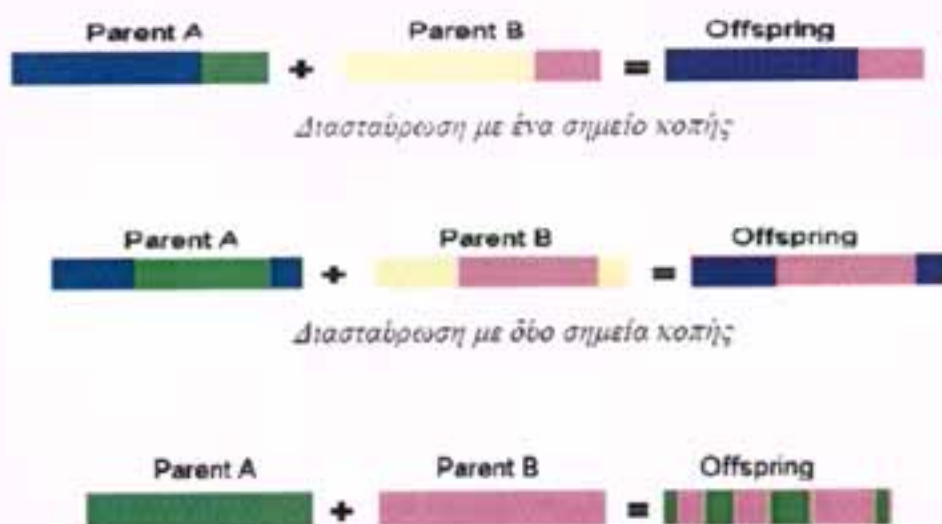
Στην πιο απλή μορφή, του ενός σημείου, επιλέγεται τυχαία ένα σημείο κοπής και ο απόγονος που δημιουργείται παίρνει τα γονίδια δεξιά από το σημείο κοπής από τον ένα γονέα και τα αριστερά από τον άλλο. Έτσι δημιουργείται ένα νέο άτομο,

διαφορετικό από τους γονείς του έχοντας όμως χαρακτηριστικά και των δύο γονέων του.

Για να αποφύγουμε το ενδεχόμενο οι ανασυνδυασμοί να δημιουργούν απογόνους με χειρότερη ποιότητα, δίνουμε τη δυνατότητα να μην υλοποιηθεί, με κάποιο μικρή συνήθως πιθανότητα και έτσι να περάσουν οι γονείς τις γενετικές πληροφορίες τους αναλλοίωτες στην επόμενη γενιά.

### Σχήμα 15

Διάφορα είδη διασταύρωσης



Uniform Crossover: Γονίδια αντιγράφονται τυχαία από τον ένα ή τον άλλο γονέα.

### Μετάλλαξη (Mutation)

Επιλέγονται κάποια γονίδια και μεταβάλλεται τυχαία (ή μή) η τιμή τους. Αυτή η λειτουργία αποτρέπει την σύγκλιση ενός πληθυσμού και οδηγεί σε νέες περιοχές με την προσδοκία καλύτερων λύσεων. Εφαρμόζεται συνήθως με μικρή πιθανότητα.

## Σχήμα 16

### Μετάλλαξη (Mutation)



## 4.5 Γιατί οι Ε.Α. είναι αποτελεσματικοί.

Οι Ε.Α. προσπαθούν να βρουν την καλύτερη λύση μέσα σε ένα χώρο τρισεκατομμυρίων πιθανών λύσεων, πολλές φορές της τάξεως του  $10^{30}$  ή και παραπάνω. Φυσικά για πρακτικούς λόγους είναι αδύνατο να χειριστούν ένα τόσο μεγάλο μέγεθος δεδομένων απευθείας και περιορίζονται σε πληθυσμούς της τάξεως των μερικών εκατοντάδων (συνήθως 100 έως 200). Το πώς γίνεται και ένας τόσο μικρός πληθυσμός (αριθμός πιθανών λύσεων), συνήθως ύστερα από μερικές χιλιάδες δοκιμές, να είναι αποτελεσματικός οφείλεται στο γεγονός ότι κάθε άτομο δεν αναπαριστά ένα μόνο συγκεκριμένο σημείο στο χώρο λύσεων του προβλήματος αλλά πολύ περισσότερα. Αυτό εξηγείται με το θεώρημα των σχημάτων και των δομικών τμημάτων (schemata theorem and building blocks hypothesis).

Ας υποθέσουμε ότι έχουμε δυαδική αναπαράσταση των ατόμων. Κάθε θέση στο άτομο αντιστοιχεί στην τιμή 0 ή 1. Αν προσθέσουμε και ένα τρίτο σύμβολο, το «αδιάφορο» και το παραστήσουμε σαν \*, τότε κάθε άτομο θα δημιουργείται από το συνδυασμό των τριών αυτών στοιχείων. Η σημασία του στοιχείου «αδιάφορο» (\*) είναι καθοριστική και ερμηνεύεται ως εξής: «στη συγκεκριμένη θέση δεν μας ενδιαφέρει αν η τιμή είναι 0 ή 1», δηλαδή η παράσταση 00\*\*10 αντιστοιχεί στο σύνολο των {000010, 000110, 001010, 001110}. Οι παραστάσεις που περιέχουν το σύμβολο «αδιάφορο» λέγονται σχήματα (schemata) ή όμοια πρότυπα

Η επιτυχία των Ε.Α. προέρχεται από το γεγονός ότι στην πραγματικότητα ψάχνουν για σχήματα με καλύτερη ποιότητα. Το πόσο σημαντική είναι η έννοια των σχημάτων φαίνεται με το ακόλουθο παράδειγμα.

Έστω ότι έχουμε τις παρακάτω παραστάσεις με τις αντίστοιχες τιμές ποιότητας:

Παράσταση	Τιμή Ποιότητας
10011	361
00110	36
11000	576
01110	196

Χωρίς να ξέρουμε τις τιμές που αντιστοιχούν στις παραστάσεις των 0 και 1 (που χρησιμοποιούνται για τον υπολογισμό της τιμής ποιότητας), παρατηρούμε ότι σημαντικά καλύτερη τιμή ποιότητας έχουν οι παραστάσεις:

- που αρχίζουν από 1
- που έχουν περισσότερα 1

με την πρώτη παρατήρηση να φαίνεται περισσότερο σημαντική.

Έτσι, μελετώντας μόνο μερικές παραστάσεις από τις συνολικά  $2^5 = 32$ , βγάζουμε σημαντικά συμπεράσματα.

- Η παράσταση που περιγράφεται από το σχήμα  $1****$  θα είναι καλύτερη από την  $0****$ , και
- Τα σχήματα που ταιριάζουν με αυτά του συνόλου  $\{1111*, 111*1, 11*11, 1*111, *1111, 111**, 11**1, 1**11, **111\}$  θα έχουν καλύτερη μέση τιμή ποιότητας από αυτά του συνόλου  $\{0000*, 000*0, 00*00, 0*000, *0000, 000**, 00**0, 0**00, **000\}$ .

Κάθε φορά που ο αλγόριθμος επεξεργάζεται ένα άτομο, στην πραγματικότητα επεξεργάζεται πολλά όμοια πρότυπα. Μια παράσταση 5 ψηφίων αναπαριστά  $2^5 = 32$  διαφορετικές περιπτώσεις αλλά και  $3^5 = 243$  όμοια πρότυπα.

Η δυνατότητα επεξεργασίας μεγάλου αριθμού όμοιων προτύπων σε κάθε γενιά είναι το χαρακτηριστικό του έμμεσου παραλληλισμού (implicit parallelism). Αυτό, σε συνδυασμό με το γεγονός ότι τα μη ποιοτικά σχήματα, με τη διαδικασία της επιλογής γρήγορα απορρίπτονται, κάνουν τους αλγόριθμους αυτούς ισχυρούς και αποτελεσματικούς μηχανισμούς αναζήτησης.

Αν και ο ανασυνδυασμός και η μετάλλαξη με τις μεταβολές που επιφέρουν καταστρέφουν υπάρχοντα όμοια πρότυπα, εντούτοις είναι απαραίτητα για την δημιουργία καλύτερων. Ο βαθμός καταστροφής τους εξαρτάται από την τάξη και το μήκος των σχημάτων. «Μήκος» είναι η απόσταση μεταξύ του πρώτου 0 ή 1 και του τελευταίου 0 ή 1, δηλαδή η παράσταση  $1^{***}0^*$  έχει μήκος 4 και η  $**111^*$  έχει μήκος 2. «Τάξη» είναι ο αριθμός των σταθερών θέσεων 0 ή 1. Έτσι, σχήματα με μικρή τάξη αναπαριστούν μεγαλύτερο αριθμό ατόμων απ' ό,τι αυτά με μεγαλύτερη τάξη.

Καθώς οι γενετικοί τελεστές αλλοιώνουν τα άτομα, είναι προφανές ότι τα μικρού μήκους σχήματα έχουν μεγαλύτερες πιθανότητες επιβίωσης, δηλαδή να μείνουν ανέπαφα. Χαμηλής τάξης σχήματα, που διαχειρίζονται αποτελεσματικά περισσότερα άτομα είναι επίσης πιθανό να μείνουν ανέπαφα. Οι καλύτερες λύσεις προσεγγίζονται με τον συνδυασμό των παραπάνω: με μικρού μήκους χαμηλής τάξης και υψηλής ποιότητας όποια πρότυπα. Αυτά λέγονται δομικά τμήματα (building blocks).

Αυτή η θεωρία επηρεάζει τον σχεδιασμό νέων ειδών γενετικών τελεστών και τον τρόπο κωδικοποίησης του προβλήματος.

#### **4.6 Πλεονεκτήματα – Μειονεκτήματα**

Τα πλεονεκτήματα των E.A. γενικά είναι τα παρακάτω:

- Είναι ευέλικτοι και αποτελεσματικοί και μπορούν να εφαρμοστούν σε διαφορετικών ειδών προβλήματα με την ανάλογη αναπαράσταση.
- Είναι σχετικά απλοί στη σύλληψη και την υλοποίησή τους.
- Το χαρακτηριστικό του έμμεσου παραλληλισμού (implicit parallelism) τους κάνει πολύ αποτελεσματικούς σε προβλήματα βελτιστοποίησης.
- Μπορούν να βρουν σχετικά γρήγορα ικανοποιητικές λύσεις σε προβλήματα με υπερβολικά μεγάλο αριθμός λύσεων.
- Μπορούν να χειριστούν δεδομένα με θόρυβο που μεταβάλλονται συνεχώς.
- Μπορούμε να χρησιμοποιούμε υβριδικές μορφές, υιοθετώντας επιμέρους τεχνικές από άλλες εξειδικευμένες μεθόδους για την επίλυση συγκεκριμένων προβλημάτων, βελτιώνοντας την αποτελεσματικότητά τους.



Τα μειονεκτήματά τους είναι:

- Δεν εγγυώνται ότι η λύση του προβλήματος είναι η βέλτιστη αν και αυτό δεν μπορεί να επαληθευθεί πάντα γιατί η καλύτερη λύση πολλές φορές δεν μπορεί να υπολογιστεί με άλλους τρόπους.
- Αν ο αλγόριθμος εγκλωβιστεί σε τοπικό βέλτιστο (μέγιστο ή ελάχιστο) δεν είναι σίγουρο ότι θα μπορεί πάντα να ξεφύγει από αυτό.
- Τέλος, δυσκολία υπάρχει στην επιλογή της κατάλληλης αναπαράστασης του κάθε προβλήματος και της επιλογής των παραμέτρων του αλγορίθμου επειδή οι διαδικασίες επιλογής, ανασυνδιασμών και μετάλλαξης διαφέρουν ανάλογα με τη φύση του προβλήματος.

#### **4.7 Εφαρμογές**

Οι Ε.Α. μπορούν να δώσουν λύσεις σε πάρα πολλούς τομείς και οι εφαρμογές τους αυξάνονται συνεχώς. Παρακάτω αναφέρονται ενδεικτικά μερικοί τομείς που βρίσκουν εφαρμογή οι Ε.Α.

##### ***Χρονοδρομολόγηση (scheduling)***

Στα προβλήματα χρονοδρομολόγησης ζητείται να βρεθεί η βέλτιστη σειρά εκτέλεσης ενός πλήθους εργασιών με τέτοιο τρόπο ώστε να μην παραβιάζονται κάποιοι κανόνες. Συνήθως ζητείται να βρεθεί η βέλτιστη αξιοποίηση διαθέσιμων πόρων ή η ελαχιστοποίηση του χρόνου που απαιτείται για να ολοκληρωθεί μια σειρά εργασιών. Προβλήματα παρουσιάζονται όταν παραβιάζονται συγκεκριμένοι περιορισμοί, όπως π.χ. όταν μια μηχανή καλείται να εκτελέσει δύο διαφορετικές εργασίες την ίδια χρονική στιγμή. Άλλοι περιορισμοί έχουν να κάνουν με τις προτεραιότητες, με το γεγονός δηλαδή ότι για να εκτελεστεί μια εργασία πρέπει απαραίτητα να έχει ολοκληρωθεί η εκτέλεση μίας άλλης.

Τέτοιου είδους προβλήματα αντιμετωπίζονται στη βιομηχανία (π.χ. γραμμές παραγωγής), στις συγκοινωνίες κ.ά.

### *Επεξεργασία εικόνας*

Στο χώρο αυτόν οι Ε.Α. χρησιμοποιούνται για να συγκρίνουν εικόνες που πάρθηκαν σε διαφορετικές χρονικές στιγμές, όπως ακτινογραφίες, εικόνες από δορυφόρους κ.ά., για τον εντοπισμό αντικειμένων σε μια εικόνα, για την αυτόματη κατάταξη αντικειμένων σε μια εικόνα με βάση κάποια κριτήρια, όπως σε εφαρμογές CAD.

### *Νευρωνικά Δίκτυα*

Τα Νευρωνικά Δίκτυα είναι προγράμματα που προσομοιώνουν, με μια πολύ απλοποιημένη μορφή, την λειτουργία του νευρικού συστήματος με στόχο την επίλυση διαφόρων προβλημάτων. Οι Ε.Α. μπορούν να βοηθήσουν τόσο στο στάδιο της εκπαίδευσης ενός Νευρωνικού Δικτύου όσο και στο βέλτιστο σχεδιασμό της δομής του.

### *Προβλέψεις*

Σαν προβλήματα πρόβλεψης θεωρούνται αυτά όπου ζητείται η εκτίμηση μια μελλοντικής συμπεριφοράς με βάση κάποια ιστορικά δεδομένα. Σαν τέτοια αναφέρονται προβλήματα πρόβλεψης καιρού, σεισμών, του αποτελέσματος ενός αγώνα κ.ά.

### *Οικονομία*

Οι Ε.Α. στην οικονομία μπορούν να δώσουν λύσεις σε ένα ευρύ φάσμα εφαρμογών όπως στις οικονομικές προβλέψεις, σε διαδικασίες έγκρισης πίστωσης, σε επενδυτικές αναλύσεις, στις χρηματαγορές κ.ά. Ειδικότερα στο Χρηματιστήριο οι Ε.Α. χρησιμοποιούνται από μεγάλους επενδυτικούς οργανισμούς σε συνδυασμό με Νευρωνικά Δίκτυα και τη Θεωρία του Χάους.

Οι Ε.Α. χρησιμοποιούνται επίσης για την βελτιστοποίηση δύσκολων και σύνθετων συναρτήσεων, στις τηλεπικοινωνίες και τα δίκτυα, στο βιομηχανικό σχεδιασμό, στη ρομποτική και σε πάρα πολλούς άλλους τομείς.

## **4.8 Εξελικτικοί Αλγόριθμοι και Χρηματιστήριο**

Η προσπάθεια πρόβλεψης της μελλοντικής πορείας των τιμών των μετοχών ήταν πάντα μια μεγάλη πρόκληση γιατί έχει σαν αποτέλεσμα σημαντικά οικονομικά οφέλη.

Στην προσπάθεια αυτή χρησιμοποιήθηκαν πολλές διαφορετικές προσεγγίσεις, από τη χρήση εξελιγμένων τεχνικών Τεχνητής Νοημοσύνης έως προβλέψεις με βάση αστρολογικές παραμέτρους.

Οι Εξελικτικοί Αλγόριθμοι, τα τελευταία ειδικά χρόνια, λόγω της αποτελεσματικότητά τους, χρησιμοποιούνται όλο και περισσότερο στο χώρο του Χρηματιστηρίου, κυρίως στην προσπάθεια εκτίμησης της πορείας των τιμών των μετοχών και όχι για την πρόβλεψη των ακριβών μελλοντικών τιμών. Η ακριβής πρόβλεψη των τιμών είναι πολύ δύσκολη αν όχι αδύνατη.

#### 4.9 Πράξεις (θέσεις) στο Χρηματιστήριο Αξιών Αθηνών και στο

##### Χρηματιστήριο Παραγώγων

Ένας επενδυτής ανάλογα με την εκτίμησή του για την πορεία της αγοράς γενικά ή μιας συγκεκριμένης μετοχής ειδικότερα, μπορεί να κάνει τις παρακάτω κινήσεις:

Στο Χ.Α.Α. (όπου δραστηριοποιούνται και οι περισσότεροι επενδυτές) μπορεί να κερδίσει κάποιος μόνο από την άνοδο μιας μετοχής. Έτσι όταν εκτιμήσει ότι επίκειται άνοδος μπορεί να αγοράσει μια μετοχή για να την πουλήσει στη συνέχεια σε υψηλότερη τιμή αποκομίζοντας κέρδος. Σ' αυτή την περίπτωση οι κινήσεις είναι: Αγορά και Πώληση.

Στο Χ.Π.Α. η κατάσταση είναι πιο σύνθετη γιατί υπάρχουν πολλές επιλογές. Ανάλογα με την εκτίμηση που υπάρχει για την πορεία της αγοράς ο επενδυτής μπορεί:

- Να επιλέξει τα Futures, όπου μπορεί να κερδίσει και στην άνοδο αλλά και στην πτώση. Ειδικότερα:
  - ο Αν η εκτίμηση είναι για άνοδο της αγοράς, θα πρέπει να πάρει θέση Long σε κάποιο συμβόλαιο των δεικτών FTSE20 ή FTSE40. Δηλαδή να αγοράσει ένα (ή περισσότερα) συμβόλαια σε κάποια τιμή για να πουλήσει αργότερα (να κλείσει την θέση του) σε τιμή υψηλότερη και να αποκομίσει κέρδος. Σ' αυτή την περίπτωση οι θέσεις του (κινήσεις) είναι : Enter Long (Αγορά συμβολαίου) και Exit Long (Πώληση συμβολαίου)
  - ο Αν η εκτίμηση είναι για πτώση της αγοράς, θα πρέπει να πάρει θέση Short σε κάποιο συμβόλαιο των δεικτών FTSE20 ή FTSE40. Δηλαδή να

πουλήσει ένα (ή περισσότερα) συμβόλαια σε κάποια τιμή (αν και δεν τα έχει) για να το αγοράσει αργότερα (να κλείσει την θέση του) σε τιμή χαμηλότερη και να αποκομίσει κέρδος. Σ' αυτή την περίπτωση οι θέσεις του (κινήσεις) είναι : Enter Short (Πώληση συμβολαίου) και Exit Short (Αγορά συμβολαίου)

- Να επιλέξει τα Options όπου μπορεί να κερδίσει με περισσότερους τρόπους.

Εκτός της ανόδου ή της πτώσης, μπορεί να κερδίσει αν εκτιμήσει σωστά την διακύμανση της αγοράς και εφαρμόσει την ανάλογη στρατηγική (άλλη στρατηγική για περιόδους με έντονες διακυμάνσεις, άλλη για περιόδους χωρίς τάση, κλπ.). Οι θέσεις στην αγορά των Options είναι πολύ σύνθετες και η παρούσα εργασία δεν αναφέρεται σε αυτές.

Φυσικά άλλοι κανόνες και προϋποθέσεις ισχύουν για την Αγορά και άλλοι για την Πώληση. Το ζητούμενο είναι να βρεθούν τα καλύτερα σημεία που θα πρέπει να γίνουν οι κινήσεις από την μεριά του επενδυτή και έχουν εφαρμογή στην αγορά του Χ.Α.Α. και των Futures του Χ.Π.Α.

Στη συνέχεια της εργασίας, ο όρος Αγορά θα είναι ισοδύναμος και με την θέση EnterLong και ο όρος Πώληση με τη θέση EnterShort.

#### **4.10 Επενδυτικά μοντέλα**

Επενδυτικό μοντέλο είναι ένα σύστημα δημιουργίας σημάτων Αγοράς / Πώλησης μετοχών με βάση την εκτίμηση για την μελλοντική πορεία των τιμών. Βασίζεται σε κάποιες παραδοχές και ανάλογα το είδος του μοντέλου επεξεργάζεται διαφορετικά δεδομένα, όπως ιστορικά αρχεία τιμών, θεμελιώδη δεδομένα, οικονομικούς δείκτες ή και συνδυασμούς αυτών κ.ά.

Η χρήση ενός επενδυτικού μοντέλου, εκτός από τον καθορισμό των καλύτερων χρονικών σημείων για την εκτέλεση των πράξεων, αποτρέπει την υποκειμενικότητα και τον συναισθηματισμό, παράγοντες καθοριστικοί στην επίτευξη των στόχων ενός επενδυτή. Ακόμα και οι λανθασμένες προτάσεις του μοντέλου είναι αποδεκτές, στα όρια του στατιστικού λάθους, και αναμενόμενες γιατί δεν μπορεί να υπάρξει σύστημα χωρίς λάθη. Ένα καλό μοντέλο θα πρέπει να έχει λίγα λάθη ή οι ζημιές που θα προκύπτουν από

τέτοια λάθη να είναι όσο το δυνατόν μικρότερες. Ο κυριότερος λόγος της αποτυχίας πολλών μικροεπενδυτών στην προσπάθεια επίτευξης ικανοποιητικών κερδών, είναι το γεγονός ότι η πλειοψηφία των πράξεων που διενεργούν γίνεται με ψυχολογικά κυρίως κριτήρια, βασίζεται σε φήμες και δεν είναι πειθαρχημένη ακολουθώντας πιστά, γενικά αποδεκτούς κανόνες επενδυτικής συμπεριφοράς.

Η επένδυση στο Χρηματιστήριο θεωρείται ότι είναι ένα παιχνίδι θετικού αθροίσματος (positive sum game), όπου μακροπρόθεσμα όλοι κερδίζουν από τα αυξανόμενα κέρδη των εταιριών και γενικότερα την βελτίωση της οικονομικής κατάστασης με μόνη προϋπόθεση την Αγορά μετοχών και τη διακράτησή τους για μεγάλο χρονικό διάστημα. Στα περισσότερα Χρηματιστήρια, μεταξύ αυτών και του Χρηματιστηρίου Αξιών Αθηνών, οι αποδόσεις την τελευταία δεκαετία ήταν εντυπωσιακές, σε μερικές όμως χώρες, όπως στην Ιαπωνία, ήταν δραματικές. Το χρονικό διάστημα διακράτησης των μετοχών μπορεί να είναι από μερικά λεπτά έως και μερικές δεκαετίες και είναι καθοριστικό για το σχεδιασμό του επενδυτικού μοντέλου.

Η αποτελεσματικότητα ενός επενδυτικού μοντέλου μπορεί να εκτιμηθεί με πολλούς διαφορετικούς τρόπους. Μερικοί χρησιμοποιούν τη σύγκριση με την τακτική της διακράτησης, όπου τα κέρδη που προκύπτουν από τις πράξεις Αγοράς / Πώλησης που προτείνει το μοντέλο συγκρίνονται με τις τιμές στην αρχή και στο τέλος της εξεταζόμενης περιόδου, θεωρώντας ότι έγινε πράξη Αγοράς στην αρχή και Πώλησης στο τέλος της περιόδου. Άλλοι συγκρίνουν τα κέρδη του μοντέλου με αυτά των αποδόσεων των κρατικών τίτλων σταθερού εισοδήματος και μηδενικού ρίσκου. Οι περισσότεροι όμως ψάχνουν για μοντέλα με τη υψηλότερη δυνατή απόδοση, στο μικρότερο χρόνο και η αξιολόγησή τους γίνεται συγκριτικά μεταξύ τους.

Τα δεδομένα που επεξεργάζεται ένα επενδυτικό μοντέλο είναι κυρίως ιστορικά αρχεία τιμών και μακροοικονομικά ή μικροοικονομικά στοιχεία. Από αυτά, τα αρχεία τιμών πρακτικά είναι διαθέσιμα σε όλους ενώ τα οικονομικά όχι. Επίσης, τα οικονομικά ή θεμελιώδη δεδομένα συνήθως χρειάζονται και πρόσθετη επεξεργασία που απαιτεί ειδικές γνώσεις Οικονομικών και Λογιστικής. Για τους λόγους αυτούς οι περισσότεροι μικροεπενδυτές χρησιμοποιούν δεδομένα ιστορικών τιμών

#### 4.10.1 Είδη επενδυτικών μοντέλων

Υπάρχουν πολλές διαφορετικές προσεγγίσεις στη δημιουργία επενδυτικών μοντέλων. Κριτήρια επιλογής του καταλληλότερου μοντέλου είναι το ύψος των κεφαλαίων για επένδυση, ο χρονικός ορίζοντας της επένδυσης και το ρίσκο που προτίθεται να πάρει ο κάθε επενδυτής. Με βάση τα παραπάνω κριτήρια μπορούμε να ξεχωρίσουμε δύο διαφορετικά επενδυτικά μοντέλα που απευθύνονται σε διαφορετικές κατηγορίες επενδυτών.

Το πρώτο αφορά κυρίως θεσμικούς επενδυτές, που είναι μεγάλοι επενδυτικοί, ασφαλιστικοί ή τραπεζικοί οργανισμοί. Αυτοί διαχειρίζονται συνήθως πολύ μεγάλα ποσά και για το λόγο αυτό οι αποφάσεις του επηρεάζουν την γενικότερη πορεία της αγοράς. Συνήθως ο χρονικός ορίζοντας της επένδυσής τους είναι μερικά χρόνια και κύριο μέλημά τους είναι η αποκόμιση κερδών με το μικρότερο δυνατό ρίσκο. Χρησιμοποιούν μοντέλα που βασίζονται κυρίως σε μακροοικονομικά δεδομένα, όπως πορεία επιτοκίων, προσφορά χρήματος, πληθωρισμός, ρυθμός οικονομικής ανάπτυξης κ.ά. καθώς και σε μικροοικονομικά δεδομένα των εταιριών όπως ο λόγος Τιμή/Κέρδη ανά μετοχή (P/E), ο λόγος Τιμή/Λογιστική Αξία (P/BV), ο ρυθμός αύξησης των κερδών (Growth) κ.ά..

Το δεύτερο αφορά τους μικρούς ιδιώτες επενδυτές, οι οποίοι γενικά επενδύουν μικρά ποσά σε σχέση με τον συνολικό τζίρο του Χρηματιστηρίου. Αυτό τους επιτρέπει να μπορούν να κάνουν περισσότερες πράξεις τον χρόνο χωρίς οι αποφάσεις τους να επηρεάζουν την γενικότερη εικόνα της αγοράς. Γενικά προτίθενται να αναλάβουν μεγαλύτερο ρίσκο προκειμένου να πετύχουν υψηλότερα κέρδη. Η Τεχνική Ανάλυση είναι το σημαντικότερο εργαλείο που χρησιμοποιούν για την εξαγωγή σημάτων Αγοράς / Πώλησης.

#### 4.11 Τεχνική Ανάλυση

Είναι μια μέθοδος πρόβλεψης της πορείας των τιμών των μετοχών με τη χρήση διαγραμμάτων και τεχνικών δεικτών. Βασίζεται σε τρεις βασικές προϋποθέσεις:

1. Η αγορά προεξοφλεί τα πάντα, δηλαδή οτιδήποτε επηρεάζει την τιμή μιας μετοχής (πολιτικοί λόγοι, οικονομικοί, ψυχολογικοί κλπ.) αντικατοπτρίζονται άμεσα στην

τιμή της μετοχής. Αυτό περιγράφεται με τη φράση «η αγορά δεν κάνει ποτέ λάθος».

2. Οι τιμές κινούνται σε τάσεις ή σχηματισμούς και όχι με τυχαίο τρόπο. Συνεπώς το ζητούμενο είναι ο γρήγορος και σωστός εντοπισμός της τάσης, ώστε κάνοντας τις ανάλογες πράξεις να αποκομίσουμε περισσότερα κέρδη.
3. Η ιστορία επαναλαμβάνεται. Οι γεωμετρικοί σχηματισμοί στα διαγράμματα τιμών των μετοχών διαμορφώνονται από τους κανόνες της προσφοράς και της ζήτησης και απεικονίζουν την ανοδική ή καθοδική ψυχολογία της αγοράς.

Αν και δεν υπάρχει επιστημονική τεκμηρίωση για την ορθότητα των παραπάνω και παρά το γεγονός ότι έχουν διατυπωθεί πολλές ενστάσεις για την ορθότητά τους, εντούτοις στην πράξη αποδεικνύεται ότι τα παραπάνω επαληθεύονται ικανοποιητικά στην πορεία του χρόνου.

Στην προσπάθεια εντοπισμού του καλύτερου σημείου Αγοράς / Πώλησης στο διάγραμμα τιμών, έχουν αναπτυχθεί πάρα πολλοί τεχνικοί δείκτες. Γενικά μπορούμε να τους κατατάξουμε σε δύο κατηγορίες, τους κινητούς μέσους και τους ταλαντωτές.

Οι Κινητοί Μέσοι ακολουθούν την τάση με κάποια καθυστέρηση πάντα ενώ οι Ταλαντωτές χρησιμοποιούνται κυρίως σε περιόδους χωρίς ξεκάθαρη τάση και ανιχνεύουν «υπεραγορασμένες» ή «υπερπουλημένες» καταστάσεις, όπου είναι πιθανή μια αντιστροφή της κατάστασης και ταυτόχρονα στο ξεκίνημα μιας νέας τάσης είναι ιδιαίτερα χρήσιμοι για τον καθορισμό της ακριβούς χρονικής στιγμής εισόδου-εξόδου από την αγορά.

Υπάρχουν ακόμη άλλοι δείκτες που εφαρμόζονται πάνω στους όγκους των μετοχών και άλλοι που περιγράφουν την διακύμανση των τιμών.

Με τη βοήθεια τέτοιων δεικτών μπορούμε είτε να εντοπίσουμε τα ακριβή σημεία Αγοράς / Πώλησης είτε να κατανοήσουμε την κατάσταση της αγοράς (ανοδική, πτωτική, ουδέτερη).

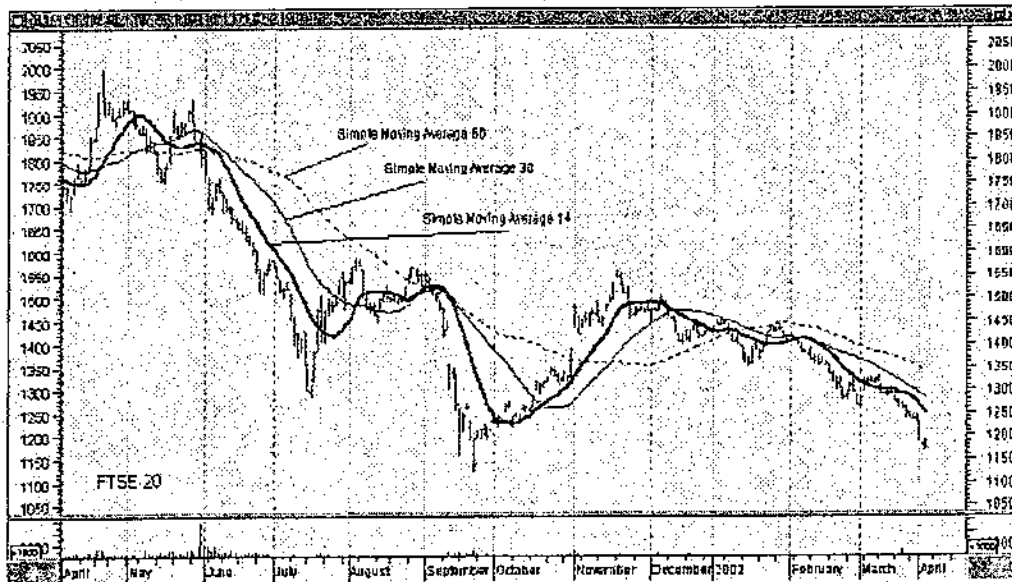
Παρακάτω παρουσιάζονται οι δείκτες που χρησιμοποιούνται με κάποιον τρόπο στην εργασία αυτή.

#### 4.11.1 Απλός Κινητός Μέσος (Simple Moving Average)

Υπολογίζει την Μέση Τιμή για μια καθορισμένη περίοδο. Στο παρακάτω σχήμα φαίνεται ο Κ.Μ. 14 ημερών, 30 ημερών και 60 ημερών. Είναι δείκτης που ακολουθεί την τάση με καθυστέρηση ανάλογη με τον αριθμό ημερών που υπολογίζει. Δίνει σήματα Αγοράς όταν ο τιμή κλεισίματος είναι μεγαλύτερη από τον Κ.Μ. και σήματα Πώλησης όταν η τιμή κλεισίματος είναι μικρότερη από αυτόν. Συνήθως χρησιμοποιείται και ένα φίλτρο λίγων ημερών που όπου η τιμή κλεισίματος πρέπει να παραμείνει πάνω (ή κάτω) από τον δείκτη για να θεωρηθεί το σήμα έγκυρο. Τα ζητούμενα είναι ο αριθμός των ημερών που θα υπολογίζονται ώστε να έχουμε τα λιγότερα λανθασμένα σήματα και ο αριθμός των ημερών που θα χρησιμεύσει σαν επιβεβαίωση του σήματος..

#### Σχήμα 17

Ο Κινητός Μέσος διαφόρων περιόδων





#### 4.11.2 Εκθετικός Κινητός Μέσος (Exponential Moving Average)

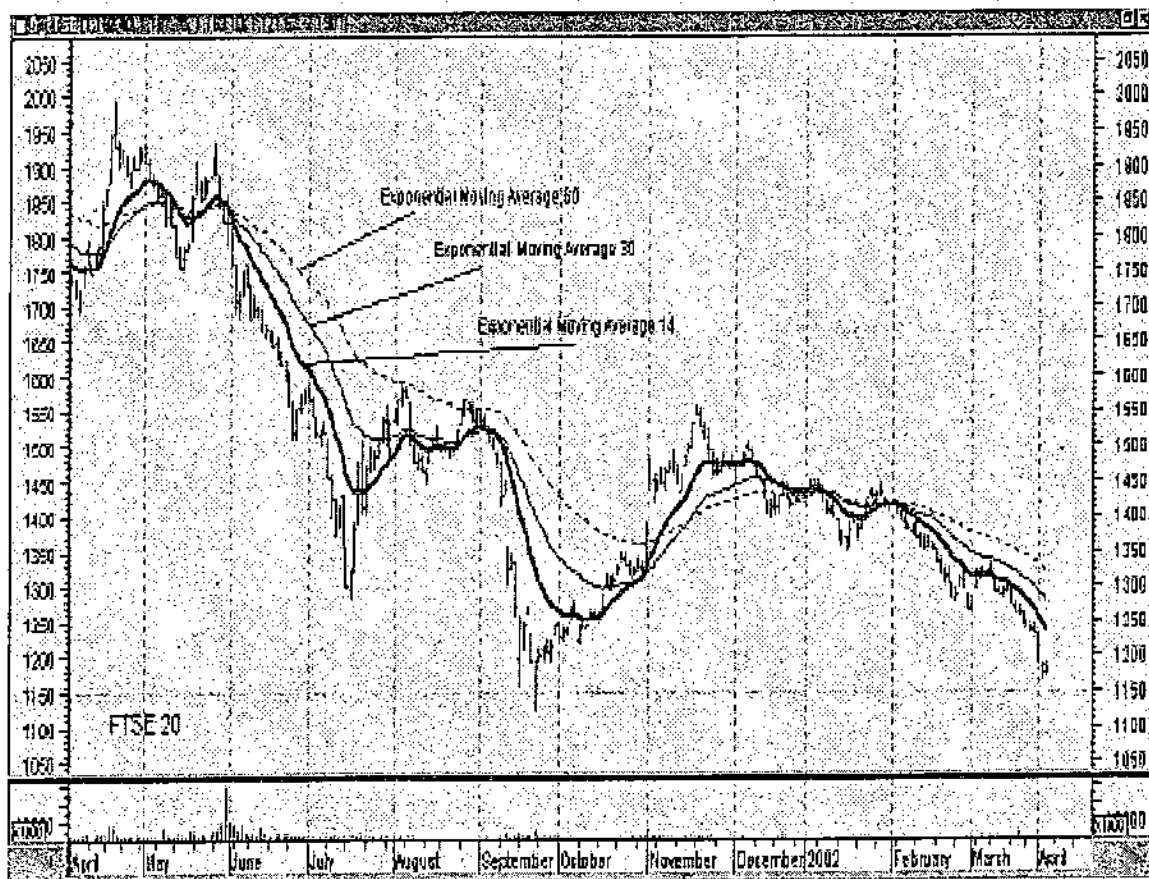
Η χρήση του είναι όμοια με αυτήν του Απλού Κινητού Μέσου με τη διαφορά ότι εδώ δίνεται μεγαλύτερο βάρος στα πιο πρόσφατα δεδομένα. Για την υπολογισμό του Εκθετικού Κινητού Μέσου  $n$  ημερών χρησιμοποιείται η παρακάτω φόρμουλα:

$E.K.M. = (\text{Τελευταία τιμή} * x) + (\text{η τιμή του E.K.M. χθες} * (1-x))$ , όπου το  $x$  υπολογίζεται από τον τύπο:

$$x = \frac{2}{\text{αριθμός } n \text{ ημερών} + 1}$$

**Σχήμα 18**

Ο Εκθετικός Κινητός Μέσος 14, 30 και 60 ημερών



### 4.11.3 Γραμμική Παλινδρόμηση (Linear Regression) και r<sup>2</sup>

Η Γραμμική Παλινδρόμηση προσαρμόζει μια γραμμή στο σύνολο των τιμών μια χρονικής περιόδου. Με τη μέθοδο των ελαχίστων τετραγώνων επιλύει την παρακάτω εξίσωση:

$$Y = m * X + C \text{ όπου}$$

Y: η τιμή, η εξαρτημένη μεταβλητή

X: ο χρόνος, η ανεξάρτητη μεταβλητή

m: κλίση της γραμμής στις μεταβολές της τιμής στη μονάδα χρόνου και

C: το σημείο που τέμνει η ευθεία τον άξονα των τιμών στην αρχή της περιόδου.

Η κλίση της ευθείας δείχνει την κατεύθυνση της τάσης (ανοδική ή καθοδική) και υπολογίζεται από τον τύπο:

$$m = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

Η τιμή  $r^2$  (συντελεστής σταθερότητας) μας δείχνει τον ισχύ της τάσης ή πόσο καλά η ευθεία παλινδρόμησης αναπαριστάει τα δεδομένα. Η τιμή του συντελεστή σταθερότητας για να είναι στατιστικά σημαντική σε ποσοστό 95% συγκρίνεται με μια «κρίσιμη» τιμή, που προκύπτει από τον πίνακα της κατανομής Student και εξαρτάται από το πλήθος των δεδομένων. Η τιμή τους  $r$  υπολογίζεται από τον τύπο:

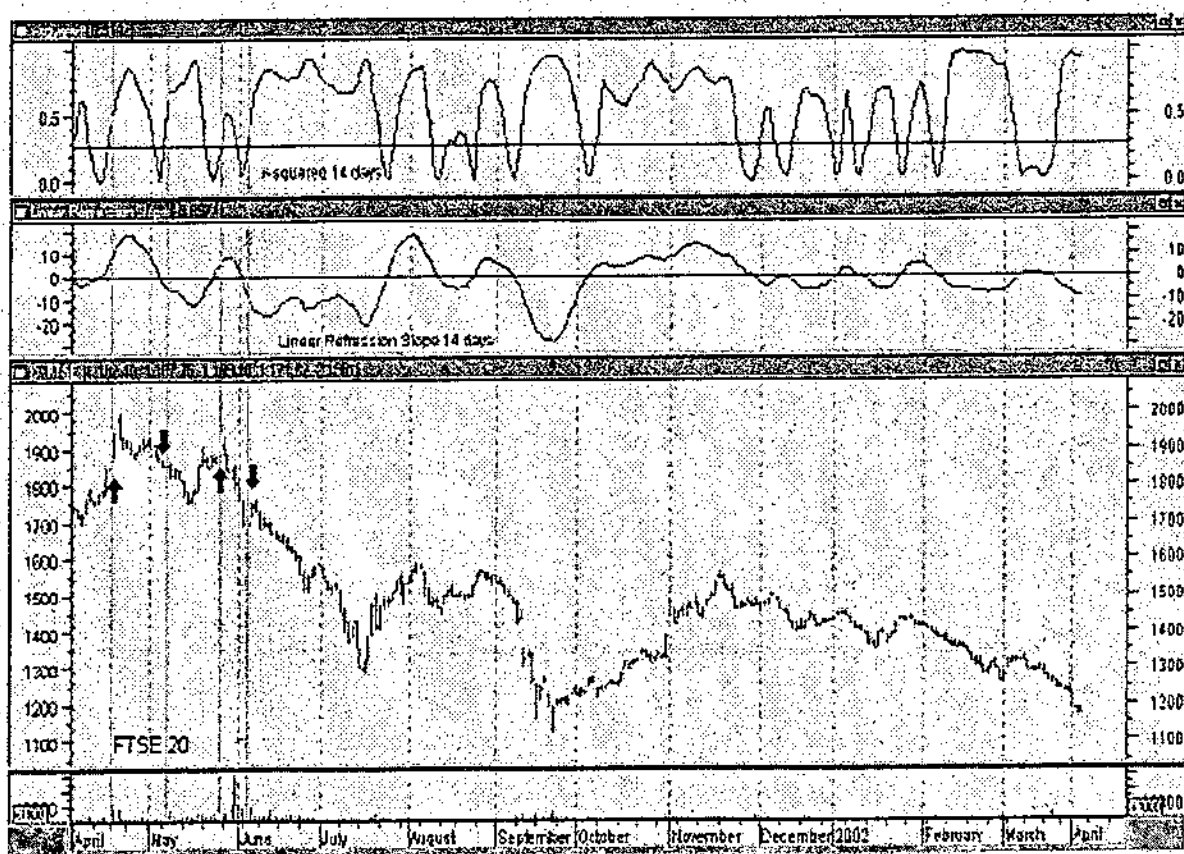
$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \sqrt{n\sum y^2 - (\sum y)^2}}$$

όπου n το πλήθος των δεδομένων στην εξεταζόμενη χρονική περίοδο.

Με βάση τα παραπάνω και για διάστημα 14 ημερών προκύπτουν τα σήματα που βλέπουμε στο επόμενο σχήμα, με κρίσιμη τιμή για τον συντελεστή σταθερότητας για 14 ημέρες το 0,27.

### Σχήμα 18

#### Γραμμική Παλινδρόμηση και Συντελεστής Σταθερότητας.



#### 4.11.4 Μέση Πραγματική Διακύμανση (Average True Range)

Μέση Πραγματική Διακύμανση είναι ο Μέσος Όρος της πραγματικής διακύμανση στη διάρκεια μιας χρονικής περιόδου. Η πραγματική διακύμανση ορίζεται σαν η μεγαλύτερη τιμή μιας εκ των παρακάτω:

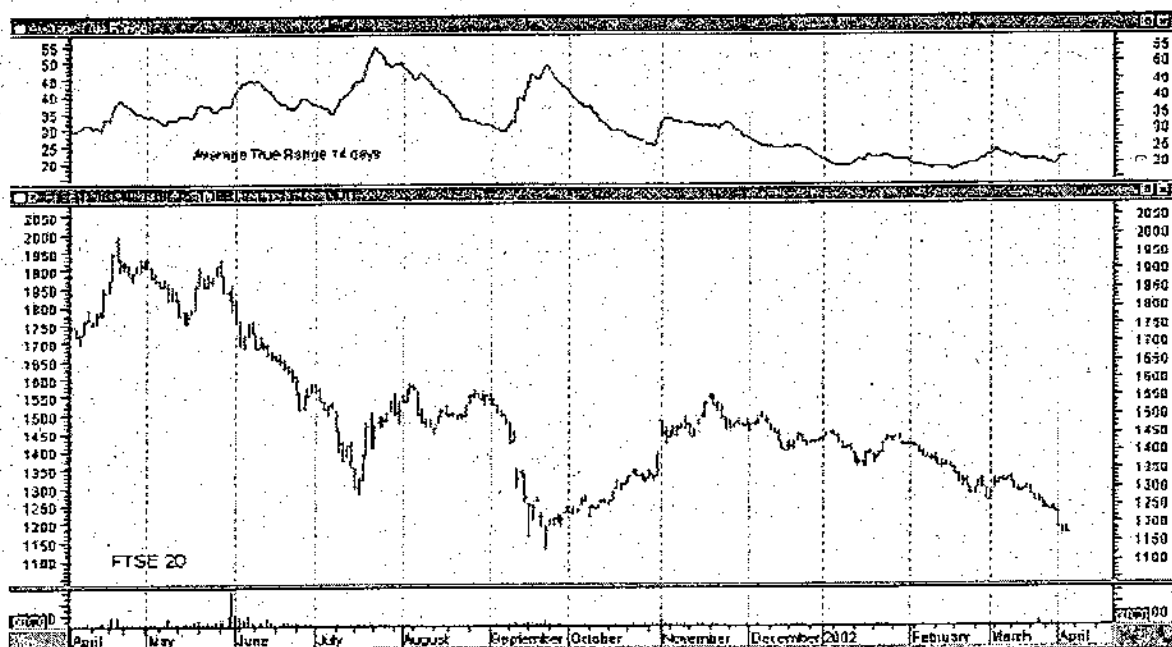
- Η διαφορά μεταξύ υψηλού ημέρας και χαμηλού ημέρας

- ο Η διαφορά μεταξύ χθεσινού κλεισίματος και υψηλού ημέρας
- ο Η διαφορά μεταξύ χθεσινού κλεισίματος και χαμηλού ημέρας

Χρησιμοποιείται για μέτρηση της αστάθειας (volatility) των τιμών γιατί έχει παρατηρηθεί ότι σε περιόδους αντιστροφής τάσεων η μεταβλητότητα των τιμών αυξάνεται.

### Σχήμα 19

#### Μέση Πραγματική Διακύμανση



#### 4.11.5 Δείκτης Σχετικής Ισχύος (Relative Strength Index) – Stochastic

Υπολογίζει την εσωτερική μεταβολή της έντασης μιας μετοχής. Καταγράφει υπεραγορασμές – υπερπουλημένες καταστάσεις και υπολογίζεται από τον παρακάτω τύπο:

$$RSI = 100 - \frac{100}{1 + \frac{UpSum}{DownSum}}$$

όπου,

**UpSum** = Το άθροισμα των (Κλείσιμο - Χθεςινό Κλείσιμο) για ν ημέρες, αν το το κλείσιμο είναι μεγαλύτερο από το χθεςινό Κλείσιμο και

**DownSum** = Το άθροισμα των (Χθεςινό Κλείσιμο - Κλείσιμο) για ν ημέρες, αν το το Κλείσιμο είναι μικρότερο από το χθεςινό Κλείσιμο.

Είναι ένας δείκτης ορμής και καταγράφει άμεσα σημαντικές μεταβολές στην τιμή μιας μετοχής.

Ο Stochastic δείχνει τη θέση της τιμής σε σχέση με πρόσφατα υψηλά / χαμηλά.

Υπολογίζεται από τον παρακάτω τύπο:

$$\text{StochRSI} = (\text{Today's RSI} - \text{RSI\_Low}) / (\text{RSI\_High} - \text{RSI\_Low})$$

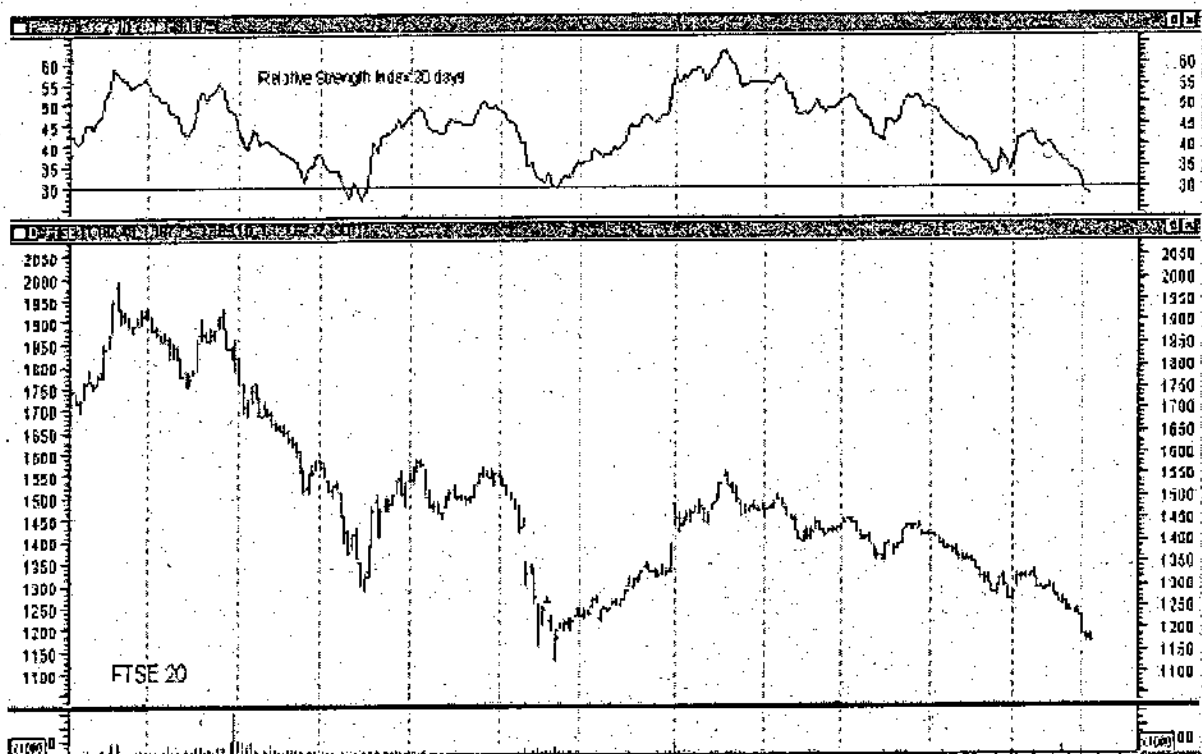
όπου,

**RSI\_Low** = Η χαμηλότερη τιμή του RSI σε μια εξεταζόμενη περίοδο

**RSI\_High** = Η υψηλότερη τιμή του RSI σε μια εξεταζόμενη περίοδο

## Σχήμα 20

### Δείκτης Σχετικής Ισχύος

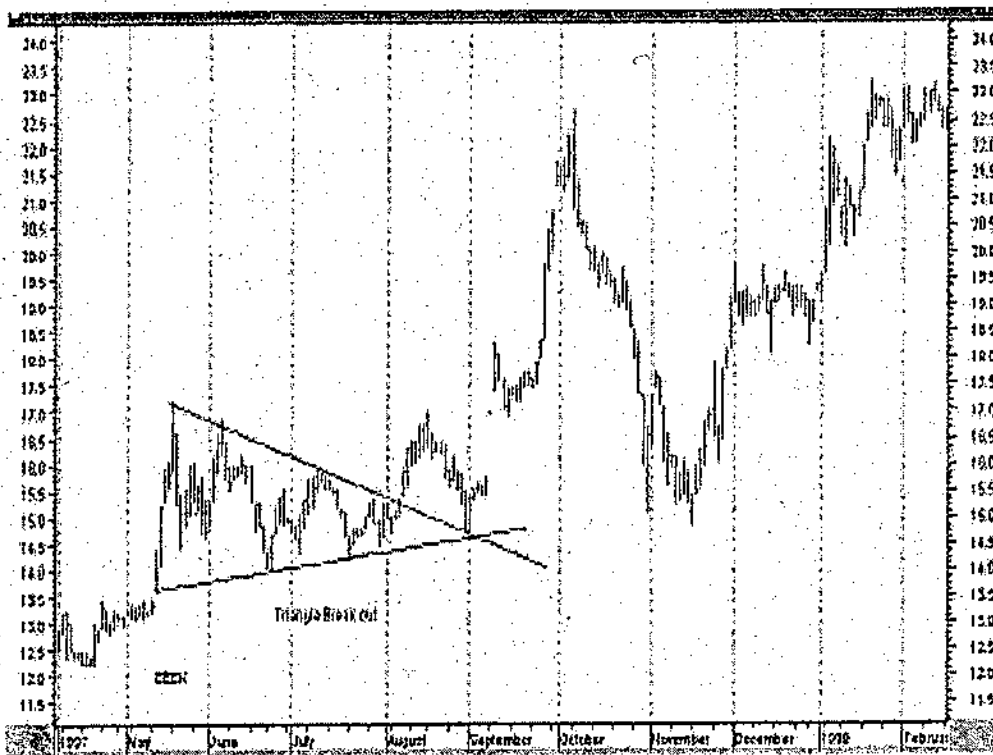


#### 4.11.6 Συμμετρικά Τρίγωνα

Εκτός από τους παραπάνω δείκτες, υπάρχουν και διαγραμματικοί που αποδεικνύονται πολύ αποτελεσματικοί. Ένας από αυτούς είναι και ο σχηματισμός συμμετρικού τριγώνου, όπως φαίνεται στο παρακάτω σχήμα. Ανάλογα με την πλευρά του τριγώνου που θα διασπαστεί, συνήθως ακολουθεί σημαντική άνοδο ή πτώση.

Σχήμα 21

Σχηματισμός Συμμετρικού Τριγώνου



#### 4.12 Το πρόβλημα στη δημιουργία επενδυτικών μοντέλων

Η αξιόπιστη πρόβλεψη της πορείας των τιμών των μετοχών είναι πολύ δύσκολη εξ' αιτίας της πολυπλοκότητας που παρουσιάζει ο χώρος. Η μεταβολή των τιμών επηρεάζεται από πολλούς παράγοντες, όπως τα θεμελιώδη δεδομένα των εταιριών (π.χ.

ανακοίνωση αποτελεσμάτων χρήσης), η πορεία των επιτοκίων, το διεθνές περιβάλλον, οι πολιτικές εξελίξεις, η ψυχολογία αγοράς, η γενικότερη οικονομική κατάσταση και πολλούς άλλους. Η δυσκολία γίνεται ακόμη μεγαλύτερη γιατί πολλοί από αυτούς τους παράγοντες αλληλοεπηρεάζονται. Ο υψηλός βαθμός της στατιστικής συσχέτισης των τιμών των μετοχών με τις ιστορικές τιμές τους και το γεγονός της δημιουργίας κερδοφόρων μοντέλων δείχνει ότι είναι εφικτό να γίνει εκτίμηση της μελλοντικής πορείας των τιμών των μετοχών βασιζόμενη σε ιστορικά αρχεία τιμών.

Το ζητούμενο από ένα επενδυτικό μοντέλο είναι διαφορετικό για κάθε κατηγορία επενδυτών και εξαρτάται από τις απαιτήσεις και τα ιδιαίτερα χαρακτηριστικά τους. Τα επενδυτικά μοντέλα που δημιουργούνται από την παρούσα εργασία αφορούν κυρίως μικροεπενδυτές όπου τα διαθέσιμα για επένδυση κεφάλαια είναι σχετικά μικρά, ο χρονικός ορίζοντας της επένδυσης δεν είναι αυστηρά καθορισμένος και ο στόχος είναι η επίτευξη όσο το δυνατόν περισσότερων κερδών χωρίς να δίνεται ιδιαίτερη βαρύτητα στο ρίσκο της επένδυσης.

Τα δεδομένα που χρησιμοποιούνται είναι ιστορικά αρχεία τιμών γιατί μπορούν να βρεθούν εύκολα, χωρίς ιδιαίτερο κόστος και δεν χρειάζονται πρόσθετη επεξεργασία.

Η ερμηνεία και ο συνδυασμός δεικτών Τεχνικής Ανάλυσης δημιουργεί σύνθετες συναρτήσεις, τους κανόνες. Τα σήματα Αγορών / Πωλήσεων δημιουργούνται από αυτούς τους κανόνες.

Η εργασία αυτή χρησιμοποιεί έναν Εξελικτικό Αλγόριθμο ώστε να επιλέξει τους καλύτερους κανόνες, με τις παραμέτρους του, που θα δίνουν τέτοια σήματα Αγορών / Πωλήσεων ώστε να επιτυγχάνεται το μέγιστο δυνατό κέρδος, με τις λιγότερες απώλειες, για το εξεταζόμενο χρονικό διάστημα.

Είναι ένα πρόβλημα βελτιστοποίησης και προκειμένου να έχει πρακτική εφαρμογή πρέπει να δίνει κερδοφόρο αποτέλεσμα και μεγαλύτερο έναντι της τακτικής της διακράτησης και της απόδοσης των τίτλων του Δημοσίου.

Πολλές παράμετροι του προβλήματος, κυρίως στο σχεδιασμό του, είναι υποκειμενικοί και δημιουργούνται με βάση κυρίως την εμπειρία. Αυτό καθιστά το όλο πρόβλημα εξαιρετικά πολύπλοκο.

Μερικά ιδιαίτερα χαρακτηριστικά του προβλήματος είναι:

- Η ύπαρξη πάρα πολλών διαφορετικών δεικτών Τεχνικής Ανάλυσης που κάνει την δημιουργία κανόνων εξαιρετικά δύσκολη.
- Η ερμηνεία των δεικτών πολλές φορές εξαρτάται από την κατάσταση της αγοράς και μπορεί να είναι διαφορετική
- Η αξιολόγηση της αποτελεσματικότητας του μοντέλου μπορεί να διαφέρει ανάλογα τις απαιτήσεις του κάθε επενδυτή. Έτσι, για άλλους το αποκλειστικό ζητούμενο μπορεί να είναι το μέγιστο κέρδος, για άλλους οι μικρότερες δυνατές απώλειες, για άλλους οι συχνές πράξεις με μικρό κέρδος, για άλλους λίγες πράξεις με μεγάλο κέρδος κλπ.
- Δεν είναι γνωστή η τιμή της βέλτιστης λύσης.
- Ο χώρος λύσεων μπορεί να γίνει εξαιρετικά μεγάλος και περίπλοκος.

#### **4.13 Προηγούμενες Μελέτες για τη δημιουργία επενδυτικών μοντέλων**

Το πρόβλημα της δημιουργίας επενδυτικών μοντέλων είναι πολύ σύνθετο και οι προσεγγίσεις που γίνονται για τη λύση του διαφέρουν. Στη συνέχεια αναφέρονται μερικές αντιπροσωπευτικές μελέτες που έχουν ανακοινωθεί διεθνώς.

Μια εργασία που μελετά και αναλύει αρκετές παραμέτρους του προβλήματος είναι αυτή με τίτλο “Using Genetic Algorithms to find Technical Trading Rules” των F. Allen και R. Karjalainen, 1998 [1]. Με χρήση Γενετικού Προγραμματισμού μελετάται η δημιουργία ενός επενδυτικού μοντέλου με βάση τις ημερήσιες τιμές του δείκτη S&P για μια πολύ μεγάλη χρονική περίοδο, από το 1928 έως το 1995. Οι κανόνες δημιουργήθηκαν έτσι ώστε να εντοπίζεται πότε θα πρέπει να γίνει επένδυση στο δείκτη και πότε όχι, με βάση την λογική ότι σε περιόδους με ανοδική τάση και μικρή αστάθεια τιμών (volatility) επενδύουμε ενώ σε αντίθετη περίπτωση όχι. Το ζητούμενο είναι να βρεθεί η καλύτερη δομή και οι καλύτερες παράμετροι των κανόνων.

Ο χώρος αναζήτησης αποτελείται από λογικούς συνδυασμούς απλών κανόνων, Κινητών Μέσων και Μεγίστων / Έλαχίστων χρησιμοποιώντας ιστορικές τιμές του δείκτη



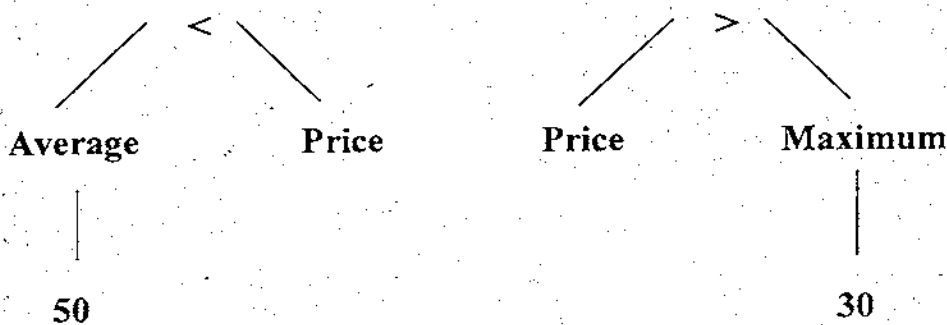
S&P. Οι κανόνες των Κινητών Μέσων χρησιμοποιούνται για τον καθορισμό βραχυχρόνιας ή μακροχρόνιας τάσης και οι κανόνες των Μέγιστων/Ελάχιστων για τον εντοπισμό “πρότυπων σχηματισμών” (patterns) στο αρχείο τιμών.

Οι κανόνες περιέχουν απλές συναρτήσεις ιστορικών τιμών, αριθμητικές και λογικές σταθερές και συναρτήσεις Boolean, που επιτρέπουν τον συνδυασμό χαμηλού επιπέδου δομικών τμημάτων για τη δημιουργία πιο σύνθετων παραστάσεων.

Το σύνολο των συναρτήσεων περιλαμβάνει δύο είδη, πραγματικές και Boolean. Οι συναρτήσεις πραγματικών τιμών περιλαμβάνουν συναρτήσεις υπολογισμού, όπως Κινητού Μέσου και Μεγίστων / Ελάχιστων μιας περιόδου και αριθμητικούς τελεστές (+, -, \*, /). Οι συναρτήσεις Boolean περιλαμβάνουν λογικές συναρτήσεις (If – then – else, and, or, not) και συγκρίσεις δύο τιμών (>, <).

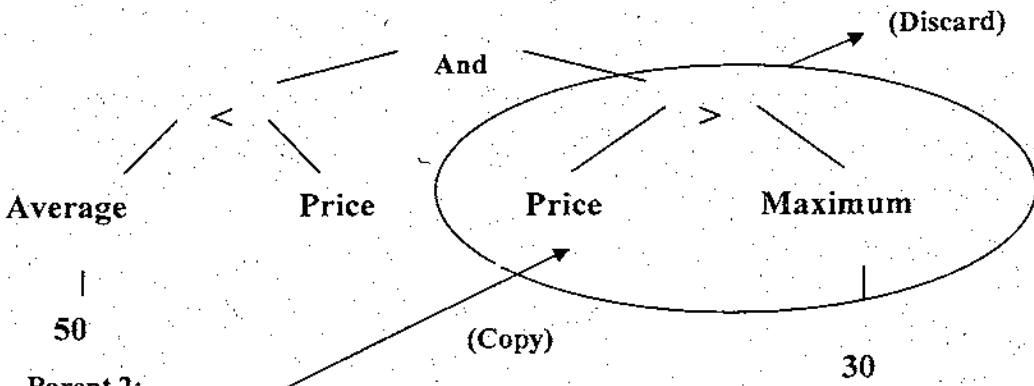
Οι σταθερές είναι είτε πραγματικές τιμές είτε λογικές (true, false).

Η δομή δύο απλών κανόνων φαίνεται στο επόμενο σχήμα.

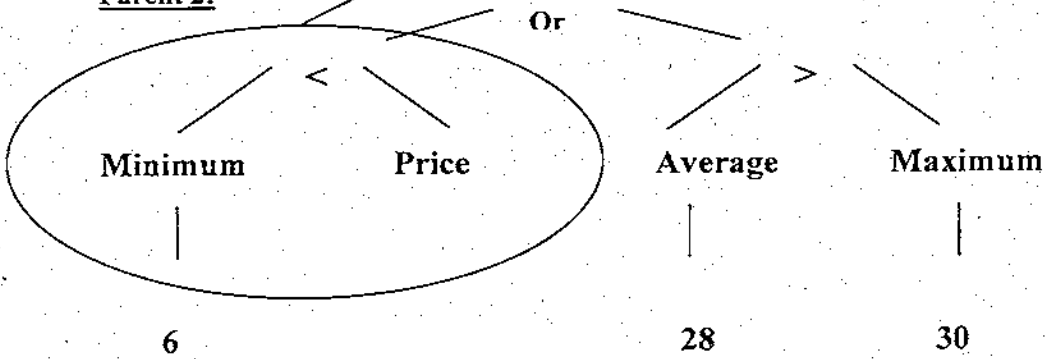


Οι ανασυνδυασμοί γίνονται όπως φαίνεται στο παρακάτω σχήμα.

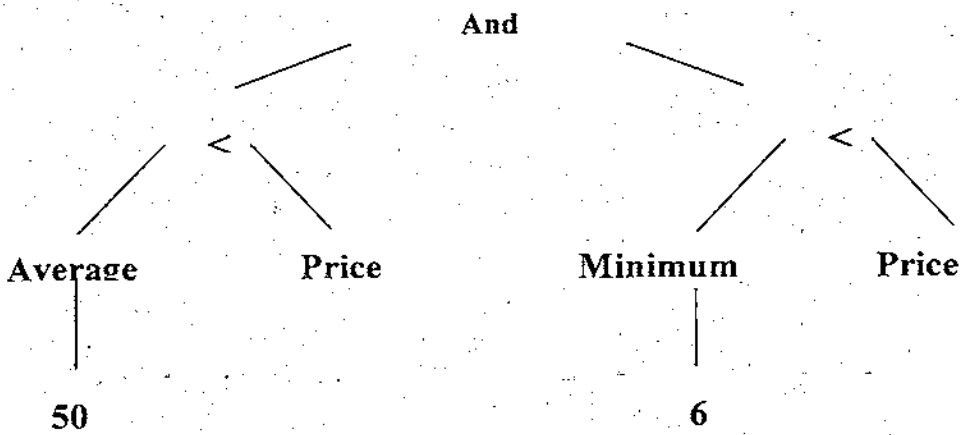
Parent 1:



Parent 2:



Offspring:



Επιλέγεται τυχαία ένας κόμβος από τον ένα γονέα και ένας κόμβος ίδιου τύπου (σ' αυτή την περίπτωση Boolean) από το δεύτερο.

Ο υπολογισμός της τιμής ποιότητας γίνεται με βάση τη σύγκριση των κερδών που αποδίδει το μοντέλο έναντι της στρατηγικής της διακράτησης. Ειδικότερα, σε περιόδους εντός αγοράς υπολογίζονται τα κέρδη ή οι ζημιές και σε περιόδους εκτός αγοράς ένα σταθερό επιτόκιο. Στους υπολογισμούς λαμβάνεται υπόψη και το κόστος συναλλαγών, που είναι μια πολύ σημαντική παράμετρος του προβλήματος και επηρεάζει σημαντικά το τελικό αποτέλεσμα.

Για τις δοκιμές χρησιμοποιήθηκε πληθυσμός 500 ατόμων, κόμβοι έως 100 και μέγιστο πλήθος επιπέδων 10. Η εξέλιξη συνεχιζόταν για 50 γενιές η τερματιζόταν αν δεν υπήρχε καλύτερη λύση για 25 συνεχείς γενιές.

Η δομή και η πολυπλοκότητα των κανόνων που βρέθηκαν από τον αλγόριθμο ποικίλει, με μήκος κανόνων από 9 έως 94 κόμβους και βάθος από 5 έως 10 επίπεδα. Πολλές φορές οι πολύπλοκοι κανόνες μπορούν να αναχθούν σε απλούστερους χωρίς να επηρεάζεται το αποτέλεσμα.

Από τη μελέτη αυτή προκύπτει ότι σε δεδομένα εκτός δείγματος (out of sample data), η χρήση κανόνων, με τον υπολογισμό του κόστους συναλλαγών, δεν έχει σημαντικά κέρδη έναντι της απλής στρατηγικής της διακράτησης (Buy and Hold).

Αυτή συνέβη κυρίως γιατί οι δείκτες των βασικών κανόνων ήταν πολύ απλοί, οι πράξεις γίνονταν με μια μέρα καθυστέρηση και το μοντέλο δεν περιελάμβανε θέσεις Short.

Μια πολύ ενδιαφέρουσα προσέγγιση γίνεται στην εργασία "The Main Ingredients of Simple Trading Models for Use in Genetic Algorithm Optimization" M. Dacorogna, 1993 [2].

Σ' αυτό αναλύονται διάφορα συστατικά ενός απλού επενδυτικού μοντέλου ώστε να μπορούν να χρησιμοποιηθούν σε έναν Εξελικτικό Αλγόριθμο και γίνονται κάποιες ενδιαφέρουσες προτάσεις.

Η δομή ενός κανόνα για το μοντέλο αυτό είναι:

If  $|M| > K$  then  $G = \text{sign}(I) * \text{sign}(C)$  else  $G=0$  όπου,

M είναι κάποιος δείκτης ή συνάρτηση δύο δεικτών, I είναι ένας δείκτης που δείχνει την κατεύθυνση της αγοράς, C ένας δείκτης που δείχνει την στρατηγική ( αν θα ακολουθείται η τάση ή αν η θέση θα είναι ενάντια στην τάση), K είναι μια σταθερά και G το ύψος του κεφαλαίου που θα επενδύεται σε κάθε πράξη.

Το μοντέλο μπορεί να είναι ένας συνδυασμός τέτοιου είδους κανόνων με χρήση λογικών τελεστών AND και OR. Χρησιμοποιώντας διαφορετικούς δείκτες δημιουργούνται πιο σύνθετες παραστάσεις, με πολλούς κανόνες. Η αναπαράσταση των κανόνων μπορεί να γίνει με δομή δυαδικών δένδρων (binary tree).

Στην εργασία "Combining Rules Learnt Using Genetic Algorithms for Financial Forecasting", K. Mehta, S. Bhattacharyya [3], παρουσιάζεται μια άλλη, διαφορετική προσέγγιση στο πρόβλημα.

Με στόχο τη μείωση της επίδρασης του θορύβου των δεδομένων, σε πρώτο επίπεδο γίνεται μια αξιολόγηση και κατάταξη των βασικών κανόνων που περιέχουν απλούς δείκτες. Αυτό επιτυγχάνεται με δοκιμές σε διαφορετικά σύνολα δεδομένων. Οι βασικοί κανόνες συνδυάζονται σε ζεύγη για να αξιολογηθούν για την ακρίβεια των προβλέψεών τους. Ο συνδυασμός σε ζεύγη γίνεται με τέτοιο τρόπο ώστε να βελτιώνεται η απόδοσή τους είτε για σήματα αγορών είτε για σήματα πωλήσεων.

Σε δεύτερο επίπεδο, που δημιουργείται το μοντέλο και γίνεται η τελική αξιολόγηση, οι βασικοί κανόνες συντίθενται σχηματίζοντας από απλές έως πολύ σύνθετες παραστάσεις, με τη χρήση λογικών τελεστών AND, OR.

Το μοντέλο κάθε μέρα δίνει ένα σήμα για παραμονή στη θέση που είχαμε (in the market) ή έξοδο από την αγορά και κλείσιμο θέσης (out of market). Οι θέσεις μπορεί να είναι Long ή Short.

Η δομή του κανόνα είναι:

If <συνθήκη> then <σήμα>

όπου, <συνθήκη> είναι ένας λογικός συνδυασμός όρων (terms), κάθε ένας από τους οποίους έχει τη μορφή <δείκτης, μέρες πίσω><αριθμητικός τελεστής><δείκτης, μέρες πίσω>, που είναι η δομή των βασικών κανόνων που αξιολογήθηκαν σε προηγούμενο επίπεδο. Μια συνθήκη μπορεί να περιέχει συνδυασμό τέτοιων όρων χρησιμοποιώντας τελεστές AND και OR, δημιουργώντας έτσι τις σύνθετες παραστάσεις του κανόνα.

Κάθε όρος αποτελείται από όμοιους δείκτες που μετρούν ίδια πράγματα, όπως ένδειξη τάσης (trend), αστάθεια τιμών (volatility), ορμή (momentum) κλπ.

Το μοντέλο αξιολογείται ανάλογα με το συνολικό κέρδος που αποδίδει και το μέσο κέρδος ανά πράξη. Η απόδοσή του εξαρτάται από την πολυπλοκότητα (complexity)

και την ιδιαιτερότητα (specificity) των κανόνων. Σαν πολυπλοκότητα ορίζεται ο αριθμός των πιθανών συνδυασμών των όρων με τους οποίους ο κανόνας επιστρέφει σήματα αληθή ενώ σαν ιδιαιτερότητα ορίζεται ο ελάχιστος αριθμός όρων που χρειάζεται να ικανοποιηθούν για έναν κανόνα ώστε να ενεργοποιηθεί.

Από την εργασία αυτή προκύπτει ότι η χρήση πολλών κανόνων, που συνδυάζουν άλλους απλούστερους αυξάνοντας την πολυπλοκότητα και την ιδιαιτερότητα του μοντέλου, είτε δίνει πολλά και συχνά σήματα, εξαιτίας του θορύβου των δεδομένων, είτε πολύ λίγα, με αποτέλεσμα, σε κάθε περίπτωση, την μειωμένη απόδοση του μοντέλου.

Σε άλλη μελέτη των Adrian E. Drake και Robert E. Marks, "Genetic Algorithms in Economics and Finance" [4], γίνεται βελτιστοποίηση ενός ή περισσότερων δεικτών Τεχνικής Ανάλυσης και στη συνέχεια, με βάση τους βελτιστοποιημένους αυτούς δείκτες και την χρήση Γενετικού Προγραμματισμού, εισάγονται λογικοί τελεστές κι έτσι δημιουργούνται σύνθετοι, μεταβλητού μήκους κανόνες. Για παράδειγμα, έστω ότι έχουμε τις παρακάτω προτάσεις:

Πρόταση 1. If (10-days Moving Average > 0) then buy else sell

Πρόταση 2. If (15-days Moving Average > 0) then buy else sell

Πρόταση 3. If (25-days Moving Average > 0) then buy else sell

Με τη χρήση Γενετικού Προγραμματισμού ζητούνται προτάσεις της μορφής:

If (Πρόταση 1 says Buy) then Buy else Sell

If (Πρόταση 1 and Πρόταση 2 says Buy) then Buy else Sell

If (Πρόταση 1 and Πρόταση 2 says Buy) OR (Πρόταση 3 says Buy) then Buy else Sell.

Το συμπέρασμα στο οποίο καταλήγει η μελέτη είναι ότι η εισαγωγή κανόνων μεταξύ των δεικτών, που καθορίζουν την επενδυτική συμπεριφορά, δείχνει να είναι πιο αποτελεσματική από την εφαρμογή των δεικτών μόνων τους χωρίς κάποια σχέση μεταξύ τους.

Στη μελέτη "Genetic Algorithms in Finance", C.Tan, D. Chou, 1997, [5] εφαρμόζεται το ίδιο μοντέλο με τη διαφορά ότι οι προτάσεις δεν αναφέρονται σε ιστορικές τιμές αλλά σε άλλα οικονομικά μεγέθη όπως η ρευστότητα του συστήματος, το επιτόκιο των εντόκων γραμματίων, ο δείκτης οικονομικής ανάπτυξης κλπ. όπου ζητούμενο είναι αν η επένδυση θα γίνει σε μετοχές ή σε ομόλογα.

Ένα τέτοιο παράδειγμα θα μπορούσε να είναι η έκφραση:

Αν (ο ρυθμός οικονομικής ανάπτυξης) είναι μεγαλύτερος από X

Και (ο δείκτης ΜΙ-προσφοράς χρήματος) είναι μεγαλύτερος από Y

ΚΑΙ (το επιτόκιο 1 Χρόνου) είναι μικρότερο από Z

Τότε επένδυσε σε μετοχές ΑΛΛΙΩΣ επένδυσε σε Ομόλογα.

Το ζητούμενο σε αυτές τις περιπτώσεις είναι το καλύτερο σύνολο κανόνων. Ο αριθμός των δεικτών, που χρησιμοποιούνται σαν μεταβλητές στους κανόνες, είναι μικρός, συνήθως όχι πάνω από τρεις μιας και ζητούμενο είναι η μεταξύ τους σχέσεις.

Στο άρθρο "Developing Systems With A Rule-Based Approach" του J.O.Katz στο περιοδικό "Technical Analysis of Stock And Commodities", V15:1 23-34, [6] ο συγγραφέας χρησιμοποιεί δύο πολύ απλούς και όμοιους κανόνες, με τέσσερις (4) παραμέτρους ο καθένας, για σήματα Αγορών και Πωλήσεων.

Με τη βοήθεια Εξελικτικού Αλγορίθμου ψάχνει για τις βέλτιστες παραμέτρους των κανόνων αυτών. Οι δοκιμές έγιναν στο δείκτη S&P για χρονική περίοδο 6 ετών.

Οι κανόνες που χρησιμοποιήθηκαν ήταν οι εξής:

Κανόνας 1. If Price (m1) > Price (m2) & (m3) \* Avgdiff1 then (buy/sell)

Κανόνας 2. If Price (n1) > Price (n2) & (n3) \* Avgdiff2 then (buy/sell)

όπου, m1, m2 (n1, n2) οι τιμές που συγκρίνονται, m3 (n3) ένας συντελεστής και avgdiff1 (avgdiff2) η μέση απόλυτη διαφορά των δύο τιμών τις τελευταίες 25 μέρες.

Αν και οι δύο κανόνες είναι True τότε δημιουργείται σήμα Αγορά (Πώλησης) στην τιμή ανοίγματος της επόμενης μέρας.

Λεπτομέρειες για την υλοποίηση του Εξελικτικού Αλγορίθμου δεν αναφέρονται, παρά μόνο ότι σαν συνάρτηση ποιότητας χρησιμοποιείται μόνο το καθαρό κέρδος χωρίς ποινές για λάθη και ζημιές.

Αυτοί οι πολύ απλοί δείκτες, που δεν χρησιμοποιούν δείκτες Τεχνικής Ανάλυσης αλλά μόνο σύγκριση τιμών, μπορούν να δώσουν αποδεκτά αποτελέσματα με λογικές απώλειες. Ακόμη και με οριζόντια διακράτηση την μία μέρα, δηλαδή Αγορά στο άνοιγμα της συνεδρίασης και Πώληση στο κλείσιμο ή το αντίστροφο, όπου το κόστος συναλλαγών είναι υψηλό, βρέθηκαν μοντέλα που χωρίς να είναι εντυπωσιακά σε αποδόσεις εντούτοις σε βάθος χρόνου δίνουν σταθερά ετήσια κέρδη.

Από τις δοκιμές προέκυψε ότι καλοί κανόνες μπορεί να μην είναι αυτοί που περιμένουμε. Γενικά, τα αποτελέσματα κρίνονται ενθαρρυντικά και με τη χρήση βελτιωμένων και πιο πολύπλοκων κανόνων καθώς και καλύτερων μεταβλητών θα ήταν δυνατόν να προκύψουν καλύτερα αποτελέσματα.

Τέλος, σε συνέχεια των δημοσιεύσεων του J.O.Katz στο περιοδικό S&C, στο άρθρο "Genetic Algorithms and Rule-Based System", V15:2, (73-82) [7] γίνεται η παρακάτω προσέγγιση. Επιλέγεται ένα σύνολο δεικτών με μεταβλητές παραμέτρους και με αυτούς δημιουργούνται κανόνες σταθερού μήκους. Με χρήση Ε.Α. επιλέγονται οι καλύτεροι δείκτες από το σύνολο, με τις καλύτερες παραμέτρους. Αυτοί δημιουργούν κανόνες που δίνουν σήματα Αγορών και Πωλήσεων.

#### 4.14 Σχεδιασμός της εφαρμογής

Το επενδυτικό μοντέλο στην παρούσα εργασία σχεδιάστηκε ώστε να ανταποκρίνεται στα ιδιαίτερα χαρακτηριστικά των μικροεπενδυτών του Χρηματιστηρίου. Οι δείκτες Τεχνικής Ανάλυσης που χρησιμοποιήθηκαν περιγράφουν τα κύρια χαρακτηριστικά της αγοράς, την τάση (trend), την αστάθεια τιμών (volatility) και την ορμή (momentum). Αν και υπάρχουν πολλοί δείκτες για την κάθε κατηγορία, επιλέχθηκαν μερικοί γνωστοί και δοκιμασμένοι για την αποτελεσματικότητά τους.

## 4.15 Οι παράμετροι του επενδυτικού μοντέλου

### 4.15.1 Κανόνες

Ένας κανόνας είναι μια συνάρτηση με κάποιες παραμέτρους που δημιουργεί μια συνθήκη, η οποία όταν επαληθεύεται προτείνει ένα σήμα Αγοράς ή Πώλησης. Για παράδειγμα έστω ο παρακάτω κανόνας:

“If Τιμή ( $v1$ ) > Μ.Α. ( $v2$ ) για  $v3$  ημέρες και για  $v3 + 1$  ημέρες πριν Τιμή ( $v1$ ) < Μ.Α. ( $v2$ ), then αγόρασε στο Open της επόμενης μέρας.”

αν υποθέσουμε ότι:

$v1$  = τιμή που αντιστοιχεί στην τιμή Κλεισίματος

$v2$  = 14

$v3$  = 3, τότε ο παραπάνω κανόνας μεταφράζεται ως εξής:

Αν η Τιμή Κλεισίματος είναι μεγαλύτερη της τιμής του Απλού Κινητού Μέσου (της τιμής Κλεισίματος) των 14 ημερών για 3 ημέρες και 4 ημέρες πριν η Τιμή Κλεισίματος ήταν μικρότερη της τιμής του Απλού Κινητού Μέσου (της τιμής Κλεισίματος) των 14 ημερών, τότε αγόρασε στην τιμή Ανοίγματος της επόμενης μέρας.

### 4.15.2 Σήματα Αγορών/Πωλήσεων

Αν επιλέξουμε να χρησιμοποιήσουμε π.χ. 3 κανόνες για σήματα Αγορών, τότε για να προκύψει αποδεκτό σήμα θα πρέπει 3 κανόνες, που έχουν σχεδιαστεί για να δίνουν προτάσεις Αγοράς, να δώσουν ταυτόχρονα πρόταση Αγοράς. Τότε, ανάλογα με τη θέση που είχαμε εκείνη τη στιγμή, θα προβούμε σε Αγορά ή όχι, δηλαδή αν προκύψει σήμα Αγοράς και βρισκόμαστε με μετρητά τότε θα αγοράσουμε, αν βρισκόμαστε ήδη αγορασμένοι το αγνοούμε ενώ αν είμαστε “πουλημένοι” (Short) τότε το αγνοούμε περιμένοντας να πάρουμε πρώτα σήμα να “κλείσουμε” τη θέση πώλησης που είχαμε.

Η σχέση σημάτων, θέσεων και πράξεων δίνεται στον παρακάτω πίνακα.



Σήμα	Σήμα	Σήμα	Σήμα	Προηγούμενη θέση	Πράξη
Αγορά (Enter Long)	Πώληση (Exit Long)	Enter Short	Exit Short		
True	False	False	True	Εκτός αγοράς	Enter Long
True	False	False	True	Long	Ignore
True	False	False	True	Short	Exit Short & Enter Long
True	False	False	False	Εκτός αγοράς	Enter Long
True	False	False	False	Long	Ignore
True	False	False	False	Short	Ignore
False	False	True	False	Εκτός αγοράς	Enter Short
False	False	True	False	Long	Ignore
False	False	True	False	Short	Ignore
False	True	True	False	Εκτός αγοράς	Enter Short
False	True	True	False	Long	Exit Long -Enter Short
False	True	True	False	Short	Ignore
False	True	False	False	Εκτός αγοράς	Ignore
False	True	False	False	Long	Exit Long
False	True	False	False	Short	Ignore
False	True	True	True	Εκτός αγοράς	Ignore
False	True	True	True	Long	Ignore
False	True	True	True	Short	Ignore
False	False	False	True	Εκτός αγοράς	Ignore
False	False	False	True	Long	Ignore
False	False	False	True	Short	Exit Short

#### 4.16 Προϋποθέσεις – Παραδοχές

Οι παρούσα μελέτη για πρακτικούς λόγους σχεδιάστηκε με βάση τις παρακάτω παραδοχές:

- Δεν υπάρχουν κανόνες που να αναφέρονται στο Money / Risk Management. Σε κάθε πράξη τοποθετούμε το σύνολο του διαθέσιμου κεφαλαίου εκείνη τη στιγμή, θεωρώντας ότι αυτό δεν επηρεάζει την ομαλή πορεία της αγοράς και ακόμα ότι όλες οι πράξεις γίνονται στο Άνοιγμα της επόμενης μέρας. Στην πράξη αυτό βέβαια δεν είναι πάντα εφικτό.
- Οι προμήθειες των συναλλαγών υπολογίζονται σταθερές στο 2% ανά συναλλαγή (Αγορά & Πώληση). Δεν υπολογίζεται το κόστος μετακύλισης θέσεων στα παράγωγα (Roll over).

#### 4.17 Οι κανόνες

Ακολουθεί η περιγραφή των κανόνων, οι οποίοι είναι προκαθορισμένοι και δεν μεταβάλλονται από το χρήστη. Οι κανόνες για σήματα Αγοράς (εκτίμηση για άνοδο των τιμών), είναι ίδιοι και για σήματα Enter Long ή Exit Short. Αντίστοιχα, οι κανόνες για τα σήματα Πώλησης (εκτίμηση για πτώση των τιμών), είναι ίδιοι και για σήματα Exit Long & Enter Short.

##### **Κανόνας 0**

Κενό.

Δεν περιέχει τίποτε και επιστρέφει πάντα True. Χρησιμοποιείται στην περίπτωση που ορίσουμε περισσότερους τους ενός κανόνες για ένα σήμα ώστε αν επιλεγεί να υποβιβάσει το σύνολο των κανόνων κατά έναν εφόσον αυτό αποδεικνύεται πιο αποτελεσματικό.

### **Κανόνας 1.**

Απλός Κινητός Μέσος για Τιμή  $>$  M.A. για  $n$  μέρες.

Χρησιμοποιείται ο Απλός Κινητός Μέσος (Simple Moving Average M.A.) για να δώσει σήμα Αγοράς

Κανόνας

**Αν Τιμή ( $v_1$ )  $>$  M.A. ( $v_2$ ) για  $v_3$  μέρες και για  $v_3+1$  μέρες πριν Τιμή ( $v_1$ )  $<$  M.A. ( $v_2$ ), τότε Αγόρασε στο Ανοίγμα της επόμενης μέρας.**

Όπου,

$v_1$ : η τιμή Κλεισίματος, Ανοίγματος, Υψηλού ή Χαμηλού

$v_2$ : η περίοδος του M.A., από 0 – 150 με βήμα 2

$v_3$  : οι μέρες για σύγκριση, από 0-10 με βήμα 1

### **Κανόνας 2.**

Απλός Κινητός Μέσος για Τιμή  $<$  M.A. για  $n$  μέρες.

Χρησιμοποιείται ο Απλός Κινητός Μέσος (Simple Moving Average M.A.) για να δώσει σήμα Πώλησης .

Κανόνας

**Αν Τιμή ( $v_1$ )  $<$  M.A. ( $v_2$ ) για  $v_3$  μέρες και για  $v_3+1$  μέρες πριν Τιμή ( $v_1$ )  $>$  M.A. ( $v_2$ ), τότε Πούλησε στο Ανοίγμα της επόμενης μέρας.**

Όπου,

$v_1$ : η τιμή Κλεισίματος, Ανοίγματος, Υψηλού ή Χαμηλού

$v_2$ : η περίοδος του M.A., από 0 – 150 με βήμα 2

$v_3$  : οι μέρες για σύγκριση, από 0-10 με βήμα 1

### **Κανόνας 3.**

Εκθετικός Κινητός Μέσος για Τιμή  $\geq$  E.M.A. για  $n$  μέρες.

Χρησιμοποιείται ο Εκθετικός Κινητός Μέσος (Exponential Moving Average E.M.A.) για να δώσει σήμα Αγοράς

Κανόνας

**Αν Τιμή (v1) > Ε.Μ.Α. (v2) για v3 μέρες και για v3+1 μέρες πριν Τιμή (v1) < Ε.Μ.Α. (v2), τότε Αγόρασε στο Άνοιγμα της επόμενης μέρας.**

Όπου,

v1: η τιμή Κλεισίματος, Ανοίγματος, Υψηλού ή Χαμηλού

v2: η περίοδος του Ε.Μ.Α., από 0 – 150 με βήμα 2

v3 : οι μέρες για σύγκριση, από 0-10 με βήμα 1

#### **Κανόνας 4.**

Εκθετικός Κινητός Μέσος για Τιμή < Ε.Μ.Α. για v μέρες.

Χρησιμοποιείται ο Εκθετικός Κινητός Μέσος (Exponential Moving Average Ε.Μ.Α.) για να δώσει σήμα Πώληση

Κανόνας

**Αν Τιμή (v1) < Ε.Μ.Α. (v2) για v3 μέρες και για v3+1 μέρες πριν Τιμή (v1) > Ε.Μ.Α. (v2), τότε Πούλησε στο Άνοιγμα της επόμενης μέρας.**

Όπου,

v1: η τιμή Κλεισίματος, Ανοίγματος, Υψηλού ή Χαμηλού

v2: η περίοδος του Ε.Μ.Α., από 0 – 150 με βήμα 2

v3 : οι μέρες για σύγκριση, από 0-10 με βήμα 1

#### **Κανόνας 5.**

Δύο Εκθετικοί Μέσοι για δημιουργία σήματος Αγοράς

Χρησιμοποιούνται δύο Εκθετικοί Κινητοί Μέσοι, ένας με μικρή χρονική περίοδο και ένας με μεγαλύτερη. Σήμα Αγοράς δημιουργείται όταν ο Ε.Κ.Μ. με τη μικρή περίοδο τμήσει από κάτω προς τα πάνω, δηλ. γίνει μεγαλύτερος από τον Ε.Κ.Μ. με την μεγαλύτερη περίοδο. Οι υπολογισμοί γίνονται με την τιμή κλεισίματος.

Κανόνας

**Αν Ε.Μ.Α. (v1) > Ε.Μ.Α. (v2) τότε Αγόρασε στο Άνοιγμα της επόμενης μέρας**

Όπου,

v1: η μικρή περίοδος του Ε.Κ.Μ., από 0-40 με βήμα 1

v2: η μεγαλύτερη περίοδος του Ε.Κ.Μ., από 0-35 με βήμα 2 &  $v2 > v1$

#### **Κανόνας 6.**

Δύο Εκθετικοί Μέσοι για δημιουργία σήματος Πώλησης

Χρησιμοποιούνται δύο Εκθετικοί Κινητοί Μέσοι, ένας με μικρή χρονική περίοδο και ένας με μεγαλύτερη. Σήμα Πώλησης δημιουργείται όταν ο Ε.Κ.Μ. με τη μικρή περίοδο τμήσει από πάνω προς τα κάτω, δηλ. γίνει μικρότερος από τον Ε.Κ.Μ. με την μεγαλύτερη περίοδο. Οι υπολογισμοί γίνονται με την τιμή κλεισίματος.

Κανόνας

**Αν Ε.Μ.Α. (v1) < Ε.Μ.Α. (v2) τότε Πούλησε στο Ανοιγμα της επόμενης μέρας.**

Όπου,

v1: η μικρή περίοδος του Ε.Κ.Μ., από 0-35 με βήμα 2

v2: η μεγαλύτερη περίοδος του Ε.Κ.Μ., από 0-35 με βήμα 2 &  $v2 > v1$

#### **Κανόνας 7.**

Κλίση γραμμής παλινδρόμησης για Αγορά

Χρησιμοποιείται η κλίση της γραμμής παλινδρόμησης και ο συντελεστής σταθερότητας. Αν η κλίση είναι θετική για την περίοδο που εξετάζουμε και η συντελεστής σταθερότητας είναι μεγαλύτερος από μια τιμή που προκύπτει από τον πίνακα της κατανομής Student, τότε το σήμα θεωρείται στατιστικά αποδεκτό και προτείνεται Αγορά. Για τους υπολογισμούς χρησιμοποιείται η τιμή Κλεισίματος.

Κανόνας

**Αν Κλίση (v1) > 0 και  $r^2 \geq$  “κρίσιμη” τιμή τότε Αγόρασε στο Ανοιγμα της επόμενης μέρας.**

Όπου,

v1: η χρονική περίοδος που εξετάζεται, από 3-35 με βήμα 1

“κρίσιμη” τιμή: από τον πίνακα κατανομής Student. Εξαρτάται από το v1

### **Κανόνας 8.**

Κλίση γραμμής παλινδρόμησης για Πώληση

Χρησιμοποιείται η κλίση της γραμμής παλινδρόμησης και ο συντελεστής σταθερότητας. Αν η κλίση είναι αρνητική για την περίοδο που εξετάζουμε και η συντελεστής σταθερότητας είναι μεγαλύτερος από μια τιμή που προκύπτει από τον πίνακα της κατανομής Student, τότε το σήμα θεωρείται στατιστικά αποδεκτό και προτείνεται Πώληση. Για τους υπολογισμούς χρησιμοποιείται η τιμή Κλεισίματος.

Κανόνας

Αν Κλίση ( $v1$ ) < 0 και  $r^2 \geq$  “κρίσιμη” τιμή τότε Πούλησε στο Άνοιγμα της επόμενης μέρας.

Όπου,

$v1$ : η χρονική περίοδος που εξετάζεται, από 0-35 με βήμα 2

“κρίσιμη” τιμή : από τον πίνακα κατανομής Student. Εξαρτάται από το  $v1$

### **Κανόνας 9.**

Stochastic RSI για Αγορά

Ο Stochastic RSI χρησιμοποιείται για τον εντοπισμό ανοδικής τάσης.

Κανόνας

Αν Stochastic RSI ( $v1, v2$ ) >  $v3$  και Stochastic RSI ( $v1, v2$ )  $\leq v3$  χθες, τότε Αγόρασε στο Άνοιγμα της επόμενης μέρας.

Όπου,

$v1$ : η χρονική περίοδος που εξετάζεται, από 1-35 με βήμα 2

$v2$  : οι μέρες πίσω για τον υπολογισμό των υψηλότερων και χαμηλότερων τιμών του RSI, από 0-14 με βήμα 1

$v3$  : μια οριακή τιμή από 0 έως 1

### **Κανόνας 10.**

Stochastic RSI για Πώληση

Ο Stochastic RSI χρησιμοποιείται για τον εντοπισμό ανοδικής τάσης.

Κανόνας

Αν  $\text{Stochastic RSI}(v1, v2) < v3$  και  $\text{Stochastic RSI}(v1, v2) \geq v3$  χθες, τότε

Πούλησε στο Άνοιγμα της επόμενης μέρας.

Όπου,

$v1$ : η χρονική περίοδος που εξετάζεται, από 0-35 με βήμα 2

$v2$ : οι μέρες πίσω για τον υπολογισμό των υψηλότερων και χαμηλότερων τιμών του RSI

$v3$ : μια οριακή τιμή από 0 έως 1

**Κανόνας 11.**

Διάσπαση Τριγωνικού Σχηματισμού για Αγορά

Αν διασπαστεί η άνω πλευρά του τριγώνου τότε δίνεται σήμα Αγοράς.

Χρησιμοποιείται και ο δείκτης Average True Range

Κανόνας

Αν το  $\text{Υψηλότερο}[\text{Υψηλού}, v1+v2] > \text{Υψηλότερου}[\text{Υψηλού}, v2]$  και  $\text{Χαμηλότερο}[\text{Χαμηλού}, v1+v2] < \text{Χαμηλότερου}[\text{Χαμηλού}, v2]$  και  $\text{Κλείσιμο} > \text{Υψηλότερο}[\text{Υψηλό}, v2] + v3 * \text{A.T.R.}(v4)$ , τότε Αγόρασε στο Άνοιγμα της επόμενης μέρας.

Όπου,

$v1$ : η μεγαλύτερη χρονική περίοδος, από 0-50 με βήμα 2

$v2$ : η μικρότερη χρονική περίοδος, από 0-50 με βήμα 2 και  $v2 < v1$

$v3$ : ένας πολλαπλασιαστής από 0 έως 5

$v4$ : η χρονική περίοδος για τον υπολογισμό του A.T.R. από 0-50 με βήμα 2.

**Κανόνας 12.**

Διάσπαση Τριγωνικού Σχηματισμού για Πώληση

Αν διασπαστεί η κάτω πλευρά του τριγώνου τότε δίνεται σήμα Πώλησης.

Χρησιμοποιείται και ο δείκτης Average True Range

Κανόνας

Αν το Υψηλότερο[Υψηλού,  $v1+v2$ ] > Υψηλότερου[Υψηλού, $v2$ ] και Χαμηλότερο[Χαμηλού,  $v1+v2$ ] < Χαμηλότερου [Χαμηλού,  $v2$ ] και Κλείσιμο < Χαμηλότερου[Χαμηλού, $v2$ ] -  $v3 * A.T.R.(v4)$ , τότε Πούλησε στο Άνοιγμα της επόμενης μέρας.

Όπου ,

$v1$  : η μεγαλύτερη χρονική περίοδος, από 0-50 με βήμα 2

$v2$  : η μικρότερη χρονική περίοδος, από 0-50 με βήμα 2 και  $v2 < v1$

$v3$  : ένας πολλαπλασιαστής από 0 έως 5

$v4$  : η χρονική περίοδος για τον υπολογισμό του A.T.R. από 0-50 με βήμα 2.

#### **4.18 Υλοποίηση - Οι παράμετροι του Εξελικτικού Αλγορίθμου**

Για τη διαχείριση των λειτουργιών του Εξελικτικού Αλγορίθμου χρησιμοποιήθηκε το control optiGA της εταιρίας Optiwater. Αυτό μπορεί να διαχειριστεί αναπαραστάσεις σε δυαδική, ακέραια και δεκαδική μορφή. Χειρίζεται ικανοποιητικά όλους τους τελεστές ενός Ε.Α. Οι παράμετροι του Ε.Α. μπορούν να μεταβληθούν κατά τη διάρκεια εκτέλεσης του προγράμματος.

##### **4.18.1 Αναπαράσταση**

Για την αναπαράσταση των χρωμοσωμάτων του συγκεκριμένου προβλήματος επιλέχθηκε ο τρόπος της ακέραιας μορφής με τιμές από 0 έως 999. Κάθε φορά που χρειάζεται να γίνει η “μετάφραση” ενός κανόνα καλείται μια συνάρτηση μετατροπής που μετατρέπει τον αρχικό αριθμό σε μια αποδεκτή τιμή για τον κανόνα ώστε να προκύπτουν ορθοί κανόνες. Κάθε γονίδιο έχει σταθερό μήκος 5 αριθμών. Από αυτούς, ο πρώτος αναφέρεται στον αριθμό του κανόνα και οι υπόλοιποι στις παραμέτρους του κανόνα αυτού. Αν κάποιος κανόνας χρησιμοποιεί λιγότερους από τέσσερις αριθμούς τότε οι υπόλοιποι αγνοούνται. Για παράδειγμα η παράσταση :



122,369,777,666,555,345,678,123,258,456,789,159,357,145,852

θα μπορούσε να αντιστοιχεί σε δύο κανόνες για πράξη Αγοράς ( $2 \times 5 = 10$  αριθμοί) και έναν για πράξη Πώλησης ( $1 \times 5 = 5$ ), δηλαδή η παραπάνω λύση απαιτεί συνολικά δεκαπέντε (15) αριθμούς. Από αυτούς οι πέντε (5) θα μπορούσαν να μεταφραστούν ως εξής:

“Αν  $Low > MovAverS(20)$  ημερών για 4 ημέρες και 5 μέρες πριν  $Low < MovAverS(20)$  τότε αγόρασε στο Open της επόμενης μέρας”.

#### 4.18.2 Γενετικοί τελεστές

Χρησιμοποιούνται και οι δύο γενετικοί τελεστές, ανασυνδυασμός (recombination) και μετάλλαξη (mutation) και εφαρμόζονται με κάποια πιθανότητα. Από το χρήστη ορίζεται αν οι πιθανότητες αυτές θα παραμένουν σταθερές σε όλη τη διάρκεια εκτέλεσης του προγράμματός ή θα προσαρμόζονται ανάλογα με το βαθμό σύγκλισης του πληθυσμού.

##### Ανασυνδυασμός (recombination)

Τα σημεία κοπής βρίσκονται ανάμεσα στους κανόνες. Για παράδειγμα η παράσταση:

123,456,789,987,741,125,123,987,568,258,145,258,159,357,682,58,249,17,985,551

↑ ↑ ↑

έχει τρία σημεία κοπής. Δοκιμές που έγιναν με διαφορετικά σημεία κοπής έδωσαν πολύ αποθαρρυντικά αποτελέσματα και ο λόγος είναι ότι με σημεία κοπής ανάμεσα στους κανόνες καταστρέφονται τα όμοια πρότυπα που είναι απαραίτητα για την πορεία εξέλιξης του αλγορίθμου και χάνεται πολύτιμο γενετικό υλικό. Για τον ίδιο λόγο και οι τιμές των τέκνων προκύπτουν με ανταλλαγή των τιμών των γονέων. Δοκιμές που έγιναν με τα τέκνα να παίρνουν νέες τιμές το Μέσο Όρος ή μικρές μεταβολές των τιμών των γονέων δεν έδωσαν ικανοποιητικά αποτελέσματα. Επίσης, από τις δοκιμές φαίνεται να επιτυγχάνονται καλύτερα αποτελέσματα με τιμές του ποσοστού ανασυνδυασμού από 0,60% έως 0,95%.

##### Μετάλλαξη (mutation)

Τα σημεία που εφαρμόζεται μετάλλαξη επιλέγονται με τυχαίο τρόπο όπως με τυχαίο τρόπο επιλέγεται και η νέα τιμή. Οι δοκιμές έδειξαν ότι γενικά με μικρές τιμές επιτυγχάνονται καλύτερα αποτελέσματα. Ο λόγος ίσως να οφείλεται και στο γεγονός

ότι αν αλλάξει ο αριθμός που αντιστοιχεί στον αριθμό κανόνα τότε και τα υπόλοιπα τέσσερα νούμερα, αν και δεν θα έχουν αλλάξει, θα ερμηνεύονται διαφορετικά.

Οι δοκιμές απέδειξαν ότι η προσαρμογή των πιθανοτήτων των γενετικών τελεστών, με βάση τον βαθμό σύγκλισης του πληθυσμού, βελτιώνει την απόδοση του αλγορίθμου. Έτσι, σε κάθε πληθυσμό υπολογίζεται το ποσοστό των ατόμων που συγκλίνουν προς το καλύτερο μέλος του πληθυσμού. Αν το ποσοστό αυτό είναι μεγαλύτερο από μια καθορισμένη από το χρήστη τιμή, τότε η πιθανότητα ανασυνδυασμού μειώνεται ενώ αυξάνεται η πιθανότητα της μετάλλαξης. Το αντίστροφο ισχύει όταν το ποσοστό των ατόμων που συγκλίνει είναι μικρότερο από μια άλλη τιμή. Ο χρήστης επιλέγει το ποσοστό για σύγκλιση, απόκλιση, το μέτρο σύγκλισης καθώς και το βήμα μεταβολής της πιθανότητας για κάθε γενετικό τελεστή. Καλύτερες τιμές για το βήμα μεταβολής φαίνεται να είναι 0,1 για το ποσοστό ανασυνδυασμού και 0,05 για το ποσοστό μετάλλαξης.

#### 4.18.3 Μηχανισμός επιλογής

Η επιλογή του μηχανισμού επιλογής γίνεται από το χρήστη. Ύστερα από δοκιμές, από τους προτεινόμενους τρόπος (top mate, roulette wheel, roulette ranking, tournament & random) χειρότερος αποδείχθηκε ο τυχαίος ενώ συγκριτικά καλύτερος ο τρόπος επιλογής με τουρνουά.

#### 4.18.4 Τοπικά βέλτιστα

Όταν η καλύτερη τιμή της συνάρτησης ποιότητας δεν έχει αλλάξει για έναν αριθμό γενεών που ορίζεται από το χρήστη, τότε υπάρχει πιθανότητα ο αλγόριθμος να βρέθηκε σε τοπικό βέλτιστο. Είναι ένα συνηθισμένο πρόβλημα στους Ε.Α. και ειδικά στην παρούσα εφαρμογή, όπου δεν γνωρίζουμε την τιμή της βέλτιστης λύσης, θα πρέπει να αναζητούμε διαρκώς καλύτερες λύσεις. Στην προσπάθεια αυτή επιχειρείται η επαναρχικοποίηση του μισού πληθυσμού διατηρώντας τα καλύτερα μέλη. Ύστερα από δοκιμές επιλέχθηκε σαν καλύτερος τρόπος, όχι η τυχαία μεταβολή των τιμών αλλά μια τυχαία μεταβολή κοντά στις τιμές των καλύτερων λύσεων.

#### 4.18.5 Αρχεία ιστορικών τιμών

Τα αρχεία ιστορικών τιμών των μετοχών είναι σε μορφή .txt, είναι comma delimited και δημιουργήθηκαν με export από το Metastock, το δημοφιλέστερο πρόγραμμα Τεχνικής Ανάλυσης. Η γραμμογράφηση των αρχείων είναι:  
<όνομα μετοχής>, <ημερομηνία>, <άνοιγμα>, <υψηλό>, <χαμηλό>, <κλείσιμο>, <όγκος>

#### 4.18.6 Αποθήκευση καλύτερων λύσεων

Με επιλογή του χρήστη, το πρόγραμμα αποθηκεύει τις καλύτερες λύσεις που δημιουργήθηκαν στην πορεία εκτέλεσης του προγράμματος. Μερικές από αυτές μπορεί να βρέθηκαν και μέσα στον ίδιο πληθυσμό. Οι λύσεις αυτές χρειάζονται γιατί μερικές φορές η καλύτερη λύση μπορεί να μην είναι πρακτικά αποτελεσματική.

#### 4.18.7 Υπολογισμός της τιμής της Συνάρτησης Ποιότητας (Fitness Function)

Η συνάρτηση ποιότητας υπολογίζεται από τον παρακάτω τύπο:

$Fitness = (Κέρδος - Ζημία) * \text{ποσοστό κερδισμένων πράξεων}$ , όπου:

Κέρδος = Συνολικό κέρδος - Μεγαλύτερο Κέρδος Πράξης

Ζημία = Συνολική Ζημία

$\text{ποσοστό κερδισμένων πράξεων} = (\text{Αριθμός κερδισμένων πράξεων} - \text{sqrt}(\text{αριθ. Κερδ.πράξεων})) / \text{Συνολικό αριθμό Πράξεων}$ .

Η διαδικασία υπολογισμού της τιμής της συνάρτησης ποιότητας είναι η εξής:

Όποτε χρειάζεται να υπολογιστεί η τιμή αυτή καλείται η fitnessFunction (), η οποία επιστρέφει έναν πίνακα με τις ακέραιες τιμές του ατόμου. Το πλήθος των αριθμών είναι σταθερό, καθορίζεται στην αρχή της εκτέλεσης του προγράμματος και είναι ίσο με τον συνολικό αριθμό των κανόνων x 5. Ο αριθμός των κανόνων για κάθε πράξη καθορίζεται από το χρήστη και παίρνει τιμές από 0 έως 5 (πρακτικά, τιμές μεγαλύτερες από 3 προσθέτουν πολυπλοκότητα στο πρόβλημα χωρίς να βελτιώνουν το αναμενόμενο αποτέλεσμα). Έτσι, αν θέλω για πράξη Αγοράς να έχω τρεις (3) κανόνες, για πράξη Πώλησης να έχω δύο (2) κανόνες, για θέση Short τρεις (3) κανόνες και για θέση Exit Short δύο (2) κανόνες, τότε το άτομο θα αποτελείται από  $(3+2+3+2) * 5 = 50$  αριθμούς, ο καθένας από τους οποίους θα έχει μια τιμή από 0 έως 999. Η πρώτη πεντάδα αριθμών, που είναι μέρος ενός εκ' των τριών κανόνων για

Αγορά, μετατρέπεται στον αντίστοιχο κανόνα με βάση τον πρώτο αριθμό της πεντάδας. Από τον κανόνα αυτόν καλείται μια συνάρτηση αναγωγής τιμών (Rescale ή BiasedScale) που μετατρέπει και τους υπόλοιπους αριθμούς σε αποδεκτές τιμές για τον κανόνα, δηλαδή στο εύρος τιμών που χρειάζεται για να υπολογιστούν οι τιμές των συναρτήσεων που χρησιμοποιεί ο κανόνας.

Η κάθε συνάρτηση διατρέχει τα δεδομένα του ιστορικού αρχείου τιμών και επιστρέφει έναν πίνακα με τις τιμές της συνάρτησης. Ο κανόνας στη συνέχεια διατρέχει πάλι το αρχείο τιμών και με βάση τις τιμές των συναρτήσεων που υπολογίστηκαν πριν, τις τιμές του αρχείου και τους λογικούς τελεστές που χρησιμοποιεί γεμίζει έναν πίνακα με σήματα Αγορών και Πωλήσεων. Αυτό επαναλαμβάνεται για κάθε πεντάδα αριθμών, δηλαδή για κάθε κανόνα. Μόλις τελειώσει η επεξεργασία των κανόνων ο πίνακας με όλα τα σήματα επεξεργάζεται και επιστρέφει έναν άλλο πίνακα με τις αποδεκτές πράξεις Αγοράς / Πώλησης. Τότε καλείται η διαδικασία TS (Trading Simulation), που με βάση τα σήματα αυτά υπολογίζει μια σειρά από μεταβλητές (όπως αριθμό πράξεων, συνολικό κέρδος, ζημία, μέγιστη ζημία πράξης, κλπ.) και από αυτές επιστρέφει τελικά την τιμή της συνάρτησης ποιότητας.

Κατά την εκτέλεση του προγράμματος ο χρήστης μπορεί να παρακολουθεί την πορεία εξέλιξης του πληθυσμού και της καλύτερης έως τώρα λύσης σε διαφορετικά διαγράμματα. Ακόμη, μπορεί να δει λεπτομέρειες που αφορούν την καλύτερη λύση καθώς και όποιας άλλης επιθυμεί από τις καλύτερες έως τώρα και την χρονική διάρκεια από την έναρξη επίλυσης του προβλήματος.

#### **4.19 Αποτελέσματα – Συμπεράσματα**

Στην προσπάθεια εύρεσης των καλύτερων παραμέτρων του αλγορίθμου που θα βελτίωναν την αποτελεσματικότητά του, έγιναν πολλές εκτελέσεις τις εφαρμογής. Κύριος στόχος ήταν να βρεθούν οι καλύτερες τιμές πιθανοτήτων για τις μεταλλάξεις και τους ανασυνδυασμούς, να αξιολογηθεί η αποτελεσματικότητα της χρήσης της προσαρμογής (adaptive) των γενετικών τελεστών και τέλος να ελεγχθεί ο τρόπος ανασυνδυασμού. Έτσι έγιναν διαφορετικές εκτελέσεις διατηρώντας όλες τις άλλες παραμέτρους αμετάβλητες εκτός της παραμέτρου που μας ενδιέφερε και που σε κάθε σύνολο εκτελέσεων είχε διαφορετικές τιμές.

## 4.20 Οι Δοκιμές

Σε όλες τις δοκιμές χρησιμοποιήθηκαν τα δεδομένα του Γενικού Δείκτη του Χρηματιστηρίου Αξιών Αθηνών από 2-1-2000 έως 8-2-2002. Το διάγραμμα τιμών του Γ.Δ. για την εξεταζόμενη περίοδο φαίνεται στο παράρτημα. Ο αριθμός των κανόνων διατηρήθηκε σταθερός και είναι δύο (2) κανόνες για κάθε θέση, συνολικά 8. Σαν κριτήριο τερματισμού χρησιμοποιήθηκαν οι 400 γενιές χωρίς μεταβολή της τιμής της συνάρτησης ποιότητας ή 2000 επαναλήψεις. Το μέγεθος του πληθυσμού ήταν πάντα 150.

### • Σημεία κοπής

Οι δυνατές επιλογές είναι ενός σημείου, δύο σημείων και ενιαίος (uniform). Για κάθε επιλογή έγιναν 10 εκτελέσεις και χρησιμοποιήθηκαν οι παρακάτω παράμετροι.

Adaptive : Yes, με αρχικές τιμές:

Πιθανότητας Ανασυνδυασμών : 0,9

Πιθανότητας Μετάλλαξης 0,1

βήμα μεταβολής 0,5 για ανασυνδυασμό και 0,2 για μετάλλαξη,

ποσοστό ατόμων του πληθυσμού που συγκλίνουν : 50 %,

ποσοστό ατόμων του πληθυσμού που αποκλίνουν : 20 %,

μέτρο σύγκλισης : 10 %.

Τρόπος επιλογής: Roulette Rank

Τα συγκεντρωτικά αποτελέσματα φαίνονται παρακάτω..

Σημεία Κοπής	Τιμή Fitness			Χρόνος
	Καλύτερη	Μέση	Χειρότερη	
1	3478	1902	714	15:34
2	3197	1866	668	16:04
Uniform	3808	1803	727	15:24

- Δοκιμές με σταθερές τιμές πιθανότητας μετάλλαξης και ανασυνδυασμού.

Οι υπόλοιπες ρυθμίσεις ήταν οι εξής:

Adaptive : Όχι.

Τρόπος επιλογής: Roulette Rank

Σημεία κοπής : 2

Τα συγκεντρωτικά αποτελέσματα φαίνονται παρακάτω.

Τιμές Τελεστών	Τιμή Fitness			Χρόνος
	Καλύτερη	Μέση	Χειρότερη	
Μετάλλαξη: 0,05 Ανασυνδυασμός: 0,9	3567	2144	1012	20:17
Μετάλλαξη: 0,10 Ανασυνδυασμός: 0,9	3004	1891	776	18:42
Μετάλλαξη: 0,15 Ανασυνδυασμός: 0,9	3284	2125	1120	16:23

- Δοκιμές με διάφορες παραμέτρους για τα την προσαρμογή των γενετικών τελεστών.

Τα συγκεντρωτικά αποτελέσματα φαίνονται παρακάτω.

Τιμές Τελεστών	Τιμή Fitness			Χρόνος
	Καλύτερη	Μέση	Χειρότερη	
Σύγκλιση 70% Απόκλιση 40% Κόστος Σύγκλισης 20%	2562	19590	776	15:57
Σύγκλιση 70% Απόκλιση 40% Κόστος Σύγκλισης 10%	3638	1596	825	16:10
Σύγκλιση 70% Απόκλιση 30% Κόστος Σύγκλισης 10%	2827	1515	856	17:45
Σύγκλιση 60% Απόκλιση 20% Κόστος Σύγκλισης 10%	3891	1290	825	17:49

## 4.21 Συμπεράσματα από τις εκτελέσεις του προγράμματος

Επειδή δεν ξέρουμε την τιμή της βέλτιστης λύσης τα αποτελέσματα των μετρήσεων αξιολογούνται μόνο συγκριτικά.

Η χρήση προσαρμοσμένων τελεστών έδωσε σαφώς καλύτερα αποτελέσματα. Αυτό θεωρείται λογικό γιατί οι απαιτήσεις για τις τιμές των τελεστών μεταβάλλονται κατά την πορεία της εξέλιξης. Η δυνατότητα μεταβολής των τιμών αυτών από το χρήστη κατά τη διάρκεια της εξέλιξης φαίνεται να βελτιώνει ακόμη περισσότερο την απόδοση του αλγορίθμου.

Οι τιμές των ποσοστών μετάλλαξης διατηρούνται γενικά σε χαμηλές τιμές και αυτό οφείλεται στο γεγονός ότι αν μεταβληθεί τιμή που αναπαριστά κανόνα, τότε και τα τέσσερα επόμενα νούμερα μεταφράζονται διαφορετικά που στην ουσία είναι σαν να άλλαξε η τιμή τους. Γενικά για τη μετάλλαξη βρίσκονται στην περιοχή του 0,1 και κάτω και για τον ανασυνδυασμό γύρω στο 0,90. Στην αρχή της εξέλιξης παρατηρήθηκαν μεγάλες τιμές στην πιθανότητα μετάλλαξης και αυτό είναι λογικό καθώς τότε η ανάγκη εξερεύνησης του χώρου λύσεων είναι μεγαλύτερη ενώ με την πάροδο του χρόνου αυξάνεται η ανάγκη εκμετάλλευσης της αποκτώμενης γνώσης.

Η επαναρχικοποίηση, που γίνεται για να ξεπεραστεί το μεγάλο πρόβλημα του εγκλωβισμού σε τοπικό βέλτιστο, γίνεται στο μισό πληθυσμό ύστερα από το όριο κολλημένων γενεών παίρνοντας είτε τυχαίες τιμές γύρω από τις μέσες τιμές είτε τιμές κοντά στα άκρα, είτε εντελώς τυχαίες. Η επιλογή με τυχαίες τιμές γύρω από τις μέσες τιμές φαίνεται να δίνει καλύτερα αποτελέσματα.

Για τους τύπους επιλογής τα χειρότερα αποτελέσματα τα έδωσε η τυχαία επιλογή και κάπως καλύτερα η επιλογή Roulette Rank.

Τέλος, σε κάθε περίπτωση απαιτούνται πάρα πολλές εκτελέσεις για να βγάλουμε ασφαλή συμπεράσματα και αυτό δεν μπορεί να γίνει στο πλαίσιο της εργασίας αυτής γιατί η διαδικασία εξέλιξης είναι αρκετά χρονοβόρα. Σε μερικές δοκιμές που έγιναν με χρόνο εκτέλεσης πάρα πολλές ώρες δεν βρέθηκαν σημαντικά καλύτερες λύσεις.

Οι καλύτερες λύσεις έδωσαν περίπου 6-8 πράξεις το χρόνο, γεγονός που ήταν αναμενόμενο για δεδομένα End of Day και για βραχυμεσοπρόθεσμο ορίζοντα επένδυσης, που ήταν και ο αρχικός στόχος στη σχεδίαση της εφαρμογής.

## 4.22 Προτάσεις βελτίωσης

Για την αντιμετώπιση του προβλήματος της δημιουργίας αποτελεσματικών επενδυτικών μοντέλων μπορούν να αναφερθούν πάρα πολλές προτάσεις, από τις οποίες άλλες μπορούν να υλοποιηθούν σχετικά εύκολα ενώ άλλες πολύ δυσκολότερα. Ενδεικτικά αναφέρονται οι παρακάτω:

### **Αντιμετώπιση του time frame**

Το time frame είναι η χρονική μονάδα στην οποία εξάγονται οι τιμές για επεξεργασία. Είναι ίσως το σημαντικότερο πρόβλημα στο χώρο και έχει να κάνει με την “απόσταση” που επεξεργαζόμαστε τα δεδομένα. Συνήθως χρησιμοποιούνται αρχεία τιμών End Of Day (EOD), δηλαδή τιμές που προκύπτουν στο κλείσιμο της ημέρας. Το time frame μπορεί να είναι από πολύ μικρό, όπως ανά πράξη, ανά 1', 5' κλπ. έως αρκετά μεγάλο όπως ανά εβδομάδα, μήνα κλπ. Η επιλογή του κατάλληλου time frame εξαρτάται από το πόσο συχνά πρακτικά μπορούμε να κάνουμε πράξεις, τον χρονικό ορίζοντα της επένδυσης κ.ά. Παρακολουθώντας την πορεία των τιμών από πολύ κοντά θα εντοπίζουμε πιο γρήγορα τα σήματα Αγοράς και Πώλησης, μοιραία όμως τα δεδομένα θα έχουν περισσότερο θόρυβο με αποτέλεσμα να κάνουμε περισσότερα λάθη. Αντίθετα, παρακολουθώντας την μεταβολή των τιμών σε εβδομαδιαία ή μηνιαία βάση, αυτόματα μειώνουμε το θόρυβο στο δεδομένα αλλά επιμηκύνουμε τον χρονικό ορίζοντα της επένδυσης και οι πράξεις μας θα γίνονται με αρκετή καθυστέρηση σε σχέση με τις καλύτερες τιμές Αγοράς / Πώλησης. Ζητούμενο είναι να βρεθεί το βέλτιστο time frame ώστε να παίρνουμε έγκαιρα τα σήματα και να έχουμε τα λιγότερα λάθη.

### **Κανόνες διαχείρισης χαρτοφυλακίου (Money Management)**

Με την έννοια διαχείριση χαρτοφυλακίου εννοούμε γενικά, το ποσοστό του διαθέσιμου κεφαλαίου που θα επενδύουμε κάθε φορά λαμβάνοντας υπόψη κανόνες διασποράς του κινδύνου και ειδικά μοντέλα εκτίμησης ρίσκου και απόδοσης. Επίσης, μπορούμε να εφαρμόσουμε σχετικά απλές τεχνικές προστασίας του κεφαλαίου (stop loss) και προστασίας των κερδών (trailing stops). Μελέτες έχουν δείξει ότι μια αποτελεσματική στρατηγική διαχείρισης ενός χαρτοφυλακίου μπορεί να βελτιώσει δραματικά την απόδοση ενός μέτριου επενδυτικού μοντέλου αλλά το αντίστροφο.



### Σχέσεις μεταξύ κανόνων

Μπορούμε να εισάγουμε κανόνες που δεν θα δίνουν σήματα Αγορών / Πωλήσεων αλλά θα περιγράφουν απλά την κατάσταση της Αγοράς και κανόνες που θα ορίζουν πότε τα σήματα θα είναι αποδεκτά. Για παράδειγμα, αν υπάρχει κανόνας που να ορίζει αν η αγορά είναι σε ανοδική τάση, τότε να αγνοούνται όλα τα σήματα πώλησης έως ότου ο κανόνας που ορίζει την ανοδική τάση μεταβληθεί.

### Νέοι κανόνες

Επειδή υπάρχει ένας πολύ μεγάλος αριθμός δεικτών Τεχνικής Ανάλυσης που συνεχώς διευρύνεται, θα μπορούσαμε να εμπλουτίσουμε την εφαρμογή με νέους κανόνες που χρησιμοποιούν άλλους δείκτες. Ακόμη, θα μπορούσαμε να συνδυάσουμε τη χρήση κανόνων με δείκτες Τεχνικής Ανάλυσης με άλλους κανόνες που δεν θα αναφέρονται σε ιστορικά αρχεία τιμών αλλά θα χρησιμοποιούν μακροοικονομικά και μικροοικονομικά δεδομένα. (π.χ. πληθωρισμός, επιτόκια, θεμελιώδεις δείκτες όπως P/E, P/BV, G/P κλπ.)

### Η Συνάρτηση Ποιότητας

Μπορούν να χρησιμοποιηθούν πολλοί διαφορετικοί τρόποι για τον υπολογισμό της τιμής της συνάρτησης ποιότητας από πολύ απλοί έως εξαιρετικά σύνθετοι. Ένας από αυτούς θα μπορούσε να είναι και η στατιστική αξιολόγηση της βιωσιμότητας της λύσης, με την αξιολόγηση να γίνεται χρησιμοποιώντας δεδομένα τα οποία δεν χρησιμοποιήθηκαν στη φάση της αξιολόγησης της λύσης (out of sample data). Μια τέτοια προσέγγιση θα μεγάλωνε σημαντικά το χρόνο εκτέλεσης γιατί θα απαιτούσε πολύ περισσότερους υπολογισμούς.

Τέλος, θα μπορούσαν να χρησιμοποιηθούν κανόνες για τον εντοπισμό γεωμετρικών σχηματισμών στο διάγραμμα τιμών, όπως διασπάσεις γραμμών τάσης, στήριξης και αντίστασης καθώς και άλλων σχηματισμών, όπως candlesticks, που η εμπειρία δείχνει ότι είναι αποτελεσματικοί.

## ΕΠΙΛΟΓΟΣ

Με βάση τα όσα αναφέρθηκαν στα προηγούμενα κεφάλαια, συμπερασματικά μπορούμε να πούμε ότι ως επιχειρηματική ευφυΐα (BI) θεωρείται η χρησιμοποίηση μεθόδων οι οποίες αναλύουν τα δεδομένα τα οποία είναι απομονωμένες νησίδες πληροφοριών σε ένα DW ή σε μία πολύ μεγάλη βάση δεδομένων και μετατρέποντας τα δεδομένα αυτά σε περιεκτική γνώση είτε προτείνουν είτε βοηθούν στην επιχειρηματική απόφαση. Πρόκειται για στατιστικούς κανόνες ή/και αναλυτικές μεθόδους που βοηθούν στην εξαγωγή συμπερασμάτων με την βοήθεια των εργαλείων της BI.

Για την ανάπτυξη των ΣΥΑ χρησιμοποιούνται τέσσερα πανίσχυρα εργαλεία τα οποία είναι οι αποθήκες δεδομένων (DW), η επεξεργασία δεδομένων σε πραγματικό χρόνο (OLAP), η εξόρυξη γνώσης (Data Mining) και η τεχνολογία που σχετίζεται με τον παγκόσμιο ιστό. Τα τρία πρώτα εργαλεία είναι η διαδικασία την οποία ακολουθεί μια εταιρεία για να προσδώσει στο σύστημα υποστήριξης αποφάσεων τις δυνατότητες εξόρυξης γνώσης. Στο DW σωρεύονται τα δεδομένα τα οποία είναι συνήθως κατάλληλα επεξεργασμένα και αποτελούν δομημένες πληροφορίες. Οι αποθήκες δεδομένων παρέχουν την δυνατότητα για συνεχή αναλυτική επεξεργασία των δεδομένων περιέχοντας ιστορικά και συγκεντρωτικά δεδομένα για υποστήριξη αποφάσεων. Με τα εργαλεία OLAP παρέχονται περισσότερες δυνατότητες και βοηθούν τα υψηλόβαθμα στελέχη των επιχειρήσεων στην ταχεία πρόσβαση και πολυδιάστατη επεξεργασία των δεδομένων με σκοπό τη παρουσίαση και την λύση των προβλημάτων των επιχειρήσεων στις πραγματικές τους διαστάσεις. Η εξόρυξη γνώσης από αποθήκη δεδομένων είναι ότι πιο σύγχρονο χρησιμοποιούν οι αναλυτές σήμερα και έχει την ικανότητα να φανερώνει σχέσεις και εξαρτήσεις που δεν είναι ορατές βοηθώντας έτσι σε τεράστιο βαθμό στην σωστή λειτουργία των επιχειρήσεων.

## **ΒΙΒΛΙΟΓΡΑΦΙΑ**

### **1. Data Mining Techniques**

**Συγγραφέας:** Berry, Michael, Linoff, Gordon

**Εκδότης:** John Wiley and Sons Ltd

**Έτος Έκδοσης:** 1997

### **2. Predictive Data Mining**

**Συγγραφέας:** Weiss, Sholom (Rutgers University, USA), Indurkha, Nitin (University of South Australia)

**Εκδότης:** Elsevier Health Sciences

**Έτος Έκδοσης:** 1997

### **3. Εξόρυξη Γνώσης από Βάσεις Δεδομένων**

**Συγγραφέας:** Βαζιργιάης, Χαλκίδη

**Εκδότης:** Gutenberg

**Έτος Έκδοσης:** 2004

### **4. Using genetic algorithms to find technical trading rules**

**Συγγραφέας:** Franklin Allen, Risto Karjalainen

**Εκδότης:** Financial Economics

**Έτος Έκδοσης:** 1999

### **5. Business Intelligence Roadmap**

**Συγγραφέας:** Moss, Larissa Terpeluk, Atré, Shakuntala

**Εκδότης:** Pearson Professional Education

**Έτος Έκδοσης:** 2003

### **6. Πληροφορίες στο Διαδύκτιο**

- [business.auc.dk](http://business.auc.dk)
- [cdgroup.com](http://cdgroup.com)
- [ibm.com](http://ibm.com)
- [iemagazine.com](http://iemagazine.com)
- [dmreview.com](http://dmreview.com)

