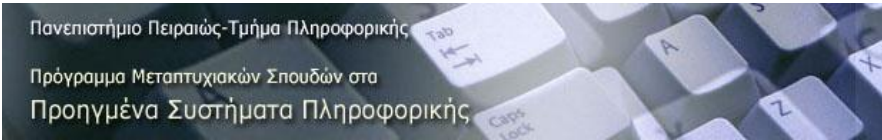




Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής  
Πρόγραμμα Μεταπτυχιακών Σπουδών  
«Προηγμένα Συστήματα Πληροφορικής»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	«Εξόρυξη γνώσης από δεδομένα και εφαρμογή της σε επιχειρησιακές διαδικασίες»
Όνοματεπώνυμο Φοιτητή	Παπαδόπουλος Κων/νος
Αριθμός Μητρώου	ΜΠΣΠ/07021
Κατεύθυνση	Συστήματα Υποστήριξης Αποφάσεων
Επιβλέπων	Δημήτρης Αποστόλου, Λέκτορας



Πανεπιστήμιο Πειραιώς-Τμήμα Πληροφορικής  
Πρόγραμμα Μεταπτυχιακών Σπουδών στα  
Προηγμένα Συστήματα Πληροφορικής

Ημερομηνία Παράδοσης

Ιανουάριος 2010

---



Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

(υπογραφή)

(υπογραφή)

Όνομα Επώνυμο  
Βαθμίδα

Όνομα Επώνυμο  
Βαθμίδα

Όνομα Επώνυμο  
Βαθμίδα

## Περιεχόμενα

<b>1. Εισαγωγή .....</b>	<b>1</b>
1.1 Σκοπός .....	1
1.2 Περιγραφή της εργασίας.....	1
1.3 Οργάνωση της εργασίας .....	1
<b>2. Εξόρυξη Γνώσης από Δεδομένα .....</b>	<b>3</b>
2.1 Τεχνικές εξόρυξης γνώσης από δεδομένα .....	4
2.1.1 Κατηγοριοποίηση (Classification) .....	5
2.1.2 Κανόνες συσχέτισης (Association Rules).....	9
2.1.3 Συσταδοποίηση (Clustering).....	10
2.2 Εφαρμογές Εξόρυξης Γνώσης.....	11
<b>3. Customer Relationship Management.....</b>	<b>14</b>
3.1 Δημοφιλείς CRM εφαρμογές.....	15
3.2 CRM εφαρμογές και σύγχρονες τάσεις.....	16
3.3 CRM και εξόρυξη γνώσης .....	16
3.3.1 Απόκτηση νέων πελατών μέσω εξόρυξης γνώσης .....	17
3.3.2 Αύξηση της αξίας των πελατών μέσω εξόρυξης γνώσης .....	18
3.3.3 Διατήρηση «καλών» πελατών μέσω εξόρυξης γνώσης .....	18
3.4 Εφαρμόζοντας την εξόρυξη γνώσης στο CRM.....	19
3.5 Γνωριμία με το Oracle Siebel 8.1 CRM.....	21
3.5.1 Κλάδοι εφαρμογών.....	21
3.5.2 Βασικές οντότητες .....	22
3.5.3 Η αρχιτεκτονική του Oracle Siebel 8.1 CRM .....	23
3.5.4 Η παραμετροποίηση του Oracle Siebel 8.1 CRM.....	24
3.5.5 Το περιβάλλον του Oracle Siebel 8.1 CRM.....	24
<b>4. Παρουσίαση Εφαρμογής .....</b>	<b>26</b>
4.1 Εξόρυξη γνώσης με τη βοήθεια του WEKA.....	26

4.1.1 Το εργαλείο WEKA.....	26
4.1.2 Χρησιμοποιώντας το WEKA.....	28
4.1.3 Κατηγοριοποίηση με χρήση του WEKA.....	29
4.1.4 Το δένδρο απόφασης.....	33
4.2 Εισαγωγή του δένδρου απόφασης στο Oracle Siebel 8.1 CRM.....	34
4.2.1 Δένδρα Απόφασης, Δομές και Βάσεις Δεδομένων.....	34
4.2.2 Το σχήμα αποθήκευσης μέσα στο Oracle Siebel 8.1 CRM.....	36
4.2.3 Το δένδρο απόφασης μέσα στο Oracle Siebel 8.1 CRM.....	40
4.3 Siebel Smartscripts.....	43
4.3.1 Τι είναι τα Siebel Smartscripts.....	44
4.3.2 Το περιβάλλον των Siebel Smartscripts.....	44
4.3.3 Siebel Smartsript και δένδρο απόφασης.....	46
4.4 Υλοποίηση της επιχειρησιακής διαδικασίας.....	52
4.4.1 Εισαγωγή.....	52
4.4.2 “Front End” διαδικασία.....	53
4.4.3 “Back End” διαδικασία.....	55
4.4.4 Υλοποίηση.....	55
<b>5. Επίδειξη Συστήματος.....</b>	<b>57</b>
5.1 Περιγραφή παραδείγματος.....	57
5.2 Παρουσίαση παραδείγματος.....	58
5.3 Τεχνική επεξήγηση παραδείγματος.....	61
<b>6. Επίλογος.....</b>	<b>64</b>
6.1 Σύνοψη – Συμπεράσματα.....	64
6.2 Δυσκολίες.....	64
6.3 Μελλοντικές Επεκτάσεις.....	65
<b>Βιβλιογραφία - Αναφορές.....</b>	<b>66</b>

## Κατάλογος Εικόνων

ΕΙΚΟΝΑ 1: ΔΙΑΔΙΚΑΣΙΑ ΕΞΟΥΡΕΗΣ ΓΝΩΣΗΣ ΑΠΟ ΔΕΔΟΜΕΝΑ ([3W1]) .....	4
ΕΙΚΟΝΑ 2: Η ΑΠΑΙΤΟΥΜΕΝΗ ΠΡΟΣΠΑΘΕΙΑ ΓΙΑ ΚΑΘΕ ΕΝΑ ΑΠΟ ΤΑ ΣΤΑΔΙΑ ΤΗΣ ΕΞΟΥΡΕΗΣ ΓΝΩΣΗΣ ΑΠΟ ΔΕΔΟΜΕΝΑ ([3W1]) .....	4
ΕΙΚΟΝΑ 3: ΜΕΡΙΚΕΣ ΑΠΟ ΤΙΣ ΠΙΟ ΔΗΜΟΦΙΛΕΙΣ ΤΕΧΝΙΚΕΣ ΕΞΟΥΡΕΗΣ ΓΝΩΣΗΣ ΑΠΟ ΔΕΔΟΜΕΝΑ ([1], [3W1]).....	5
ΕΙΚΟΝΑ 4: ΚΑΤΗΓΟΡΙΕΣ ΤΩΝ ΑΛΓΟΡΙΘΜΩΝ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ([1]) .....	6
ΕΙΚΟΝΑ 5: ΚΑΤΗΓΟΡΙΕΣ ΑΛΓΟΡΙΘΜΩΝ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ ([1]).....	11
ΕΙΚΟΝΑ 6: Η ΠΟΡΕΙΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΟΝ ΧΡΟΝΟ ([3W12]).....	13
ΕΙΚΟΝΑ 7: ΤΑ ΣΤΑΔΙΑ ΤΟΥ ΚΥΚΛΟΥ ΖΩΗΣ ΕΝΟΣ ΠΕΛΑΤΗ ΣΕ ΜΙΑ ΕΠΙΧΕΙΡΗΣΗ .....	16
ΕΙΚΟΝΑ 8: ΣΥΓΚΡΙΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ BB&CC .....	17
ΕΙΚΟΝΑ 9: Η «ΕΞΙΣΩΣΗ» ΠΟΥ ΠΡΟΣΦΕΡΕΙ ΑΝΤΑΓΩΝΙΣΤΙΚΑ ΠΛΕΟΝΕΚΤΗΜΑΤΑ .....	19
ΕΙΚΟΝΑ 10: ΤΑ ΒΗΜΑΤΑ ΓΙΑ ΑΠΟΔΟΤΙΚΕΣ CRM ΕΦΑΡΜΟΓΕΣ.....	20
ΕΙΚΟΝΑ 11: SIEBEL CRM ENTERPRISE ([12]).....	21
ΕΙΚΟΝΑ 12: SIEBEL CRM PROFESSIONAL ENTERPRISE ([12]).....	22
ΕΙΚΟΝΑ 13: SIEBEL CRM ONDEMAND ([12], [3W18]).....	22
ΕΙΚΟΝΑ 14: ΜΕΡΙΚΕΣ ΑΠΟ ΤΙΣ ΒΑΣΙΚΕΣ ΟΝΤΟΤΗΤΕΣ ΤΟΥ ORACLE SIEBEL 8.1 CRM.....	23
ΕΙΚΟΝΑ 15: Η ΑΡΧΙΤΕΚΤΟΝΙΚΗ 3-ΕΠΙΠΕΔΩΝ ΤΟΥ ORACLE SIEBEL 8.1 CRM ([12]) .....	23
ΕΙΚΟΝΑ 16: ΤΑ SIEBEL TOOLS ΠΟΥ ΠΑΡΕΧΕΙ Η ORACLE ΓΙΑ ΤΗΝ ΠΑΡΑΜΕΤΡΟΠΟΙΗΣΗ ΤΗΣ ΕΦΑΡΜΟΓΗΣ.....	24
ΕΙΚΟΝΑ 17: Ο ΚΑΝΟΝΑΣ 80/20 ΠΟΥ ΑΝΑΦΕΡΕΙ Η ORACLE ([12]).....	24
ΕΙΚΟΝΑ 18: Η ΚΕΝΤΡΙΚΗ ΟΘΟΝΗ ΤΟΥ ORACLE SIEBEL 8.1 CRM.....	25
ΕΙΚΟΝΑ 19: Η "ΤΑΥΤΟΤΗΤΑ" ΤΟΥ WEKA ΟΠΩΣ ΠΑΡΟΥΣΙΑΖΕΤΑΙ ΣΤΟ [3W19].....	27
ΕΙΚΟΝΑ 20: ΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΟΥ WEKA.....	27
ΕΙΚΟΝΑ 21: ΤΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ.....	29
ΕΙΚΟΝΑ 22: ΓΡΑΜΜΟΓΡΑΦΗΣΗ ΕΝΟΣ *.ARFF ΑΡΧΕΙΟΥ.....	29
ΕΙΚΟΝΑ 23: ΕΙΣΑΓΩΓΗ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ ΣΤΟ WEKA.....	30
ΕΙΚΟΝΑ 24: ΕΠΙΛΟΓΗ ΑΛΓΟΡΙΘΜΟΥ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ .....	31
ΕΙΚΟΝΑ 25: ΕΚΤΕΛΕΣΗ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ.....	31
ΕΙΚΟΝΑ 26: ΟΠΤΙΚΟΠΟΙΗΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ.....	32
ΕΙΚΟΝΑ 27: ΤΟ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ ΠΟΥ ΠΑΡΗΓΑΓΕ ΤΟ WEKA.....	32
ΕΙΚΟΝΑ 28: ΤΟ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ .....	34
ΕΙΚΟΝΑ 29: ΑΠΟΘΗΚΕΥΣΗ ΔΥΑΔΙΚΟΥ ΔΕΝΔΡΟΥ ΣΕ ΜΟΝΟΔΙΑΣΤΑΤΟ ΠΙΝΑΚΑ (ARRAY) ([15]) .....	35
ΕΙΚΟΝΑ 30: ΔΥΑΔΙΚΟ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ ΣΕ ΠΙΝΑΚΕΣ ([16]) .....	36
ΕΙΚΟΝΑ 31: ΑΠΕΙΚΟΝΙΣΗ ΤΟΥ ΔΕΝΔΡΟΥ ΜΕ ΤΗ ΒΟΗΘΕΙΑ ΣΥΝΔΕΔΕΜΕΝΩΝ ΛΙΣΤΩΝ .....	37
ΕΙΚΟΝΑ 32: Η ΔΟΜΗ ΑΠΟΘΗΚΕΥΣΗΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΕ .....	37
ΕΙΚΟΝΑ 33: ΤΟ ΜΟΝΤΕΛΟ ΟΝΤΟΤΗΤΩΝ-ΣΥΣΧΕΤΙΣΕΩΝ.....	38
ΕΙΚΟΝΑ 34: Η ΚΑΤΗΓΟΡΙΑ ΚΑΙ ΤΟ ΕΝΕΡΓΟ ΣΕΝΑΡΙΟ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ .....	41
ΕΙΚΟΝΑ 35: ΟΙ ΚΛΑΣΕΙΣ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ .....	42
ΕΙΚΟΝΑ 36: ΤΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΓΝΩΡΙΣΜΑΤΑ ΚΑΙ ΟΙ ΤΙΜΕΣ ΤΟΥΣ .....	42
ΕΙΚΟΝΑ 37: ΟΙ ΚΟΜΒΟΙ ΤΟΥ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ .....	43
ΕΙΚΟΝΑ 38: ΟΙ ΜΕΤΑΒΑΣΕΙΣ ΤΟΥ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ.....	43
ΕΙΚΟΝΑ 39: Η ΙΕΡΑΡΧΙΚΗ ΣΧΕΣΗ ΤΩΝ ΔΟΜΩΝ ΤΩΝ SIEBEL SMARTSCRIPTS ([10]).....	44
ΕΙΚΟΝΑ 40: Η ΔΙΑΧΕΙΡΙΣΤΙΚΗ ΟΘΟΝΗ ΤΩΝ SIEBEL SMARTSCRIPTS .....	45
ΕΙΚΟΝΑ 41: ΤΟ ΓΡΑΦΙΚΟ ΠΕΡΙΒΑΛΛΟΝ ΓΙΑ ΤΟΝ ΚΑΘΟΡΙΣΜΟ ΤΩΝ ΜΕΤΑΒΑΣΕΩΝ .....	45
ΕΙΚΟΝΑ 42: ΟΘΟΝΗ SIEBEL SMARTSCRIPT .....	46
ΕΙΚΟΝΑ 43: ΑΠΕΙΚΟΝΙΣΗ ΕΝΟΣ SIEBEL SMARTSCRIPT ΩΣ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ.....	47

ΕΙΚΟΝΑ 44: ΕΝΔΕΙΚΤΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΤΩΝ ΔΥΟ ΠΡΩΤΩΝ ΕΠΙΠΕΔΩΝ ΤΟΥ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ ΜΕ ΤΗ ΒΟΗΘΕΙΑ ΤΩΝ SIEBEL SMARTSCRIPTS.....	48
ΕΙΚΟΝΑ 45: ΠΑΡΑΔΕΙΓΜΑ ΑΠΕΙΚΟΝΙΣΗΣ ΤΟΥ ΚΟΜΒΟΥ "SPONGIOSIS" .....	49
ΕΙΚΟΝΑ 46: ΕΝΔΕΙΚΤΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ ΕΝΟΣ ΔΕΝΔΡΟΥ ΑΠΟΦΑΣΗΣ ΑΠΟ ΕΝΑ SIEBEL SMARTSCRIPT .....	50
ΕΙΚΟΝΑ 47: Ο ΚΩΔΙΚΑΣ ΣΤΟ EVENT HANDLER: QUESTION_PRELEAVE() ΤΩΝ ΕΙΚΟΝΙΚΩΝ «ΕΡΩΤΗΜΑΤΩΝ» .....	51
ΕΙΚΟΝΑ 48: ΓΕΝΙΚΗ ΑΠΟΨΗ ΤΗΣ ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΔΙΑΔΙΚΑΣΙΑΣ .....	53
ΕΙΚΟΝΑ 49: ΤΟ ΔΙΑΓΡΑΜΜΑ ΡΟΗΣ ΤΗΣ ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΔΙΑΔΙΚΑΣΙΑΣ .....	54
ΕΙΚΟΝΑ 50: Η "ΚΥΚΛΙΚΗ" ΔΙΕΚΠΕΡΑΙΩΣΗ ΤΩΝ "BACK END" ΛΕΙΤΟΥΡΓΙΩΝ .....	55
ΕΙΚΟΝΑ 51: ΤΟ WORKFLOW OBJECT ΓΙΑ ΤΗΝ ΚΛΗΣΗ ΤΟΥ SIEBEL SMARTSCRIPT ΚΑΙ ΤΗΝ ΚΑΤΑΧΩΡΗΣΗ ΡΑΝΤΕΒΟΥ .	56
ΕΙΚΟΝΑ 52: Η ΑΝΤΙΣΤΟΙΧΗΣΗ ΙΑΤΡΩΝ - ΕΙΔΙΚΟΤΗΤΩΝ ΣΤΟ ORACLE SIEBEL 8.1 CRM .....	57
ΕΙΚΟΝΑ 53: Ο ΥΠΑΛΛΗΛΟΣ ΑΝΑΖΗΤΑ ΤΟΝ ΑΣΘΕΝΗ ΣΤΗΝ CRM ΕΦΑΡΜΟΓΗ.....	58
ΕΙΚΟΝΑ 54: Η ΔΙΑΔΙΚΑΣΙΑ ΚΑΤΑΧΩΡΗΣΗΣ ΝΕΟΥ ΡΑΝΤΕΒΟΥ ΞΕΚΙΝΑ ΜΕ ΤΟ ΠΑΤΗΜΑ ΕΝΟΣ ΚΟΥΜΠΙΟΥ.....	58
ΕΙΚΟΝΑ 55: Η ΕΠΙΒΕΒΑΙΩΣΗ ΤΩΝ ΣΤΟΙΧΕΙΩΝ ΤΟΥ ΑΣΘΕΝΟΥΣ .....	59
ΕΙΚΟΝΑ 56: ΤΟ ΠΡΩΤΟ "ΕΡΩΤΗΜΑ" .....	59
ΕΙΚΟΝΑ 57: ΤΟ ΔΕΥΤΕΡΟ "ΕΡΩΤΗΜΑ".....	59
ΕΙΚΟΝΑ 58: ΤΟ ΤΡΙΤΟ "ΕΡΩΤΗΜΑ" .....	60
ΕΙΚΟΝΑ 59: ΤΟ ΤΕΤΑΡΤΟ "ΕΡΩΤΗΜΑ" .....	60
ΕΙΚΟΝΑ 60: ΤΟ ΠΕΜΠΤΟ "ΕΡΩΤΗΜΑ" .....	60
ΕΙΚΟΝΑ 61: ΤΟ ΑΠΟΤΕΛΕΣΜΑ ΤΗΣ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ .....	61
ΕΙΚΟΝΑ 62: Η ΟΛΟΚΛΗΡΩΣΗ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΜΕ ΤΗΝ ΚΑΤΑΧΩΡΗΣΗ ΡΑΝΤΕΒΟΥ .....	61
ΕΙΚΟΝΑ 63: ΤΑ ΣΤΟΙΧΕΙΑ ΤΗΣ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗΣ ΟΠΩΣ ΚΑΤΑΧΩΡΗΘΗΚΑΝ ΣΤΗΝ ΔΡΑΣΤΗΡΙΟΤΗΤΑ .....	61
ΕΙΚΟΝΑ 64: Η ΠΟΡΕΙΑ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΜΕΣΑ ΣΤΟ ΔΕΝΔΡΟ ΑΠΟΦΑΣΗΣ .....	62
ΕΙΚΟΝΑ 65: Η ΠΛΗΡΟΦΟΡΙΑ ΤΟΥ ΚΟΜΒΟΥ - ΡΙΖΑ ΤΟΥ ΔΕΝΔΡΟΥ .....	62
ΕΙΚΟΝΑ 66: ΥΠΟΛΟΓΙΣΜΟΣ ΕΠΟΜΕΝΟΥ "ΕΡΩΤΗΜΑΤΟΣ" .....	63
ΕΙΚΟΝΑ 67: ΤΕΡΜΑΤΙΣΜΟΣ ΣΕ ΚΑΠΟΙΑ ΚΛΑΣΗ.....	63

## 1. Εισαγωγή

Στον σύγχρονο κόσμο των επιχειρήσεων και των πληροφοριακών συστημάτων, ο πλέον πολύτιμος πόρος για κάθε επιχείρηση, πέρα από το ανθρώπινο δυναμικό, είναι τα δεδομένα και η γνώση που μπορεί να εξαχθεί από αυτά.

Μεγάλο μέρος του προϋπολογισμού κάθε επιχείρησης διατίθεται πλέον στη κατεύθυνση της συλλογής δεδομένων με στόχο το ανταγωνιστικό πλεονέκτημα που τους προσφέρει η αξιοποίηση της γνώσης που αυτά περιέχουν. Η εποχή κατά την οποία κάθε επιχείρηση διέθετε «χρόνο – χρήμα – ανθρώπους» στην βελτιστοποίηση των εσωτερικών της διαδικασιών έχει παρέλθει. Πλέον, έχοντας εξασφαλιστεί ένα πολύ υψηλό επίπεδο αυτοματισμού σε όλες τις “back-office” διαδικασίες, η προσοχή έχει στραφεί σε όλες τις πελατοκεντρικές δραστηριότητες, καθιστώντας τον «πελάτη» μιας επιχείρησης ως το κέντρο όλων των διαδικασιών.

Τα τελευταία χρόνια, οι επιχειρήσεις προσπαθούν να μάθουν τα πάντα για τους πελάτες τους. Ενδιαφέρονται για τα δημογραφικά δεδομένα τους, τις προτιμήσεις και τις συνήθειες τους στοχεύοντας σε εξατομικευμένες σχέσεις μαζί τους. Ενδεικτικό παράδειγμα της τάσης αυτής αποτελεί η άνθηση επιχειρήσεων τα τελευταία χρόνια που ειδικεύεται στην συλλογή και στην πώληση τέτοιων δεδομένων.

Εύκολα επομένως εξάγεται το συμπέρασμα ότι δύο σημαντικοί στόχοι για κάθε σύγχρονη επιχείρηση είναι η συλλογή δεδομένων και η εκμετάλλευσή τους. Ο πρώτος από τους δύο επιτυγχάνεται σχετικά εύκολα μέσω εμπορικών εφαρμογών που έχουν αναπτυχθεί ειδικά για το σκοπό αυτό. Ο δεύτερος αποτελεί αντικείμενο μελέτης για την επιστημονική περιοχή της εξόρυξης γνώσης από δεδομένα.

### 1.1 Σκοπός

Σκοπός της διπλωματικής εργασίας είναι να συνδυάσει μια σύγχρονη «πελατοκεντρική» εφαρμογή με τις τεχνικές εξόρυξης γνώσης από δεδομένα στα πλαίσια της υλοποίησης μιας επιχειρησιακής διαδικασίας.

### 1.2 Περιγραφή της εργασίας

Στα πλαίσια της διπλωματικής εργασίας θα κινηθούμε αρχικά σε δύο παράλληλες κατευθύνσεις οι οποίες στη πορεία θα συγκλίνουν στον κοινό σκοπό της αποδοτικότερης υλοποίησης μιας επιχειρησιακής διαδικασίας.

Αρχικά θα γνωρίσουμε μερικές από τις πιο δημοφιλείς τεχνικές εξόρυξης γνώσης από δεδομένα όπως η κατηγοριοποίηση, οι κανόνες συσχέτισης κα. Στη συνέχεια θα παρουσιάσουμε την πιο διαδεδομένη «πελατοκεντρική» εμπορική εφαρμογή, το Oracle Siebel 8.1 CRM επικεντρώνοντας το ενδιαφέρον μας σε ένα ιδιαίτερο συστατικό στοιχείο της εφαρμογής, τα Siebel Smartscripts. Ολοκληρώνοντας την εργασία αυτή, θα προσπαθήσουμε να συνδυάσουμε τις δύο αυτές κατευθύνσεις στα πλαίσια της υλοποίησης μιας συγκεκριμένης επιχειρησιακής διαδικασίας.

### 1.3 Οργάνωση της εργασίας

Στο πρώτο κεφάλαιο, γίνεται μια σύντομη εισαγωγή και περιγραφή του αντικείμενου της διπλωματικής εργασίας, καθώς και η οργάνωση του τόμου.

Στο *δεύτερο κεφάλαιο*, περιγράφεται το θεωρητικό υπόβαθρο της εξόρυξης γνώσης από δεδομένα. Παρουσιάζονται μερικές από τις πιο δημοφιλείς τεχνικές, επικεντρώνοντας το ενδιαφέρον μας στην κατηγοριοποίηση και τα δένδρα απόφασης που θα χρησιμοποιηθούν στην πλαίσια της παρούσης διπλωματικής εργασίας.

Στο *τρίτο κεφάλαιο*, περιγράφονται τα συστήματα Customer Relationship Management δίνοντας έμφαση στο Oracle Siebel 8.1 CRM το οποίο και χρησιμοποιήσαμε. Βλέπουμε πώς οι τεχνικές εξόρυξης γνώσης μπορούν να συνδυαστούν με CRM διαδικασίες μέσα από συγκεκριμένα παραδείγματα της παγκόσμιας αγοράς.

Στο *τέταρτο κεφάλαιο*, περιγράφεται λεπτομερώς η υλοποίηση της επιχειρησιακής διαδικασίας. Ερχόμαστε σε επαφή με το εργαλείο WEKA που χρησιμοποιήσαμε για την εξόρυξη γνώσης, γνωρίζουμε σε αρκετό βάθος τα Siebel Smartscrips και παρουσιάζουμε τον τρόπο με τον οποίο όλα τα προηγούμενα μπορούν να συνδυαστούν για την αποδοτική υλοποίηση μιας συγκεκριμένης επιχειρησιακής διαδικασίας.

Στο *πέμπτο κεφάλαιο*, παρουσιάζεται μια επίδειξη του συστήματος. Η επιχειρησιακή διαδικασία που υλοποιήσαμε αναλύεται βήμα προς βήμα με την βοήθεια συγκεκριμένου παραδείγματος.

Η διπλωματική εργασία ολοκληρώνεται με το *έκτο κεφάλαιο* στο οποίο γίνεται μια σύνοψη, παρουσιάζουμε κάποια από τα προβλήματα που προέκυψαν κατά την υλοποίηση και κάνουμε μια σύντομη αναφορά σε μελλοντικές επεκτάσεις του συστήματος.



## 2. Εξόρυξη Γνώσης από Δεδομένα

Διανύουμε μια εποχή όπου τα δεδομένα παίζουν κυρίαρχο ρόλο. Ο όγκος τους αυξάνεται με ταχύτατους ρυθμούς, κάτι που δεν επιτρέπει την άμεση εκμετάλλευσή τους από τους χρήστες. Όλες οι εφαρμογές υπολογιστών, από τις πιο απλές μέχρι τις πιο εξειδικευμένες, «γεμίζουν» βάσεις δεδομένων με πληροφορίες σε μεγάλο βαθμό άχρηστες, οι οποίες κρύβουν ωστόσο πιθανή γνώση που μπορεί να χρησιμοποιηθεί με ποικίλους τρόπους και για διάφορους σκοπούς.

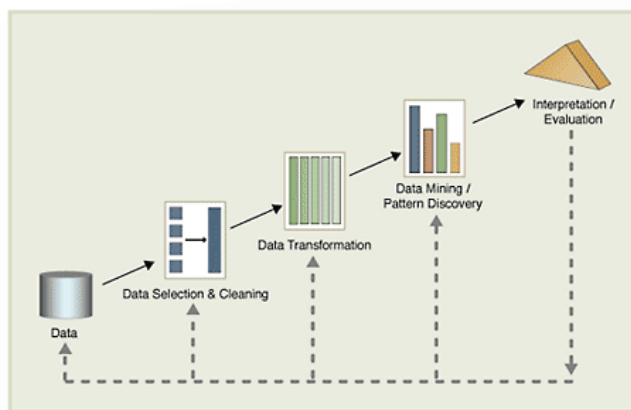
Το διαδίκτυο και η ταχύτατη εξέλιξή του έχει βοηθήσει σε σημαντικό βαθμό ώστε δεδομένα μεγάλου όγκου να παράγονται και να διακινούνται καθημερινώς σε παγκόσμια κλίμακα. Το ζητούμενο επομένως στις μέρες μας δεν είναι τα δεδομένα αυτά καθ' αυτά, αλλά η γνώση που ενδεχομένως περιέχουν. Για το ευρωπαϊκό "Very Long Baseline Interferometry" (VLBI), λόγω χάρη, τα 16 gigabit αστρονομικών δεδομένων που παράγουν τα 16 τηλεσκόπια του ανά δευτερόλεπτο ([2]) μπορεί να μην έχουν καμία πρακτική αξία εάν δεν ανακαλυφθεί η γνώση που αυτά περιέχουν.

Ο τομέας της εξόρυξης γνώσης από δεδομένα γνωρίζει ιδιαίτερη ανάπτυξη στις μέρες μας. Έπαψε να αποτελεί μια καθαρά ερευνητική περιοχή έχοντας πλέον αντίκτυπο και στον πραγματικό κόσμο. Οι επιχειρήσεις, άλλες λιγότερο και άλλες περισσότερο, έχουν αρχίσει να δαπανούν σημαντικό μέρος του προϋπολογισμού τους με σκοπό την εξόρυξη γνώσης από δεδομένα που διαθέτουν, καθώς καθημερινά παράγουν μεγάλο όγκο πληροφοριών τις οποίες και διατηρούν στις «αποθήκες» τους (βάσεις δεδομένων) για πολλά χρόνια. Χαρακτηριστικό παράδειγμα τέτοιων επιχειρήσεων αποτελούν οι εταιρείες τηλεπικοινωνιών και οι τράπεζες, που συνηθίζουν να έρχονται σε επαφή με μεγάλο αριθμό ανθρώπων καθημερινά παράγοντας χιλιάδες bytes δεδομένων για τις βάσεις τους.

Άμεσο συμπέρασμα όλων των παραπάνω είναι ότι ο άνθρωπος πλέον δεν είναι σε θέση να μπορεί να κοιτά και να αναλύει τέτοιες ποσότητες δεδομένων. Ωστόσο, η ανάλυση αυτή είναι απαραίτητη ώστε τα «πολλά» δεδομένα να μετατραπούν σε «λίγη» ποιοτική γνώση η οποία θα είναι άμεσα εκμεταλλεύσιμη από τον άνθρωπο.

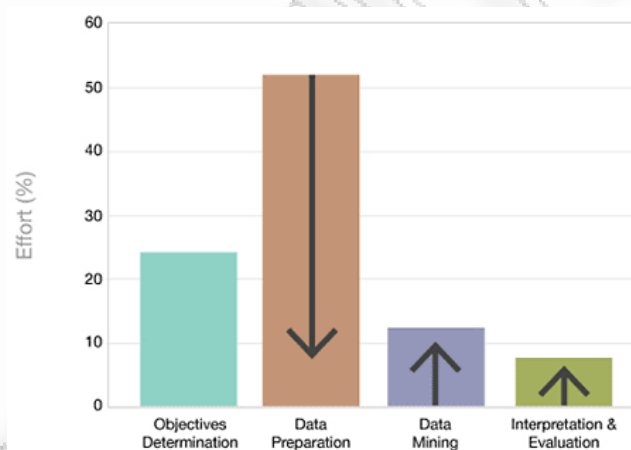
Η εξόρυξη γνώσης από τα δεδομένα μπορεί να οριστεί ως «μια μη τετριμμένη διαδικασία εύρεσης έγκυρων, πρωτότυπων, πιθανώς χρήσιμων και οπωσδήποτε κατανοητών προτύπων μέσα στα δεδομένα» ([2]). Με άλλα λόγια θα μπορούσαμε να αναφέρουμε την εξόρυξη γνώσης από δεδομένα ως «τη σύνθετη διαδικασία εξαγωγής συγκεκριμένης, προηγούμενης άγνωστης και δυνητικά ωφέλιμης, γνώσης από δεδομένα» ([3]).

Στην επόμενη εικόνα παρουσιάζεται η διαδικασία εξόρυξης γνώσης από δεδομένα σε πέντε διακριτά στάδια.



Εικόνα 1: Διαδικασία εξόρυξης γνώσης από δεδομένα ([3W1])

Η διαδικασία αυτή στο σύνολό της είναι συνήθως επαναληπτική ενώ κάθε ένα από τα στάδια της εκτελείται πολλαπλές φορές. Στην επόμενη εικόνα παρουσιάζεται η «προσπάθεια» που απαιτείται συγκριτικά για κάθε ένα από τα πέντε στάδια. Ιδεατά, το στάδιο της προετοιμασίας των δεδομένων απαιτεί προσπάθεια το μέγεθος της οποίας θα πρέπει να μειώνεται σε αντίθεση με τα στάδια της εξόρυξης γνώσης και της ερμηνείας – αποτίμησης των αποτελεσμάτων στα οποία θα πρέπει να αυξάνεται.

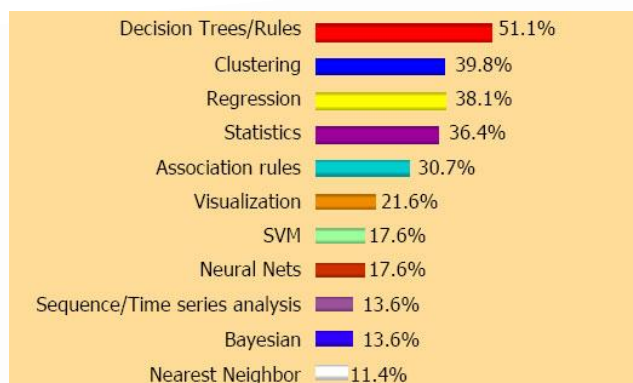


Εικόνα 2: Η απαιτούμενη προσπάθεια για κάθε ένα από τα στάδια της εξόρυξης γνώσης από δεδομένα ([3W1])

Στη συνέχεια θα παρουσιάσουμε μερικές τεχνικές εξόρυξης γνώσης από δεδομένα. Θα δούμε τις πιο διαδεδομένες και θα επικεντρωθούμε στην τεχνική της κατηγοριοποίησης (classification) την οποία και χρησιμοποιήσαμε στα πλαίσια της παρούσης εργασίας.

## 2.1 Τεχνικές εξόρυξης γνώσης από δεδομένα

Υπάρχουν πάρα πολλές τεχνικές που χρησιμοποιούνται για εργασίες εξόρυξης γνώσης. Πολλές από αυτές απαιτούν ειδικούς τύπους δομών δεδομένων και αλγοριθμικών προσεγγίσεων. Στη επόμενη εικόνα παρουσιάζονται οι πιο δημοφιλείς τεχνικές εξόρυξης γνώσης ([1], [3W2]).



**Εικόνα 3: Μερικές από τις πιο δημοφιλείς τεχνικές εξόρυξης γνώσης από δεδομένα ([1]; [3W1])**

Θα μπορούσαμε να διαχωρίσουμε την εξόρυξη γνώσης από δεδομένα σε δύο βασικές κατηγορίες. Την «πρόβλεψη τάσεων και συμπεριφορών» και την «αυτόματη ανακάλυψη, προηγουμένως άγνωστης, γνώσης και προτύπων» ([3W4]).

Στην πρώτη κατηγορία όπου κυριαρχούν τα «παραμετρικά μοντέλα», οι διάφορες τεχνικές βασίζονται στον συσχετισμό των δεδομένων εισόδου – εξόδου μέσω αλγεβρικών εξισώσεων βασισμένων σε παραμέτρους που δε προκαθορίζονται αλλά προσδιορίζονται μέσω συγκεκριμένων παραδειγμάτων εισόδου. Αυτό έχει ως αποτέλεσμα μοντέλα θεωρητικής χρήσιμα, άχρηστα ωστόσο στη πράξη μιας και είτε είναι ιδιαίτερος απλουστευμένα είτε απαιτούν εκ των προτέρων περισσότερη γνώση από αυτή που είναι διαθέσιμη στα δεδομένα εισόδου ([1]).

Οι τεχνικές της δεύτερης κατηγορίας είναι περισσότερο αποτελεσματικές για εφαρμογές εξόρυξης γνώσης. Αποτελούνται από «μη παραμετρικά μοντέλα» που καθοδηγούνται αποκλειστικά και μόνο από τα δεδομένα. Με άλλα λόγια, τεχνικές της κατηγορίας αυτής αν και απαιτούν μεγαλύτερο αριθμό δεδομένων εισόδου, δημιουργούν μοντέλα προσαρμοσμένα 100% στα δεδομένα εισόδου τα οποία και χρησιμοποιούν για την παραμετροποίησή τους. Στην κατηγορία αυτή ανήκουν τα νευρωνικά δίκτυα, τα δένδρα απόφασης αλλά και οι γενετικοί αλγόριθμοι ([1]).

Στις επόμενες παραγράφους, θα αναφερθούμε στη τεχνική της κατηγοριοποίησης (classification), την οποία και χρησιμοποιήσαμε στα πλαίσια της παρούσας εργασίας, στους κανόνες συσχέτισης (association rules) και θα ολοκληρώσουμε με τη τεχνική της συσταδοποίησης (clustering)..

### 2.1.1 Κατηγοριοποίηση (Classification)

Η κατηγοριοποίηση ορίζεται ως η τεχνική εκείνη μέσω της οποίας μπορούμε να κατατάξουμε ένα στοιχείο σε μια συγκεκριμένη ομάδα – κλάση επιλέγοντας μέσα από προκαθορισμένες ομάδες.

Πρόκειται ίσως για την πιο δημοφιλή τεχνική με πλήθος εφαρμογών στην αναγνώριση προτύπων και εικόνας, στην ιατρική διάγνωση, στην έγκριση τραπεζικών δανείων, στην ανίχνευση λαθών βιομηχανικών εφαρμογών κα ([1]).

Ο επιστημονικός ορισμός της κατηγοριοποίησης περιγράφεται στο [1] και είναι ο ακόλουθος:

*Δεδομένης μιας βάσης δεδομένων  $D = \{t_1, t_2, \dots, t_n\}$  πλειάδων και ένα σύνολο από κατηγορίες  $C = \{C_1, C_2, \dots, C_m\}$ , το πρόβλημα της κατηγοριοποίησης είναι να ορίσουμε μια απεικόνιση  $f: D$*

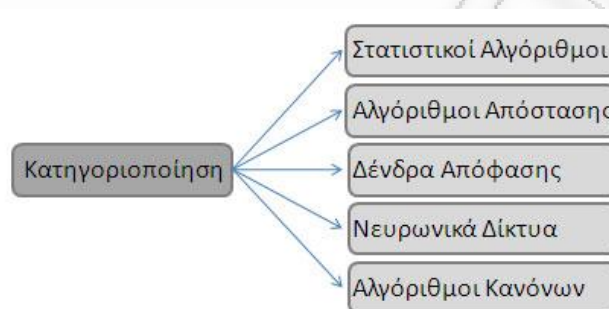
→  $C$  όπου κάθε πλειάδα  $t_i$  εκχωρείται σε μία κατηγορία. Μια κατηγορία ή κλάση,  $C_i$ , περιέχει ακριβώς τις πλειάδες που έχουν οριστεί σε αυτή.

Στην πράξη, μια διαδικασία κατηγοριοποίησης μπορεί να οριστεί ως η εκτέλεση δύο συγκεκριμένων βημάτων:

- Δημιουργία μοντέλου βασιζόμενου σε δεδομένα εκπαίδευσης.
- Εφαρμογή του μοντέλου στο σύνολο των δεδομένων.

Αν και βάσει του επιστημονικού ορισμού το δεύτερο από τα παραπάνω βήματα είναι αυτό της κατηγοριοποίησης, το πρώτο είναι το βήμα που απαιτεί την μεγαλύτερη προσπάθεια ([1]).

Οι αλγόριθμοι κατηγοριοποίησης μπορούν να διαχωριστούν στις ακόλουθες κατηγορίες ([1]):



Εικόνα 4: Κατηγορίες των αλγορίθμων κατηγοριοποίησης ([1])

Η επίδοση των αλγορίθμων κατηγοριοποίησης συνήθως υπολογίζεται με το ποσοστό των πλειάδων που τοποθετούνται στη σωστή κατηγορία. Συνηθισμένες τεχνικές αποτίμησης της επίδοσης ενός αλγορίθμου κατηγοριοποίησης είναι:

- Η μήτρα σύγχυσης (confusion matrix): Με δεδομένες  $m$  κατηγορίες, μια μήτρα σύγχυσης είναι ένα πίνακας  $m \times m$  όπου κάθε καταχώρηση  $c_{i,j}$  δείχνει τον αριθμό των πλειάδων από το  $D$  οι οποίες καταχωρήθηκαν στην κατηγορία  $C_j$  αλλά η πραγματική τους κατηγορία είναι η  $C_i$ . ([2])
- Η χαρακτηριστική καμπύλη λειτουργίας (operating characteristic): Πρόκειται για μια «καμπύλη» που δείχνει τη σχέση των αναληθώς θετικών και των αληθώς θετικών αποτελεσμάτων ([2]).

Ολοκληρώνοντας την αναφορά μας στην εξόρυξη γνώσης από δεδομένα μέσω της κατηγοριοποίησης, αξίζει να κάνουμε μια σύντομη αναφορά στη κατηγοριοποίηση μέσω δένδρων απόφασης τα οποία και χρησιμοποιήσαμε στα πλαίσια της παρούσης εργασίας.

Μέσω της τεχνικής αυτής, ένα δένδρο απόφασης μοντελοποιεί την διαδικασία κατηγοριοποίησης. Η πρώτη φάση της διαδικασίας, αναλώνεται στο «χτίσιμο» του δένδρου, ενώ η δεύτερη στη χρήση του δένδρου αυτού με σκοπό την κατηγοριοποίηση κάθε πλειάδας της βάσης δεδομένων. Ασφαλώς, από τα δύο παραπάνω βήματα (φάσεις) το πρώτο είναι αυτό που χρίζει μεγαλύτερης έρευνας και προσπάθειας.

Στο [1] παρατίθεται ο ορισμός ενός δένδρου απόφασης ή κατηγοριοποίησης – όπως συνήθίζεται να αποκαλείται στο χώρο της εξόρυξης γνώσης από δεδομένα:

*Δοθείσης μιας βάσης δεδομένων  $D = \{t_1, t_2, \dots, t_n\}$  με  $t_i = t_{i1}, \dots, t_{in}$ , του συνόλου γνωρισμάτων ενός σχήματος βάσης  $\{A_1, A_2, \dots, A_n\}$  και του συνόλου κατηγοριών  $C = \{C_1, C_2, \dots, C_m\}$ , δένδρο απόφασης ή κατηγοριοποίησης είναι ένα δένδρο που σχετίζεται με το  $D$  και έχει τις ακόλουθες ιδιότητες:*

- Κάθε εσωτερικός κόμβος παίρνει το όνομα του από ένα γνώρισμα  $A_i$ .
- Κάθε τόξο παίρνει το όνομά του από ένα κατηγορήμα το οποίο μπορεί να εφαρμοστεί στο γνώρισμα που συνδέεται με το πατέρα κόμβο.
- Κάθε φύλλο έχει ως όνομα μια κατηγορία  $C_j$ .

Οι παράγοντες που επηρεάζουν άμεσα την απόδοση ενός αλγορίθμου που δημιουργεί ένα δένδρο απόφασης είναι δύο ([1]):

- Το μέγεθος του συνόλου δεδομένων εκπαίδευσης.
- Η επιλογή των γνωρισμάτων διάσπασης.

Μερικά από τα κρίσιμα ζητήματα που αφορούν τους αλγορίθμους δημιουργίας δένδρων απόφασης ή κατηγοριοποίησης είναι τα ακόλουθα ([1]):

- **Η επιλογή των γνωρισμάτων διάσπασης:** Τα γνωρίσματα του συνόλου δεδομένων που θα επιλεγούν για τη δημιουργία του δένδρου είναι κρίσιμης σημασίας. Αναμφισβήτητα, κάποια γνωρίσματα είναι σημαντικότερα από κάποια άλλα. Η επιλογή των καταλληλότερων είναι πολλές φορές όχι μόνο θέμα εξέτασης των δεδομένων εκπαίδευσης αλλά και εμπειριστατωμένης άποψης ειδικών πάνω στη φύση των δεδομένων.
- **Η διάταξη των γνωρισμάτων διάσπασης:** Εκτός από τα ποια γνωρίσματα είναι τα πλέον κατάλληλα, κρίσιμη απόφαση είναι και η διάταξη των καταλληλότερων. Λάθος διάταξη των γνωρισμάτων διάσπασης μπορεί να σημαίνει επανέλεγχος γνωρισμάτων αρκετές φορές.
- **Οι διασπάσεις:** Γνωρίσματα με μικρό πεδίο τιμών οδηγούν σε φανερό αριθμό διασπάσεων, σε αντίθεση με γνωρίσματα συνεχών πεδίων όπου ο αριθμός των διασπάσεων κάθε άλλο παρά εύκολος μπορεί να θεωρηθεί.
- **Η δομή του δένδρου:** Η δομή του δένδρου απόφασης ασφαλώς και παίζει σημαντικό ρόλο στη δεύτερη από τις δύο φάσεις της κατηγοριοποίησης, αυτή της εφαρμογής του δένδρου πάνω στις πλειάδες της βάσης δεδομένων. Ισοζυγισμένα δένδρα λίγων επιπέδων ασφαλώς και βοηθούν στην αποδοτικότερη κατηγοριοποίηση.
- **Τα κριτήρια του τερματισμού:** Η δημιουργία του δένδρου απόφασης ολοκληρώνεται όταν τα δεδομένα κατηγοριοποιούνται με απόλυτη ακρίβεια. Αυτό ωστόσο κρύβει κινδύνους δημιουργίας μεγάλων δένδρων. Ο συμβιβασμός μεταξύ ακρίβειας της κατηγοριοποίησης και αποδοτικότητας του δένδρου είναι αναγκαίος.
- **Τα δεδομένα εκπαίδευσης:** Η δημιουργία του δένδρου βασίζεται αποκλειστικά στα δεδομένα εκπαίδευσης. Μικρό σύνολο τέτοιων δεδομένων ίσως οδηγήσει σε δένδρο μη κατάλληλο για το σύνολο δεδομένων που διαθέτουμε. Μεγάλο σύνολο δεδομένων εκπαίδευσης μπορεί να προκαλέσει υπερπροσαρμογή.
- **Το «κλάδεμα» του δένδρου:** Η ολοκλήρωση της δημιουργίας ενός δένδρου απόφασης πολλές φορές απαιτεί αφαίρεση περιττών συγκρίσεων ή και διαγραφή ολόκληρων υπο-δένδρων με στόχο την καλύτερη απόδοση της κατηγοριοποίησης. Η φάση του «κλαδέματος» έχει ως σκοπό τη βελτίωση της απόδοσης του δένδρου.

Τα πλεονεκτήματα από τη χρήση δένδρων κατηγοριοποίησης είναι πολλά ([1]):

- Τα δένδρα απόφασης είναι εύκολα στη χρήση και αποτελεσματικά, με κανόνες κατανοητούς και βατούς ως προς την ερμηνεία τους.
- Λειτουργούν πάρα πολύ καλά σε μεγάλες βάσεις δεδομένων λόγω του γεγονότος ότι το μέγεθος του δένδρου είναι ανεξάρτητο από το μέγεθος της βάσης.
- Δένδρα απόφασης μπορούν να κατασκευαστούν και για δεδομένα με πολλά γνωρίσματα.

Δε λείπουν ωστόσο και τα μειονεκτήματα ([1]):

- Δε χειρίζονται εύκολα δεδομένα τα γνωρίσματα των οποίων αποτελούνται από συνεχείς τιμές.
- Υπάρχει η πιθανότητα υπερπροσαρμογής ενός δένδρου στα σύνολα δεδομένων εκπαίδευσης.

Στη συνέχεια αναφέρουμε συνοπτικά κάποιες δημοφιλείς προσεγγίσεις για τη δημιουργία δένδρων απόφασης.

### 2.1.1.1 ID3

Η τεχνική που χρησιμοποιεί ο αλγόριθμος ID3 (ο οποίος εφευρέθηκε το 1979 από τον J. Ross Quinlan, [3W7]) βασίζεται στη θεωρία της πληροφορίας και προσπαθεί να ελαχιστοποιήσει τον αναμενόμενο αριθμό των συγκρίσεων ([1]). Η βασική του ιδέα είναι η χρήση ερωτήσεων οι απαντήσεις των οποίων παρέχουν τη μέγιστη δυνατή πληροφορία. Για το σκοπό αυτό ως γνωρίσματα διάσπασης επιλέγονται πρώτα τα γνωρίσματα εκείνα με το υψηλότερο κέρδος πληροφορίας.

Για τη μέτρηση της πληροφορίας χρησιμοποιείται η εντροπία, η οποία χρησιμοποιείται για να μετρήσει την ποσότητα της αβεβαιότητας, της έκπληξης, ή της τυχαιότητας σε ένα σύνολο δεδομένων.

Στόχος της κατηγοριοποίησης ενός δένδρου απόφασης είναι να διαχωρίσει επαναληπτικά το υπό εξέταση σύνολο δεδομένων σε υποσύνολα, έτσι ώστε όλα τα στοιχεία σε κάθε τελικό υποσύνολο να ανήκουν στην ίδια κατηγορία ([1])

Μερικά από τα πλεονεκτήματα και μειονεκτήματα του ID3 περιγράφονται στο [3W7]:

- Δημιουργεί κατανοητούς κανόνες.
- Χτίζει πολύ γρήγορα και μικρά δένδρα απόφασης.
- Με μικρά σύνολα δεδομένων εκπαίδευσης υπάρχει μεγάλος κίνδυνος δημιουργίας υπερπροσαρμοσμένου δένδρου απόφασης.
- Δεν ενδείκνυται για δεδομένα με γνωρίσματα συνεχών τιμών μιας και απαιτείται η δημιουργία πολλών δένδρων ούτως ώστε να αποσαφηνιστεί πώς θα «διασπαστούν» οι τιμές των γνωρισμάτων.

### 2.1.1.2 C4.5 και C5.0

Ο αλγόριθμος C4.5 αποτελεί βελτίωση του ID3. Στο ([1]) αναφέρονται οι τρόποι με τους οποίους η βελτίωση αυτή επιτυγχάνεται:

- Τα ελλιπή δεδομένα απλά αγνοούνται. Αυτό σημαίνει ότι το κλάσμα του κέρδους υπολογίζεται εξετάζοντας μόνο εκείνες τις εγγραφές που έχουν κάποια τιμή για ένα γνώρισμα.
- Τα συνεχή δεδομένα χωρίζονται σε διαστήματα με βάση τις τιμές των γνωρισμάτων για εκείνα τα στοιχεία τα οποία ανήκουν στο δείγμα εκπαίδευσης.
- Χρήση «κλαδέματος» για βελτίωση της απόδοσης του δένδρου. Αυτό γίνεται ακολουθώντας δύο στρατηγικές: α) την αντικατάσταση ενός ολόκληρου υποδένδρου από ένα φύλλο εάν και εφόσον η αντικατάσταση αυτή έχει σαν αποτέλεσμα ένα σφάλμα κοντά σε αυτό του αρχικού δένδρου, β) την αντικατάσταση ενός υποδένδρου με το περισσότερο χρησιμοποιούμενο υποδένδρο του.

Ο αλγόριθμος C5.0 είναι μια εμπορική έκδοση του C4.5 που χρησιμοποιείται εκτενέστατα σε πολλά πακέτα εξόρυξης γνώσης από δεδομένα, όπως για παράδειγμα το Clementine και το RuleQuest ([1]). Αν και οι ακριβείς αλγόριθμοι που χρησιμοποιεί δεν έχουν αποκαλυφθεί από τους δημιουργούς τους, η βελτίωση στην ακρίβειά του βασίζεται στην ενίσχυση (boosting), μια προσέγγιση που συνδυάζει διάφορους μηχανισμούς κατηγοριοποίησης. Τα αποτελέσματα δείχνουν ότι ο C5.0 βελτιώνει τη χρήση της μνήμης κατά 90%, τρέχει μεταξύ 5.7 και 240 φορές πιο γρήγορα από το C4.5 παράγοντας παράλληλα ακριβέστερους κανόνες ([1]).

### 2.1.1.3 CART

Η τεχνική CART (Classification And Regression Trees) δημιουργεί δυαδικά δένδρα απόφασης. Παρ' όλο που, και σε αυτήν την περίπτωση, μέτρο για την επιλογή γνώρισμάτων διάσπασης είναι η εντροπία, η βασική διαφορά με τον ID3 έγκειται στο ότι ενώ ο τελευταίος δημιουργεί παιδιά κόμβους για κάθε υποκατηγορία, ο CART δημιουργεί κάθε φορά μονάχα δύο παιδιά ([1]).

Πρόκειται για μια αρκετά «δυνατή» τεχνική που δουλεύει το ίδιο καλά με ημιτελή σύνολα δεδομένων διαφόρων τύπων, παράγοντας κανόνες εύκολα κατανοητούς από τον άνθρωπο ([3W10]).

Κάθε βήμα του αλγορίθμου απαιτεί μια πλήρη αναζήτηση στο σύνολο δεδομένων, με σκοπό την εύρεση της καλύτερης δυνατής διάσπασης, ολοκληρώνοντας την ανάπτυξη του δένδρου όταν καμία περαιτέρω διάσπαση δεν μπορεί να βελτιώσει την απόδοση ([1]).

### 2.1.2 Κανόνες συσχέτισης (Association Rules)

Οι κανόνες συσχέτισης συχνά χρησιμοποιούνται από καταστήματα λιανικής πώλησης (πχ supermarkets) για να βοηθούν στο μάρκετινγκ, στη διαφήμιση, ακόμα και στην χωροθέτηση των προϊόντων πάνω σε ράφια.

Χρησιμοποιούνται για να καταδείξουν μη έμφυτες συσχετίσεις ανάμεσα στα δεδομένα χωρίς να αντιπροσωπεύουν κανένα είδος αιτιότητας ή συσχέτισης ([1]).

Σύμφωνα με το [3W5], το πρόβλημα εξόρυξης κανόνων συσχέτισης από δεδομένα ορίζεται ως εξής:

*Δεδομένου ενός συνόλου στοιχείων  $I = \{i_1, i_2, \dots, i_n\}$  και μίας βάσης δεδομένων  $D = \{t_1, t_2, \dots, t_n\}$  που περιέχει δοσοληψίες (transactions) κάθε μία εκ των οποίων περιέχει ένα υποσύνολο του  $I$ , ως κανόνας συσχέτισης ορίζουμε κάθε σχέση της μορφής  $X \rightarrow Y$  όπου  $X, Y \subseteq I$  και  $X$  ένωση  $Y = \emptyset$ . Το  $X$  αποκαλείται «προηγούμενο» (antecedent) του κανόνα ενώ το  $Y$  «επακόλουθο» (consequent) αυτού.*

Η σπουδαιότητα ενός κανόνα συσχέτισης, όπως εξάγεται βάσει του ορισμού, υπολογίζεται από δύο βασικά χαρακτηριστικά ([2]):

- Την υποστήριξη (support), που ορίζεται για έναν κανόνα συσχέτισης  $X \rightarrow Y$  ως το ποσοστό των συναλλαγών στη βάση δεδομένων που περιέχουν το « $X$  ένωση  $Y$ ».
- Την εμπιστοσύνη (confidence), που ορίζεται για έναν κανόνα συσχέτισης  $X \rightarrow Y$  ως ο λόγος του αριθμού των συναλλαγών που περιέχουν το « $X$  ένωση  $Y$ » προς τον αριθμό των συναλλαγών που περιέχουν το  $X$ .

Μερικοί από τους αλγορίθμους που χρησιμοποιούνται συχνότερα για την αποτελεσματική ανακάλυψη στοιχειοσυνόλων είναι οι ακόλουθοι ([1]):

- *Αλγόριθμος Apriori*. Βασίζεται στην ιδιότητα πως κάθε υποσύνολο ενός συχνού συνόλου είναι και αυτό συχνό. Βασική ιδέα του αλγορίθμου είναι η δημιουργία υποψήφια συνόλων συγκεκριμένου μεγέθους και η εν συνεχεία σάρωση της βάσης δεδομένων, με σκοπό τον χαρακτηρισμό του συνόλου ως συχνού ή μη. Κατά την  $i$ -οστή σάρωση της βάσης καταμετρούνται υποψήφια στοιχειοσύνολα μεγέθους  $i$ . Στην επόμενη  $(i+1)$ -οστή σάρωση, χρησιμοποιούνται μονάχα τα συχνά στοιχειοσύνολα μεγέθους  $i$ .
- *Αλγόριθμος Δειγματοληψίας*. Έρχεται να «βοηθήσει» τον αλγόριθμο Apriori (ή και οποιονδήποτε άλλο) μειώνοντας αισθητά το μέγεθος μεγάλων βάσεων δεδομένων εκτελώντας διαδικασίες δειγματοληψίας. Με τον τρόπο αυτό ο αριθμός των απαιτούμενων σαρώσεων της βάσης δεδομένων μειώνεται σε μία (καλύτερη περίπτωση) ή δύο (χειρότερη περίπτωση).
- *Αλγόριθμος Διαμέρισης*. Και σε αυτήν την περίπτωση σκοπός του αλγορίθμου είναι η αισθητή μείωση των σαρώσεων, της βάσης δεδομένων που εκτελεί ο αλγόριθμος Apriori (ή οποιοσδήποτε άλλος). Ο συγκεκριμένος αλγόριθμος μειώνει τον αριθμό των σαρώσεων της βάσης σε δύο και διαιρεί τη τελευταία σε τμήματα ικανά να χωρούν στην κύρια μνήμη του συστήματος. Η διαμέριση αυτή, μπορεί και βελτιώνει την απόδοση της εύρεσης συχνών στοιχειοσυνόλων, εκμεταλλευόμενη κανόνες που πηγάζουν από αυτή καθ' αυτή τη διαμέριση: χρήση παράλληλων ή/και κατανεμημένων αλγορίθμων, μέγεθος διαμερίσεων ικανό να χωρέσει στη κύρια μνήμη κλπ. Ο βασικός κανόνας του αλγορίθμου διαμέρισης υποστηρίζει ότι ένα στοιχειοσύνολο για να είναι συχνό, θα πρέπει να είναι συχνό σε τουλάχιστον μια από τις διαμερίσεις της βάσης δεδομένων.

### 2.1.3 Συσταδοποίηση (Clustering)

Η συσταδοποίηση (clustering) αποτελεί παρόμοια τεχνική με αυτή της κατηγοριοποίησης αφού σκοπός και των δύο τεχνικών είναι η οργάνωση των δεδομένων σε ομάδες. Στην συσταδοποίηση, ωστόσο, η οποία επιτυγχάνεται με την εύρεση ομοιοτήτων μεταξύ των διαφόρων χαρακτηριστικών των δεδομένων, οι ομάδες (clusters) δεν είναι προκαθορισμένες.

Σύμφωνα με τον επιστημονικό ορισμό που περιγράφεται στο [1]:

*Δοθείσης μιας βάσης δεδομένων  $D = \{t_1, t_2, \dots, t_n\}$  που αποτελείται από πλειάδες και μιας ακέραιης τιμής  $k$ , το πρόβλημα της συσταδοποίησης είναι να οριστεί μια αντιστοίχιση  $f: D \rightarrow \{1, 2, \dots, k\}$  όπου κάθε  $t_i$  ανατίθεται σε μια πλειάδα  $K_j$ ,  $1 \leq j \leq k$ . Μια συστάδα,  $K_j$ , περιέχει ακριβώς εκείνες τις πλειάδες που της ανατέθηκαν.*

Από τον παραπάνω ορισμό προκύπτει άμεσα και η βασική διαφορά μεταξύ κατηγοριοποίησης και συσταδοποίησης: στην κατηγοριοποίηση ο αριθμός και η ουσία των συστάδων αποτελεί πληροφορία εκ των προτέρων γνωστή. Εξαιτίας αυτού, στη συσταδοποίηση εφαρμόζεται πάντα μη εποπτευόμενη μάθηση, εν αντιθέσει με την κατηγοριοποίηση όπου λόγω της πρότερης γνώσης των κλάσεων κάνουμε χρήση της εποπτευόμενης μάθησης ([2]).

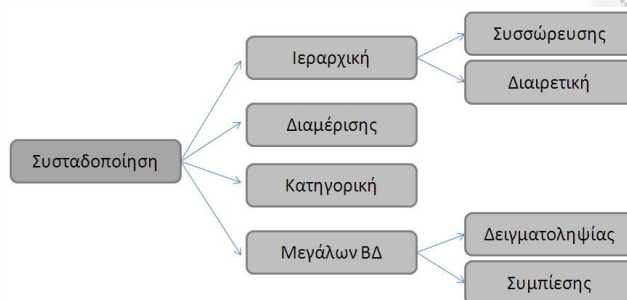
Μερικά βασικά ζητήματα που προκύπτουν στην συσταδοποίηση είναι τα ακόλουθα ([1]):

- Ο χειρισμός των ακραίων σημείων: πρόκειται στη ουσία για δεδομένα που στην πράξη δεν ανήκουν σε καμία συστάδα. Μπορούν από μόνα τους να θεωρηθούν ως ξεχωριστές συστάδες κάτι που καθιστά κάθε προσπάθεια συσταδοποίησης φτωχή.
- Τα δυναμικά δεδομένα: που μπορεί να υπάρχουν σε βάσεις δεδομένων και τα οποία καθιστούν και τις ίδιες τις συστάδες δυναμικά μεταβαλλόμενες στο χρόνο.



- Το είδος των δεδομένων που χρησιμοποιούνται: και για το οποίο δεν είμαστε σε θέση να έχουμε καμία περαιτέρω πληροφορία όσον αφορά τα γνωρίσματα των ομάδων.
- Η μοναδικότητα της λύσης του προβλήματος: που πολλές φορές δε γίνεται να επιτευχθεί σε πολλά από τα προβλήματα συσταδοποίησης, μιας και το ακριβές πλήθος των ομάδων που απαιτούνται δεν είναι και τόσο εύκολο να προσδιοριστεί.

Οι αλγόριθμοι συσταδοποίησης μπορούν να ταξινομηθούν στις ακόλουθες κατηγορίες ([1]):



Εικόνα 5: Κατηγορίες αλγορίθμων συσταδοποίησης ([1])

Έτσι, ανάλογα με το πλήθος των επιπέδων συστάδων που παράγονται, διακρίνουμε τους ιεραρχικούς αλγορίθμους κατηγοριοποίησης (με πολλά επίπεδα εμφωλιασμένων συστάδων) και τους αλγορίθμους διαμέρισης (ένα επίπεδο  $k$  συστάδων). Υπάρχουν αλγόριθμοι που δημιουργούν επικαλύπτοντες ή μη συστάδες, αλγόριθμοι που χειρίζονται σειριακά ή ταυτόχρονα τα στοιχεία εισόδου και αλγόριθμοι που αφορούν μικρές ή μεγάλες βάσεις δεδομένων ([2]).

Αξίζει, τέλος, να αναφέρουμε ότι σε πολλούς από τους αλγορίθμους αυτούς σημαντικό ρόλο παίζει η απόσταση μεταξύ των συστάδων, η οποία θα πρέπει να ορίζεται καταλλήλως. Αντιπροσωπευτική λίστα μεθόδων υπολογισμού αποστάσεων είναι αυτή που περιγράφεται στο [1]:

- *Απόσταση απλού συνδέσμου*: η μικρότερη απόσταση μεταξύ δύο στοιχείων των δύο συστάδων.
- *Απόσταση πλήρους συνδέσμου*: η μεγαλύτερη απόσταση μεταξύ δύο στοιχείων των δύο συστάδων.
- *Μέση απόσταση*: η μέση απόσταση μεταξύ των στοιχείων των δύο συστάδων.
- *Απόσταση κέντρων βάρους*: η απόσταση μεταξύ των κέντρων βάρους των δύο συστάδων.

## 2.2 Εφαρμογές Εξόρυξης Γνώσης

Τα παραδείγματα εξόρυξης γνώσης από δεδομένα ποικίλουν ανάλογα με τον τομέα στον οποίο εφαρμόζονται. Στην σημερινή εποχή όπου τα δεδομένα υπάρχουν σχεδόν παντού και τις περισσότερες φορές και σε ηλεκτρονική μορφή, η σωστή ανάλυσή τους οδηγεί πάντα στην ανάδειξη και οργάνωση της πληροφορίας η γνώση της οποίας είναι ο σημαντικότερος παράγοντας για τη χάραξη στρατηγικής και την ορθολογική λήψη αποφάσεων.

Ο χρηματοοικονομικός τομέας, ο τομέας των τηλεπικοινωνιών, της υγείας και της εκπαίδευσης, ο δημόσιος τομέας καθώς και αυτός της βιομηχανίας και της έρευνας αποτελούν ίσως το μεγαλύτερο δείγμα εφαρμογών των τεχνολογιών εξόρυξης γνώσης από δεδομένα.

Στο χρηματοοικονομικό τομέα η παρουσία αρκετών τραπεζικών και ασφαλιστικών ιδρυμάτων σε συνδυασμό με το νέο οικονομικό κλίμα έχει αυξήσει ιδιαίτερα τον ανταγωνισμό. Πολλά από

τα επιχειρηματικά ζητήματα του κλάδου, όπως είναι η προσέλκυση νέων επικερδών πελατών, η πώληση επιπρόσθετων προϊόντων/υπηρεσιών, η διατήρηση πελατών, ο προσδιορισμός οικονομικού δόλου και η ανάλυση του πιστωτικού κινδύνου ενδεχόμενων πελατών, μπορούν να αντιμετωπιστούν με την εξόρυξη γνώσης.

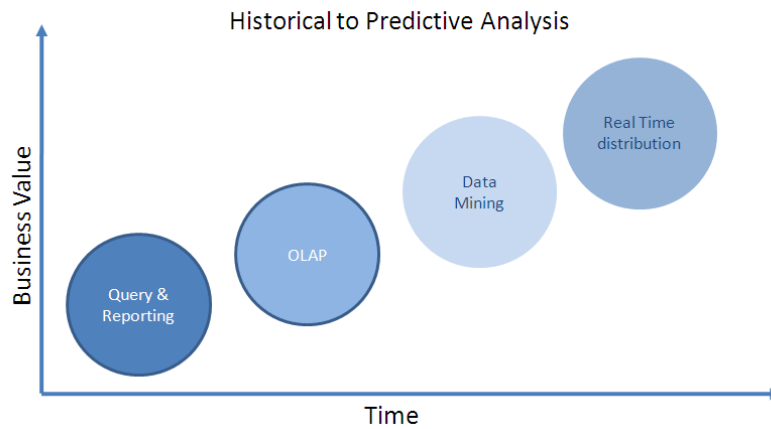
Στον κλάδο των τηλεπικοινωνιών το διαρκώς μεταβαλλόμενο επιχειρηματικό περιβάλλον με τον αστείρευτο ανταγωνισμό, αναγκάζουν τις εταιρείες τηλεπικοινωνιών να αναζητούν συνεχώς τρόπους ενίσχυσης της θέσης τους. Έχοντας γίνει πλήρως αντιληπτό από τη πλευρά τους το συγκριτικό πλεονέκτημα που τους προσφέρει η εξόρυξη γνώσης από δεδομένα, που ούτως ή άλλως διαθέτουν, εφαρμόζουν, μεταξύ άλλων, τεχνικές έγκαιρης πρόβλεψης διακοπής υπηρεσιών από πελάτες, κατηγοριοποίηση των απαιτήσεων τους, ομαδοποίηση των συνθηγιών τους, με τελικό σκοπό την συγκράτηση των πελατών τους και την προσέλκυση νέων.

Οι επαγγελματίες στο χώρο της υγείας πάντα αντιμετωπίζουν την ανάγκη να συλλέγουν, αποθηκεύουν και να αναλύουν μεγάλες ποσότητες δεδομένων που μπορεί να περιλαμβάνουν καρτέλες ασθενών, δοκιμές νέων φαρμάκων, εξάρσεις συγκεκριμένων ασθενειών κ.α.

Ο τομέας του λιανικού εμπορίου είναι ένας άλλος κλάδος ιδιαίτερα ανταγωνιστικός όπου οι εφαρμογές εξόρυξης γνώσης βρίσκουν μεγάλη ανταπόκριση. Οι συνεχείς αλλαγές των καταναλωτικών προτιμήσεων και οι τεράστιοι όγκοι δεδομένων πωλήσεων κρύβουν πολύτιμα στοιχεία εκ των οποίων ελάχιστα μπορούν να αξιοποιηθούν από τα συμβατικά συστήματα ανάλυσης πληροφορίας. Αντιθέτως, οι εφαρμογές εξόρυξης γνώσης δίνουν μια νέα διάσταση στην παλαιότερη και βασικότερη επιχειρηματική διαδικασία: αναλύοντας ό,τι έγινε στο παρελθόν και κατανοώντας τα αποτελέσματα, μπορούμε να γίνουμε αποτελεσματικότεροι στο μέλλον. Επιπροσθέτως, οι εφαρμογές εξόρυξης γνώσης κάνουν εφικτή μια προσωποποιημένη σχέση με κάθε ένα πελάτη ξεχωριστά, κάτι που εξασφαλίζει την διαχρονική σχέση και την μεγιστοποίηση του κέρδους ανά πελάτη.

Όλα τα προαναφερθέντα καθιστούν την εξόρυξη γνώσης απαραίτητη σε πολλούς τομείς της σύγχρονης κοινωνίας. Η ραγδαία αύξηση του όγκου δεδομένων έχει καταστήσει σαφές ότι παραδοσιακές τεχνικές στατιστικής ανάλυσης του παρελθόντος, δεν μπορούν πλέον να βοηθήσουν στην ανάλυση και οργάνωση της πληροφορίας. Πολύ περισσότερο, δε μπορούν να φέρουν στην επιφάνεια γνώση που τα δεδομένα περιέχουν καλά κρυμμένη και η οποία απαιτεί εφαρμογή ειδικών τεχνικών για να αποκαλυφθεί.

Στην επόμενη εικόνα φαίνεται χαρακτηριστικά η αξία της εξόρυξης γνώσης σε σχέση με παλαιότερες τεχνικές ανάλυσης. Με τη πάροδο των χρόνων, ο όγκος των δεδομένων και οι σύγχρονες ανάγκες (ανθρώπων, επιχειρήσεων, κοινωνίας) μεγάλωσαν, η ιστορική ανάλυση των δεδομένων έδωσε τη θέση της σε τεχνικές πρόβλεψης μελλοντικών γεγονότων βασιζόμενες στη γνώση που προσφέρουν τα υπάρχοντα δεδομένα.



Εικόνα 6: Η πορεία της ανάλυσης δεδομένων στον χρόνο. ([3W12])

Και το ερώτημα που τίθεται τώρα είναι απλό: πώς πραγματοποιείται η συλλογή των δεδομένων πάνω στα οποία μπορούμε να εφαρμόσουμε τις τεχνικές εξόρυξης γνώσης;

### 3. Customer Relationship Management

Τα συστήματα διαχείρισης των σχέσεων μιας εταιρείας με τους πελάτες της (Customer Relationship Management Systems, CRMS) άρχισαν να πρωτοεμφανίζονται στην αγορά στις αρχές της δεκαετίας του '90 και κατάφεραν να αποκτήσουν ένα σημαντικό ρόλο στη ζωή των εταιρειών μόλις στα τέλη της δεκαετίας αυτής.

Για πολλά χρόνια, οι επιχειρήσεις είχαν επικεντρώσει τις προσπάθειες τους στην περικοπή των λειτουργικών εξόδων και στην αύξηση της απόδοσης στα πλαίσια του ίδιου του οργανισμού. Προσπάθησαν κατά καιρούς να απλοποιήσουν και να αυξήσουν την αποδοτικότητα των εσωτερικών διαδικασιών, αυτοματοποιώντας κάποια από τα στοιχεία των καλούμενων "back-office" λειτουργιών όπως είναι το κατασκευαστικό κομμάτι, η διαχείριση αποθήκης και τα οικονομικά. Τα ERP συστήματα αποτέλεσαν τον αρωγό στην προσπάθεια αυτή. Αντιθέτως, οι επενδύσεις σε συστήματα που σχετίζονταν με τις πελατοκεντρικές δραστηριότητες (συχνά καλούμενων ως "front-office") όπως το marketing και οι πωλήσεις, συχνά παραμελούνταν διακριτικά προς όφελος των προηγούμενων.

Όλες αυτές οι προσπάθειες καρποφόρησαν, με αποτέλεσμα, οι εταιρείες να μπορούν να «παράγουν» προϊόντα/υπηρεσίες με το ελάχιστο δυνατό κόστος. Ολοένα και περισσότερες επιχειρήσεις βελτιώναν την αποδοτικότητά τους στους συγκεκριμένους τομείς, με αποτέλεσμα να γίνεται όλο και πιο δύσκολη η διαφοροποίηση από τους ανταγωνιστές τους, καθιστώντας δύσκολη την απόκτηση συγκριτικού πλεονεκτήματος έναντι της υπόλοιπης αγοράς.

Στο σημείο αυτό, στην αρχή λίγες και στη συνέχεια όλο και περισσότερες, οι εταιρείες κατάλαβαν πως θα έπρεπε να επενδύσουν ανάλογα στις πωλήσεις, το marketing και στις σχέσεις τους με τους πελάτες. Απαραίτητη προϋπόθεση για αυτό ήταν η κατανόηση των αναγκών των πελατών και η προσαρμογή των προϊόντων και υπηρεσιών τους στην όσο το δυνατόν καλύτερη ικανοποίησή τους. Οι επιχειρήσεις συνειδητοποίησαν πως το δύσκολο είναι να κερδίσεις ένα νέο ή να διατηρήσεις ικανοποιημένο έναν υπάρχον πελάτη σου παρέχοντας του προϊόντα και υπηρεσίες όσο το δυνατόν πιο καλά προσαρμοσμένα στις ανάγκες του.

Στο [5] παρουσιάζεται ένας ενδιαφέρον «ορισμός» για τα CRM συστήματα:

*Τα CRM συστήματα είναι μια επιχειρησιακή προσέγγιση με την οποία οι άνθρωποι, οι διαδικασίες και η τεχνολογία συνεργάζονται (ολοκληρώνονται) με σκοπό την μεγιστοποίηση των σχέσεων της επιχείρησης με τους πελάτες της.*

Η συστηματική χρήση εφαρμογών CRM παρέχει ένα σύνολο οφελών στις επιχειρήσεις ([5]) που επιθυμούν αυτοματοποίηση των πωλήσεών τους, βελτίωση του marketing τους και παροχή ολοκληρωμένων υπηρεσιών προς τους πελάτες τους:

- Ένα μεγάλο πλήθος πληροφοριών που αφορά στους πελάτες είναι πλέον διαθέσιμο στις επιχειρήσεις. Τα στοιχεία ενός πελάτη, οι ανάγκες και οι συνήθειες του είναι μερικά από τα στοιχεία που συλλέγει μια εφαρμογή CRM.
- Η παραγωγικότητα βελτιώνεται. Με πολλές πληροφορίες διαθέσιμες για κάθε πελάτη, μια επιχείρηση είναι σε θέση να εκτελέσει στοχευμένο marketing, να απευθυνθεί σε συγκεκριμένους ανθρώπους για συγκεκριμένα προϊόντα και να μειώσει τον κύκλο ζωής μιας πώλησης στο ελάχιστο.
- Τα τμήματα πωλήσεων μπορούν πλέον να «ξοδέψουν» περισσότερο χρόνο δίπλα στους πελάτες τους. Να καταγράψουν την ικανοποίησή τους, να προβλέψουν τα προβλήματά τους και τελικά να αντιδράσουν πιο άμεσα στις ανάγκες τους.

Με τη γλώσσα των αριθμών οι εφαρμογές CRM είναι το νούμερο ένα λογισμικό παγκοσμίως σε άδειες χρήσης ενώ το 2006 τα χρήματα που δαπανήθηκαν στην αγορά του CRM ανέρχονταν σε 31 δισεκατομμύρια δολάρια ([5]).

Γενικά μιλώντας θα μπορούσαμε να πούμε ότι μια CRM εφαρμογή διαθέτει 11 βασικά συστατικά (components) τα οποία παρέχονται ως ξεχωριστές αυτοτελείς λογισμικές μονάδες (software modules) από τους διάφορους προμηθευτές λογισμικού (software vendors) ([8]):

- *Time management*: το συστατικό αυτό αφορά την οργάνωση του χρόνου των χρηστών προσφέροντας ολοκλήρωση (integration) με διάσημες σουίτες γραφείου όπως είναι το Microsoft Outlook και το Lotus Notes.
- *Sales/Sales management*: το συστατικό αυτό αφορά την διαχείριση των πελατών και των πληροφοριών που σχετίζονται με αυτούς όπως είναι οι παραγγελίες και οι δραστηριότητες. Οι χρήστες μπορούν πολύ εύκολα να παραμετροποιούν προϊόντα και υπηρεσίες ανάλογα με τις ανάγκες των πελατών τους.
- *Telemarketing and telesales*: το συστατικό αυτό αφορά κλασσικές διαδικασίες τηλεφωνικού κέντρου όπως είναι η καταγραφή κλήσεων, οι τηλεπωλήσεις και η παραγγελιοληψία.
- *Customer contact center*: το συστατικό αυτό αφορά την διαχείριση των επικοινωνιών του πελάτη που σχετίζονται με αιτήματα εξυπηρέτησης (service requests), συμβόλαια και εγγυήσεις κλπ.
- *eMarketing*: το συστατικό αυτό αφορά το τμήμα marketing μιας επιχείρησης και όλες τις διαδικασίες που αυτό μπορεί να εκτελέσει με τρόπο αποδοτικό.
- *Business analytics*: το συστατικό αυτό αφορά τη δυνατότητα δημιουργίας εκτεταμένων και εύχρηστων αναφορών παρέχοντας ολοκλήρωση (integration) με διάσημες εφαρμογές reporting όπως είναι τα Crystal Reports και το Actuate.
- *Field service support*: το συστατικό αυτό αφορά στη διαχείριση της τεχνικής υποστήριξης που είναι υποχρεωμένη μια επιχείρηση να παρέχει για προϊόντα ή υπηρεσίες της.
- *eBusiness*: το συστατικό αυτό αφορά το κομμάτι και τις διαδικασίες μιας CRM εφαρμογής όπως προβάλλονται μέσα από το διαδίκτυο.
- *Supply chain management*: το συστατικό αυτό αφορά τις διαδικασίες διαχείρισης αποθηκών και συνήθως παρέχει ολοκλήρωση (integration) με μεγάλες εφαρμογές του είδους όπως είναι τα Teradata, JDA Software.
- *Multimodal access*: το συστατικό αυτό αφορά τη δυνατότητα που προσφέρει μια επιχείρηση στους πελάτες της για επικοινωνία με ποικίλους τρόπους όπως είναι το τηλέφωνο, το fax, το email και το διαδίκτυο.
- *Data sharing tools*: το συστατικό αφορά εργαλεία με τα οποία η πληροφορία που διαθέτει μια CRM εφαρμογή μπορεί να διαμοιραστεί σε διάφορα κανάλια πληροφόρησης. Για το σκοπό αυτό, υπάρχουν CRM εφαρμογές που διαθέτουν ενσωματωμένα μικρά ERP συστήματα.

### 3.1 Δημοφιλείς CRM εφαρμογές

Στο [5] παρουσιάζεται μια ενδιαφέρουσα λίστα με πολλές εφαρμογές CRM που κυκλοφορούν παγκοσμίως:

- Amdocs CRM v.6 – Amdocs Limited
- C2 CRM v.8.0 – Clear Technologies, Inc.
- Infor CRM – Infor
- mySAP CRM – SAP AG
- PeopleSoft CRM – Oracle's PeopleSoft

- Siebel 8.1 – Oracle’s Siebel
- Microsoft CRM 3.0 – Microsoft Corp.
- NetSuite CRM v.11.0 – NetSuite, Inc.

### 3.2 CRM εφαρμογές και σύγχρονες τάσεις

Τα τελευταία χρόνια έχουν σημειωθεί σημαντικές αλλαγές όσον αφορά τις τεχνολογίες και αρχιτεκτονικές που χρησιμοποιούν οι CRM εφαρμογές. Στο [8] περιγράφονται εκτενώς τεχνολογικές τάσεις που επηρέασαν και συνεχίζουν να επηρεάζουν τα συστήματα CRM.

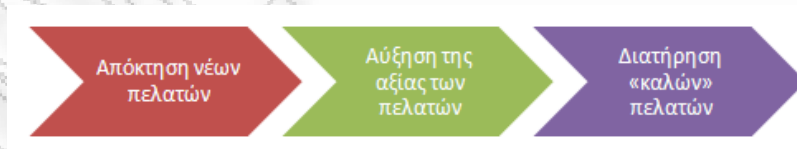
- Η αρχιτεκτονική πελάτη/εξυπηρετητή (client/server architecture) θεωρείται πλέον ξεπερασμένη για τα περισσότερα συστήματα CRM τα οποία πλέον κινούνται στην σύγχρονη τάση των διαδικτυακών εφαρμογών (web based applications).
- Ακόμα και οι πρόσφατες αρχιτεκτονικές τριών επιπέδων (3-tier architectures) έχουν δώσει την θέση τους στις πολύ-επίπεδες αρχιτεκτονικές (N-tier architectures).
- Η τεχνολογία XML και τα Web Services έχουν βασικό πλέον ρόλο στις περισσότερες εφαρμογές CRM. Αυτό τις καθιστά πολύ ευέλικτες στην ολοκλήρωσή τους (integration) με άλλα συστήματα.
- Οι περισσότεροι προμηθευτές CRM ενσωματώνουν στα συστήματά τους λογισμικά επιχειρηματικής ευφυΐας (business intelligence) ή προσφέρουν έτοιμες λύσεις ολοκλήρωσης με αποθήκες δεδομένων (data warehouses).
- Οι αρχιτεκτονικές προσανατολισμένες στις υπηρεσίες (service oriented architectures) ακολουθούνται σε πολλά συστήματα CRM.
- Οι διαδικτυακές πύλες (portals) αποτελούν αναπόσπαστο κομμάτι πολλών συστημάτων CRM.

Όλα τα παραπάνω αποδεικνύουν με τον πλέον κατηγορηματικό τρόπο ότι τα συστήματα CRM είναι «ζωντανές εφαρμογές» στον κόσμο της τεχνολογίας. Εξελίσσονται, βελτιώνονται συνεχώς, ενσωματώνουν νέες και απορρίπτουν παλιές τεχνολογίες.

### 3.3 CRM και εξόρυξη γνώσης

Με την βοήθεια των συστημάτων CRM, οι επιχειρήσεις είναι πλέον σε θέση να διαχειρίζονται με τρόπο αποτελεσματικό και αποδοτικό τις αλληλεπιδράσεις με τους πελάτες τους σε όλα τα στάδια του κύκλου ζωής των τελευταίων μέσα σε μία επιχείρηση (customer life cycle).

Αυτός ο κύκλος μπορεί να διαιρεθεί σε τρία στάδια ([3W14]):



Εικόνα 7: Τα στάδια του κύκλου ζωής ενός πελάτη σε μια επιχείρηση

Στο [3W14] περιγράφονται τρία χαρακτηριστικά παραδείγματα όπου ο συνδυασμός της ύπαρξης ενός συστήματος CRM και η εφαρμογή τεχνολογιών εξόρυξης γνώσης βοήθησαν τρεις διαφορετικές επιχειρήσεις να αυξήσουν την αποτελεσματικότητα και αποδοτικότητά τους σε κάθε ένα από αυτά τα στάδια.

### 3.3.1 Απόκτηση νέων πελατών μέσω εξόρυξης γνώσης

Ο πρώτος και βασικός στόχος όλων των συστημάτων CRM είναι η απόκτηση νέων πελατών. Η αναγνώριση δυνητικών πελατών και η μετατροπή τους σε ενεργούς πελάτες είναι η μεγαλύτερη πρόκληση για κάθε επιχείρηση.

Η “Big Bank and Credit Card Company” (BB&CC), μια εταιρεία που «πουλάει» πιστωτικές κάρτες, δαπανά ετησίως 1,000,000\$ για να εκτελεί τις 25 εκστρατείες της μέσω ταχυδρομείου δίνοντας την δυνατότητα σε 1,000,000 ανθρώπους να αποκτήσουν την δική τους πιστωτική κάρτα. Από τους ανθρώπους που απαντούν θετικά στις εκστρατείες αυτές (ένα ποσοστό που κυμαίνεται γύρω στο 6%) μόλις το 16% αποτελούν χαμηλού ρίσκου πελάτες για την εταιρεία με αποτέλεσμα μόλις 10,000 (δηλαδή το 1% του συνολικού αριθμού δυνητικών πελατών) να αποκτούν τελικά μια πιστωτική κάρτα από την εταιρεία. Σε βάθος χρόνου διείσδυσης, οι 10,000 αυτοί πελάτες «επιστρέφουν» στην εταιρεία 125\$ ο καθένας αυξάνοντας το όφελος της BB&CC κατά 250,000\$.

Με την βοήθεια των τεχνολογιών εξόρυξης γνώσης η BB&CC κατόρθωσε να αυξήσει τα κέρδη της από τις συγκεκριμένες εκστρατείες με τον ακόλουθο τρόπο.

Χρησιμοποίησε αρχικά μια λίστα από 50,000 δυνητικούς πελάτες τους οποίους και εισήγαγε σε μια ταχυδρομική εκστρατεία με σκοπό την ανάλυση των αποτελεσμάτων της και την εξαγωγή χρήσιμων συμπερασμάτων (εξόρυξη γνώσης). Από τις αποκρίσεις που πήρε, κατάφερε με την βοήθεια δένδρων απόφασης και νευρωνικών δικτύων να «χτίσει» ένα μοντέλο πρόβλεψης των χαρακτηριστικών που θα πρέπει να έχουν οι «καλοί» - χαμηλού ρίσκου πελάτες της. Το μοντέλο αυτό εφαρμόστηκε στους εναπομείναντες 950,000 πελάτες μειώνοντας τη λίστα σε 700,000 εγγραφές. Από αυτούς τους (συνολικά) 750,000 δυνητικούς πελάτες κατάφερε να πουλήσει πιστωτικές κάρτες στους 9,000. Βελτίωσε επομένως το ποσοστό απόκρισης των «καλών» πελατών από 1% σε 1.2% (αύξηση της τάξης του 20%). Παίρνοντας την απόφαση να μην «ασχοληθεί» με τους 250,000 δυνητικούς πελάτες της λίστας της, η BB&CC κατάφερε να αυξήσει τα κέρδη της σε βάθος διείσδυσης από 250,000\$ σε 335,000\$ θεωρώντας ως επιπλέον ζημία το κόστος ανάπτυξης του μοντέλου πρόβλεψης.

Τα παραπάνω αποτελέσματα παρουσιάζονται συγκριτικά στην επόμενη εικόνα.

	CRM χωρίς εξόρυξη γνώσης	CRM με εξόρυξη γνώσης
• Αριθμός ταχυδρομικών επιστολών	• 1,000,000	• 750,000
• Κόστος εκστρατείας	• 1,000,000 \$	• 750,000 \$
• Αριθμός αποκρίσεων	• 10,000	• 9,000
• Κέρδος ανά πελάτη	• 125 \$	• 125 \$
• Συνολικό κέρδος	• 1,250,000 \$	• 1,125,000 \$
• Κόστος μοντέλου πρόβλεψης	• 0	• 40,000 \$
• Τελικό κέρδος	• 250,000 \$	• 335,000 \$

Εικόνα 8: Συγκριτικά αποτελέσματα της BB&CC

### 3.3.2 Αύξηση της αξίας των πελατών μέσω εξόρυξης γνώσης

Ο όρος “cross-selling” αναφέρεται στο [3W15] ως η απόπειρα πώλησης επιπλέον προϊόντων κάθε φορά που κάποιος πελάτης έρχεται σε επαφή με μια επιχείρηση.

Η “Guns and Roses Company” (G&R) είναι μια εταιρεία που πουλάει αντίκες (στρατιωτικούς όλμους, κανόνια κτλ) ως γλάστρες εξωτερικών χώρων με καταλόγους προϊόντων που αποστέλλονται σε πάνω από δώδεκα εκατομμύρια νοικοκυριά ετησίως. Η CRM εφαρμογή που η επιχείρηση διαθέτει, διασυνδέεται με τηλεφωνικό κέντρο μέσω ενός συστήματος ολοκλήρωσης, πιο γνωστού ως “Computer Telephony Integration” (CTI). Κάθε φορά που κάποιος υποψήφιος πελάτης καλέσει την G&R με σκοπό κάποια αγοροπωλησία, το σύστημα αναγνωρίζει τον πελάτη αυτόματα μέσω του τηλεφωνικού του αριθμού. Εξαιτίας ακριβώς του τρόπου λειτουργίας της εταιρείας, οι αγοροπωλησίες μέσω τηλεφώνου ευνοούν τις περισσότερες φορές το cross-selling.

Έρευνες ωστόσο που διεξήγαγε η C&R είχαν αποδείξει ότι μια αποτυχημένη σύσταση για αγορά ενός διαφορετικού προϊόντος από το πελάτη, τις περισσότερες φορές (αν όχι όλες) κατέληγε σε εκνευρισμό και άμεσο τερματισμό της επικοινωνίας από τον πελάτη. Αυτός ήταν και ο λόγος, ο οποίος «απαγόρευε» στην C&R οποιαδήποτε προσπάθεια για cross-selling.

Λύση στο σημαντικό αυτό πρόβλημα ήρθε να δώσει η τεχνολογία της εξόρυξης γνώσης από δεδομένα. Για το σκοπό αυτό αναπτύχθηκαν δύο διαφορετικά μοντέλα πρόβλεψης. Το πρώτο, προέκυψε από μια σύντομη τηλεφωνική έρευνα που διεξήγαγε η εταιρεία, και σκοπός του ήταν η πρόβλεψη των δεκτικών, σε συστάσεις νέων προϊόντων, πελατών. Το δεύτερο μοντέλο που χρησιμοποιήθηκε αφορούσε την πρόβλεψη προσφορών κατάλληλων για κάθε πελάτη της G&R.

Αποτέλεσμα των παραπάνω ήταν η αύξηση κατά 2% των πωλήσεων της επιχείρησης και η επιτυχής είσοδος της στο «χώρο» του cross-selling.

### 3.3.3 Διατήρηση «καλών» πελατών μέσω εξόρυξης γνώσης

Είναι ευρέως γνωστό στον κόσμο των επιχειρήσεων ότι η εύρεση νέων πελατών έχει πολύ μεγαλύτερο κόστος για μια εταιρεία σε σχέση με τη προσπάθεια διατήρησης υπαρχόντων πελατών.

Αυτό ήταν γνωστό και στην εταιρεία “Know Service”, έναν πάροχο υπηρεσιών διαδικτύου (internet service provider) που έβλεπε τους πελάτες της να μειώνονται με έναν ρυθμό κοντά στο 8% ανά μήνα. Αυτό σήμαινε ότι, για τη συγκεκριμένη εταιρεία με πελατειακή βάση περίπου 1,000,000 ανθρώπων, κάθε μήνα έχανε 80,000 πελάτες με συνολικό κόστος 16,000,000\$ (200\$ ανά πελάτη).

Η πρώτη ενέργεια που αποφάσισε να εκτελέσει για να αντιστρέψει την παραπάνω κατάσταση ήταν να προσπαθήσει να προβλέψει τους πελάτες που αποχωρούν κάθε φορά. Μέσα από τη CRM εφαρμογή που χρησιμοποιούσε σαν βασικό εργαλείο είχε διαθέσιμο ένα πλούσιο σύνολο δεδομένων που περιλάμβανε στοιχεία πελατών (πχ δημογραφικά δεδομένα), στοιχεία συμβολαίων (πχ πόσες ηλεκτρονικές διευθύνσεις διέθετε κάθε πελάτης), ιστορικά στοιχεία πληρωμών και υπηρεσιών.

Η δεύτερη επιτακτική ανάγκη για την εταιρεία ήταν η προσδιορισμός των «καλών» πελατών. Αν και πρόκειται περισσότερο για μια επιχειρηματική ερώτηση παρά ερώτηση εξόρυξης γνώσης, κατάφερε και «έχτισε» ένα μοντέλο που όχι μόνο διαχώριζε τους «καλούς» από τους «κακούς» πελάτες, αλλά αναγνώριζε και τους πελάτες εκείνους που θα μπορούσαν στο μέλλον να αποδειχτούν «καλοί» για την εταιρεία.



Η τεχνολογία εξόρυξης γνώσης βοήθησε στο να παραχθεί ένα μοντέλο πρόβλεψης των «καλών» πελατών που προτίθενται να «εγκαταλείψουν» την εταιρεία. Η πρόβλεψη αυτή δεν ήταν ωστόσο αρκετή. Το να γνωρίζει μια επιχείρηση τους πελάτες της που πρόκειται να αποχωρήσουν, δε βοηθά σε τίποτα αν δε μπορεί να κάνει κάτι για να τους κρατήσει. Αποτέλεσμα αυτής της ανάγκης ήταν η δημιουργία ενός τρίτου μοντέλου το οποίο με τη σειρά του θα μπορούσε να προβλέψει τα προϊόντα ή τις υπηρεσίες εκείνες που θα δελεάσουν έναν υποψήφιο προς αποχώρηση πελάτη στο να μείνει.

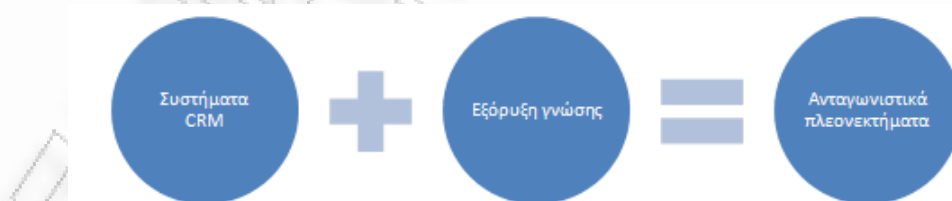
Αποτέλεσμα των τριών παραπάνω μοντέλων ήταν η “Know Service” να μειώσει το ποσοστό των αποχωρήσεων ανά μήνα από 8% σε 7.5%. Αν και η διαφορά μπορεί να μην φαίνεται μεγάλη, οι απόλυτοι αριθμοί λένε ότι μέσω αυτών των τριών μοντέλων, και του συνδυασμού των δυνατοτήτων που σου παρέχει μια CRM εφαρμογή με τις τεχνολογίες εξόρυξης γνώσης, η “Know Service” κατάφερε να μειώσει τη ζημία της κατά 1,000,000\$ μηνιαίως!

### 3.4 Εφαρμόζοντας την εξόρυξη γνώσης στο CRM

Από τα παραπάνω χαρακτηριστικά παραδείγματα αλλά και τις πληροφορίες που αφορούν τα συστήματα CRM μπορούμε να εξάγουμε τα εξής χρήσιμα συμπεράσματα:

- Οι εφαρμογές CRM είναι πλέον αναγκαίες σε κάθε σύγχρονη επιχείρηση που επιθυμεί να διαχειριστεί αποδοτικά και αποτελεσματικά τις σχέσεις της με τους πελάτες της.
- Οι εφαρμογές CRM «καταγράφουν» ολόκληρο το κύκλο ζωής ενός πελάτη μέσα σε μια επιχείρηση. Φτάνουν στο σημείο μάλιστα, να μη χρειάζεται κάποιος να γίνει πελάτης για να εισαχθεί μέσα σε μια τέτοια εφαρμογή. Όλοι είναι δυνητικοί πελάτες.
- Τα συστατικά που διαθέτουν οι εφαρμογές CRM τις καθιστούν ιδιαίτερα ισχυρές σε πολλούς και διαφορετικούς επιχειρηματικούς κλάδους.
- Από την άλλη, οι τεχνολογίες εξόρυξης γνώσης αποτελούν ισχυρά εργαλεία για το σύγχρονο marketing αφού βοηθούν στην εξαγωγή πολύτιμης γνώσης από «τόνους δεδομένων».
- Ο συνδυασμός των δύο παραγόντων, των συστημάτων CRM και της εξόρυξης γνώσης, είναι αυτός που καθιστά μια επιχείρηση ανταγωνιστική στον παγκόσμιο χάρτη και προσφέρει ανταγωνιστικά πλεονεκτήματα έναντι όλων των άλλων.

Συμπερασματικά, η εξόρυξη γνώσης προσφέρει τους κανόνες και τα μοντέλα εκείνα ώστε η εκμετάλλευση των δυνατοτήτων των CRM εφαρμογών να είναι η μέγιστη δυνατή.



Εικόνα 9: Η «εξίσωση» που προσφέρει ανταγωνιστικά πλεονεκτήματα

Στο [3W14] περιγράφονται τα βήματα μέσω των οποίων μια εφαρμογή CRM μπορεί να γίνει αποδοτικότερη με χρήση τεχνολογιών εξόρυξης γνώσης από δεδομένα.



Εικόνα 10: Τα βήματα για αποδοτικές CRM εφαρμογές

Αν και η εικόνα παρουσιάζει τα επτά βήματα σειριακά, η πορεία προς την επίτευξη μέγιστης αποδοτικότητας τις περισσότερες φορές (αν όχι όλες) εμπεριέχει «κύκλους». Έτσι, πχ, η εξερεύνηση των δεδομένων μπορεί μερικές φορές να προσφέρει τέτοια γνώση, ώστε να απαιτηθεί ο επανασχεδιασμός της βάσης δεδομένων προς εμπορική χρήση (marketing database).

- *Ορισμός του προβλήματος:* Κάθε CRM εφαρμογή έχει έναν ή περισσότερους επιχειρηματικούς στόχους τους οποίους προσπαθεί να εκπληρώσει. Ανάλογα με τον στόχο, τα μοντέλα που «χτίζονται» κάθε φορά μπορεί να διαφέρουν. Ο αποτελεσματικός ορισμός του προβλήματος ωστόσο βοηθά στην μετέπειτα αξιολόγηση των μοντέλων.
- *Δημιουργία ΒΔ (marketing):* Τα επόμενα 3 βήματα είναι αυτά που απαιτούν το περισσότερο «κόπο» και χρόνο. Αποτελούν ουσιαστικά υποσύνολα του ενιαίου βήματος «Προετοιμασία δεδομένων» και φυσικά εμπεριέχουν πολλούς «κύκλους» προς τα πίσω, δεδομένου ότι η γνώση που προσφέρουν μεταγενέστερα βήματα βοηθά στην αναθεώρηση προηγούμενων βημάτων. Στο βήμα αυτό, ουσιαστικά «χτίζεται» μια ειδική βάση δεδομένων (marketing database) που διαφοροποιείται από την αντίστοιχη λειτουργική (operational database) αλλά πιθανώς και από τις αποθήκες δεδομένων (data warehouses) στο ότι τα δεδομένα εδώ περιέχονται ακριβώς στην μορφή που μας ενδιαφέρει. Φυσικά σε μια τέτοια βάση, τα δεδομένα που περιέχονται θα πρέπει εκτός των άλλων να είναι «καθαρά» (clean data), εξαλείφοντας διαφορές που προκύπτουν από τις διαφορετικές βάσεις δεδομένων από τις οποίες έχουν προκύψει.
- *Εξερεύνηση δεδομένων:* Απαραίτητη για την δημιουργία αποδοτικών μοντέλων πρόβλεψης είναι η πλήρης κατανόηση των δεδομένων. Η οπτικοποίηση των τελευταίων συχνά προσφέρει γνώση κρυμμένη που έχει σαν αποτέλεσμα την αναθεώρηση των δεδομένων.
- *Προετοιμασία δεδομένων:* Στο τελευταίο από τα τρία βήματα της προετοιμασίας των δεδομένων, εμπεριέχονται τέσσερις κατευθύνσεις όσον αφορά τα δεδομένα:
  - Επιλογή των μεταβλητών πάνω στα οποία θα στηριχθούν τα μοντέλα.
  - Δημιουργία νέων μεταβλητών από δεδομένα παραγόμενα από τα υπάρχοντα.
  - Επιλογή κατάλληλου υποσυνόλου δεδομένων που θα χρησιμοποιηθεί για τη παραγωγή των μοντέλων.
  - Μετατροπή των μεταβλητών έτσι ώστε να μπορούν να εισαχθούν σαν παράμετροι στους αλγορίθμους εξόρυξης γνώσης.
- *Δημιουργία μοντέλου:* Το πιο σημαντικό στοιχείο του συγκεκριμένου βήματος είναι η επαναληπτικότητά του. Πολλές φορές χρειάζεται να εξερευνηθούν πολλά εναλλακτικά μοντέλα με σκοπό να βρεθεί το καταλληλότερο για τη λύση ενός συγκεκριμένου επιχειρηματικού προβλήματος.
- *Αξιολόγηση μοντέλου:* Η αξιολόγηση των μοντέλων συνήθως γίνεται με μέτρα την ακρίβεια (accuracy) και την ανύψωση (lift) ([3W16]).

- *Εφαρμογή μοντέλου:* Κατά την δημιουργία CRM εφαρμογών, η εξόρυξη γνώσης τις περισσότερες φορές αποτελεί ένα μικρό κομμάτι του τελικού προϊόντος. Ο τρόπος με τον οποίο η εξόρυξη γνώσης εφαρμόζεται μέσα σε μια CRM εφαρμογή εξαρτάται από την φύση της αλληλεπίδρασης των επιχειρήσεων με τους πελάτες. Γενικά μιλώντας θα μπορούσαμε να ταξινομήσουμε τις αλληλεπιδράσεις αυτές σε δύο κατηγορίες: τις εισερχόμενες (inbound interaction) και τις εξερχόμενες (outbound interaction). Η βασική διαφοροποίηση των δύο, έγκειται στην φύση της επικοινωνίας. Στις εξερχόμενες αλληλεπιδράσεις, η επιχείρηση έχει όλο το χρόνο να μελετήσει και να αποφασίσει την ταυτότητα των πελατών με τους οποίους θα επικοινωνήσει. Στην αντίθετη περίπτωση των εισερχόμενων αλληλεπιδράσεων, οι κανόνες εξόρυξης γνώσης θα πρέπει να εμπεριέχονται στην CRM εφαρμογή προσφέροντας άμεσες ενέργειες.

### 3.5 Γνωριμία με το Oracle Siebel 8.1 CRM

Στα πλαίσια της παρούσας διπλωματικής εργασίας θα χρησιμοποιήσουμε το Siebel 8.1 CRM που παρέχει η Oracle και στην ενότητα αυτή θα επιχειρήσουμε μια πρώτη γνωριμία με την εφαρμογή.

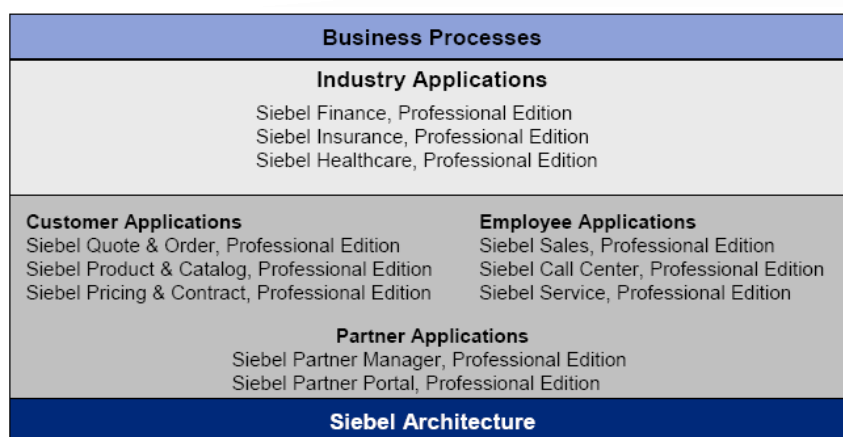
#### 3.5.1 Κλάδοι εφαρμογών

Το Oracle Siebel 8.1 CRM διαθέτει έναν μεγάλο αριθμό εφαρμογών ειδικευμένων σε διάφορους επιχειρηματικούς κλάδους. Στο [12] αναφέρεται πως μια εγκατεστημένη λύση (Enterprise) παρέχει μια ολοκληρωμένη σουίτα με λειτουργικότητες προσαρμοσμένες σε περισσότερες από είκοσι βιομηχανίες (industries).

Business Processes		
Industry Applications		
Siebel Consumer Goods	Siebel Media	
Siebel eEnergy	Siebel Pharma	
Siebel Finance	Siebel Public Sector	
Siebel Industrial Manufacturing	Siebel Travel & Hospitality	
<b>Employee Applications</b>	<b>Customer Applications</b>	<b>Siebel Mobile Solutions</b>
Siebel Call Center	Siebel Advisor	Siebel Sales Handheld
Siebel Field Service	Siebel Configurator	Siebel Remote
Siebel Marketing	Siebel eSales	Siebel Wireless
Siebel Sales	Siebel eService	
	Siebel Orders	<b>Partner Applications</b>
	Siebel Pricer	Siebel Partner Manager
	Siebel Sales Catalog	Siebel Partner Portal
<b>Siebel Analytics</b>		<b>Siebel Architecture</b>

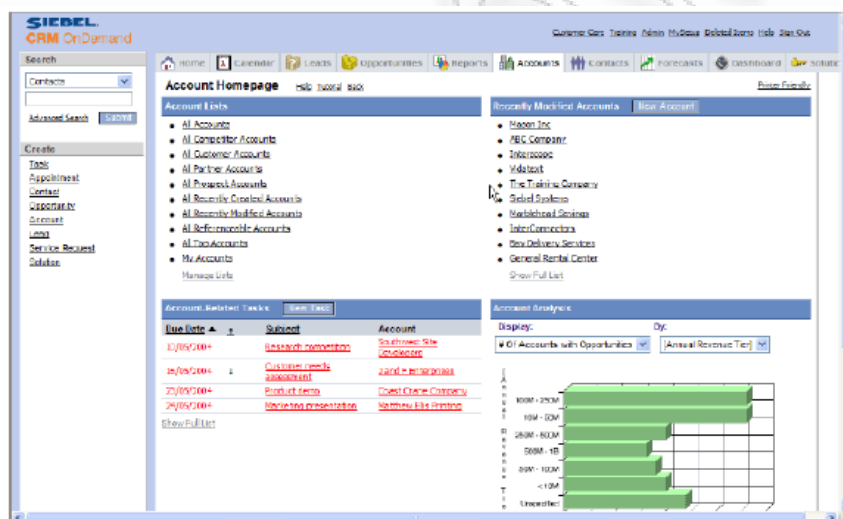
Εικόνα 11: Siebel CRM Enterprise ([12])

Ειδικά για επιχειρήσεις με λιγότερους από 100 χρήστες η Oracle παρέχει το Siebel CRM Professional Enterprise που προσφέρει ένα σύνολο από κανάλια πωλήσεων, εξυπηρέτησης πελατών, εφαρμογών marketing κα με σκοπό την μέγιστη εξυπηρέτηση των αναγκών μεσαίου επιπέδου εταιρειών (mid-sized companies).



Εικόνα 12: Siebel CRM Professional Enterprise ([12])

Πέραν των δύο παραπάνω «εκδόσεων» η Oracle παρέχει και τη δυνατότητα του “CRM OnDemand” μέσω του οποίου οποιαδήποτε εταιρεία μπορεί να φιλοξενηθεί από την ίδια την Oracle παρέχοντας λύσεις Siebel CRM μετά από απαίτηση (OnDemand).

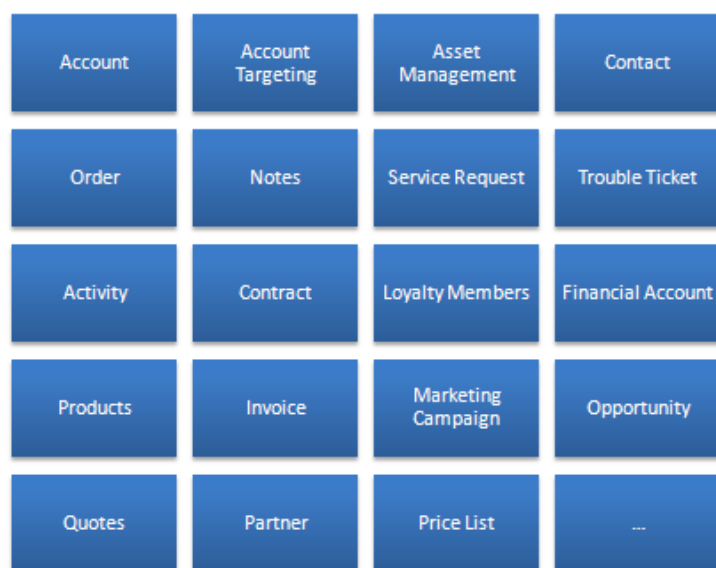


Εικόνα 13: Siebel CRM OnDemand ([12], [3W18])

### 3.5.2 Βασικές οντότητες

Όπως και σε όλες τις εφαρμογές CRM, έτσι και στο Oracle Siebel 8.1 CRM κεντρική οντότητα είναι ο πελάτης. Είτε πρόκειται για επιχείρηση είτε για φυσικό πρόσωπο, στο Siebel απεικονίζονται όλες οι αλληλεπιδράσεις των πελατών με την επιχείρηση μέσω διαφόρων οντοτήτων όπως είναι οι δραστηριότητες, τα αιτήματα εξυπηρέτησης, οι παραγγελίες, οι υπηρεσίες κ.α.

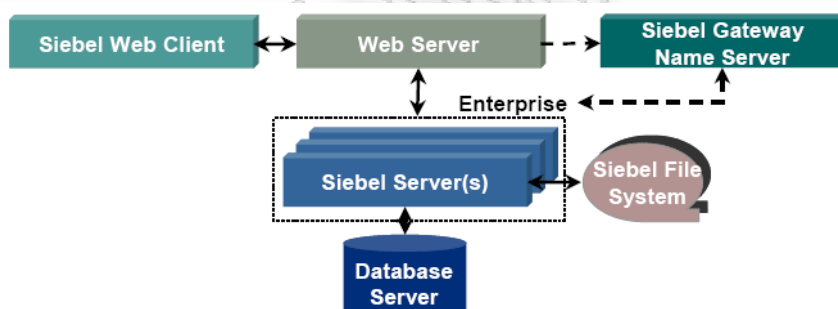
Στην επόμενη εικόνα παρουσιάζουμε ενδεικτικά μερικές από τις οντότητες που περιγράφονται λεπτομερώς στο [11].



Εικόνα 14: Μερικές από τις βασικές οντότητες του Oracle Siebel 8.1 CRM

### 3.5.3 Η αρχιτεκτονική του Oracle Siebel 8.1 CRM

Η αρχιτεκτονική που χρησιμοποιεί η Oracle για το Siebel 8.1 CRM, η οποία είναι μια σύγχρονη αρχιτεκτονική τριών επιπέδων (3-tier architecture), παρουσιάζεται στην επόμενη εικόνα.



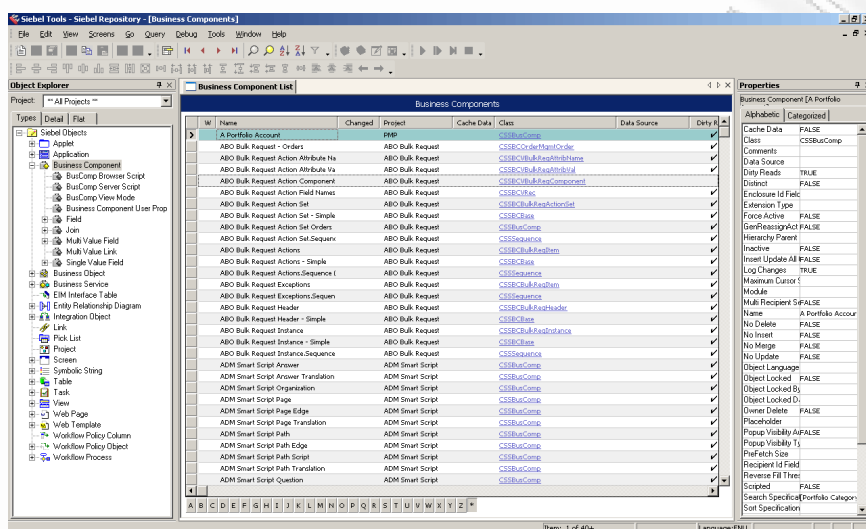
Εικόνα 15: Η αρχιτεκτονική 3-επιπέδων του Oracle Siebel 8.1 CRM ([12])

Η αρχιτεκτονική αυτή περιλαμβάνει:

- Έναν ή περισσότερους web clients οι οποίοι προσπελαίνουν την εφαρμογή μέσω ενός web browser.
- Έναν web server που διαχειρίζεται τις αλληλεπιδράσεις με τους web clients.
- Έναν ή περισσότερους Siebel servers στους οποίους εκτελείται η εφαρμογή και οι διάφορες διαδικασίες προκειμένου τα αιτήματα των web clients να εξυπηρετούνται κατάλληλα.
- Μια σχεσιακή βάση δεδομένων και ένα file system στα οποία αποθηκεύονται όλα τα επιχειρησιακά δεδομένα. Τα δεδομένα της εφαρμογής (business data) καθώς και αυτά που αφορούν την παραμετροποίηση και τον εκτελέσιμο κώδικα (repository data) βρίσκονται στην βάση δεδομένων. Η εφαρμογή μπορεί να χρησιμοποιήσει τις περισσότερες από τις εμπορικές εφαρμογές βάσεων δεδομένων (SQL Server, Oracle, DB2).

### 3.5.4 Η παραμετροποίηση του Oracle Siebel 8.1 CRM

Για την παραμετροποίηση της εφαρμογής η ίδια η Oracle παρέχει το εργαλείο “Siebel Tools” με πρόσβαση σε όλα τα repository data ούτως ώστε να είναι εφικτή η τροποποίηση της εφαρμογής και η απεικόνιση μέσα σε αυτήν όλων των επιχειρηματικών διαδικασιών μιας εταιρείας.



Εικόνα 16: Τα Siebel Tools που παρέχει η Oracle για την παραμετροποίηση της εφαρμογής

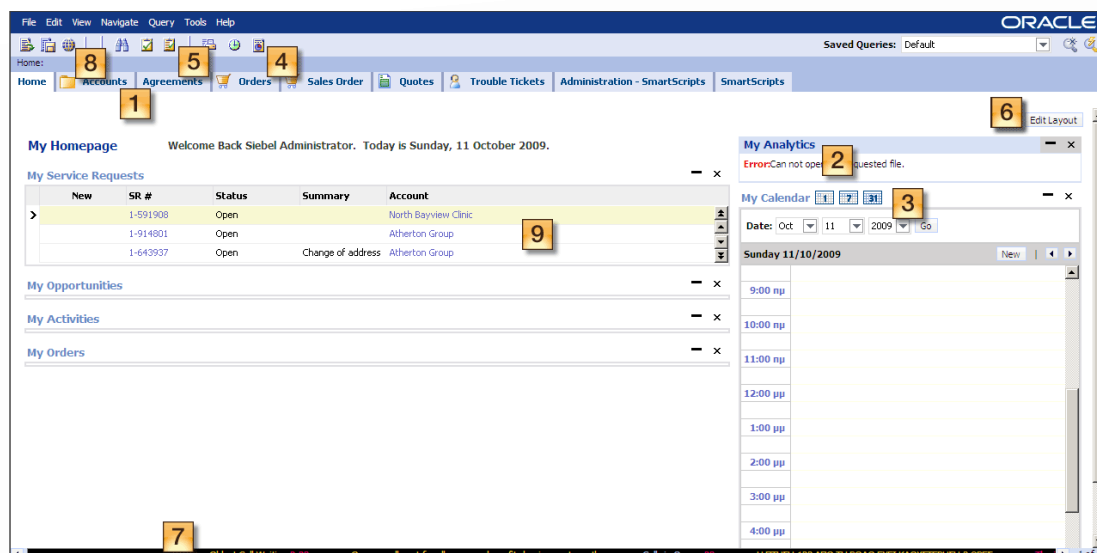
Ωστόσο, η ίδια η Oracle επισημαίνει πως οι τροποποιήσεις που πραγματοποιούνται πάνω στην εφαρμογή με σκοπό την προσαρμογή της στις ανάγκες της εκάστοτε εταιρείας, δε θα πρέπει να ξεπερνά το 20% της συνολικής παραμετροποίησης της εφαρμογής (κανόνας 80/20). Οποιαδήποτε καταπάτηση του κανόνα αυτού (σύμφωνα πάντα με την Oracle) μπορεί να βάλει σε μεγάλο κίνδυνο κάθε έργο CRM σε οποιαδήποτε επιχείρηση.



Εικόνα 17: Ο κανόνας 80/20 που αναφέρει η Oracle ([12])

### 3.5.5 Το περιβάλλον του Oracle Siebel 8.1 CRM

Ο χρήστης της εφαρμογής, αμέσως μετά την είσοδό του στο σύστημα, έρχεται σε επαφή με πολλές δυνατότητες που του παρέχει το Oracle Siebel 8.1 CRM μερικές από τις οποίες φαίνονται στην επόμενη εικόνα.



Εικόνα 18: Η κεντρική οθόνη του Oracle Siebel 8.1 CRM

- 1 Οι βασικές οντότητες (Accounts, Sales Order, Quotes κλπ) και τα δεδομένα που αυτές περιέχουν είναι άμεσα προσβάσιμες από τη κεντρική οθόνη.
- 2 Υπάρχει άμεση πρόσβαση στην πλατφόρμα εξόρυξης γνώσης που παρέχει η Oracle, το Siebel Analytics.
- 3 Ο χρήστης μπορεί άμεσα να παρακολουθήσει το ημερολόγιο του για εργασίες που ολοκληρώθηκαν ή εκκρεμούν.
- 4 Ο χρήστης έχει άμεση πρόσβαση σε εργαλείο αναφορών (reporting tool).
- 5 Ο χρήστης έχει άμεση πρόσβαση στο εργαλείο εργασίες (tasks) μέσω των οποίων μπορεί να εκτελέσει συγκεκριμένες διαδικασίες (πχ καταχώρηση ενός δυνητικού πελάτη).
- 6 Παρέχεται η δυνατότητα προσαρμογής της εφαρμογής ανάλογα με τις ανάγκες κάθε χρήστη (personalization).
- 7 Ο χρήστης από την κεντρική οθόνη μπορεί άμεσα να ενημερώνεται για γεγονότα που μπορεί να επηρεάζουν εμμέσως την εργασία του (πχ ενημέρωση σε μια εταιρεία τηλεπικοινωνιών πως κάποιος συγκεκριμένος κόμβος βρίσκεται εκτός λειτουργίας).
- 8 Πέρα από τις βασικές οντότητες, ο χρήστης μέσω του χάρτη της εφαρμογής (sitemap) έχει πρόσβαση σε όλες τις οντότητες που εμπλέκονται στη θέση εργασίας του.
- 9 Από τη κεντρική οθόνη ο χρήστης έχει άμεση πρόσβαση σε όλα τις εγγραφές συγκεκριμένων οντοτήτων (πχ αιτημάτων εξυπηρέτησης).

## 4. Παρουσίαση Εφαρμογής

Στα προηγούμενα δύο κεφάλαια ήρθαμε σε επαφή με τις δύο βασικές έννοιες πάνω στις οποίες στηρίζεται η παρούσα διπλωματική εργασία: την εξόρυξη γνώσης και τα συστήματα CRM.

Στο τρέχον κεφάλαιο θα δούμε πώς οι δύο αυτές έννοιες μπορούν να συνδυαστούν με σκοπό την «εκμετάλλευση» των χαρακτηριστικών τους στα πλαίσια μιας επιχείρησης. Θα δούμε με ποιο τρόπο μπορεί μια επιχείρηση να χρησιμοποιήσει τα συμπεράσματα που προκύπτουν μέσω της εξόρυξης γνώσης από τα δεδομένα της, και πώς αυτά μπορούν να συμβάλλουν στην βελτίωση των επιχειρησιακών της διαδικασιών.

Για το σκοπό αυτό, το συγκεκριμένο κεφάλαιο θα καλύψει τις ακόλουθες ενότητες:

- Αρχικά θα έρθουμε σε επαφή με ένα συγκεκριμένο εργαλείο εξόρυξης γνώσης, το WEKA. Θα δούμε πώς από ένα σύνολο δεδομένων μπορούμε να καταλήξουμε σε ένα δένδρο απόφασης με την βοήθεια τεχνικών κατηγοριοποίησης.
- Στη συνέχεια, θα προτείνουμε ένα τρόπο αποθήκευσης των αποτελεσμάτων του WEKA στην CRM εφαρμογή που θα χρησιμοποιήσουμε, το Oracle Siebel 8.1 CRM.
- Θα γνωρίσουμε ένα συγκεκριμένο εργαλείο που παρέχει το Oracle Siebel 8.1 CRM και το οποίο θα χρησιμοποιήσουμε για την αποδοτικότερη εκμετάλλευση των αποτελεσμάτων του WEKA, τα Siebel Smartscripsts («Έξυπνα Σενάρια»).
- Τέλος, θα παρουσιάσουμε μία πλήρη επιχειρησιακή διαδικασία όπως μπορεί να αποτυπωθεί μέσα από την CRM εφαρμογή μας, ενσωματώνοντας όλα τα προαναφερθέντα.

Στο επόμενο κεφάλαιο θα δούμε ένα πραγματικό παράδειγμα επίδειξης του παραπάνω συστήματος.

### 4.1 Εξόρυξη γνώσης με τη βοήθεια του WEKA

#### 4.1.1 Το εργαλείο WEKA

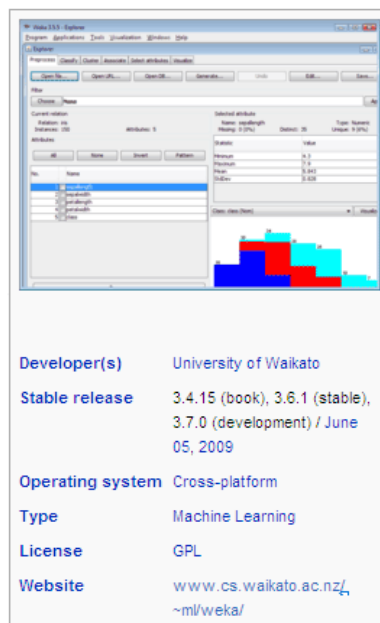
Το εργαλείο WEKA (Waikato Environment for Knowledge Analysis) είναι μια δημοφιλής σουίτα λογισμικού με σκοπό τη μηχανική μάθηση, ανεπτυγμένη σε Java από το πανεπιστήμιο του Waikato ([3W19]). Περιλαμβάνει μια συλλογή από εργαλεία οπτικοποίησης και αλγόριθμους για ανάλυση δεδομένων, εξόρυξη γνώσης και ανάπτυξη προγνωστικών μοντέλων (predictive modeling).

Από το 1997 – οπότε και δημιουργήθηκε η πρώτη έκδοση του εργαλείου πλήρως βασισμένη σε Java – μέχρι σήμερα, το WEKA χρησιμοποιείται ευρέως, και ιδιαίτερα για εκπαιδευτικούς ή/και ερευνητικούς σκοπούς.

Οι λόγοι για τους οποίους το WEKA αποτελεί μια ιδιαίτερως δημοφιλή εφαρμογή παγκοσμίως μπορούν να συνοψιστούν στους ακόλουθους:

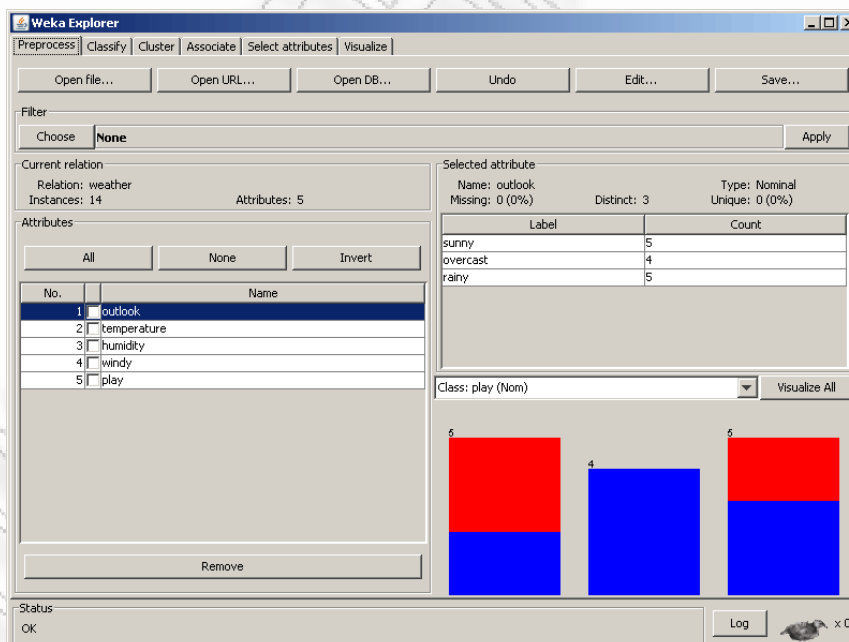
- Πρόκειται για ελεύθερο λογισμικό (open source software).
- Μπορεί να χρησιμοποιηθεί σε όλες τις σύγχρονες πλατφόρμες λογισμικού εξαιτίας του ότι είναι ανεπτυγμένο σε Java.
- Περιέχει μια εκτενή συλλογή τεχνικών ανάλυσης δεδομένων και εξόρυξης γνώσης.
- Είναι ιδιαίτερως εύχρηστο χάρη στο γραφικό περιβάλλον που διαθέτει.





Εικόνα 19: Η "ταυτότητα" του WEKA όπως παρουσιάζεται στο [3W19]

Με τη βοήθεια του WEKA μπορούμε να εκτελέσουμε διάφορες τεχνικές εξόρυξης γνώσης από δεδομένα όπως είναι η κατηγοριοποίηση (classification), η συσταδοποίηση (clustering) και οι κανόνες συσχέτισης (association rules). Για κάθε μία από τις παραπάνω τεχνικές το εργαλείο παρέχει ένα σύνολο αλγορίθμων που μπορούν να χρησιμοποιηθούν ανάλογα με την περίπτωση, δίνοντας πλήρη δυνατότητα παραμετροποίησης τους από το χρήστη.



Εικόνα 20: Το περιβάλλον του WEKA

### 4.1.2 Χρησιμοποιώντας το WEKA

Στα πλαίσια της παρούσας διπλωματικής εργασίας χρησιμοποιήσαμε το εργαλείο WEKA με σκοπό την κατηγοριοποίηση ενός συνόλου δεδομένων.

Το σύνολο δεδομένων που χρησιμοποιήθηκε είναι διαθέσιμο στο [3W22] και αφορά ιατρικά δεδομένα δερματολογικής φύσεως. Πιο συγκεκριμένα, πρόκειται για ένα σετ δεδομένων με τα ακόλουθα χαρακτηριστικά.

Κατηγορία δεδομένων	Ιατρικά δεδομένα δερματολογικής φύσης
Πλήθος δεδομένων	366
Αριθμός χαρακτηριστικών	13
Πηγή	<a href="http://www.kdnuggets.com/datasets/">http://www.kdnuggets.com/datasets/</a>

Πίνακας 1: Η "ταυτότητα" του συνόλου δεδομένων

Τα 13 χαρακτηριστικά μπορούν να χωριστούν σε 3 υποομάδες ανάλογα με το πεδίο τιμών τους:

- 1<sup>η</sup> ομάδα: περιλαμβάνει 11 από τα 13 χαρακτηριστικά με πεδίο τιμών το σύνολο {0, 1, 2, 3} που αντιστοιχεί στο βαθμό εμφάνισης κάθε χαρακτηριστικού σε κάθε μία από τις 366 περιπτώσεις.
  - 0 – μηδενικός βαθμός εμφάνισης
  - 1 – μικρός βαθμός εμφάνισης
  - 2 – μεγάλος βαθμός εμφάνισης
  - 3 – μέγιστος βαθμός εμφάνισης
- 2<sup>η</sup> ομάδα: περιλαμβάνει το χαρακτηριστικό της ηλικίας με γραμμικό πεδίο τιμών (linear).
- 3<sup>η</sup> ομάδα: περιλαμβάνει την κλάση στην οποία ανήκει κάθε ένας από τους 366 ασθενείς και μπορεί να παίρνει τιμές από το σύνολο {1, 2, 3, 4, 5, 6} κάθε μία από τις οποίες αντιστοιχεί σε μία από τις ακόλουθες έξι δερματολογικές ασθένειες.
  - 1 – Psoriasis
  - 2 – Seboric Dermatitis
  - 3 – Lichen Planus
  - 4 – Pityriasis Rosea
  - 5 – Cronic Dermatitis
  - 6 – Pityriasis Rubra Pilaris

Στην επόμενη εικόνα μπορούμε να δούμε δειγματοληπτικά ένα υποσύνολο του συνόλου δεδομένων:

- Η πρώτη γραμμή περιλαμβάνει τα 13 χαρακτηριστικά του συνόλου.
- Οι πρώτες 11 στήλες αφορούν τα χαρακτηριστικά της ομάδας 1.
- Η 12<sup>η</sup> στήλη αφορά το χαρακτηριστικό της ηλικίας (ομάδα 2).
- Η 13<sup>η</sup> στήλη αφορά την ομάδα (κατηγορία) στην οποία ανήκει κάθε μία από τις 366 περιπτώσεις.

#	koebner phenomenon	eosinophils in the infiltrate	PNL infiltrate	fibrosis of the papillary dermis	hyperkeratosis	parakeratosis	elongation of the rete ridges	disappearance of the granular layer	vacuolisation and damage of basal layer	spongiosis	perifollicular parakeratosis	agelinear	class code	
1	0	0	0	0	0	0	0	0	0	0	3	0	55	2
2	1	0	1	0	0	0	2	2	0	0	0	0	8	1
3	1	0	0	0	0	0	2	0	0	2	3	0	26	3
4	0	0	3	0	0	0	3	2	3	0	0	0	40	1
5	2	0	0	0	0	0	0	0	2	3	2	0	45	3
6	0	2	1	0	0	0	2	0	0	0	2	0	41	2
7	0	0	0	0	3	0	0	2	0	0	0	0	18	5
8	3	0	0	0	0	0	0	0	2	2	3	0	57	3
9	2	0	0	0	0	0	1	0	0	0	2	2	22	4
10	1	0	0	0	0	0	2	0	0	0	2	0	30	4

Εικόνα 21: Το σύνολο δεδομένων

Επιγραμματικά, παρατηρώντας το πιο πάνω υποσύνολο δεδομένων, μπορούμε να πούμε ότι ο 4ος ασθενής, είναι ηλικίας 40 ετών, ανήκει στην 1η κατηγορία (ομάδα) ασθενειών και παρουσιάζει (μεταξύ άλλων) μέγιστη εμφάνιση (βαθμός 3) του στοιχείου “PNL Infiltrate”, μηδενική εμφάνιση (βαθμός 0) του στοιχείου “Koebner Phenomenon” και μεγάλη εμφάνιση (βαθμός 2) του στοιχείου “Elongation of the rete Ridges”.

Το παραπάνω σύνολο δεδομένων χρειάστηκε να τροποποιηθεί καταλλήλως ώστε να είναι δυνατή η επεξεργασία του από το WEKA καθώς το συγκεκριμένο εργαλείο δέχεται μόνο αρχεία συγκεκριμένου τύπου (με επέκταση \*.arff) και γραμμογράφησης.

Επικεφαλίδα αρχείου	@relation dermatology finaldata
Γνωρίσματα συνόλου δεδομένων και πεδία τιμών	@attribute koebnerphenomenon {0, 1, 2, 3}
	@attribute eosinophilsintheinfiltrate {0, 1, 2, 3}
	@attribute PNLinfiltrate {0, 1, 2, 3}
	@attribute fibrosisofthepapillarydermis {0, 1, 2, 3}
	@attribute hyperkeratosis {0, 1, 2, 3}
	@attribute parakeratosis {0, 1, 2, 3}
	@attribute elongationofthereteridges {0, 1, 2, 3}
	@attribute disappearanceofthegranularlayer {0, 1, 2, 3}
	@attribute vacuolisationanddamageofbasallayer {0, 1, 2, 3}
	@attribute spongiosis {0, 1, 2, 3}
@attribute perifollicularparakeratosis {0, 1, 2, 3}	
@attribute Agelinear real	
@attribute ClassCode {1, 2, 3, 4, 5, 6}	
Δεδομένα	@data 0,0,0,0,0,0,0,0,0,3,0,55,2 1,0,1,0,0,2,2,0,0,0,0,8,1 1,0,0,0,0,2,0,0,2,3,0,26,3 0,0,3,0,0,3,2,3,0,0,0,40,1 2,0,0,0,0,0,0,2,3,2,0,45,3

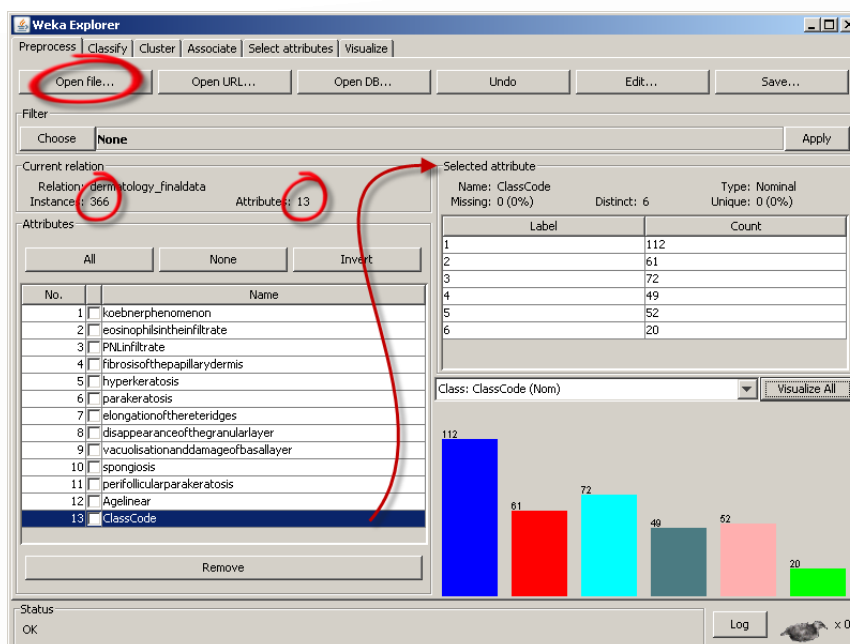
Εικόνα 22: Γραμμογράφηση ενός \*.arff αρχείου

Με δεδομένο το παραπάνω αρχείο \*.arff μπορούμε πλέον να χρησιμοποιήσουμε το εργαλείο WEKA με σκοπό την κατηγοριοποίηση των δεδομένων και την εξαγωγή ενός δένδρου απόφασης.

#### 4.1.3 Κατηγοριοποίηση με χρήση του WEKA

Για την εξαγωγή του επιθυμητού δένδρου απόφασης χρησιμοποιήσαμε την τεχνική της κατηγοριοποίησης με τη βοήθεια του WEKA.

Αρχικά φορτώσαμε στο εργαλείο το \*.arff αρχείο που δημιουργήσαμε από το σύνολο δεδομένων μας.



Εικόνα 23: Εισαγωγή συνόλου δεδομένων στο WEKA

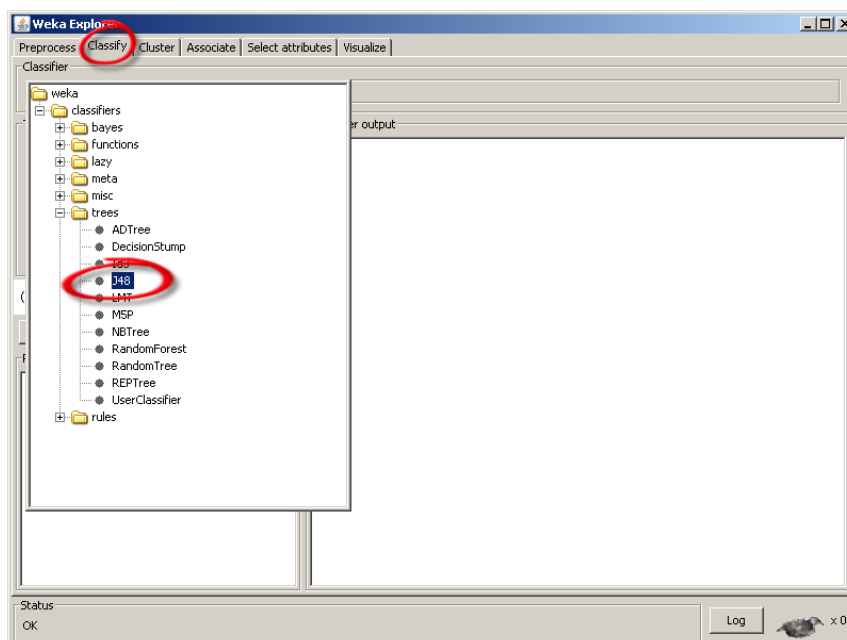
Από την καρτέλα “Preprocess” του WEKA επιλέξαμε και ανοίξαμε το αρχείο `dermatology_data.arff` που περιλαμβάνει το σύνολο των δεδομένων μας στη μορφή που απαιτεί το WEKA. Ανοίγοντας το αρχείο, το εργαλείο αναγνωρίζει το πλήθος των χαρακτηριστικών γνωρισμάτων (attributes) που περιλαμβάνει καθώς και το πλήθος των περιπτώσεων (instances) που είναι καταχωρημένες σε αυτό. Επιλέγοντας το γνώρισμα της ομάδας (κατηγορίας) είμαστε σε θέση να δούμε την κατανομή των περιπτώσεων (instances) για το συγκεκριμένο χαρακτηριστικό. Έτσι πχ παρατηρούμε ότι το 30.6% (112/366) των περιπτώσεων αφορούν την ασθένεια 1 – Psoriasis ενώ μόλις το 5.46% (20/366) αφορά την ασθένεια 6 – Pityriasis Rubra Pilaris.

Για την κατηγοριοποίηση των δεδομένων και την εξαγωγή του δένδρου απόφασης χρησιμοποιήσαμε τον αλγόριθμο J48 που παρέχει το εργαλείο και ο οποίος είναι παρόμοιος με τον C4.5 που περιγράψαμε σε προηγούμενο κεφάλαιο.

Στο [3W21] αναφέρεται πως ο J48 εισάγει δύο μεθόδους «κλαδέματος» των δένδρων απόφασης με απώτερο σκοπό την απλοποίηση των αποτελεσμάτων:

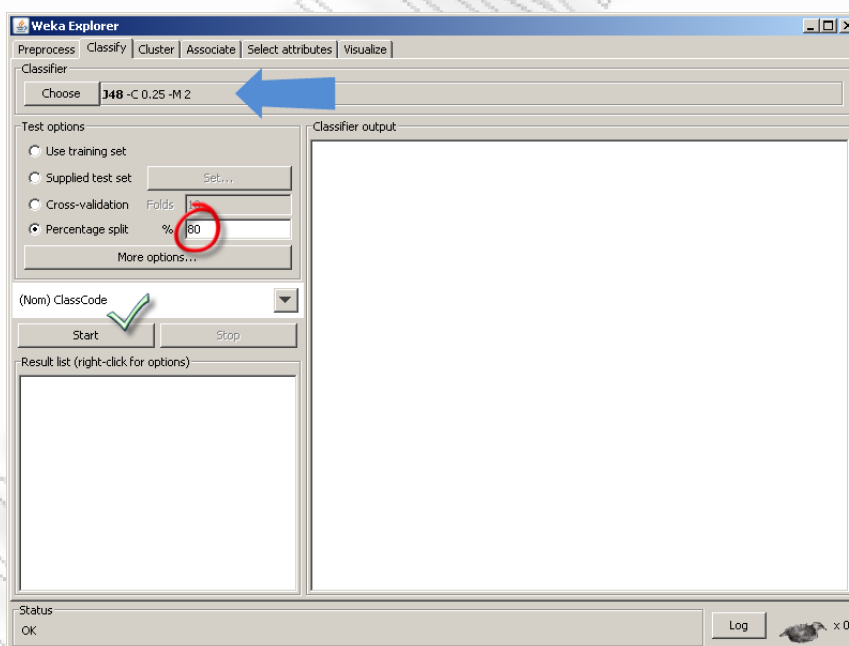
- Ο πρώτος τρόπος «κλαδέματος» αφορά την αντικατάσταση ολόκληρων υποδένδρων από φύλλα του συνολικού δένδρου.
- Ο δεύτερος τρόπος «κλαδέματος» του J48 αφορά την «μετακίνηση» υποδένδρων προς την ρίζα του συνολικού δένδρου με άμεση συνέπεια την διαγραφή αρκετών εσωτερικών κόμβων του αρχικού δένδρου.

Από την καρτέλα “Classify” και το κουμπί “Choose” στο τμήμα “Classifier” του εργαλείου WEKA επιλέγουμε τον επιθυμητό αλγόριθμο (J48) για την εξαγωγή του δένδρου απόφασης.



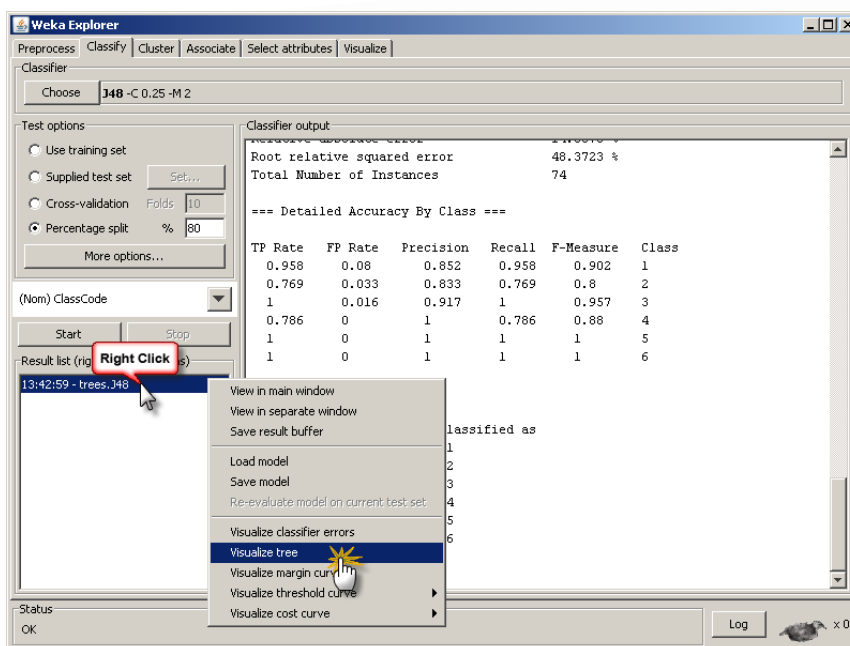
Εικόνα 24: Επιλογή αλγορίθμου κατηγοριοποίησης

Εφόσον επιλέξουμε τον αλγόριθμο κατηγοριοποίησης (J48), επιλέγουμε το ποσοστό του συνόλου των δεδομένων που θα χρησιμοποιηθούν για την εκμάθηση του μοντέλου κατηγοριοποίησης (80%). Μετά και από αυτό είμαστε έτοιμοι να εκτελέσουμε τον αλγόριθμο κατηγοριοποίησης πατώντας το κουμπί “Start”.



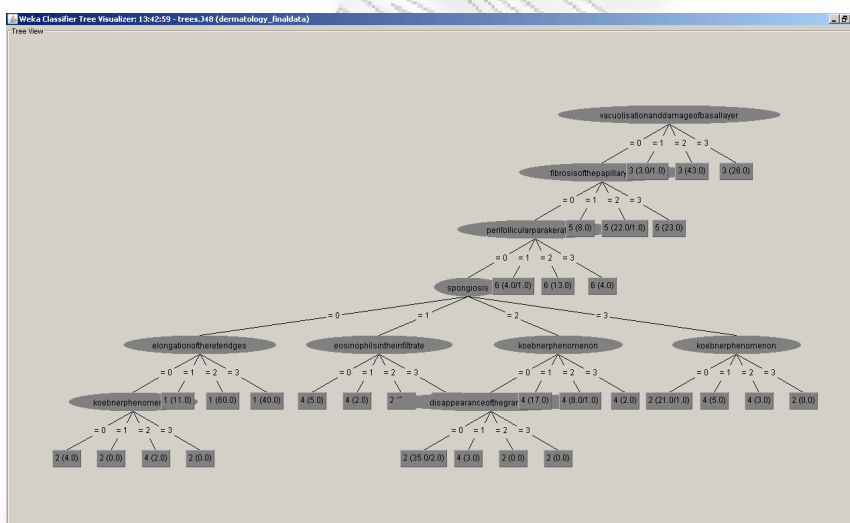
Εικόνα 25: Εκτέλεση της διαδικασίας κατηγοριοποίησης

Η εκτέλεση της διαδικασίας μας επιστρέφει ένα σύνολο αποτελεσμάτων διαθέσιμων στο τμήμα “Classifier output” του WEKA. Η οπτικοποίηση ωστόσο του δένδρου απόφασης μας βοηθάει περισσότερο στην κατανόηση των αποτελεσμάτων.



Εικόνα 26: Οπτικοποίηση των αποτελεσμάτων

Έτσι, από το υπομενού “Visualize tree” έχουμε πρόσβαση στο δένδρο απόφασης που παρήγαγε το WEKA.



Εικόνα 27: Το δένδρο απόφασης που παρήγαγε το WEKA

Στο τμήμα “Classifier output” της εφαρμογής του WEKA παρουσιάζεται όλη η πληροφορία που αφορά τη διαδικασία της κατηγοριοποίησης. Στον επόμενο πίνακα περιγράφεται συνοπτικά αυτή η πληροφορία.

Τμήμα	Πληροφορία
Run information	Παρουσιάζεται η πληροφορία σχετικά με το σύνολο των δεδομένων, του αλγορίθμου και των παραμέτρων που χρησιμοποιήθηκαν, το πλήθος των περιπτώσεων (instances), των γνωρισμάτων (attributes) κα.

Τμήμα	Πληροφορία
<b>Classifier model (full training set)</b>	Το δένδρο απόφασης έτσι όπως το δημιούργησε το WEKA.
<b>Summary</b>	Περιλαμβάνει στατιστικά στοιχεία της διαδικασίας όπως ποσοστά σφαλμάτων, πλήθος επιτυχώς κατηγοριοποιημένων περιπτώσεων κα.
<b>Detailed Accuracy By Class</b>	Περιλαμβάνει με μεγαλύτερη λεπτομέρεια την ανάλυση της ακρίβειας πρόβλεψης του εργαλείου ανά ομάδα (κλάση).
<b>Confusion Matrix</b>	Περιλαμβάνει τη μήτρα σύγχυσης όπως αυτή έχει περιγραφεί σε προηγούμενη ενότητα της παρούσας εργασίας. Το άθροισμα των στοιχείων της διαγωνίου του πίνακα αναφέρεται στο πλήθος των περιπτώσεων του συνόλου δεδομένων που κατηγοριοποιήθηκαν ορθώς από το εργαλείο.

Πίνακας 2: Τα αποτελέσματα εξόδου του WEKA

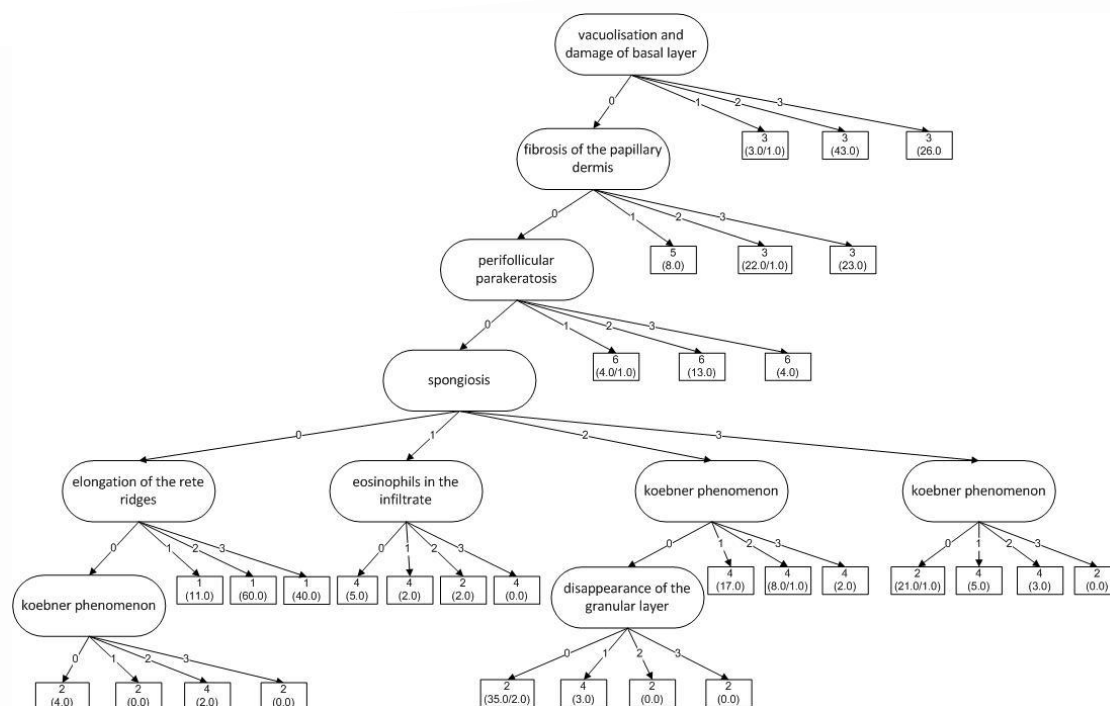
#### 4.1.4 Το δένδρο απόφασης

Το δένδρο απόφασης όπως το δημιούργησε το WEKA είναι αυτό της επόμενης εικόνας.

- Κάθε κόμβος του δένδρου είναι ένα από τα γνωρίσματα του συνόλου δεδομένων.
- Κάθε φύλλο του δένδρου είναι μία από τις ομάδες (κλάσεις) στις οποίες κατηγοριοποιήθηκαν τα δεδομένα.
- Κάθε μετάβαση από ένα κόμβο A σε ένα κόμβο B γίνεται βάση της τιμής του γνωρίσματος που βρίσκεται στο κόμβο A.

Όπως παρατηρούμε και στην εικόνα του δένδρου, κάθε φύλλο του δε περιέχει μονάχα την ομάδα της κατηγοριοποίησης. Περιλαμβάνει δύο επιπλέον αριθμούς στη μορφή X/Y. Ο αριθμός X αναπαριστά το πλήθος των περιπτώσεων που κατηγοριοποιήθηκαν στην ομάδα την οποία αντιπροσωπεύει το συγκεκριμένο φύλλο, ενώ ο αριθμός Y το πλήθος των περιπτώσεων που κατηγοριοποιήθηκαν εσφαλμένως στην ομάδα αυτή.

Στην επόμενη ενότητα θα δούμε πώς μπορούμε να καταχωρήσουμε το δένδρο απόφασης που παρήγαγε το WEKA μέσα στο Oracle Siebel 8.1 CRM με σκοπό την αποδοτικότερη εκμετάλλευσή του μέσα από τα Siebel Smartscripts.



Εικόνα 28: Το δένδρο απόφασης

## 4.2 Εισαγωγή του δένδρου απόφασης στο Oracle Siebel 8.1 CRM

Στην προηγούμενη ενότητα παρουσιάσαμε την εξαγωγή ενός δένδρου απόφασης ως αποτέλεσμα μιας ευρύτερης διαδικασίας κατηγοριοποίησης που εκτελέσαμε εφαρμόζοντας τον αλγόριθμο J48 πάνω σε ένα συγκεκριμένο σύνολο δεδομένων (μέσω του εργαλείου WEKA).

Στην ενότητα αυτή θα δούμε τον τρόπο που επιλέξαμε ώστε να αποθηκεύσουμε αυτό το δένδρο απόφασης στην βάση δεδομένων του Oracle Siebel 8.1 CRM με σκοπό την εκμετάλλευση της γνώσης που αυτό διαθέτει.

### 4.2.1 Δένδρα Απόφασης, Δομές και Βάσεις Δεδομένων

Με τον όρο «επαγωγικές βάσεις δεδομένων» (inductive databases – IDB) ([17]) αναφερόμαστε σε βάσεις δεδομένων στις οποίες αποθηκεύονται όχι μόνο «απλά» δεδομένα αλλά και μοντέλα/πρότυπα (models/patterns) αυτών. Σε τέτοιου είδους βάσεις συναντάμε δύο είδη ερωτημάτων SQL ([16]):

- Τα συνήθη ερωτήματα SQL με τα οποία έχουμε πρόσβαση και διαχειριζόμαστε δεδομένα,
- Τα επαγωγικά ερωτήματα SQL με τα οποία παράγουμε, διαχειριζόμαστε και εφαρμόζουμε τα διάφορα μοντέλα/πρότυπα πάνω στα δεδομένα.

Για την απεικόνιση των ερωτημάτων της δεύτερης κατηγορίας χρησιμοποιούνται δύο γενικές προσεγγίσεις:

- Η πρώτη αφορά την προσπάθεια επέκτασης των διαφόρων γλωσσών ερωτημάτων (query languages) με ειδικές δομές ή/και εντολές, με σκοπό την «υλοποίηση» των επαγωγικών ερωτημάτων.
- Η δεύτερη προσέγγιση αφορά τη προσπάθεια αποθήκευσης των μοντέλων/πρωτύπων μιας επαγωγικής βάσης δεδομένων, με τρόπο τέτοιο ώστε η



καθιερωμένη γλώσσα SQL να είναι ικανή να αναπαραστήσει, αποθηκεύσει και να χρησιμοποιήσει όλα τα μοντέλα/πρότυπα πάνω στα δεδομένα.

Βασικό μειονέκτημα της πρώτης προσέγγισης – αν και σαν εργαλείο μια εκτεταμένη SQL γλώσσα είναι σαφώς πιο ισχυρή – είναι ότι η χρήση τέτοιων γλωσσών απαιτεί βαθιά γνώση από τη μεριά του χρήστη τόσο της ίδιας της SQL όσο και των αλγορίθμων εξόρυξης γνώσης ([16]).

Από την άλλη πλευρά, η δεύτερη προσέγγιση αν και απαιτεί αρκετή προσπάθεια όσον αφορά στην αποδοτικότερη αποθήκευση των μοντέλων και των προτύπων στη σχεσιακή βάση δεδομένων, μετατρέπει ωστόσο την εξόρυξη γνώσης σε μια διαφανή διαδικασία που δεν απαιτεί την συμβολή του χρήστη, ο οποίος από τη πλευρά του έρχεται σε επαφή με τις γνωστές σε αυτόν σχεσιακές βάσεις δεδομένων. Δε χρειάζεται να μάθει πώς οι πίνακες της βάσης του γεμίζουν από τα δεδομένα που αφορούν τα μοντέλα και τα πρότυπα της εξόρυξης γνώσης και ασφαλώς δεν χρειάζεται να έρθει σε επαφή με γλώσσες ερωτημάτων ειδικού σκοπού για να μπορέσει να τα διαχειριστεί.

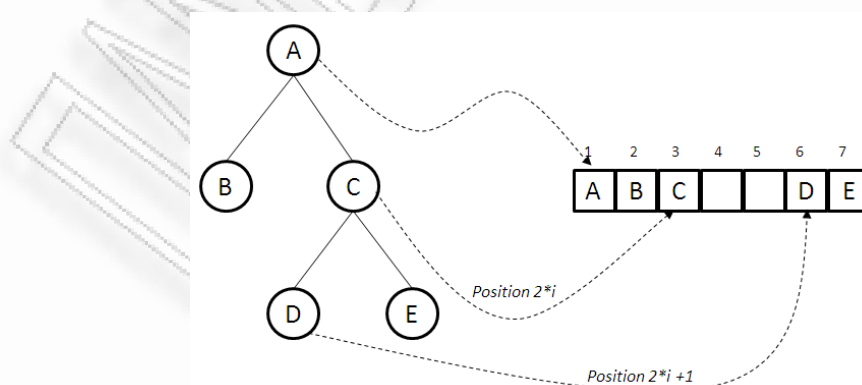
Στη παρούσα διπλωματική εργασία θα ακολουθήσουμε τη δεύτερη προσέγγιση και θα προσπαθήσουμε να αποθηκεύσουμε το δένδρο απόφασης σε μια σχεσιακή βάση δεδομένων με σκοπό στη συνέχεια να μπορούμε να εκμεταλλευτούμε τη γνώση του χωρίς να χρειάζεται να εισάγουμε μια νέα γλώσσα ερωτημάτων ή να επεκτείνουμε κάποια υπάρχουσα.

Η απεικόνιση ενός δένδρου απόφασης σε μια δομή δεδομένων ή σε ένα πίνακα κάποιας βάσης δεδομένων είναι ένα αρκετά κρίσιμο ζήτημα με άμεση συνέπεια το βαθμό αποδοτικότητας με τον οποίο μια διαδικασία μπορεί να εκμεταλλευτεί τη γνώση που η δομή ή ο πίνακας δεδομένων περιέχει.

Για παράδειγμα, η προσπάθεια αποθήκευσης ενός δυαδικού δένδρου σε έναν μονοδιάστατο πίνακα (array) όπως περιγράφεται στο [15] κρύβει αρκετά μειονεκτήματα με σημαντικότερα τη μη πλήρη εκμετάλλευση της δομής δεδομένων καθώς και την ανάγκη αναδιάταξης της δομής κάθε φορά που νέοι κόμβοι προστίθενται ή παλιοί αφαιρούνται από το δένδρο.

Η βασική ιδέα του συγκεκριμένου τρόπου αποθήκευσης του δυαδικού δένδρου που παρουσιάζεται στην επόμενη εικόνα είναι:

- Η ρίζα του δένδρου αποθηκεύεται πάντα στη θέση 0 της δομής του μονοδιάστατου πίνακα.
- Κάθε αριστερό παιδί του κόμβου  $i$  αποθηκεύεται στη θέση  $2^*i$  της δομής.
- Κάθε δεξί παιδί του κόμβου  $i$  αποθηκεύεται στη θέση  $2^*i + 1$  της δομής.



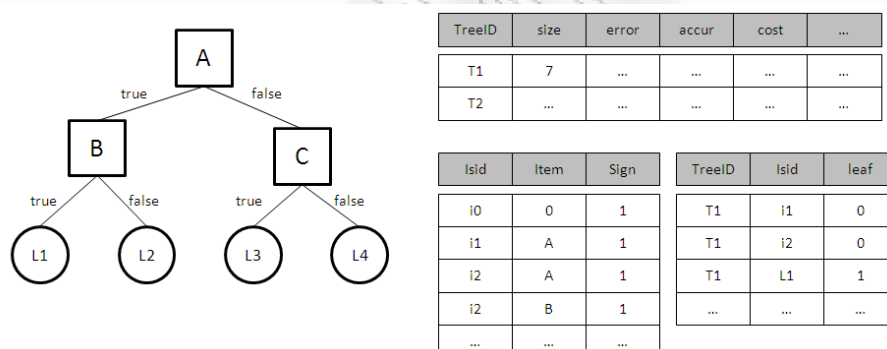
Εικόνα 29: Αποθήκευση δυαδικού δένδρου σε μονοδιάστατο πίνακα (array) ([15])

Ένας άλλος τρόπος αποθήκευσης δυαδικών δένδρων σε πίνακες περιγράφεται στο [16]. Στο άρθρο αυτό γίνεται σαφές πως κάθε φύλλο ενός δένδρου απόφασης που έχει προκύψει από την εκτέλεση κάποιας διαδικασίας κατηγοριοποίησης, θα μπορούσε να θεωρηθεί ως μια μορφή κανόνα συσχέτισης όπου όλοι οι ενδιάμεσοι κόμβοι αναπαριστούν τα «προηγούμενα» (antecedents) του κανόνα ενώ η κλάση που αντιστοιχεί στο φύλλο αφορά το «επακόλουθο» (consequent) αυτού.

Ωστόσο, μια απεικόνιση ενός δένδρου απόφασης με την μορφή κανόνων συσχέτισης (πιο εύκολα αποθηκεύσιμων σε μορφή πίνακα μιας βάσης δεδομένων<sup>1</sup>), αν και βοηθά στην ευκολότερη απεικόνιση των προβλέψεων, αποκρύπτει σημαντική πληροφορία που σχετίζεται αποκλειστικά με την δομή των δένδρων: όσο υψηλότερα εμφανίζεται ένα χαρακτηριστικό γνώρισμα στην δομή ενός δένδρου, τόσο μεγαλύτερη πληροφοριακή αξία για ενδεχόμενη πρόβλεψη έχει.

Στο άρθρο αυτό παρουσιάζεται η ιδέα της αποθήκευσης ενός δυαδικού δένδρου απόφασης στα πλαίσια μιας βάσης δεδομένων με την βοήθεια τριών πινάκων:

- Στο πρώτο πίνακα διατηρούνται στατιστικά στοιχεία των υποδένδρων του αρχικού δένδρου όπως η ακρίβεια, το μέγεθος του δένδρου, το κόστος κλπ.
- Στο δεύτερο πίνακα διατηρούνται όλα τα πιθανά μονοπάτια που παρουσιάζονται στο δένδρο.
- Στο τρίτο πίνακα διατηρούνται όλα τα υποδένδρα του αρχικού δένδρου συνδεδεμένα με τα μονοπάτια του δεύτερου πίνακα.



Εικόνα 30: Δυαδικό δένδρο απόφασης σε πίνακες ([16])

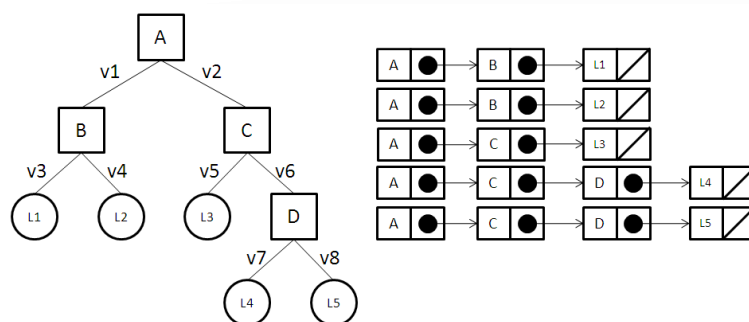
## 4.2.2 Το σχήμα αποθήκευσης μέσα στο Oracle Siebel 8.1 CRM

### 4.2.2.1 Η βασική ιδέα

Η βασική ιδέα που χρησιμοποιήσαμε για την αποθήκευση του δένδρου απόφασης μέσα στη σχεσιακή βάση δεδομένων μας, ήταν οι συνδεδεμένες λίστες (linked lists). Όπως περιγράφεται και στο [3W24], ως συνδεδεμένη λίστα ορίζουμε μια δομή δεδομένων που αποτελείται από μια ακολουθία δεδομένων, κάθε μέλος της οποίας περιέχει μια αναφορά (reference ή link) προς το επόμενο μέλος της.

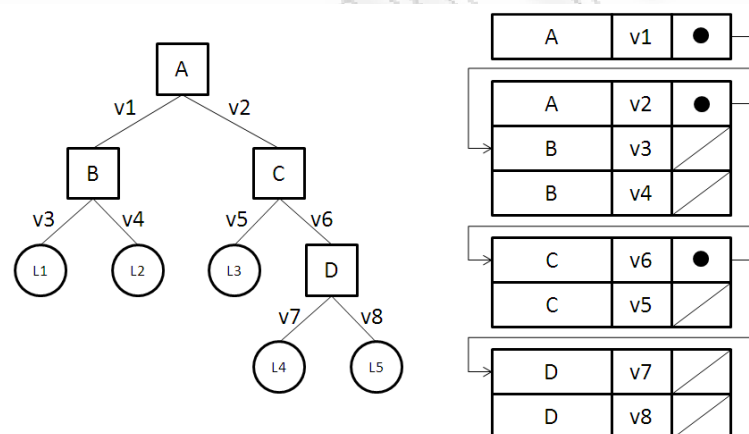
Ακολουθώντας, λοιπόν, τη λογική των συνδεδεμένων λιστών, αποφασίσαμε κάθε διαδρομή από τη ρίζα του δένδρου μέχρι κάποιο φύλλο του να την αναπαραστήσουμε ως μια συνδεδεμένη λίστα.

<sup>1</sup> Στο [16] περιγράφεται η διαδικασία απεικόνισης κανόνων συσχέτισης με τρεις πίνακες μιας σχεσιακής βάσης δεδομένων.



Εικόνα 31: Απεικόνιση του δένδρου με τη βοήθεια συνδεδεμένων λιστών

Όπως φαίνεται και στην εικόνα, οι δύο πρώτες συνδεδεμένες λίστες διαφοροποιούνται μονάχα στις τιμές των μεταβάσεων και στο φύλλο στο οποίο καταλήγουν. Το ίδιο ισχύει και για τις δύο τελευταίες λίστες. Έχοντας αυτό υπόψη μας, εύκολα μπορούμε να συμπεράνουμε πως για τη διαφοροποίηση δύο κόμβων της λίστας μας δεν αρκεί απλά η τιμή του κόμβου αυτή καθαυτή αλλά απαιτείται και η τιμή της σύνδεσης του δένδρου που ξεκινά από το κόμβο αυτό. Με τη σκέψη αυτή ως αφηρητία και έχοντας κατά νου το που επιθυμούμε να καταλήξουμε (σε πίνακες μιας σχεσιακής βάσης δεδομένων), το δένδρο απόφασης μπορεί να απεικονιστεί με την δομή της επόμενης εικόνας.



Εικόνα 32: Η δομή αποθήκευσης που χρησιμοποιήθηκε

Κάθε «γραμμή» της πιο πάνω δομής διαφοροποιείται με βάση το κόμβο του δένδρου και την τιμή της μετάβασης (σύνδεσης) του δένδρου που ξεκινάει από τον κόμβο αυτό.

Ο πίνακας στον οποίο θα απεικονιστεί ένα δένδρο απόφασης θα πρέπει να έχει τα ακόλουθα χαρακτηριστικά:

- Κάθε εγγραφή του πίνακα πρέπει να συνδέεται με ακριβώς μια σύνδεση του δένδρου (ή μια μετάβαση από τις λίστες μας).
- Κάθε εγγραφή του πίνακα αυτού πρέπει να περιέχει ως πληροφορία:
  - Το κόμβο του δένδρου από τον οποίο η μετάβαση ξεκινά, και
  - Το κόμβο ή το φύλλο του δένδρου στον οποίο η μετάβαση καταλήγει.

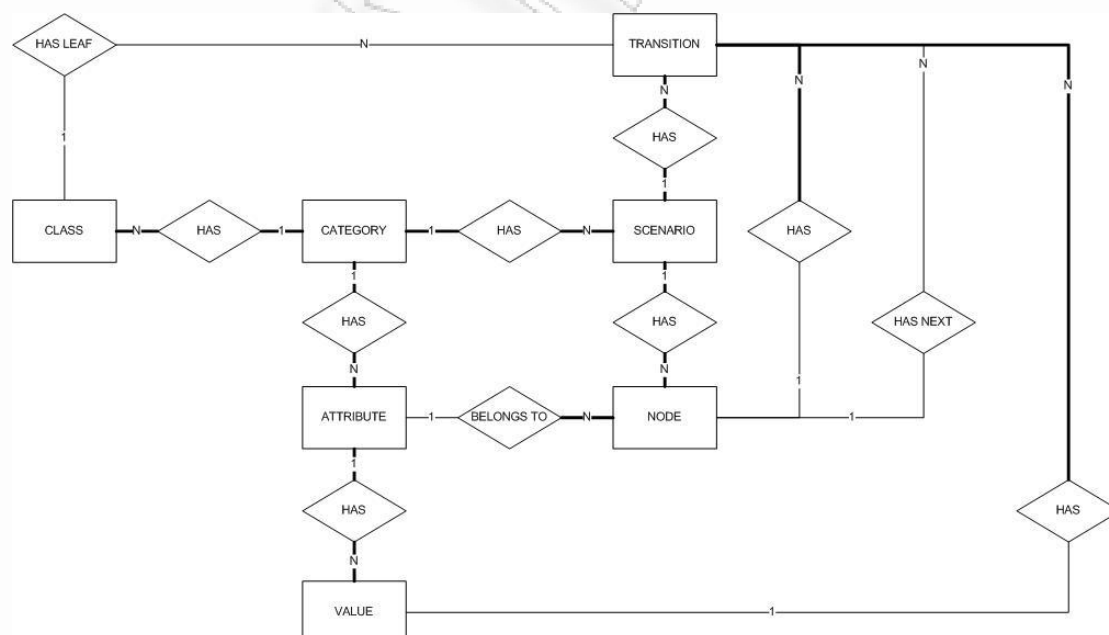
Στην επόμενη ενότητα θα παρουσιάσουμε το σύνολο των πινάκων που χρησιμοποιήσαμε για την αποδοτικότερη απεικόνιση ενός δένδρου απόφασης.

#### 4.2.2.2 Το σχήμα απεικόνισης

Για την πλήρη απεικόνιση του δένδρου απόφασης σε μια σχεσιακή βάση δεδομένων χρησιμοποιήσαμε 7 πίνακες:

- Έναν πίνακα *κατηγοριών* (CATEGORY table), στον οποίο αποθηκεύεται η βασική πληροφορία για όλα τα δένδρα απόφασης που επιθυμούμε να απεικονίσουμε στο σχήμα μας. Κάθε εγγραφή του πίνακα αυτού συνδέεται με ακριβώς ένα δένδρο απόφασης.
- Έναν πίνακα *σεναρίων* (SCENARIO table), στον οποίο αποθηκεύεται κάθε δυνατό αποτέλεσμα του αλγορίθμου κατηγοριοποίησης. Για παράδειγμα, δύο δένδρα απόφασης που έχουν προκύψει σε δύο διαφορετικές χρονικές στιγμές από διαφορετικά σύνολα δεδομένων του ίδιου πεδίου (της ίδιας κατηγορίας) θα απεικονιστούν στον πίνακα αυτό ως δύο ξεχωριστές εγγραφές.
- Έναν πίνακα *κλάσεων* (CLASS table), στον οποίο αποθηκεύονται οι κλάσεις (ομάδες) της διαδικασίας κατηγοριοποίησης.
- Έναν πίνακα *χαρακτηριστικών γνωρισμάτων* (ATTRIBUTE table), στον οποίο αποθηκεύονται τα γνωρίσματα του συνόλου δεδομένων πάνω στα οποία εκτελέστηκε ένας αλγόριθμος κατηγοριοποίησης.
- Έναν πίνακα *τιμών* (VALUE table), στον οποίο αποθηκεύονται οι διακριτές τιμές των χαρακτηριστικών γνωρισμάτων.
- Έναν πίνακα *κόμβων* (NODE table), στον οποίο αποθηκεύονται οι κόμβοι του δένδρου απόφασης.
- Έναν πίνακα *μεταβάσεων* (TRANSITION table), στον οποίο αποθηκεύονται όλες οι μεταβάσεις του δένδρου απόφασης, με βάση την ιδέα που αναπτύχθηκε στην προηγούμενη ενότητα.

Το μοντέλο οντοτήτων συσχετίσεων του παραπάνω σχήματος απεικονίζεται στην επόμενη εικόνα.



Εικόνα 33: Το μοντέλο Οντοτήτων-Συσχετίσεων

Σε κάθε έναν από του προαναφερθέντες πίνακες διατηρείται η ακόλουθη πληροφορία:

- **CATEGORY**
  - Category Id, το id της κατηγορίας.
  - Name, το όνομα της κατηγορίας.
  - Dataset, το σύνολο δεδομένων στο οποίο αναφέρεται η κατηγορία.
  - Description, η περιγραφή της κατηγορίας.
- **SCENARIO**
  - Scenario Id, το id του σεναρίου.
  - Name, το όνομα του σεναρίου.
  - Active, flag που υποδηλώνει το ενεργό σενάριο για μια κατηγορία (κάθε χρονική στιγμή ένα σενάριο μπορεί να είναι ενεργό για μία κατηγορία).
  - Comment, σχόλια που αφορούν το σενάριο.
- **CLASS**
  - Class Id, το id της κλάσης.
  - Name, το όνομα της κλάσης.
  - Code, ο κωδικός της κλάσης.
  - Description, η περιγραφή της κλάσης.
- **ATTRIBUTE**
  - Attribute Id, το id του γνωρίσματος
  - Name, το όνομα του γνωρίσματος.
  - Description, η περιγραφή του γνωρίσματος.
- **VALUE**
  - Value Id, το id της τιμής του γνωρίσματος.
  - Value, η τιμή του γνωρίσματος.
  - Active, flag που υποδηλώνει εάν κάποια τιμή του γνωρίσματος είναι ενεργή (διαθέσιμη) ή όχι.
  - Description, η περιγραφή της τιμής ενός γνωρίσματος.
- **NODE**
  - Node Id, το id του κόμβου.
  - Name, το όνομα του κόμβου του δένδρου απόφασης.
  - Attribute Id, το id του γνωρίσματος με τον οποίο συνδέεται ένας κόμβος του δένδρου.
  - Node, flag που υποδηλώνει τον κόμβο – ρίζα του δένδρου.
  - Description, η περιγραφή του κόμβου.
- **TRANSITION**
  - Node Id, το id του κόμβου του δένδρου από τον οποίο ξεκινά μια μετάβαση.
  - Value Id, το id της τιμής της μετάβασης.
  - Next Node Id, το id του κόμβου του δένδρου στον οποίο ενδεχομένως να καταλήγει μια μετάβαση.
  - Class Id, το id της κλάσης (ομάδας) στην οποία ενδεχομένως να καταλήγει μια μετάβαση.
  - Description, η περιγραφή της μετάβασης.

Στη συνέχεια περιγράφονται οι συσχετίσεις μεταξύ των οντοτήτων όπως παρουσιάζονται στο διάγραμμα οντοτήτων – συσχετίσεων.



Κάθε κατηγορία περιέχει τουλάχιστον ένα σενάριο.



#### 4.2.3 Το δένδρο απόφασης μέσα στο Oracle Siebel 8.1 CRM

Για την δημιουργία του παραπάνω σχήματος στη βάση δεδομένων της CRM εφαρμογής μας, χρησιμοποιήσαμε τα "Siebel Tools". Μέσω του εργαλείου αυτού δημιουργήσαμε τους πίνακες, τις σθόνες και γενικά όλα τα αντικείμενα ούτως ώστε ένας χρήστης να μπορεί μέσα από την ίδια την εφαρμογή να καταχωρήσει, να τροποποιήσει και γενικά να διαχειριστεί κάθε δένδρο απόφασης.

Στις επόμενες σελίδες θα δούμε πώς το δένδρο απόφασης που παρήγαγε το WEKA σε προηγούμενη ενότητα καταχωρήθηκε μέσα στο Oracle Siebel 8.1 CRM.

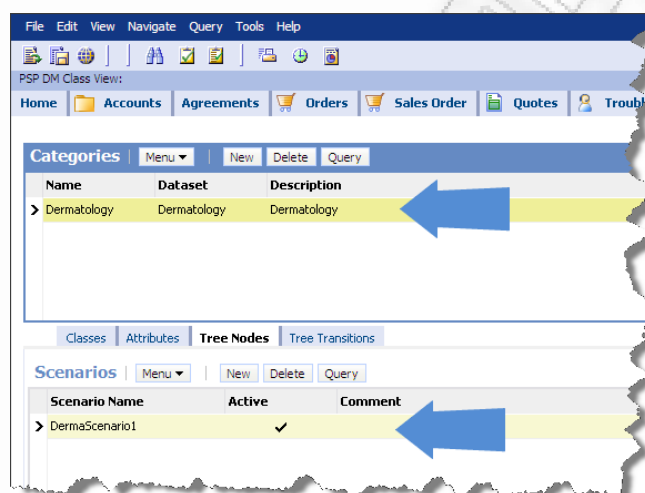
Συγκεντρώνοντας όλα τα στοιχεία που αφορούν το σύνολο δεδομένων μας αλλά και το δένδρο απόφασης που δημιουργήθηκε από αυτό έχουμε τα ακόλουθα δεδομένα:

- Εφαρμόσαμε τη μέθοδο της κατηγοριοποίησης σε ένα σύνολο δεδομένων.
- Δημιουργήσαμε ένα δένδρο απόφασης (μέσω του εργαλείου WEKA).

- Το σύνολο δεδομένων μας περιείχε 12 χαρακτηριστικά γνωρίσματα βάσει των οποίων πραγματοποιήσαμε την κατηγοριοποίηση.
- Τα 11 από τα 12 αυτά γνωρίσματα είχαν πεδίο τιμών το σύνολο {0, 1, 2, 3}.
- Ένα από τα 12 γνωρίσματα είχε συνεχές πεδίο τιμών (πρόκειται για το γνώρισμα της ηλικίας).
- Οι ομάδες στις οποίες κατατάχθηκαν τα δεδομένα του συνόλου μας ήταν 6.
- Το δένδρο απόφασης που δημιουργήθηκε περιείχε 10 κόμβους.
- Οι 10 κόμβοι του δένδρου συνδέονταν με 8 διαφορετικά χαρακτηριστικά γνωρίσματα.

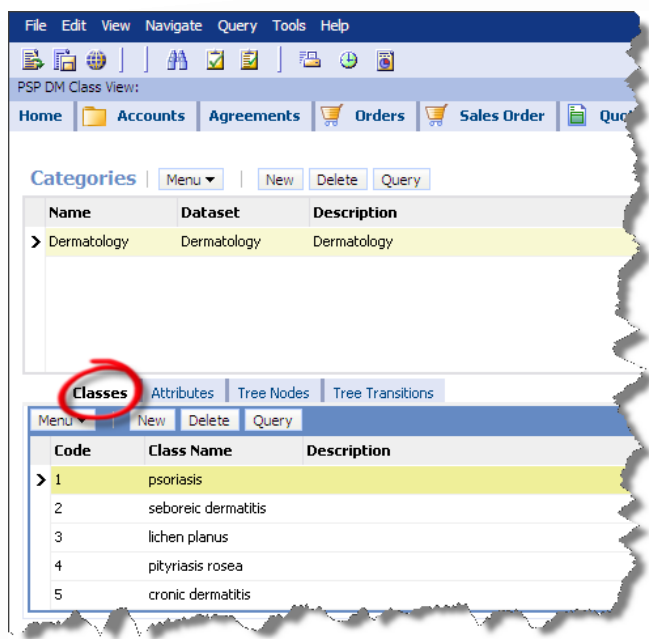
Έχοντας όλα τα παραπάνω υπόψη μας, στο Oracle Siebel 8.1 CRM καταχωρήθηκαν τα ακόλουθα δεδομένα:

- Δημιουργήθηκε 1 κατηγορία για το πρόβλημα της κατηγοριοποίησης και 1 ενεργό σενάριο για το δένδρο απόφασης που παρήγαγε το WEKA.



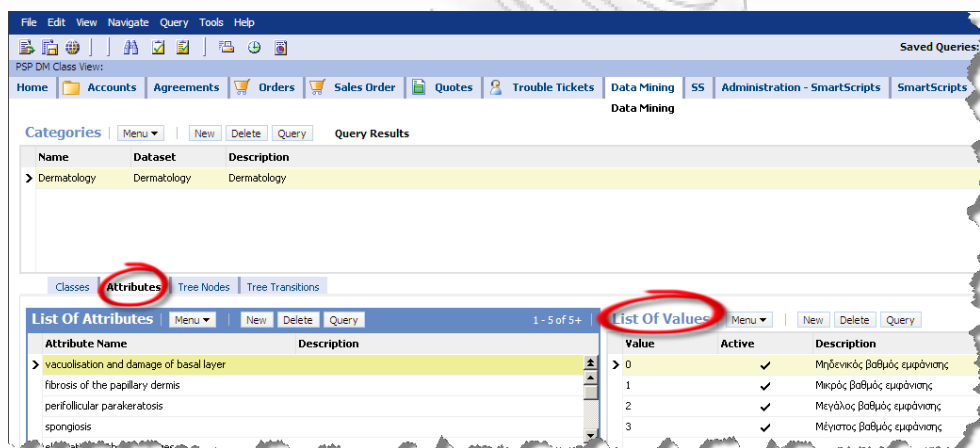
Εικόνα 34: Η κατηγορία και το ενεργό σενάριο του προβλήματος

- Δημιουργήθηκαν 6 κλάσεις για τις ομάδες της κατηγοριοποίησης.



Εικόνα 35: Οι κλάσεις του προβλήματος

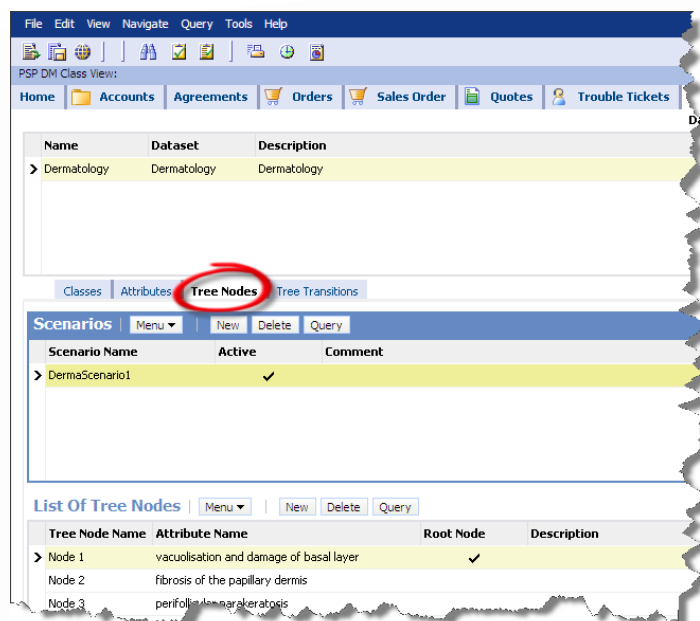
- Δημιουργήθηκαν 12 γνώρισμα με λίστες τιμών για τα 11 από αυτά.



Εικόνα 36: Τα χαρακτηριστικά γνώρισμα και οι τιμές τους

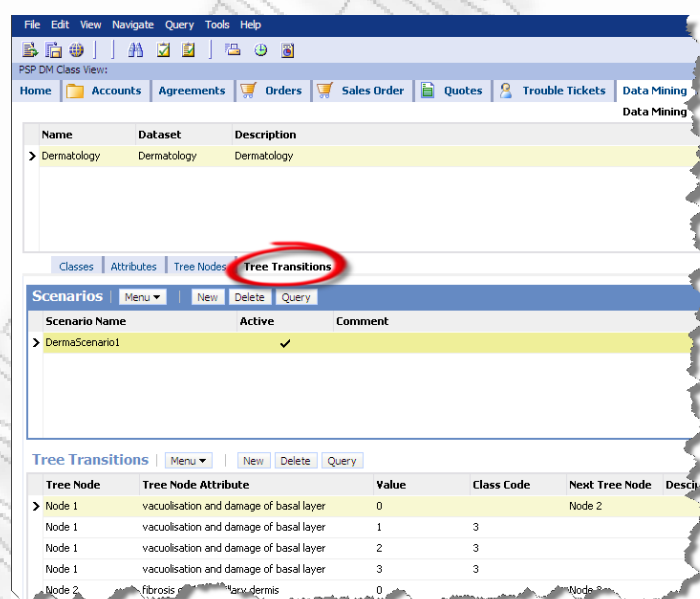
- Δημιουργήθηκαν 10 κόμβοι για το δένδρο απόφασης κάθε ένας από τους οποίους συνδέθηκε με ένα χαρακτηριστικό γνώρισμα του συνόλου δεδομένων μας. Επιπλέον ένας από τους κόμβους αυτούς ορίστηκε ως κόμβος – ρίζα του δένδρου.





Εικόνα 37: Οι κόμβοι του δένδρου απόφασης

- Δημιουργήθηκαν 40 εγγραφές στον πίνακα των μεταβάσεων όσες και οι ακμές του δένδρου απόφασης. Κάθε μια από τις εγγραφές αυτές συνδέεται με έναν κόμβο αφετηρία και μπορεί να καταλήγει σε έναν κόμβο προορισμό ή σε ένα φύλλο του δένδρου. Επιπλέον συνδέεται με μια τιμή του γνωρίσματος που χαρακτηρίζει τον κόμβο αφετηρία.



Εικόνα 38: Οι μεταβάσεις του δένδρου απόφασης

Στην ενότητα που ακολουθεί θα έρθουμε σε επαφή με τα “Siebel Smartscripts” και θα δούμε πώς με τη βοήθειά τους θα καταφέρουμε να εκμεταλλευτούμε τη γνώση του δένδρου απόφασης που καταχωρήσαμε στο Oracle Siebel 8.1 CRM.

### 4.3 Siebel Smartscripts

Στην προηγούμενη ενότητα παρουσιάσαμε έναν τρόπο αποθήκευσης του δένδρου απόφασης

που δημιουργήσαμε με το WEKA σε μια σχεσιακή βάση δεδομένων όπως αυτή του Oracle Siebel 8.1 CRM. Σκοπός της μεταφοράς του δένδρου στη βάση ήταν η εκμετάλλευση της γνώσης που περιέχει μέσα από την ίδια την CRM εφαρμογή.

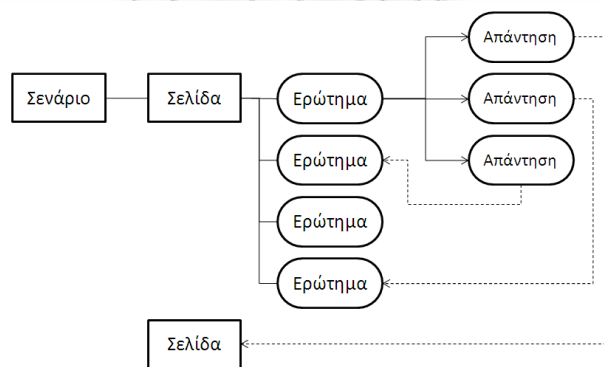
Στην ενότητα αυτή θα γνωρίσουμε τα Siebel Smartscripsts, ένα ιδιαίτερα εύχρηστο εργαλείο που παρέχει το Oracle Siebel 8.1 CRM και θα δούμε τον τρόπο με τον οποίο το module αυτό μπορεί να χρησιμοποιήσει τη γνώση του δένδρου απόφασης.

### 4.3.1 Τι είναι τα Siebel Smartscripsts

Τα Siebel Smartscripsts είναι ένα εργαλείο που παρέχει η Oracle σαν μέρος του Oracle Siebel 8.1 CRM, με την βοήθεια του οποίου μπορούν να σχεδιαστούν επιχειρησιακές διαδικασίες που περιλαμβάνουν αλληλεπιδράσεις μιας επιχείρησης με τους πελάτες της και οι οποίες καθορίζουν την ροή της εκάστοτε διαδικασίας.

Οι αλληλεπιδράσεις αυτές απεικονίζονται με την βοήθεια μίας βασικής οντότητας των «Ερωτημάτων»: ο χρήστης της εφαρμογής απευθύνει ερωτήματα στον πελάτη της επιχείρησης, οι απαντήσεις του οποίου κατευθύνουν την ροή μιας διαδικασίας δημιουργώντας νέες ερωτήσεις ή καταλήγοντας σε συγκεκριμένες καταστάσεις. Οι καταστάσεις αυτές μπορεί να είναι απλά συμπεράσματα για το προφίλ του πελάτη, δημιουργία προσφορών για συγκεκριμένα προϊόντα ή υπηρεσίες μιας επιχείρησης κα.

Τα «Ερωτήματα» περιλαμβάνουν διάφορα γνωρίσματα, όπως τις απαντήσεις τους, αλλά και δεδομένα που σχετίζονται με τον τρόπο που αυτά εμφανίζονται στην οθόνη των χρηστών. Τα «Ερωτήματα» συνδέονται μεταξύ τους με την βοήθεια «Κλαδιών» (branches) ενώ παράλληλα μπορούν να ομαδοποιούνται σχηματίζοντας ομάδες ερωτημάτων, γνωστές ως «Σελίδες» (pages). Οι «Σελίδες» με τη σειρά τους ομαδοποιούνται συγκροτώντας τα «Σενάρια» (scripts).



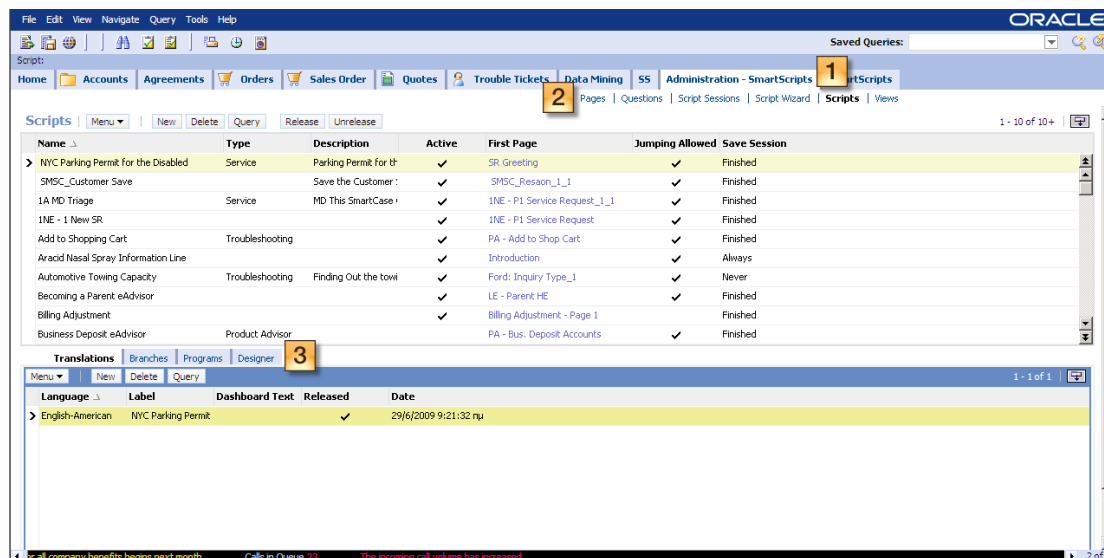
Εικόνα 39: Η ιεραρχική σχέση των δομών των Siebel Smartscripsts ([10])

Εκτός των προαναφερθέντων συστατικών στοιχείων, τα Siebel Smartscripsts διαθέτουν και τους κατάλληλους «Διαχειριστές Γεγονότων» (event handlers) με την βοήθεια των οποίων ο διαχειριστής του συστήματος είναι σε θέση να ελέγξει συγκεκριμένα γεγονότα με χρήση δύο γλωσσών προγραμματισμού, της eScript και της Visual Basic.

### 4.3.2 Το περιβάλλον των Siebel Smartscripsts

Η διαχείριση των Siebel Smartscripsts εκτελείται από το διαχειριστή του συστήματος CRM και πραγματοποιείται μέσα από την ίδια την εφαρμογή μέσω ειδικών διαχειριστικών οθονών (administration views) χωρίς να απαιτείται χρήση των Siebel Tools ή οποιουδήποτε άλλου εξωτερικού εργαλείου.

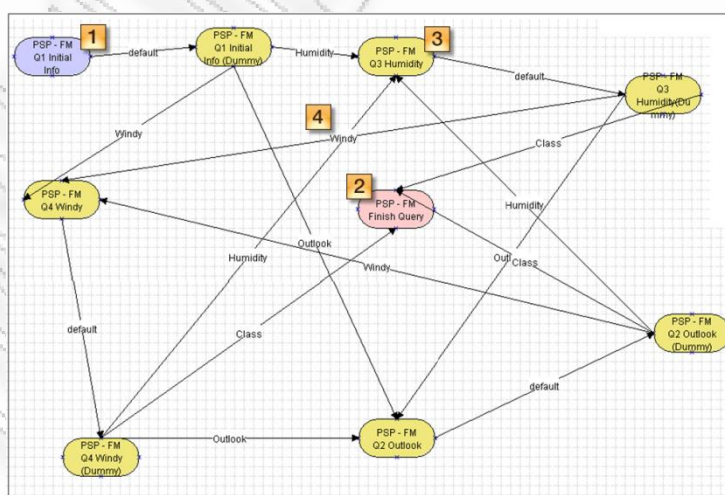
Στην επόμενη εικόνα παρουσιάζεται ενδεικτικά μια από τις οθόνες που παρέχει το Oracle Siebel 8.1 CRM για την διαχείριση των Smartscripts.



Εικόνα 40: Η διαχειριστική οθόνη των Siebel Smartscripts

- 1 Η οθόνη διαχείρισης των Siebel Smartscripts.
- 2 Τα συστατικά στοιχεία των Siebel Smartscripts (Ερωτήματα, Σελίδες, Σενάρια κλπ).
- 3 Τα στοιχεία παραμετροποίησης των συστατικών στοιχείων (Μεταφράσεις, Κλάδοι κλπ).

Για την οργάνωση και διαχείριση των μεταβάσεων από «Ερώτημα» σε «Ερώτημα» το Oracle Siebel 8.1 CRM παρέχει ένα γραφικό περιβάλλον μέσω του οποίου ο διαχειριστής του συστήματος είναι σε θέση να καθορίσει όλες τις επιθυμητές μεταβάσεις ανάλογα πάντα με την επιθυμητή λειτουργικότητα. Στην επόμενη εικόνα παρουσιάζεται το γραφικό αυτό περιβάλλον.



Εικόνα 41: Το γραφικό περιβάλλον για τον καθορισμό των μεταβάσεων

Στο συγκεκριμένο παράδειγμα, παρατηρούμε τη μοναδική «Σελίδα» ενός Siebel Smartscript αποτελούμενη από 9 «Ερωτήματα».

- 1 Πρόκειται για το «Ερώτημα» με το οποίο ξεκινά η συγκεκριμένη «Σελίδα» και επομένως και το Siebel Smartsript (εφόσον πρόκειται για «Σενάριο» μίας «Σελίδας»).
- 2 Πρόκειται για το «Ερώτημα» με το οποίο ολοκληρώνεται η «Σελίδα» και το Siebel Smartsript. Όπως φαίνεται και στην εικόνα, κανένας «Κλάδος» δε ξεκινά από το «Ερώτημα» αυτό.
- 3 Πρόκειται για ένα από τα ενδιάμεσα «Ερωτήματα» της «Σελίδας» και του Siebel Smartsript γενικότερα. Με τον όρο «ενδιάμεσα» εννοούμε όλα τα «Ερωτήματα» από τα οποία μπορείς να μεταβείς σε κάποιο άλλο «Ερώτημα».
- 4 Πρόκειται για τον κλάδο με την ετικέτα "Windy". Ο κλάδος αυτός καθορίζει την μετάβαση από ένα «Ερώτημα» (της αρχής του κλάδου) σε ένα άλλο (αυτό του τέλους) και εκτελείται όταν η «Απάντηση» που θα καταχωρήσει ο χρήστης του συστήματος στο «Ερώτημα» αφετηρίας έχει την τιμή "Windy".

Ολοκληρώνοντας την παραμετροποίηση ενός Siebel Smartsript, ο διαχειριστής του συστήματος μπορεί να το «δημοσιεύσει» (release) δίνοντας τη δυνατότητα στους χρήστες της εφαρμογής να το χρησιμοποιήσουν. Στην επόμενη εικόνα παρουσιάζεται μια ενδεικτική οθόνη Smartsript όπως την βλέπει ένας απλός χρήστης του συστήματος.

Εικόνα 42: Οθόνη Siebel Smartsript

Μέσα από την οθόνη αυτή, ο χρήστης μπορεί:

- να δει τη σελίδα «Πρόβλεψη διεξαγωγής ποδοσφαιρικού αγώνα»,
- να διαβάσει το ερώτημα «Ποιά η τιμή του χαρακτηριστικού Outlook;»,
- να επιλέξει ανάμεσα σε τρεις προκαθορισμένες απαντήσεις "Sunny", "Overcast" ή "Rainy" και
- να πατήσει το κουμπί "Next" ώστε να προχωρήσει στο επόμενο ερώτημα. Η μετάβαση αυτή εξαρτάται από την απάντηση που έχει καταχωρηθεί και καθορίζεται από την παραμετροποίηση που έχει εισάγει ο διαχειριστής του συστήματος για το συγκεκριμένο Siebel Smartsript.

Στην επόμενη ενότητα θα δούμε με ποιο τρόπο, μέσω των Siebel Smartscripts, καταφέραμε να εκμεταλλευτούμε την γνώση του δένδρου απόφασης που δημιουργήσαμε με τη βοήθεια του εργαλείου WEKA.

### 4.3.3 Siebel Smartsript και δένδρο απόφασης

Στην προηγούμενη ενότητα είδαμε πως δύο είναι τα βασικά αντικείμενα των Siebel Smartscripts, τα «Ερωτήματα» και οι «Απαντήσεις». Μέσω των «Ερωτημάτων» ο χρήστης του συστήματος μπορεί να αντλήσει συγκεκριμένη πληροφορία από τους πελάτες μιας επιχείρησης, οι οποίοι με τη σειρά τους, μέσω των «Απαντήσεων» κατευθύνουν (χωρίς να το γνωρίζουν) τη ροή (την μετάβαση) των «Ερωτημάτων» με σκοπό την εξαγωγή χρήσιμων

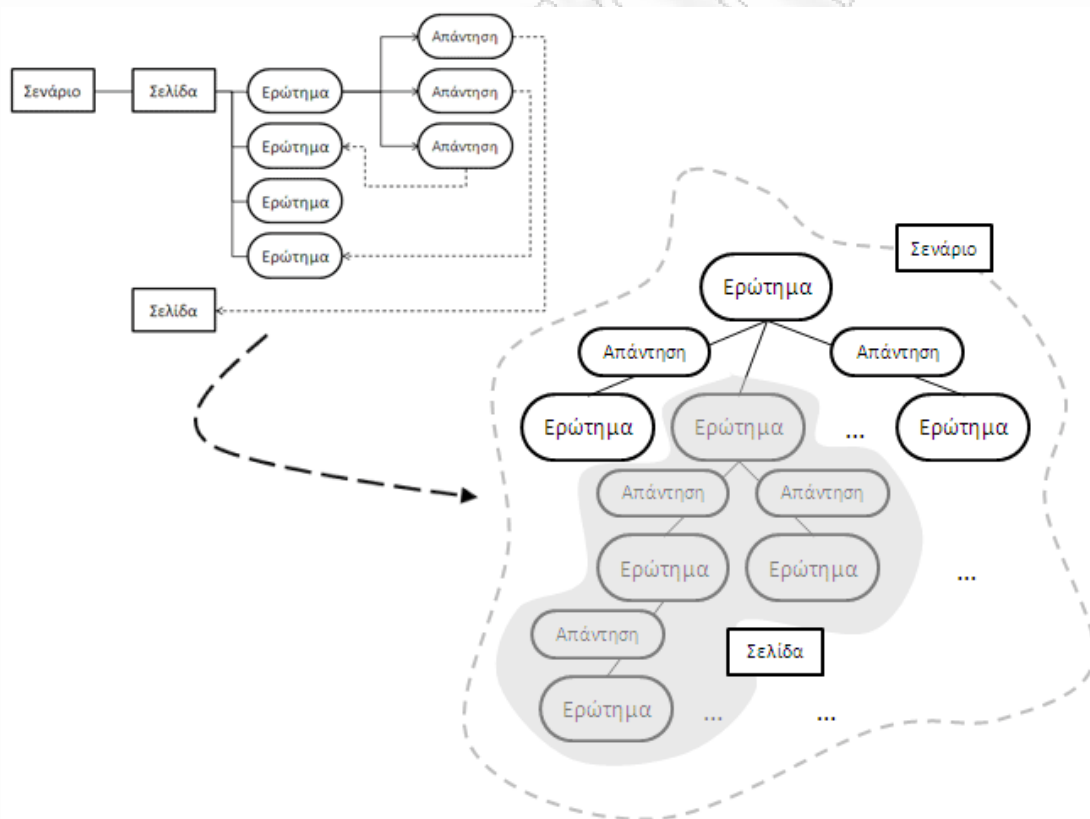
συμπερασμάτων, τη δημιουργία προσφορών κα. Θα μπορούσαμε να φανταστούμε επομένως τα «Ερωτήματα» σαν κόμβους ή/και φύλλα ενός δένδρου ενώ τις «Απαντήσεις» ως τις μεταβάσεις του.

Με την παραδοχή αυτή, ένα Siebel Smartsript θα μπορούσε να αναπαρασταθεί με ένα δένδρο απόφασης ως εξής:

- Κάθε κόμβος ή φύλλο του δένδρου να αντιστοιχεί σε ένα «Ερώτημα», και
- Κάθε μετάβαση του δένδρου να αναπαριστά μια «Απάντηση».

Η διαφοροποίηση των «Ερωτημάτων» που αντιστοιχούν σε εσωτερικούς κόμβους και αυτών που σχετίζονται με τα φύλλα του δένδρου έγκειται στο γεγονός ότι τα πρώτα αφορούν «Ερωτήματα» από τα οποία ξεκινούν μεταβάσεις (προς άλλα «Ερωτήματα») ενώ τα δεύτερα σχετίζονται με τερματικές καταστάσεις. Πρόκειται δηλαδή για καταληκτικά «Ερωτήματα» στα οποία ολοκληρώνεται ένα Siebel Smartsript χωρίς να υπάρχει καμία μετάβαση που να ξεκινά από αυτά.

Για την ολοκλήρωση της παραπάνω αναπαράστασης θα μπορούσαμε επιπλέον να θεωρήσουμε κάθε υποδένδρο του αρχικού μας δένδρου ως μία «Σελίδα» του Siebel Smartsript και ολόκληρο το δένδρο ως το «Σενάριο» αυτού.



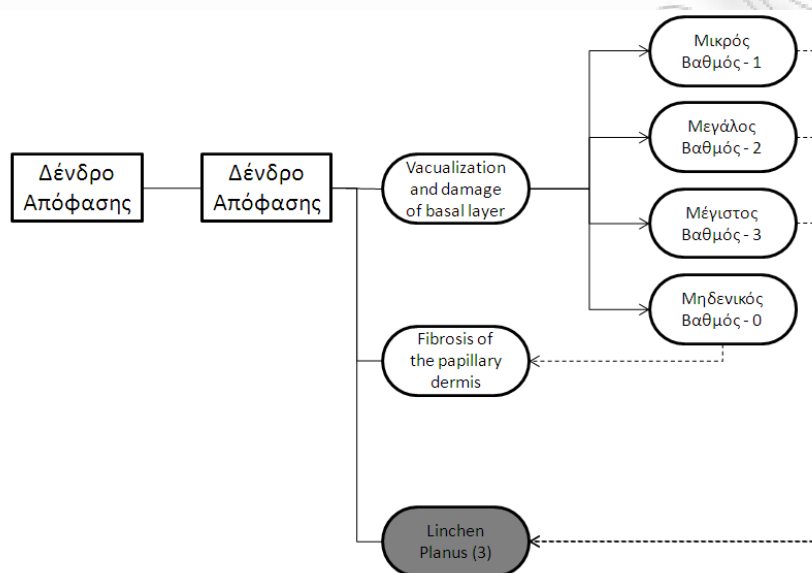
Εικόνα 43: Απεικόνιση ενός Siebel Smartsript ως Δένδρου Απόφασης

Έχοντας κατά νου την απεικόνιση αυτή, θα μπορούσαμε να σκεφτούμε πως το δένδρο απόφασης, που δημιουργήσαμε μέσω του εργαλείου WEKA εκτελώντας κατηγοριοποίηση ενός συνόλου δεδομένων, θα μπορούσε να απεικονιστεί ως ένα Siebel Smartsript το οποίο να διαθέτει:

- Ένα ακριβώς «Σενάριο».

- Τόσες «Σελίδες» (δυναμικά) όσα και τα υποδένδρα του (για απλούστευση της αναπαράστασης θα θεωρήσουμε μια ακριβώς «Σελίδα»).
- Ένα «Ερώτημα» για κάθε έναν από τους δέκα κόμβους του δένδρου.
- Μια «Απάντηση» για κάθε μία από τις σαράντα μεταβάσεις του δένδρου.
- Ένα καταληκτικό «Ερώτημα» για κάθε ένα από τα είκοσι εννέα φύλλα του δένδρου .

Ενδεικτικά, στην επόμενη εικόνα, παρουσιάζεται η απεικόνιση των δύο πρώτων επιπέδων του δένδρου απόφασης με την βοήθεια των Siebel Smartscripts.



**Εικόνα 44: Ενδεικτική απεικόνιση των δύο πρώτων επιπέδων του δένδρου απόφασης με τη βοήθεια των Siebel Smartscripts**

Η παραπάνω απεικόνιση «κρύβει» ωστόσο δύο σημαντικά μειονεκτήματα:

- Την επανάληψη της πληροφορίας, και
- Την μικρή ευελιξία του μοντέλου απεικόνισης.

Το πρώτο μειονέκτημα αφορά τα γνώρισμα που εμφανίζονται σε περισσότερους από ένα κόμβους του δένδρου με πιο χαρακτηριστικό παράδειγμα, το γνώρισμα “koebner phenomenon” (συνολικά τρεις εμφανίσεις). Σύμφωνα με την απεικόνιση που περιγράψαμε προηγουμένως, για κάθε έναν από τους τρεις κόμβους του δένδρου στους οποίους εμφανίζεται το γνώρισμα αυτό, θα έπρεπε να αντιστοιχηθεί και από ένα «Ερώτημα» δημιουργώντας τελικά τρία πανομοιότυπα ερωτήματα σχετιζόμενα με το ίδιο γνώρισμα και τις ίδιες ακριβώς «Απαντήσεις».

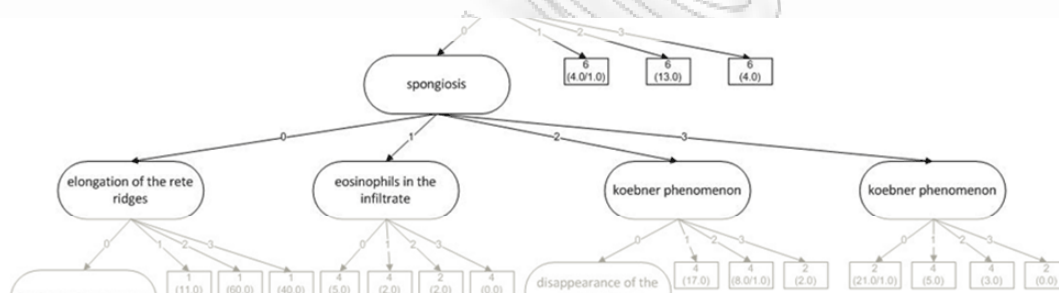
Το δεύτερο μειονέκτημα αφορά την ανάγκη συνεχούς τροποποίησης του Siebel Smartscript κάθε φορά που το δένδρο απόφασης αλλάζει. Ένα «Ερώτημα» ανά κόμβο δένδρου συνεπάγεται πως «Ερωτήματα» θα πρέπει να προσθαφαιρούνται στο Siebel Smartscript κάθε φορά που νέοι κόμβοι εισέρχονται ή παλιοί διαγράφονται από το δένδρο. Το τελευταίο, βασιζόμενο σε ένα συνεχώς επεκτεινόμενο σύνολο δεδομένων, μπορεί να τροποποιείται αρκετά συχνά δημιουργώντας την ανάγκη ενημέρωσης κάθε φορά του Siebel Smartscript. Ο μικρός βαθμός ευελιξίας του μοντέλου μας θα είχε άμεσες συνέπειες στην αποδοτικότητά του.

Για τους λόγους αυτούς, το μοντέλο απεικόνισης του δένδρου απόφασης με τα Siebel Smartscripts θα έπρεπε να έχει μικρό βαθμό ευαισθησίας στις μεταβολές του δένδρου και επιπλέον, να είναι αρκετά «έξυπνο» ώστε να μην απαιτείται η επανάληψη της πληροφορίας για την ορθή λειτουργία του.

Για να ξεπεράσουμε τα δύο παραπάνω μειονεκτήματα σκεφτήκαμε να απεικονίσουμε το δένδρο απόφασης ως ένα Siebel Smartsript με τα εξής χαρακτηριστικά:

- Κάθε διακριτό γνώρισμα του συνόλου δεδομένων μας το απεικονίσαμε χρησιμοποιώντας δύο «Ερωτήματα». Δεν περιοριστήκαμε μονάχα στα γνωρίσματα του δένδρου απόφασης αλλά συμπεριλάβαμε όλα τα γνωρίσματα του συνόλου δεδομένων μας (συνολικά 12). Από τα δύο αυτά «Ερωτήματα», το πρώτο θα χρησιμοποιείται για την καταχώρηση των προτιμήσεων ενός πελάτη ενώ το δεύτερο θα χρησιμεύει στον «υπολογισμό» της μετάβασης προς τον επόμενο κόμβου του δένδρου βάσει των δεδομένων που είναι καταχωρημένα στο σχήμα της βάσης του Oracle Siebel 8.1 CRM.
- Για όλα τα φύλλα του δένδρου μας δημιουργήσαμε μονάχα ένα καταληκτικό «Ερώτημα» που εκφράζει το αποτέλεσμα της κατηγοριοποίησης.
- Με την βοήθεια των «Απαντήσεων» των Siebel Smartscripts δεν απεικονίσαμε μονάχα τις εναλλακτικές τιμές των χαρακτηριστικών γνωρισμάτων του συνόλου μας (πχ «0 – μικρός βαθμός εμφάνισης» κλπ) αλλά και τις μεταβάσεις των ίδιων των κόμβων του δένδρου (πχ ο «κόμβος 1» στον οποίο συνδέεται το «Γνώρισμα 1» καταλήγει στο «κόμβο 2» όταν η «Απάντηση» για το «Ερώτημα» του γνωρίσματος είναι «0 – μικρός βαθμός εμφάνισης» κκ).

Για να γίνει ευκολότερα κατανοητή η παραπάνω απεικόνιση θα παρουσιάσουμε ένα παράδειγμα για κάποιον ενδεικτικό κόμβο του δένδρου απόφασης.



Εικόνα 45: Παράδειγμα απεικόνισης του κόμβου “spongiosis”

Στην παραπάνω εικόνα βλέπουμε τον κόμβο που αφορά το γνώρισμα της “spongiosis” και έχει καταχωρηθεί ως “Node 4” στο σχήμα της βάσης μας. Κατά παρόμοιο τρόπο και βάσει της αποθήκευσης του δένδρου που περιγράψαμε στην προηγούμενη ενότητα βλέπουμε επίσης τους κόμβους:

- Node 5, που σχετίζεται με το γνώρισμα “elongation of the rete ridges”
- Node 6, που σχετίζεται με το γνώρισμα “eosinophils in the infiltrate”
- Node 7, που σχετίζεται με το γνώρισμα “koebner phenomenon”
- Node 8, που σχετίζεται με το γνώρισμα “koebner phenomenon”

Όπως αναφέραμε ήδη, για κάθε ένα από τα τέσσερα διακριτά γνωρίσματα (το “koebner phenomenon” εμφανίζεται δύο φορές) δημιουργήσαμε δύο «Ερωτήματα» Siebel Smartsript: το πρώτο για την καταχώρηση της πληροφορίας της τιμής ενός γνωρίσματος με βάση τις προτιμήσεις ενός πελάτη και το δεύτερο για τον υπολογισμό της μετάβασης στο δένδρο απόφασης.

Έτσι για το κόμβο “4” δημιουργήθηκαν τα εξής δύο «Ερωτήματα»:

- Το «Ερώτημα A» με τις εξής 4 πιθανές απαντήσεις (όλες αφορούν το βαθμό εμφάνισης του χαρακτηριστικού “spongiosis” ενώ η τελική απάντηση προσδιορίζεται από τον ίδιο τον πελάτη):

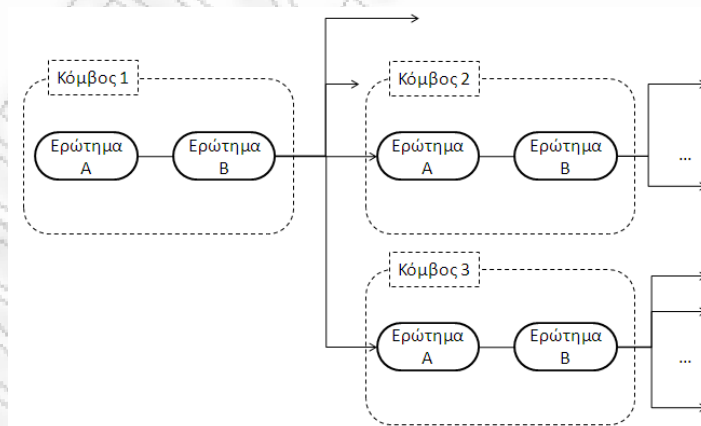
- 0 – μηδενικός βαθμός εμφάνισης
- 1 – μικρός βαθμός εμφάνισης
- 2 – μεγάλος βαθμός εμφάνισης
- 3 – μέγιστος βαθμός εμφάνισης
- Το «Ερώτημα Β» με τις εξής 13 πιθανές απαντήσεις (όλες αφορούν τα 12 γνωρίσματα που περιλαμβάνονται στο σύνολο δεδομένων μας καθώς και το γνώρισμα της κατηγοριοποίησης ενώ η τελικά απάντηση προσδιορίζεται από το ίδιο το σύστημα):
  - vacuolisation and damage of basal layer
  - fibrosis of the papillary dermis
  - perifollicular parakeratosis
  - ...
  - parakeratosis
  - Class

Το πρώτο «Ερώτημα» χρησιμοποιείται για την εισαγωγή της πληροφορίας που μας δίνει ο πελάτης στο σύστημα. Το δεύτερο «Ερώτημα» είναι αυτό που εκτελεί τις αναζητήσεις στον πίνακα μεταβάσεων του σχήματός μας (TRANSITION table) ούτως ώστε να αποφασίσει αν πρέπει να γίνει μετάβαση σε κάποιο από τα υπόλοιπα ζεύγη «Ερωτημάτων» που σχετίζονται με τα γνωρίσματα του συνόλου δεδομένων μας ή στο μοναδικό καταληκτικό «Ερώτημα» το οποίο και θα εξάγει το συμπέρασμα της κατηγοριοποίησης. Η μετάβαση από το «Ερώτημα Α» στο «Ερώτημα Β» είναι άμεση χωρίς να λαμβάνεται υπ' όψιν καμία συνθήκη.

Για κάθε ζεύγος «Ερωτημάτων» επομένως, ο κανόνας του μοντέλου μας περιγράφεται ως εξής:

*«Χρησιμοποίησε το πρώτο Ερώτημα για να αντλήσεις την πληροφορία από τον πελάτη και το δεύτερο Ερώτημα για να υπολογίσεις την μετάβαση βάσει της απάντησης του πρώτου Ερωτήματος»*

Ακριβώς επειδή το δεύτερο «Ερώτημα» δεν υποδηλώνει καμία αλληλεπίδραση μεταξύ συστήματος και πελάτη μπορούμε να το χαρακτηρίσουμε ως «εικονικό» (dummy).



**Εικόνα 46:** Ενδεικτική αναπαράσταση ενός δένδρου απόφασης από ένα Siebel Smartsript

Το δεύτερο «Ερώτημα» αν και «εικονικό» είναι αυτό που στην πραγματικότητα χρησιμοποιεί τους πίνακες αποθήκευσης του δένδρου απόφασης με σκοπό τον υπολογισμό της μετάβασης προς τον επόμενο κόμβο. Για να μπορέσουμε να πετύχουμε τη λειτουργικότητα αυτή εκμεταλλευτήκαμε τους «Διαχειριστές Γεγονότων» (event handlers) που παρέχουν τα Siebel Smartscripts.



Πιο συγκεκριμένα, στο Question\_PreLeave (που εκτελείται ακριβώς πριν ο χρήστης εγκαταλείψει ένα συγκεκριμένο εικονικό «Ερώτημα») εκτελέσαμε όλες τις απαραίτητες αναζητήσεις πάνω στους πίνακες δεδομένων μας με σκοπό τον υπολογισμό του επόμενου κόμβου μετάβασης (εσωτερικού ή φύλλου) ενώ στο Question\_PreBranch (με το οποίο μπορείς να προσδιορίσεις την «Απάντηση» σε ένα συγκεκριμένο «Ερώτημα») καθορίσαμε το όνομα του επόμενου ζεύγους «Ερωτημάτων» στο οποίο επιθυμούμε να μεταβούμε.

```
function Question_PreLeave ()
{
    var ssPage = Script().GetPage(gPage);
    var ssQuestion = ssPage.GetQuestion(gQuestion);
    var userinput = ssQuestion.GetCurrentValue();

    var catid = Script().GetParameter("CATEGORYID");
    var scenid = Script().GetParameter("SCENARIOID");

    var bo = TheApplication().GetBusObject("PSP Data Mining");
    var bc = bo.GetBusComp("PSP DM Scenario Detail");
    var answer = "";
    var nexttreenode = "";

    with(bc)
    {
        SetViewMode(AllView);
        ActivateField("Tree Node Attribute");
        ActivateField("Value");
        ActivateField("Category Id");
        ActivateField("Parent Id");
        ActivateField("Next Tree Node Attribute");
        ActivateField("Class Name");
        ActivateField("Tree Node");
        ActivateField("Next Tree Node");

        ClearToQuery();

        var mySearchSpec = "[Tree Node Attribute] = '" + attribute + "' AND
[Value] = '" + userinput + "' AND EXISTS([Category Id] = '" + catid + "') AND [Parent
Id] = '" + scenid + "' AND [Tree Node] = '" + treenode + "'";

        SetSearchExpr(mySearchSpec);

        ExecuteQuery();
        if (FirstRecord())
        {
            answer = GetFieldValue("Next Tree Node Attribute");
            nexttreenode = GetFieldValue("Next Tree Node");
            if (answer == "")
            {
                answer = "Class";
                classname = GetFieldValue("Class Name");
            }
        }
    }

    Script().SetUserParameter(QDUMMY, answer);
    Script().SetUserParameter("CLASS", classname);
    Script().SetUserParameter("TREENODE", nexttreenode );

    bc = null;
    bo = null;

    return (ContinueOperation);
}
```

**Εικόνα 47: Ο κώδικας στο event handler: Question\_PreLeave() των εικονικών «Ερωτημάτων»**

Η αναζήτηση του επόμενου κόμβου πραγματοποιείται εκτελώντας ένα ερώτημα στον πίνακα των μεταβάσεων στο οποίο καθορίζουμε τον τρέχοντα κόμβο στον οποίο βρισκόμαστε, το γνώρισμα του κόμβου και την τιμή που καταχώρησε ο πελάτης της επιχείρησης για το

γνώρισμα αυτό. Η αναζήτηση αυτή επιστρέφει είτε κάποιον εσωτερικό κόμβο του δένδρου μας με αποτέλεσμα να κινούμαστε προς το επόμενο ζεύγος «Ερωτημάτων» είτε κάποιο φύλλο του δένδρου με συνέπεια τη μετάβασή μας προς ένα από τα καταληκτικά «Ερωτήματα» του μοντέλου μας.

Με την παραπάνω απεικόνιση δημιουργήσαμε ένα αρκετά ευέλικτο μοντέλο Siebel Smartsript με δύο ιδιαίτερος σημαντικά χαρακτηριστικά:

- Δε χρειάζεται τροποποίηση του μοντέλου κάθε φορά που το δένδρο απόφασης για το σύνολο δεδομένων αλλάζει.
- Κάθε δένδρο απόφασης που βασίζεται στο σύνολο (ή και οποιοδήποτε υποσύνολο) των 13 χαρακτηριστικών γνωρισμάτων μπορεί να υποστηριχτεί από το μοντέλο αυτό.

Στην επόμενη ενότητα θα δούμε πώς όλα τα παραπάνω μπορούν να συνδυαστούν στα πλαίσια μιας επιχειρησιακής διαδικασίας.

#### 4.4 Υλοποίηση της επιχειρησιακής διαδικασίας

Στις τρεις προηγούμενες ενότητες παρουσιάσαμε τα εργαλεία με την βοήθεια των οποίων η εξόρυξη γνώσης από δεδομένα μπορεί να προσθέσει σημαντική αξία σε μία επιχειρησιακή διαδικασία.

Αρχικά γνωρίσαμε ένα εργαλείο εξόρυξης γνώσης, το WEKA, με την βοήθεια του οποίου εκτελέσαμε μια διαδικασία κατηγοριοποίησης πάνω σε ένα συγκεκριμένο σύνολο δεδομένων με στόχο την εξαγωγή ενός δένδρου απόφασης. Στην συνέχεια, παρουσιάσαμε ένα σχήμα αποθήκευσης του δένδρου στα πλαίσια της σχεσιακής βάσης δεδομένων του Oracle Siebel 8.1 CRM και τελικά γνωρίσαμε τα Siebel Smartscripts μέσω των οποίων προτείναμε μια απεικόνιση του δένδρου απόφασης με σκοπό την πλήρη εκμετάλλευση της γνώσης που περιέχει.

Στην ενότητα αυτή θα παρουσιάσουμε τις λειτουργικές προδιαγραφές μιας επιχειρησιακής διαδικασίας μέσω της οποίας τα τρία προαναφερθέντα εργαλεία θα συνεργαστούν αρμονικά με σκοπό την εκμετάλλευση της εξόρυξης γνώσης από δεδομένα.

##### 4.4.1 Εισαγωγή

Η επιχειρησιακή διαδικασία που θα παρουσιάσουμε αφορά το νοσηλευτικό κλάδο και σχετίζεται με τη διαδικασία δημιουργίας ιατρικών ραντεβού.

Ένα διαγνωστικό κέντρο που χρησιμοποιεί το Oracle Siebel 8.1 CRM σε όλα τα τμήματά του επιθυμεί να βελτιστοποιήσει τη διαδικασία «τηλεφωνικών ραντεβού» εκμεταλλευόμενο την γνώση που μπορεί να του παρέχουν δεδομένα (πχ συμπτώματα εμφάνισης δερματολογικών ασθενειών) που έχει καταχωρημένα στη βάση του.

Πιο συγκεκριμένα, διαθέτοντας μια πολύ μεγάλη βάση δεδομένων επιθυμεί να δημιουργήσει ένα μοντέλο κατηγοριοποίησης δερματολογικών συμπτωμάτων με σκοπό την προώθηση ιατρικών ραντεβού στις κατάλληλες ειδικότητες δερματολόγων που διαθέτει. Για το σκοπό αυτό, επιθυμεί οι χρήστες του τηλεφωνικού του κέντρου να είναι σε θέση μέσω κατάλληλων ερωτημάτων που απευθύνουν στους πελάτες – ασθενείς να προσδιορίσουν την ειδικότητα των ιατρών που απαιτούνται ανά περίπτωση.

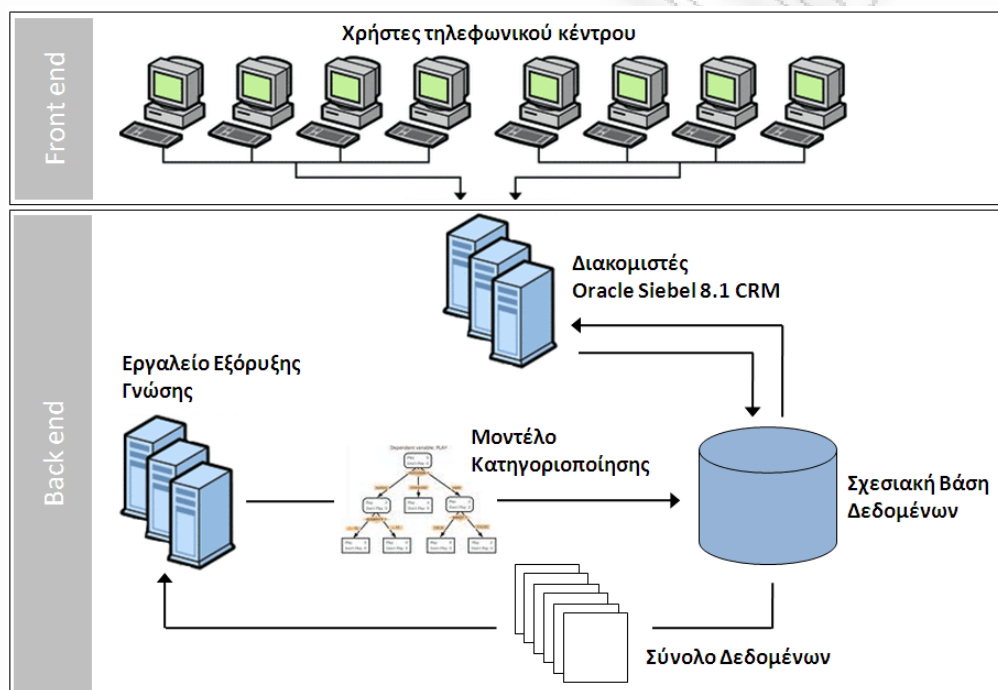
Τα ερωτήματα αυτά θα παράγονται αυτόματα από το σύστημα και, ανάλογα με τις απαντήσεις των ασθενών, θα κατευθύνουν τους χρήστες του τηλεφωνικού κέντρου στο να καταχωρήσουν τα σωστά ραντεβού ανά ιατρική ειδικότητα.

Τα μοντέλα κατηγοριοποίησης των δερματολογικών συμπτωμάτων βάσει των οποίων θα δημιουργούνται οι ερωτήσεις, θα ενημερώνονται σε τακτά χρονικά διαστήματα ούτως ώστε να περιλαμβάνουν κάθε φορά την μεγαλύτερη δυνατή πληροφορία.

Στόχος της διοίκησης του διαγνωστικού κέντρου είναι να περιοριστούν τα ραντεβού δερματολόγων με ασθενείς στις παθήσεις των οποίων δεν διαθέτουν την κατάλληλη εξειδίκευση.

Στην επόμενη εικόνα παρουσιάζεται μια γενική άποψη της επιχειρησιακής διαδικασίας. Όπως φαίνεται και από το σχήμα, η διαδικασία μπορεί να διαχωριστεί ανάλογα με το επίπεδο στο οποίο εκτελείται σε δύο βασικές συνιστώσες:

- Τη “Front End” διαδικασία που αφορά την αλληλεπίδραση του ασθενή με το διαγνωστικό κέντρο και σχετίζεται με την υποστήριξη των χρηστών του τηλεφωνικού κέντρου από το βασική εφαρμογή που χρησιμοποιούν, το Oracle Siebel 8.1 CRM.
- Τη “Back End” διαδικασία που περιλαμβάνει όλες τις λειτουργίες που εκτελούνται στο παρασκήνιο ούτως ώστε το “Front End” κομμάτι της διαδικασίας να είναι αποδοτικό.

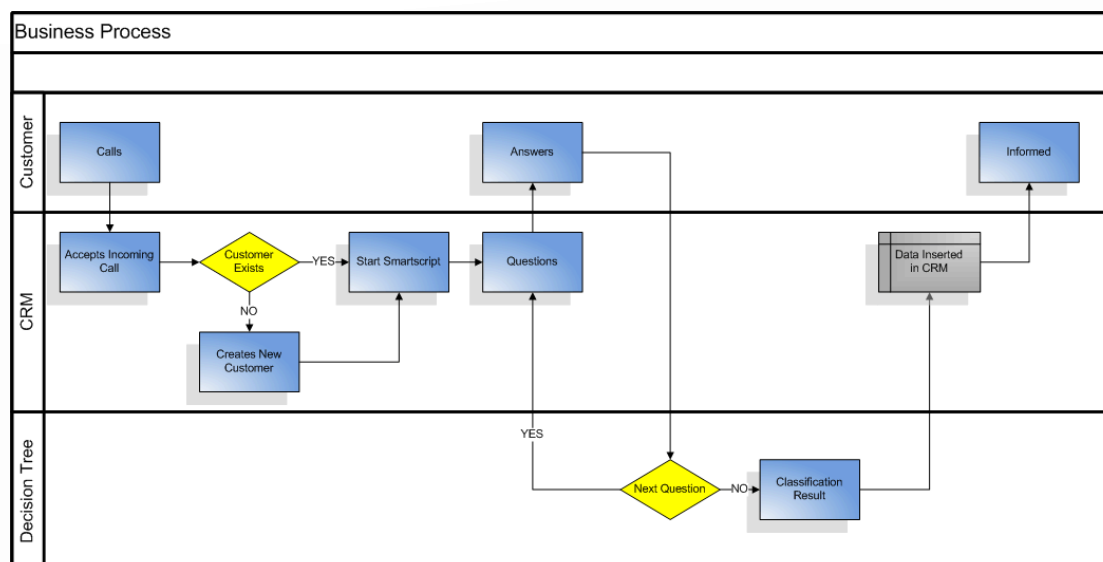


Εικόνα 48: Γενική άποψη της επιχειρησιακής διαδικασίας

Στις επόμενες παραγράφους θα περιγράψουμε τις δύο αυτές συνιστώσες της επιχειρησιακής διαδικασίας.

#### 4.4.2 “Front End” διαδικασία

Η “front end” διαδικασία αφορά τη λειτουργία του τμήματος του τηλεφωνικού κέντρου και περιγράφεται πλήρως από το ακόλουθο διάγραμμα ροής.



Εικόνα 49: Το διάγραμμα ροής της επιχειρησιακής διαδικασίας

Τα βήματα που περιγράφουν την παραπάνω διαδικασία είναι τα εξής:

- Ο πελάτης – ασθενής καλεί το τηλεφωνικό κέντρο του διαγνωστικού κέντρου.
- Ο υπάλληλος του τηλεφωνικού κέντρου λαμβάνει την κλήση και μέσω του Oracle Siebel 8.1 CRM αναζητά τα στοιχεία του πελάτη.
- Για την περίπτωση νέου πελάτη, ο υπάλληλος του τηλεφωνικού κέντρου καταχωρεί τα στοιχεία του ασθενή στην εφαρμογή CRM.
- Ο υπάλληλος του τηλεφωνικού κέντρου ξεκινά τη διαδικασία καταχώρησης νέου ραντεβού.
- Το σύστημα CRM συμβουλευόμενο το μοντέλο εξόρυξης γνώσης που διαθέτει (δένδρο απόφασης) προτείνει στο χρήστη του συστήματος συγκεκριμένες ερωτήσεις προς τον ασθενή.
- Ο ασθενής ξεκινά να απαντά στις ερωτήσεις.
- Κάθε απάντηση του ασθενή που καταχωρείται στο σύστημα CRM από τον υπάλληλο του τηλεφωνικού κέντρου καθορίζει και την επόμενη ερώτηση που χρειάζεται ενδεχομένως ο υπάλληλος να απευθύνει.
- Ο καθορισμός των ερωτήσεων από το σύστημα γίνεται με τρόπο αυτόματο και αφού προηγουμένως το σύστημα συμβουλευθεί το μοντέλο κατηγοριοποίησης που έχει καταχωρημένο στη βάση του.
- Όταν το σύστημα έχει αρκετά στοιχεία από τις απαντήσεις του ασθενή εκτελεί την κατηγοριοποίησή του τοποθετώντας σε μια από τις προκαθορισμένες ομάδες που το σύστημα γνωρίζει.
- Ανάλογα με την ομάδα κατηγοριοποίησης (κλάση) το σύστημα δημιουργεί ηλεκτρονικό ραντεβού με τον ιατρό της αντίστοιχης ειδικότητας (τα ραντεβού καταχωρούνται μέσα στην εφαρμογή ως δραστηριότητες τύπου “Appointment”). Η αντιστοίχιση ιατρών με ομάδες ασθενών (κλάσεις κατηγοριοποίησης) είναι προκαθορισμένες μέσα στο σύστημα.
- Ο ασθενής ενημερώνεται για την ημερομηνία και ώρα του επόμενου ραντεβού καθώς και για τον ιατρό που θα επισκεφτεί.
- Ο πελάτης – ασθενής ολοκληρώνει την επικοινωνία του με το διαγνωστικό κέντρο.

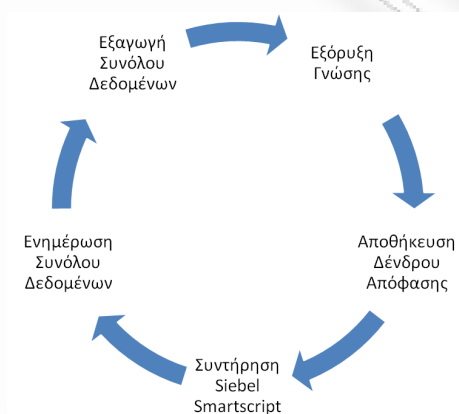
### 4.4.3 “Back End” διαδικασία

Η “back end” διαδικασία περιλαμβάνει όλες εκείνες τις λειτουργίες που θα πρέπει να εκτελούνται σε τακτά χρονικά διαστήματα ούτως ώστε να εξασφαλίζεται η ομαλή και αποδοτική λειτουργία της “front end” διαδικασίας.

Οι λειτουργίες αυτές αφορούν την ενημέρωση της σχεσιακής βάσης δεδομένων με το μοντέλο κατηγοριοποίησης όπως αυτό εξάγεται από κάποιο εξωτερικό εργαλείο εξόρυξης γνώσης.

Για τον σκοπό αυτό, το τμήμα μηχανογράφησης του διαγνωστικού κέντρου μέσω του εργαλείου WEKA αρχικά θα παράγει και στην συνέχεια θα καταχωρεί στο Oracle Siebel 8.1 CRM το δένδρο απόφασης όπως προκύπτει από την διαδικασία κατηγοριοποίησης των δεδομένων. Επιπλέον θα πρέπει να συντηρεί το Siebel Smartscrip που στόχο έχει την εκμετάλλευση της γνώσης που το δένδρο απόφασης περιέχει.

Οι παραπάνω διαδικασίες θα εκτελούνται περιοδικά σε τακτά χρονικά διαστήματα και θα ανατροφοδοτούνται από τα αποτελέσματά τους.



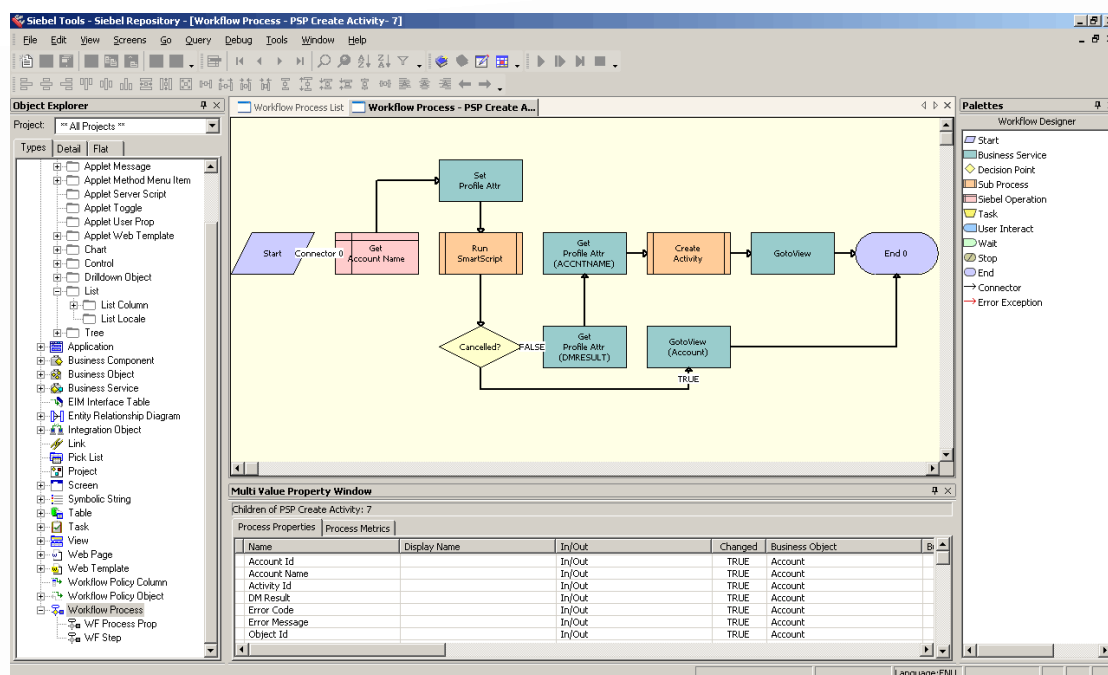
Εικόνα 50: Η “κυκλική” διεκπεραίωση των “back end” λειτουργιών

### 4.4.4 Υλοποίηση

Στις προηγούμενες ενότητες αναφερθήκαμε με αρκετή λεπτομέρεια στα θέματα της υλοποίησης που αφορούσαν την εξαγωγή του δένδρου απόφασης ως αποτέλεσμα μιας διαδικασίας κατηγοριοποίησης δεδομένων, την αποθήκευση του δένδρου απόφασης στη σχεσιακή βάση δεδομένων του Oracle Siebel 8.1 CRM και της παραμετροποίησης των Siebel Smartscripts για την αποδοτικότερη απεικόνιση του δένδρου απόφασης.

Στα πλαίσια ωστόσο της παρουσίασης της ολοκληρωμένης επιχειρησιακής διαδικασίας υλοποιήθηκε επιπλέον και το κομμάτι που αφορά την “front end” διαδικασία, την κλήση των Siebel Smartscripts και την αυτόματη δημιουργία ιατρικών ραντεβού μέσα από το περιβάλλον της εφαρμογής.

Για τη συγκεκριμένη υλοποίηση χρησιμοποιήσαμε τα “Siebel Tools” δημιουργώντας ένα καινούριο workflow object.



Εικόνα 51: Το workflow object για την κλήση του Siebel Smartsript και την καταχώρηση ραντεβού

Μέσω του συγκεκριμένου object κάθε φορά που ο χρήστης της εφαρμογής ξεκινά τη διαδικασία καταχώρησης νέου ραντεβού, τα ακόλουθα βήματα εκτελούνται σειριακά:

- Συλλέγονται τα δεδομένα που αφορούν τον πελάτη – ασθενή.
- Ξεκινά η διαδικασία του Siebel Smartsript.
- Δημιουργείται μια καινούρια δραστηριότητα (τύπου “Appointment”) με την επιτυχή ολοκλήρωση της διαδικασίας κατηγοριοποίησης του πελάτη.

Στο επόμενο κεφάλαιο θα παρουσιάσουμε ένα συγκεκριμένο παράδειγμα της επιχειρησιακής διαδικασίας.

## 5. Επίδειξη Συστήματος

Στο κεφάλαιο αυτό θα παρουσιάσουμε ένα παράδειγμα για την επίδειξη της “front end” διαδικασίας. Θεωρούμε ότι οι λειτουργίες που συνιστούν την “back end” διαδικασία έχουν εκτελεστεί επιτυχώς:

- Το τμήμα μηχανογράφησης της επιχείρησης έχει εξάγει το σύνολο δεδομένων δερματολογικής φύσης.
- Έχει εκτελέσει με τη βοήθεια του WEKA τη διαδικασία κατηγοριοποίησης και έχει εξάγει το επιθυμητό δένδρο απόφασης.
- Έχει καταχωρήσει στη σχεσιακή βάση δεδομένων του Oracle Siebel 8.1 CRM το δένδρο απόφασης με τον τρόπο που περιγράφηκε στο προηγούμενο κεφάλαιο.
- Έχει δημιουργήσει το Siebel Smartscripτ μέσω του οποίου τα αποτελέσματα της κατηγοριοποίησης θα είναι άμεσα εκμεταλλεύσιμα από τους τελικούς χρήστες της εφαρμογής.

### 5.1 Περιγραφή παραδείγματος

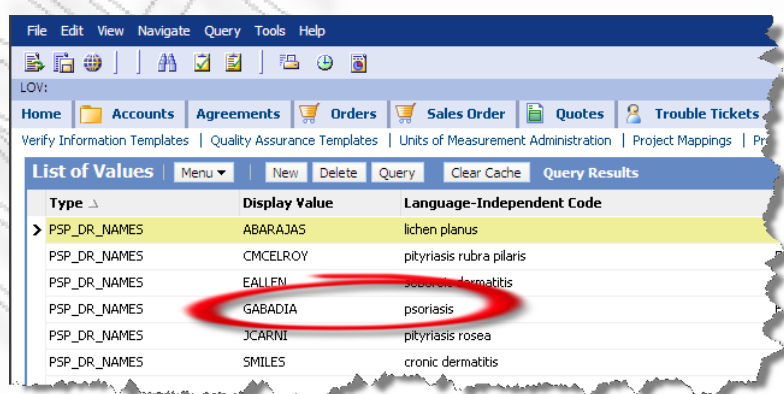
Η ασθενής *Eleanor Rigby* καλεί το τηλεφωνικό κέντρο του διαγνωστικού κέντρου “Parker Healthcare” με σκοπό να κλείσει ραντεβού με κάποιον δερματολόγο.

Το διαγνωστικό κέντρο διαθέτει 6 ιατρούς δερματολογίας ειδικευμένους σε 6 διαφορετικές δερματικές παθήσεις:

- Ο καθηγητής A. Barajas ειδικεύεται στη πάθηση “Lichen Planus”.
- Ο καθηγητής C. McElroy ειδικεύεται στη πάθηση “Pityriasis Rubra Pilaris”.
- Ο καθηγητής E. Allen ειδικεύεται στη πάθηση “Seboric Dermatitis”.
- Ο καθηγητής G. Abadia ειδικεύεται στη πάθηση “Psoriasis”.
- Ο καθηγητής J. Carni ειδικεύεται στη πάθηση “Pityriasis Rosea”.
- Ο καθηγητής S. Miles ειδικεύεται στη πάθηση “Cronic Dermatitis”.

Η ασθενής φυσικά δεν γνωρίζει κανέναν από αυτούς τους ιατρούς. Θα την ενδιέφερε ωστόσο να εξεταστεί από εκείνον που ειδικεύεται στην πάθησή της.

Η ειδίκευση των 6 ιατρών στις αντίστοιχες ειδικότητες βρίσκεται καταχωρημένη στο σύστημα CRM που χρησιμοποιεί το τηλεφωνικό κέντρο.

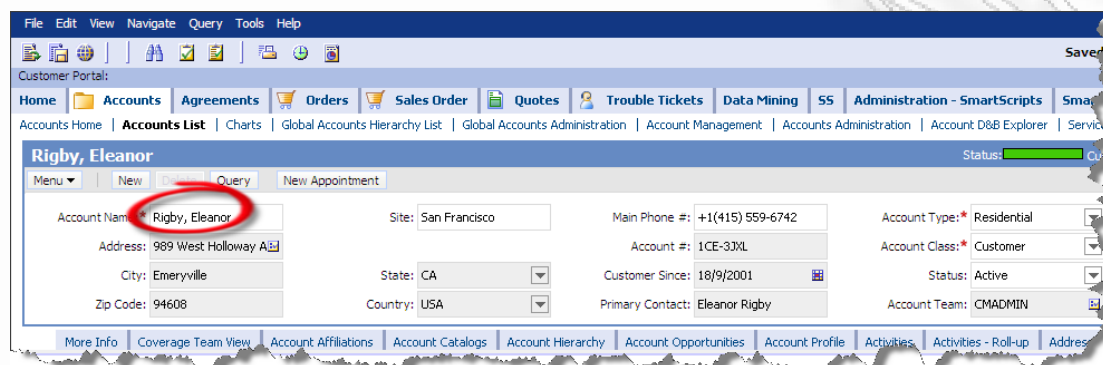


Type	Display Value	Language-Independent Code
PSP_DR_NAMES	ABARAJAS	lichen planus
PSP_DR_NAMES	CMCELROY	pityriasis rubra pilaris
PSP_DR_NAMES	EALLEN	seboric dermatitis
PSP_DR_NAMES	GABADIA	psoriasis
PSP_DR_NAMES	JCARNI	pityriasis rosea
PSP_DR_NAMES	SMILES	cronic dermatitis

Εικόνα 52: Η αντιστοίχιση ιατρών - ειδικτήτων στο Oracle Siebel 8.1 CRM

## 5.2 Παρουσίαση παραδείγματος

Η Eleanor Rigby αφού καλέσει το τηλεφωνικό κέντρο του “Parker Healthcare” επικοινωνεί με κάποιον υπάλληλο του διαγνωστικού κέντρου. Ο τελευταίος με σκοπό να ταυτοποιήσει τον ασθενή στο σύστημα CRM, ζητά κάποια προσωπικά στοιχεία (πχ τηλέφωνο επικοινωνίας, ονοματεπώνυμο κλπ). Με την ολοκλήρωση της αναζήτησης στην εφαρμογή, ο υπάλληλος του τηλεφωνικού κέντρου έχει μια πρώτη εικόνα του ασθενή.

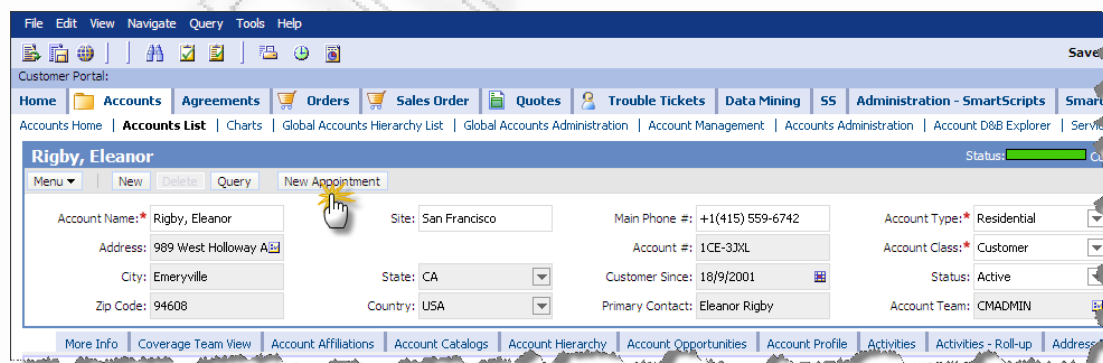


Εικόνα 53: Ο υπάλληλος αναζητά τον ασθενή στην CRM εφαρμογή

Στη συνέχεια, η Eleanor Rigby ενημερώνει τον υπάλληλο πως επιθυμεί να κλείσει ένα ραντεβού με κάποιον από τους ιατρούς δερματολογίας που συνεργάζονται με το διαγνωστικό κέντρο.

Ο υπάλληλος του τηλεφωνικού κέντρου ενημέρωνει την Eleanor πως για την καλύτερη εξυπηρέτησή της θα χρειαστεί να απαντήσει σε κάποιες ερωτήσεις όσον αφορά τα δερματολογικής φύσης συμπτώματα που αντιμετωπίζει. Ενημερώνεται επιπλέον πως αυτό γίνεται για να μπορέσει το σύστημα να καταχωρήσει το ραντεβού στον «σωστό», βάση εξειδίκευσης, ιατρό.

Η Eleanor ικανοποιημένη από την ποιότητα εξυπηρέτησης του τηλεφωνικού κέντρου δέχεται ευχαρίστως να απαντήσει στις ερωτήσεις που ο υπάλληλος προτίθεται να της απευθύνει. Ο τελευταίος, έχοντας τη συγκατάθεση της ασθενούς ξεκινά τη διαδικασία καταχώρησης νέου ραντεβού πατώντας το κουμπί “New Appointment” της οθόνης του.



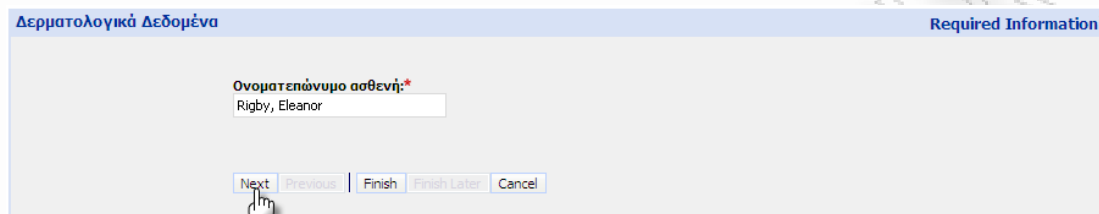
Εικόνα 54: Η διαδικασία καταχώρησης νέου ραντεβού ξεκινά με το πάτημα ενός κουμπιού

Πατώντας το κουμπί “New Appointment” η εφαρμογή CRM ξεκινά τη διαδικασία του Siebel Smartscript με σκοπό την προβολή «Ερωτημάτων» που ο χρήστης του συστήματος θα πρέπει να απευθύνει στην ασθενή. Παράλληλα οι «Απαντήσεις» της ασθενούς θα κατευθύνουν το σύστημα ώστε να διατρέξει καταλλήλως το δένδρο απόφασης που βρίσκεται



στη βάση δεδομένων του, με απώτερο στόχο την κατηγοριοποίηση της ασθενούς σε μία από τις 6 παθήσεις (κλάσεις/ομάδες) που το σύστημα γνωρίζει.

Το Siebel Smartscrip ξεκινά ζητώντας από τον χρήστη της εφαρμογής να επιβεβαιώσει το όνομα του ασθενούς.

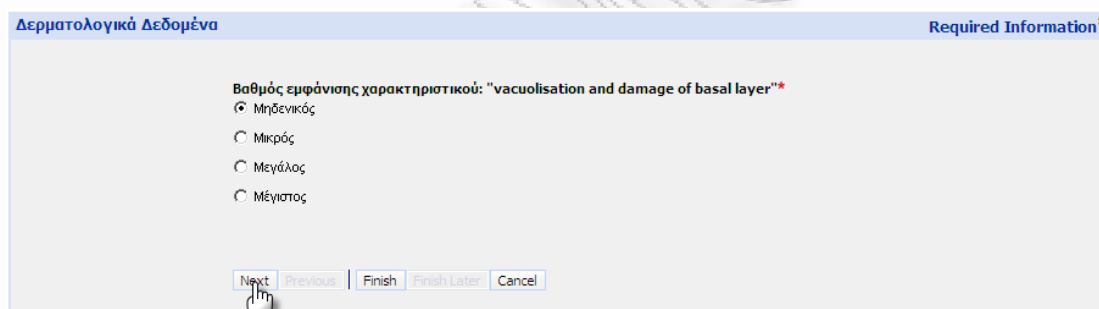


The screenshot shows a window titled "Δερματολογικά Δεδομένα" (Dermatological Data) with a "Required Information\*" label. The main content area displays "Όνοματεπώνυμο ασθενή:" (Patient Name) with the value "Rigby, Eleanor" entered in a text field. Below the text field are five buttons: "Next", "Previous", "Finish", "Finish Later", and "Cancel". A mouse cursor is pointing at the "Next" button.

Εικόνα 55: Η επιβεβαίωση των στοιχείων του ασθενούς

Πατώντας ο χρήστης το κουμπί "Next", ξεκινά το σετ «Ερωτημάτων» με την βοήθεια των οποίων το CRM σύστημα θα προσπαθήσει να κατηγοριοποιήσει τον ασθενή σε μία από τις 6 κλάσεις – ασθένειες και θα καταχωρήσει το ραντεβού με τον σωστό ιατρό.

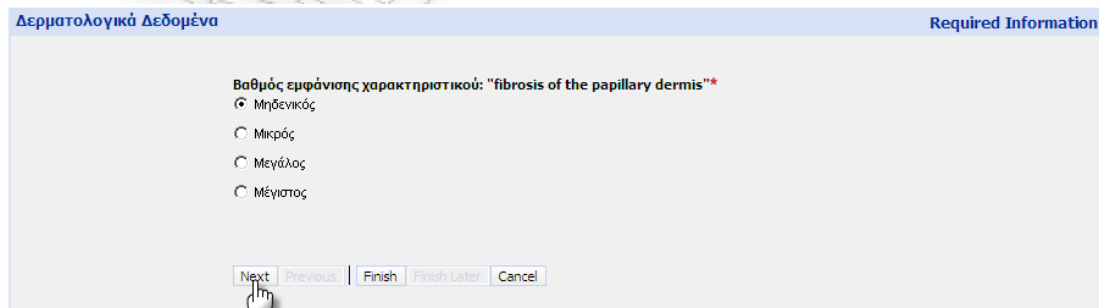
Το πρώτο «Ερώτημα» που το σύστημα απευθύνει (μέσω του χρήστη) στον ασθενή αφορά τον βαθμό εμφάνισης του χαρακτηριστικού γνωρίσματος "Vacualisation and Damage of Basal Layer". Ο χρήστης καταχωρεί την απάντηση του ασθενούς («Μηδενικός») και πατά το κουμπί "Next".



The screenshot shows the same window as Figure 55, but the question is now: "Βαθμός εμφάνισης χαρακτηριστικού: 'vacuolisation and damage of basal layer'". There are four radio button options: "Μηδενικός" (selected), "Μικρός", "Μεγάλος", and "Μέγιστος". The "Next" button is still highlighted by the mouse cursor.

Εικόνα 56: Το πρώτο "Ερώτημα"

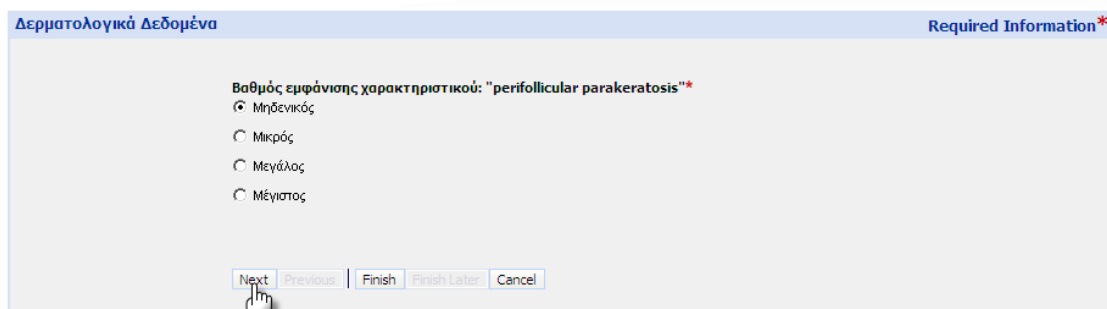
Το επόμενο «Ερώτημα» αφορά το χαρακτηριστικό γνώρισμα "Fibrosis of the Papillary Dermis". Και σε αυτό το «Ερώτημα» η απάντηση του ασθενούς είναι «Μηδενικός» βαθμός εμφάνισης.



The screenshot shows the same window, but the question is now: "Βαθμός εμφάνισης χαρακτηριστικού: 'fibrosis of the papillary dermis'". There are four radio button options: "Μηδενικός" (selected), "Μικρός", "Μεγάλος", and "Μέγιστος". The "Next" button is still highlighted by the mouse cursor.

Εικόνα 57: Το δεύτερο "Ερώτημα"

Το τρίτο κατά σειρά «Ερώτημα» σχετίζεται με τον βαθμό εμφάνισης του χαρακτηριστικού "Perifollicular Parakeratosis", και στο οποίο ο ασθενής απαντά «Μηδενικός».



Δερματολογικά Δεδομένα Required Information\*

Βαθμός εμφάνισης χαρακτηριστικού: "perifollicular parakeratosis"\*

Μηδενικός

Μικρός

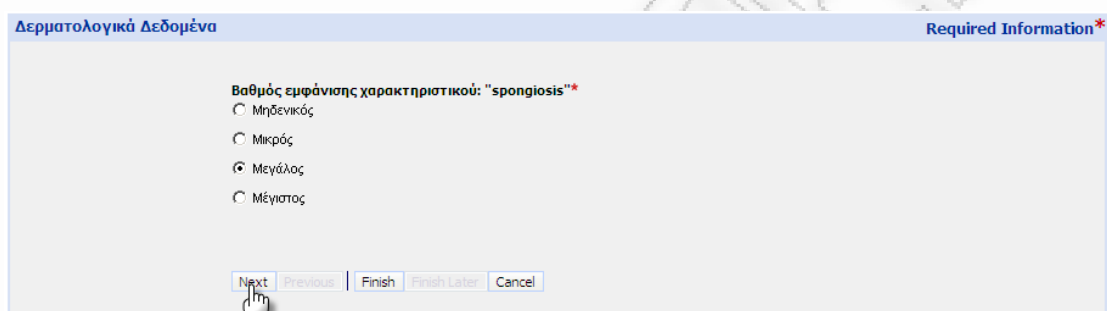
Μεγάλος

Μέγιστος

Next Previous Finish Finish Later Cancel

Εικόνα 58: Το τρίτο "Ερώτημα"

Το σύστημα συνεχίζει να απευθύνει «Ερωτήματα» προς τον ασθενή και έρχεται η σειρά του χαρακτηριστικού γνωρίσματος "Spongiosis". Στο «Ερώτημα» αυτό η απάντηση του ασθενή είναι «Μεγάλος».



Δερματολογικά Δεδομένα Required Information\*

Βαθμός εμφάνισης χαρακτηριστικού: "spongiosis"\*

Μηδενικός

Μικρός

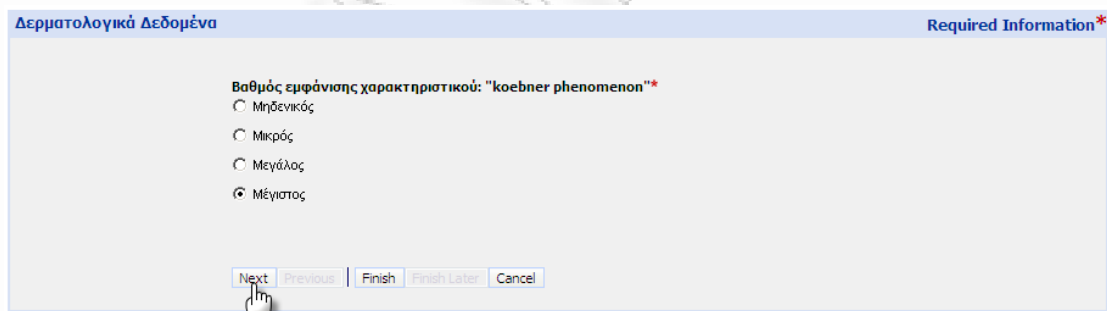
Μεγάλος

Μέγιστος

Next Previous Finish Finish Later Cancel

Εικόνα 59: Το τέταρτο "Ερώτημα"

Στο επόμενο «Ερώτημα» σχετικά με τον βαθμό εμφάνισης του χαρακτηριστικού "Koebner Phenomenon" ο ασθενής απαντά «Μέγιστος».



Δερματολογικά Δεδομένα Required Information\*

Βαθμός εμφάνισης χαρακτηριστικού: "koebner phenomenon"\*

Μηδενικός

Μικρός

Μεγάλος

Μέγιστος

Next Previous Finish Finish Later Cancel

Εικόνα 60: Το πέμπτο "Ερώτημα"

Καταχωρώντας και αυτήν την απάντηση ο χρήστης του CRM πατά το κουμπί "Next" ούτως ώστε η διαδικασία να συνεχιστεί. Ωστόσο, το σύστημα δεν έχει πλέον άλλα «Ερωτήματα» να απευθύνει ολοκληρώνοντας τη διαδικασία της κατηγοριοποίησης και κατατάσσοντας τον ασθενή στην κλάση "Pityriasis Rosea" (ομάδα 4).

Εικόνα 61: Το αποτέλεσμα της κατηγοριοποίησης

Στην οθόνη αυτή, ο χρήστης της εφαρμογής CRM το μόνο που μπορεί να εκτελέσει είναι η ολοκλήρωση της διαδικασίας πατώντας το κουμπί “Finish”. Με την εκτέλεση της ενέργειας αυτής το σύστημα δημιουργεί ένα ραντεβού καταχωρώντας τα αποτελέσματα της κατηγοριοποίησης και αναθέτοντάς το στον ιατρό J. Carni.

Εικόνα 62: Η ολοκλήρωση της διαδικασίας με την καταχώρηση ραντεβού

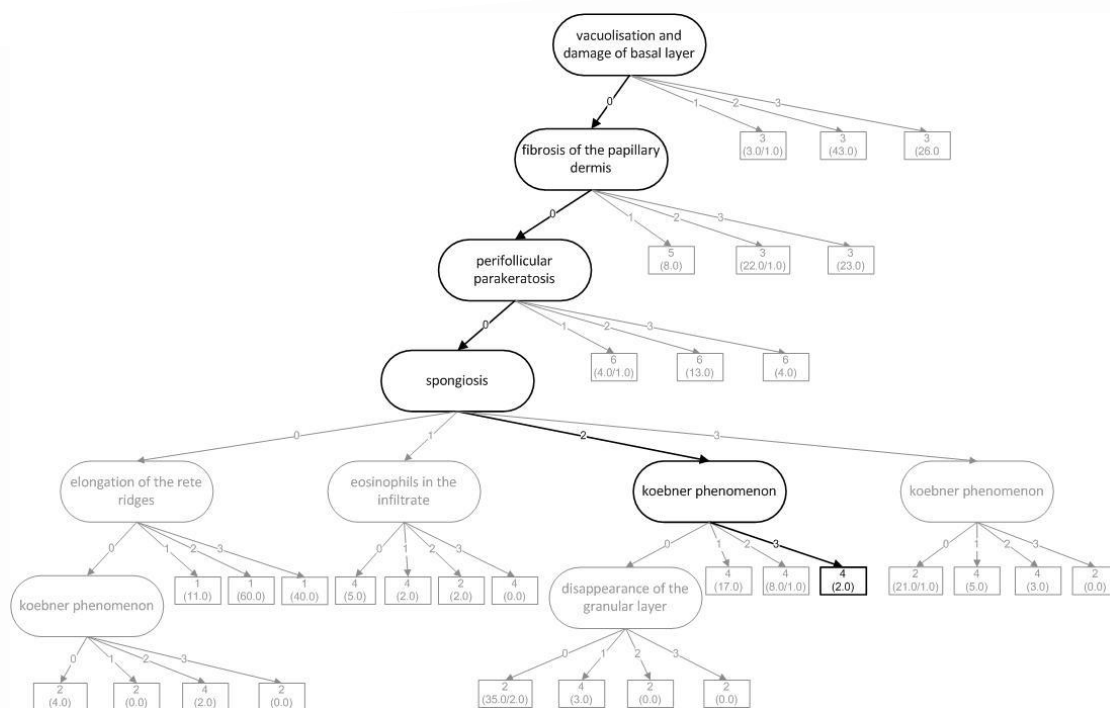
Όπως φαίνεται και από την οθόνη του Oracle Siebel 8.1 CRM το ραντεβού δημιουργήθηκε για τον ιατρό J. Carni με ειδικότητα στη πάθηση “Pityriasis Rosea”. Επιπλέον πληροφορία που καταχωρήθηκε στην δραστηριότητα τύπου “Appointment” είναι το ονοματεπώνυμο του ασθενούς και το αποτέλεσμα της κατηγοριοποίησης.

Εικόνα 63: Τα στοιχεία της κατηγοριοποίησης όπως καταχωρήθηκαν στην δραστηριότητα

Στην επόμενη ενότητα θα δούμε μια σύντομη τεχνική εξήγηση του παραδείγματος.

### 5.3 Τεχνική επεξήγηση παραδείγματος

Τα «Ερωτήματα» που απεύθυνε το σύστημα CRM στον ασθενή – πελάτη του διαγνωστικού κέντρου σε συνδυασμό με τις απαντήσεις του τελευταίου οδήγησαν στην κλάση – ασθένεια “Pityriasis Rosea”. Η πορεία που ακολούθησε το σύστημα μέσα στο δένδρο απόφασης που είχε καταχωρημένο στη βάση του φαίνεται στην επόμενη εικόνα.



Εικόνα 64: Η πορεία της διαδικασίας μέσα στο δένδρο απόφασης

Στην ενότητα αυτή αξίζει να περιγράψουμε τον τρόπο με τον οποίο το σύστημα παράγαγε τα «Ερωτήματα».

Όσον αφορά το εναρκτήριο «Ερώτημα», το σύστημα το μόνο που χρειαζότανε για να το παράγει ήταν ο κόμβος – ρίζα του δένδρου και το χαρακτηριστικό γνώρισμά της. Τη πληροφορία αυτή μπορούσε πολύ εύκολα να την υπολογίσει από τον πίνακα των κόμβων του δένδρου.

Tree Node Name	Attribute Name	Root Node	Description
> Node 1	vacuolisation and damage of basal layer	✓	
Node 2	fibrosis of the papillary dermis		
Node 3	perifollicular parakeratosis		
Node 4	spongiosis		
Node 5	elongation of the rete ridges		

Εικόνα 65: Η πληροφορία του κόμβου - ρίζα του δένδρου

Στην συνέχεια κάθε επόμενο «Ερώτημα» του συστήματος προέκυπτε εύκολα από τον πίνακα των μεταβάσεων σε συνδυασμό πάντα με τις εκάστοτε απαντήσεις του ασθενή.

Έτσι με την πρώτη καταχώρηση απάντησης για το γνώρισμα “Vacualisation and Damage of Basal Layer”, το σύστημα εύκολα μπορεί να υπολογίσει το επόμενο «Ερώτημα» ή την κλάση στην οποία πρέπει να μεταβεί.

Tree Node	Tree Node Attribute	Value	Class Code	Next Tree Node	Description
Node 1	vacuolisation and damage of basal layer	0		Node 2	
Node 1	vacuolisation and damage of basal layer	1	3		
Node 1	vacuolisation and damage of basal layer	2	3		
Node 1	vacuolisation and damage of basal layer	3	3		
Node 2	fibrosis of the papillary dermis	0		Node 3	

Εικόνα 66: Υπολογισμός επόμενου "Ερωτήματος"

Όπως φαίνεται και στην εικόνα, για τον κόμβο "Node 1" στον οποίο και βρισκόμαστε, η απάντηση 0 («Μηδενικός» βαθμός εμφάνισης) οδηγεί στον επόμενο κόμβο "Node 2" που σύμφωνα με τον πίνακα των κόμβων αντιστοιχεί στο χαρακτηριστικό "Fibrosis of the Papillary Dermis". Το νέο «Ερώτημα» επομένως θα αφορά το συγκεκριμένο χαρακτηριστικό.

Οι μεταβάσεις αυτές από «Ερώτημα» σε «Ερώτημα» συνεχίζονται έως ότου κάποια απάντηση του ασθενή για συγκεκριμένο γνώρισμα δεν οδηγήσει σε νέο κόμβο αλλά σε κάποια κλάση. Στο παράδειγμα μας, αυτό συνέβη για την απάντηση «Μέγιστος» που έδωσε η ασθενής σχετικά με τον βαθμό εμφάνισης του γνωρίσματος "Koebner Phenomenon".

Tree Node	Tree Node Attribute	Value	Class Code	Next Tree Node	Description
Node 7	koebner phenomenon	0		Node 10	
Node 7	koebner phenomenon	1	4		
Node 7	koebner phenomenon	2	4		
Node 7	koebner phenomenon	3	4		
Node 8	koebner phenomenon	0	2		

Εικόνα 67: Τερματισμός σε κάποια κλάση

Όπως φαίνεται και στην εικόνα, για τον κόμβο "Node 7" που αντιστοιχεί στο γνώρισμα "Koebner Phenomenon" η απάντηση 4 («Μέγιστος» βαθμός εμφάνισης) δεν οδηγεί σε κάποιον νέο κόμβο, άρα και σε νέο «Ερώτημα», αλλά απευθείας σε κάποια κλάση, στην ομάδα 4 ("Pityriasis Rosea").

Με λίγα λόγια, όπως περιγράψαμε εκτενώς και στην ενότητα των Siebel Smartscripts, κάθε απάντηση που καταχωρείται στο σύστημα από τον ασθενή βοηθά στον προσδιορισμό του επόμενου γνωρίσματος για το οποίο θα πρέπει να απευθυνθεί «Ερώτημα». Αυτά που έχουν συνδεθεί μεταξύ τους στο Siebel Smartscript δεν είναι οι κόμβοι του δένδρου απόφασης αλλά τα ίδια τα γνωρίσματα του συνόλου δεδομένων μας. Με τον τρόπο αυτό, η υλοποίηση του Siebel Smartscript γίνεται αρκετά γενική και ανεξαρτητοποιείται από την δομή του δένδρου απόφασης. Όσο το σύνολο δεδομένων μας περιέχει τα ίδια χαρακτηριστικά γνωρίσματα, το Siebel Smartscript δεν χρειάζεται να τροποποιείται και είναι ανεξάρτητο από το δένδρο απόφασης που παράγει κάθε αλγόριθμος κατηγοριοποίησης. Τη χρονική στιγμή εκείνη κατά την οποία ένα νέο γνώρισμα θα μπει ή ένα παλιό θα βγει από το σύνολο δεδομένων μας το Siebel Smartscript θα πρέπει να τροποποιηθεί. Αλλά σε αυτή τη περίπτωση θα έχει ήδη τροποποιηθεί το σύνολο δεδομένων.

Ολοκληρώνοντας τη σύντομη αυτή τεχνική περιγραφή του παραδείγματος, το Siebel Smartscript είναι «χτισμένο» πάνω στα γνωρίσματα του συνόλου δεδομένων μας ενώ οι κινήσεις μας μέσα σε αυτό (από ποιο «Ερώτημα» θα πάμε σε ποιο «Ερώτημα») καθορίζονται από τον πίνακα μεταβάσεων του δένδρου απόφασης.

## 6. Επίλογος

Στο κεφάλαιο αυτό θα επιχειρήσουμε μια ανακεφαλαίωση της διπλωματικής εργασίας. Θα περιγράψουμε μερικά από τα προβλήματα που αντιμετωπίσαμε και θα αναφερθούμε σε κάποιες ενδεικτικές μελλοντικές επεκτάσεις του συστήματός μας.

### 6.1 Σύνοψη – Συμπεράσματα

Στην εργασία αυτή προσπαθήσαμε να παρουσιάσουμε την ανάγκη συνεργασίας δύο διαφορετικών κόσμων με σκοπό την αποδοτικότερη υλοποίηση επιχειρησιακών διαδικασιών. Αρχικά γνωρίσαμε την περιοχή της εξόρυξης γνώσης από δεδομένα και ήρθαμε σε επαφή με μερικές από τις δημοφιλέστερες τεχνικές της. Στη συνέχεια παρουσιάσαμε το Oracle Siebel 8.1 CRM ως το πιο διαδεδομένο σύστημα CRM παγκοσμίως. Είδαμε μέσα από συγκεκριμένα παραδείγματα την αξία της εξόρυξης γνώσης από δεδομένα που διαθέτουν τα CRM συστήματα και πώς αυτή μπορεί να επηρεάσει επιχειρησιακές διαδικασίες στο σύνολό τους.

Προσπαθήσαμε να συνδυάσουμε την εξόρυξη γνώσης και το Oracle Siebel 8.1 CRM μέσα από μια συγκεκριμένη επιχειρησιακή διαδικασία, όπως είναι «η αυτοματοποίηση της διαδικασίας δημιουργίας ραντεβού ενός διαγνωστικού κέντρου». Διαχωρίσαμε τη διαδικασία αυτή σε δύο τομείς, το “front-end” και το “back-end” δίνοντας ιδιαίτερη έμφαση στο πρώτο με σκοπό να παρουσιάσουμε όλα τα πλεονεκτήματα που παρέχουν εφαρμογές CRM σε μια επιχείρηση.

Συνοψίζοντας, μπορούμε να πούμε πως μέσω της διπλωματικής αυτής εργασίας γνωρίσαμε την αξία της εξόρυξης γνώσης από δεδομένα στην πράξη, ενσωματώνοντας το θεωρητικό κομμάτι της κατηγοριοποίησης σε ένα συγκεκριμένο συστατικό στοιχείο μιας εμπορικής εφαρμογής, όπως είναι τα Siebel Smartscripts.

### 6.2 Δυσκολίες

Κατά την υλοποίηση της επιχειρησιακής διαδικασίας αντιμετωπίσαμε ιδιαίτερα προβλήματα, μερικά εκ των οποίων επιλύθηκαν ενώ κάποια άλλα ξεπεράστηκαν.

Το βασικότερο πρόβλημα που αντιμετωπίσαμε ήταν οι περιορισμένοι διαθέσιμοι πόροι σε αντιπαράθεση με τις ιδιαίτερα αυξημένες απαιτήσεις του Oracle Siebel 8.1 CRM. Μετά από πολύ σκέψη αποφασίστηκε να μην εγκατασταθεί το πλήρες περιβάλλον του CRM αλλά να χρησιμοποιηθεί μια ειδική έκδοση του πακέτου (Sample Database Environment) δεδομένου ότι αυτό δεν θα επηρέαζε ούτε κατά το ελάχιστον την υλοποίηση της επιχειρησιακής διαδικασίας.

Ένα δεύτερο και ιδιαίτερος σημαντικό πρόβλημα που κόστισε περισσότερο σε χρόνο παρά σε κόπο ήταν η εύρεση του κατάλληλου συνόλου δεδομένων στο οποίο θα μπορούσαμε να βασιστούμε τόσο για την εκτέλεση της διαδικασίας της κατηγοριοποίησης όσο και για την επιλογή της προς υλοποίηση επιχειρησιακής διαδικασίας.

Όσον αφορά το κομμάτι της υλοποίησης ιδιαίτερες δυσκολίες αντιμετωπίσαμε στην προσπάθεια εύρεσης ενός αποδοτικού τρόπου απεικόνισης του δένδρου απόφασης μέσα στη βάση δεδομένων του Oracle Siebel 8.1 CRM με σκοπό την όσο το δυνατόν γενικότερη υλοποίηση του Siebel Smartsript. Σκοπός μας ήταν να μπορέσουμε να προτείνουμε μια δομή Siebel Smartsript που να μην επηρεάζεται άμεσα αφενός από την ποσότητα των δεδομένων όσο και από την ποιότητα του δένδρου απόφασης που εξάγεται από αυτά.

### 6.3 Μελλοντικές Επεκτάσεις

Οι επεκτάσεις που θα μπορούσαν μελλοντικά να ενσωματωθούν στα πλαίσια της συγκεκριμένης υλοποίησης περιλαμβάνουν τα ακόλουθα:

- Online διασύνδεση του WEKA με το Oracle Siebel 8.1 CRM.

Υλοποίηση web services (inbound/outbound) από το Oracle Siebel 8.1 CRM και το WEKA με σκοπό την αυτοματοποιημένη μεταφορά του συνόλου δεδομένων και των αποτελεσμάτων της κατηγοριοποίησης από το ένα σύστημα στο άλλο χωρίς την ανθρώπινη παρέμβαση (human intervention).

- Υλοποίηση διαδικασίας ενημέρωσης του αρχικού συνόλου δεδομένων μέσα από το Oracle Siebel 8.1 CRM.

Με την ολοκλήρωση κάθε ραντεβού, ο αρμόδιος ιατρός θα συμπληρώνει ειδική ηλεκτρονική φόρμα μέσα στο Oracle Siebel 8.1 CRM με τα πραγματικά αποτελέσματα της διάγνωσης, τα οποία με αυτοματοποιημένο τρόπο (batch/online) θα μεταφέρονται στο ήδη καταχωρημένο σύνολο δεδομένων με σκοπό την επόμενη φορά που θα εκτελεστεί η διαδικασία της κατηγοριοποίησης να λαμβάνονται υπόψη.

- Υλοποίηση «πίνακα αποστάσεων» για τις συσχετίσεις των ιατρικών ειδικοτήτων.

Καταχώρηση και συντήρηση μέσα στη βάση του Oracle Siebel 8.1 CRM ενός πίνακα αποστάσεων μέσω του οποίου θα διατηρούνται οι συσχετίσεις των ιατρικών ειδικοτήτων όσον αφορά την ομοιότητά τους με σκοπό να λαμβάνονται υπόψη από τη διαδικασία δημιουργίας ραντεβού για τις περιπτώσεις όπου ένας ή περισσότεροι ιατροί απουσιάζουν.

## Βιβλιογραφία - Αναφορές

- Βιβλιογραφικές Αναφορές

[1] "Data Mining, Εισαγωγικά και Προηγμένα Θέματα Εξόρυξης Γνώσης από Δεδομένα", Margaret H. Dunham, Επιμέλεια Ελληνικής Έκδοσης Β. Βερούκιος & Γ. Θεοδωρίδης, Εκδόσεις Νέων Τεχνολογιών

[2] "Αποθήκες Δεδομένων και Εξόρυξη Γνώσης", Γ. Θεοδωρίδης, Ν. Πελέκης, Σημειώσεις Πανεπιστημίου Πειραιώς

[3] "Knowledge Discovery in Databases: An Overview", W. Frawley and G. Piatetsky-Shapiro and C. Matheus, ISSN 0738-4602

[4] "Decision Support and Business Intelligence Systems", E. Turban, J.E. Aronson, T. Liang, R. Sharda, Eight Edition, Prentice Hall

[5] "CRM in Real Time", B.J. Goldenberg, CyberAge Books

[6] "Υλοποιώντας Επιχειρησιακές διαδικασίες με την Αρχιτεκτονική SOA", Θ. Τσάβαλος, Μεταπτυχιακή Διατριβή, Πανεπιστήμιο Πειραιώς

[7] "Ανάλυση Δεδομένων Επιχειρησιακών Διαδικασιών σε Πραγματικό Χρόνο", Κ. Μεντή, Μεταπτυχιακή Διατριβή, Πανεπιστήμιο Πειραιώς

[8] "CRM Automation", B.J. Goldenberg, Prentice Hall

[9] "Application of Knowledge Management Technology in Customer Relationship Management", Ranjit Bose, Vijayan Sugumaran, Research Article

[10] "Oracle's Siebel Business Applications Bookshelf Documentation Library", vs.8.1, 2008 Oracle

[11] "Data Model Reference for Industry Applications", vs.8.1, 2008 Oracle

[12] "Core Consultant Course", vs.8.1, 2008 Oracle

[13] "Classification of Arrhythmia Using Machine Learning Techniques", T. Soman, P. Bobbie, School of Computing and Software Engineering, Southern Polytechnic State University ([http://cse.spsu.edu/pbobbie/SharedFile/ECGDiagnosis\\_ICOSSE\\_2005\\_VFinal.pdf](http://cse.spsu.edu/pbobbie/SharedFile/ECGDiagnosis_ICOSSE_2005_VFinal.pdf))

[14] "Data Mining: Practical Machine Learning Tools and Techniques", I.H. Witten, E. Frank, Second Edition, Elsevier

[15] "Classic Data Structures", D. Samanta, PHI, online version [http://books.google.gr/books?id=\\_IDjLLasQycC&lpg=PR1&pg=PR1#v=onepage&q=&f=false](http://books.google.gr/books?id=_IDjLLasQycC&lpg=PR1&pg=PR1#v=onepage&q=&f=false)

[16] "Integrating Decision Tree Learning into Inductive Databases", E. Fromont, H. Blockeel, J. Struyf, Universiteit Leuven, Research Article

[17] "A perspective on inductive databases", Luc De Raedt, Institut fur Informatik, Albert-Ludwigs-University, Research Article, online version <http://www.sigkdd.org/explorations/issues/4-2-2002-12/deraedt.pdf>

- Διαδικτυακές Αναφορές

[3W1] <http://alg.ncsa.uiuc.edu/tools/docs/d2k/manual/dataMining.html>

[3W2] <http://www.kdnuggets.com>



- [3W3] <http://www.datamining.gr/el/whatisdatamining.html>
- [3W4] <http://www-pub.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/10.htm>
- [3W5] [http://en.wikipedia.org/wiki/Association\\_rule\\_learning](http://en.wikipedia.org/wiki/Association_rule_learning)
- [3W6]  
[http://el.wikipedia.org/wiki/%CE%91%CE%BB%CE%B3%CF%8C%CF%81%CE%B9%CE%B8%CE%BC%CE%BF%CF%82\\_ID3](http://el.wikipedia.org/wiki/%CE%91%CE%BB%CE%B3%CF%8C%CF%81%CE%B9%CE%B8%CE%BC%CE%BF%CF%82_ID3)
- [3W7] <http://www.cs.sjsu.edu/~lee/cs157b/ID3-AllanNeymark.ppt>
- [3W8]  
<http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/algorithms/14.0/TREE-CART.pdf>
- [3W9] <http://users.auth.gr/~sgardeli/CART%20final.doc>
- [3W10] [http://festvox.org/docs/speech\\_tools-1.2.0/c16616.htm](http://festvox.org/docs/speech_tools-1.2.0/c16616.htm)
- [3W11] <http://www.spss.gr/applications/>
- [3W12]  
[http://byfiles.storage.msn.com/y1pVdIFJZdBH0DtLepUhYfKwZDenqmJWIUTEmSFKao\\_l4mXCOY3l4sNof9LP8wzBF3OCBIP5Wo9rgQ?PARTNER=WRITER](http://byfiles.storage.msn.com/y1pVdIFJZdBH0DtLepUhYfKwZDenqmJWIUTEmSFKao_l4mXCOY3l4sNof9LP8wzBF3OCBIP5Wo9rgQ?PARTNER=WRITER)
- [3W13] [http://www.go-online.gr/ebusiness/specials/article.html?article\\_id=125](http://www.go-online.gr/ebusiness/specials/article.html?article_id=125)
- [3W14] <http://www.twocrows.com/crm-dm.pdf>
- [3W15] <http://en.wikipedia.org/wiki/Cross-selling>
- [3W16] [http://en.wikipedia.org/wiki/Lift\\_\(data\\_mining\)](http://en.wikipedia.org/wiki/Lift_(data_mining))
- [3W17] <https://support.oracle.com>
- [3W18] <http://crmondemand.oracle.com/en/index.htm>
- [3W19] [http://en.wikipedia.org/wiki/Weka\\_\(machine\\_learning\)](http://en.wikipedia.org/wiki/Weka_(machine_learning))
- [3W20] <http://www.cs.waikato.ac.nz/ml/weka/>
- [3W21] <http://gautam.lis.illinois.edu/monkmiddleware/public/analytics/decisiontree.html>
- [3W22] <http://www.kdnuggets.com/datasets/>
- [3W23] <http://archive.ics.uci.edu/ml/>
- [3W24] [http://en.wikipedia.org/wiki/Linked\\_list](http://en.wikipedia.org/wiki/Linked_list)