

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΔΥΤΙΚΗΣ ΕΛΛΑΔΟΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**ΧΡΗΣΗ ΙΣΤΟΡΙΚΟΥ ΕΝΕΡΓΕΙΩΝ ΠΕΛΑΤΩΝ
ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΑΠΩΛΕΙΑΣ ΠΕΛΑΤΩΝ**

ΓΕΩΡΓΑΚΟΠΟΥΛΟΥ ΑΛΕΞΑΝΔΡΑ

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ : ΠΙΕΡΡΑΚΕΑΣ ΧΡΗΣΤΟΣ

ΠΑΤΡΑ, ΙΟΥΝΙΟΣ 2015

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ	1
1. ΑΝΑΛΥΣΗ ΑΠΩΛΕΙΑΣ ΠΕΛΑΤΩΝ	2
1.1. Διαχείριση απώλειας πελατών.....	2
1.1.1. Πρόβλεψη απώλειας πελατών.....	2
1.1.2. Διατήρηση πελατών	3
1.1.3 Αφοσίωση πελατών και η αξία διάρκειας ζωής του πελάτη	3
1.2. Αιτίες απώλειας πελατών	5
1.2.1. Διαχωρισμός πελατών	5
1.2.2. Κίνδυνοι απώλειας πελατών.....	6
1.2.3. Παράγοντες απώλειας πελατών.....	7
1.3. Διαχείριση πελατειακών σχέσεων (CRM)	9
1.3.1. Αρχιτεκτονική του CRM.....	9
1.3.2. Οφέλη του CRM.....	11
1.3.3. Στόχοι του CRM.....	11
1.3.4. Εξόρυξη δεδομένων (Data Mining) και CRM	12
1.3.4.1. Κυριότερες τεχνικές εξόρυξης δεδομένων	15
1.3.5. Κινήσεις στρατηγικής CRM.....	17
2.1. Λογιστική Παλινδρόμηση	18
2.2. Δέντρα Απόφασης.....	20
2.3. Τεχνητά Νευρωνικά Δίκτυα	22
2.4. Μηχανές Διανυσμάτων Υποστήριξης.	27
2.5. Κριτήρια αξιολόγησης μοντέλων	29
2.6. Ταξινομητές Συνόλων	31
2.7. Άλλες μέθοδοι	32
3. Βιβλιογραφική ανασκόπηση	34
3.1. Βιβλιογραφική ανασκόπηση εφαρμογών των μεθόδων ταξινόμησης.....	34
3.2. Βιβλιογραφική ανασκόπηση μελέτης: The Relevant Length of Customer Event History for Churn Prediction: How long is long enough? Working Paper of Faculty of Economics and Business Administration, Department of Marketing, Ghent University	35
3.2.1. Δεδομένα	35
3.2.2. Τεχνικές.....	36
3.2.3. Μεταβλητές	37
3.2.4. Απόδοση Μοντέλου	40
3.2.5. Αποτελέσματα	40
3.2.6. Συμπεράσματα.....	41

3.2.7. Περιορισμοί και Μελλοντικές Κατευθύνσεις	41
4. ΣΥΜΠΕΡΑΣΜΑΤΑ – ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ	43
4.1. Συμπεράσματα.....	43
4.2. Μελλοντικές Κατευθύνσεις.....	45
5. ΒΙΒΛΙΟΓΡΑΦΙΑ.....	46

ΕΙΣΑΓΩΓΗ

Την τελευταία δεκαετία η απώλεια πελατών (customer churn), δηλαδή η τάση των πελατών να διακόπτουν τις σχέσεις και τις συναλλαγές τους με μια εταιρία σε κάποια χρονική περίοδο (Chandar et al. 2006), αποτελεί μια από τις κυριότερες προκλήσεις που οι εταιρίες παγκοσμίως πρέπει να αντιμετωπίσουν προκειμένου να επιβιώσουν σε μια όλο και περισσότερο ανταγωνιστική αγορά προσφέροντας εξατομικευμένες υπηρεσίες ώστε να είναι επιτυχημένη και ανταγωνιστική εταιρία. Έτσι επικεντρώνονται στην χρήση εργαλείων από το χώρο της πληροφορικής επιστήμης, δηλαδή σε αναλυτικά συστήματα διαχείρισης πελατειακών σχέσεων (Customer Relationship Management, CRM).

Τα συστήματα αυτά έχουν αντικαταστήσει τις παραδοσιακές στρατηγικές μαζικού μάρκετινγκ (mass marketing) με εξατομικευμένες πρακτικές (one-to-one marketing) και δίνουν προσοχή στην διατήρηση των πελατών καθώς είναι αποδεδειγμένο ότι η διατήρηση παλαιών πελατών είναι πολύ πιο κερδοφόρα από την απόκτηση καινούριων (Reinarth et al. 2003).

Το επίκεντρο αυτής της εργασίας είναι να αναθεωρήσει ορισμένες από τις πιο δημοφιλείς τεχνολογίες που έχουν εντοπιστεί στη βιβλιογραφία της πρόβλεψης απώλειας πελατών, για την ανάπτυξη ενός μοντέλου διαχείρισης πελατειακών σχέσεων ώστε οι αναλυτές να μειώσουν σημαντικά το μέγεθος των δεδομένων του πελατειακού ιστορικού των μεγάλων βάσεων δεδομένων όπου η υπολογιστική πολυπλοκότητα είναι σημαντική.

1. ΑΝΑΛΥΣΗ ΑΠΩΛΕΙΑΣ ΠΕΛΑΤΩΝ

Η απώλεια πελατών (customer churn) ορίζεται ως η κλίση των πελατών να διακόπτουν τις σχέσεις και συναλλαγές τους με μια εταιρία σε κάποια χρονική περίοδο και αποτελεί μία από τις κύριες προκλήσεις που οι επιχειρήσεις σήμερα καλούνται να αντιμετωπίσουν (Chandar et al 2006).

Σήμερα, είναι γνωστό από έρευνες στον τομέα του μάρκετινγκ ότι κοστίζει περίπου 5 φορές περισσότερο σε μια επιχείρηση να κερδίσει έναν νέο πελάτη, από το να διατηρήσει έναν ήδη υπάρχοντα (Siber, 1997).

Οι εταιρίες τείνουν να χάνουν ένα ποσοστό της τάξης του 25% του πελατολογίου τους κάθε χρόνο. Συνεπώς η μείωση του ποσοστού αυτού κατά 5% σημαίνει αύξηση των κερδών τους ακόμη και κατά 100% σε κάποιες περιπτώσεις.

Άρα είναι φανερή η σημαντικότητα της διαχείρισης απώλειας πελατών (churn management), ώστε οι επιχειρήσεις να σταθεροποιήσουν τη θέση τους στην αγορά και να διατηρήσουν ή και να αυξήσουν την αποδοτικότητα και κερδοφορία τους.

1.1. Διαχείριση απώλειας πελατών

Η διαχείριση απώλειας πελατών (churn management) περιλαμβάνει α) προβλέψεις για το ποιοι πελάτες είναι πιο πιθανόν να διαφύγουν και β) εκτιμήσεις σχετικά με το ποιες είναι οι καταλληλότερες κινήσεις για να διατηρηθούν οι πελάτες αυτοί (Hung, Yen & Wang, 2006). Ειδικά το πρώτο λαμβάνει την μεγαλύτερη προσοχή από τους ερευνητές. Υπάρχουν δυο βασικές προσεγγίσεις στη διαχείριση της απώλειας πελατών (churn management), οι στοχευμένες (targeted) και οι μη στοχευμένες (untargeted) (Burez et al. 2007), όπου από αυτές τις δύο προσεγγίσεις προκύπτουν εύλογοι προβληματισμοί, οι οποίοι και αποτελούν παραμέτρους που πρέπει να ληφθούν υπόψη κατά τη διαχείριση απώλειας πελατών:

- Γνωρίζει μια εταιρία ότι ένας πελάτης πρόκειται να την εγκαταλείψει ;
- Είναι όλοι πελάτες συμφέρον να διατηρηθούν ;
- Και πως μπορεί να τους διατηρήσει ;

1.1.1. Πρόβλεψη απώλειας πελατών

Ένα σημαντικό εργαλείο στην διατήρηση των πελατών αποτελεί η τεχνική πρόβλεψης απώλειας (churn prediction), που αποσκοπεί στο να αναγνωρίσει τους πελάτες εκείνους που είναι πιο πιθανό να χαθούν (Nelson et al. 2006). Μια τυπική τεχνική πρόβλεψης απώλειας γενικεύει την σχέση μεταξύ πελατειακών χαρακτηριστικών και συμπεριφορών με βάση ιστορικά δεδομένα από την άλλη πλευρά, έτσι ώστε η εταιρία να μπορέσει να χρησιμοποιήσει αυτές τις πληροφορίες για να προβλέψει την μελλοντική συμπεριφορά των πελατών της.

Η πρόβλεψη απώλειας προσεγγίζεται γενικά ως δυαδικής κατηγοριοποίησης. Σε αυτό το πλαίσιο οι εταιρίες συνήθως κάνουν εξόρυξη δεδομένων για να αναλύσουν την πελατειακή απώλεια (Xie et al 2009). Οι αλγόριθμοι δυαδικής κατηγοριοποίησης είναι κατάλληλοι για να γενικεύσουν τη σχέση μεταξύ του αποτελέσματος, δηλαδή το αν ένας πελάτης θα διαφύγει ή όχι, με διάφορες μεταβλητές πρόβλεψης που περιγράφουν τα χαρακτηριστικά και την συμπεριφορά του πελάτη. Η ποιότητα ενός μοντέλου πρόβλεψης απώλειας πελατών εξαρτάται από δύο πολύ σημαντικούς παράγοντες: τα διαθέσιμα δεδομένα και τον αλγόριθμο εξόρυξης δεδομένων που θα χρησιμοποιηθεί για να υπολογιστεί η διαφυγή των πελατών. Ο

πρώτος παράγοντας αφορά τα διαθέσιμα δεδομένα που έχουν να κάνουν με την περιγραφή των πελατών και την σχέση τους με την εταιρία. Πληροφορίες που μπορεί να είναι σχετικές εδώ περιλαμβάνουν τα δημογραφικά χαρακτηριστικά των πελατών (Van den Poel et al. 2007), οικονομικές πληροφορίες (Van den Poel et al. 2005), λεκτικές πληροφορίες από e-mails με τους πελάτες (Van den Poel et al. 2008), την ανάλυση δεδομένων (Van den Poel et al. 2004) και λοιπά. Ο δεύτερος παράγοντας αφορά στην τεχνική κατηγοριοποίησης που θα χρησιμοποιηθεί από μια εταιρία για να μοντελοποιήσει την απώλεια των πελατών. Στην παρούσα εργασία, θα αναλυθούν οι πιο διαδεδομένα στην πρόβλεψη της απώλειας πελατών, όπως είναι οι τεχνικές ανάλυσης με Λογιστική Παλινδρόμηση και τα δέντρα αποφάσεων. αλλά και πιο σύγχρονες τεχνικές όπως τα Νευρωνικά δίκτυα και οι μηχανές υποστήριξης διανυσμάτων.

1.1.2. Διατήρηση πελατών

Με τον όρο διατήρηση πελατών εννοούμε τον βαθμό εκείνο στον οποίο μια εταιρία είναι ικανή να διατηρεί τους πελάτες της προσφέροντας τους υψηλό επίπεδο εξυπηρέτησης και ικανοποίησης και γενικά θεωρείται ακρογωνιαίος λίθος στην επιτυχημένη διαχείριση πελατειακών σχέσεων (Customer Relationship Management ,CRM) (Payne et al. 2005 ; Reinartz et al 2004 ; Winer et al. 2001)

Η αναγνώριση ότι κάποιοι πελάτες είναι πιθανότερο να απολεσθούν μπορεί να αποβεί ωφέλιμη για την κερδοφορία της εταιρίας, αφού επιτρέπει στους υπεύθυνους του μάρκετινγκ να κάνουν στοχευμένες κινήσεις. Όπως για παράδειγμα, οι καμπάνιες για την διατήρηση των πελατών μπορούν να περιορίζονται μόνο σε κάποιους πελάτες, αλλά καλύπτοντας τους πελάτες εκείνους που είναι πιο πιθανό να εγκαταλείψουν την εταιρία. Δεύτερον, υψηλά ποσοστά διατήρησης πελατών μειώνουν την πίεση εύρεσης νέων πελατών, το οποίο στοιχίζει πολύ περισσότερο από το να κρατά κανείς το ήδη υπάρχον πελατολόγιο του ικανοποιημένο (Reinartz et al. 2003).

Η έμφαση στη διατήρηση βασίζεται στην υπόθεση ότι υπάρχει μεγάλη σχέση μεταξύ αυτής και της αποδοτικότητας: οι μακροχρόνιοι πελάτες αγοράζουν περισσότερο και είναι λιγότερο δαπανηρό να εξυπηρετηθούν (Hwang et al. 2006), ενώ η αντικατάσταση υπάρχοντων πελατών από νέους είναι πιο ακριβή και ριψοκίνδυνη στρατηγική, αφού είναι εύκολο να υποτεθεί ότι οι νέοι αυτοί πελάτες είναι πιο επιρρεπείς στο να συνεχίσουν την συμπεριφορά μετακίνησής τους στο κοντινό μέλλον, και άρα χαρακτηρίζονται από υψηλή πιθανότητα απώλειας (Colgate et al. 2000).

1.1.3 Αφοσίωση πελατών και η αξία διάρκειας ζωής του πελάτη

Δεν έχουν όλοι οι πελάτες την ίδια σημασία για την επιχείρηση και για αυτό όταν αποφασίζεται η εφαρμογή μιας στρατηγικής διατήρησης πελατών, πρέπει να καθοριστούν οι πελάτες που πρέπει να δοθεί προσοχή ,ποιοι είναι οι αφοσιωμένοι πελάτες στην εταιρία και ποιοι πελάτες με υψηλό ρίσκο απώλειας θα ήταν πλήγμα αν χαθούν.

Οι δύο έννοιες που συνδέονται στενά με την απώλεια πελατών είναι η αφοσίωση των πελατών (custom loyalty) και η αξία διάρκειας ζωής του πελάτη (Customer Lifetime Value,CLV).

➤ Αφοσίωση πελατών

Η αφοσίωση των πελατών, φανερώνει τη τάση συστηματικής επαναγοράς ενός προϊόντος ή υπηρεσίας στο μέλλον και δεν αφορά τις προσπάθειες του μάρκετινγκ που στοχεύουν σε αλλαγή της αγοραστικής συμπεριφοράς (Oliver et al. 1990), η οποία απασχολεί τις επιχειρήσεις, αφού συνδέεται άμεσα με την ικανοποίηση και διατήρηση των πελατών. Οι αφοσιωμένοι πελάτες παραμένουν σε μια εταιρία περισσότερο και εμφανίζουν πρόθεση επαναγοράς (Dick et al. 1994).

Για να γίνει πιο αντιληπτή η σχέση μεταξύ αφοσίωσης και απώλειας πελατών αρκεί να ληφθεί υπόψη η σχέση της μορφής (Hwang et al. 2004 και Kim et al. 2006):

$$\text{Αφοσίωση πελατών} = 1 - \text{ποσοστό απώλειας πελατών}$$

Οι επιχειρήσεις γνωρίζοντας το επίπεδο αφοσίωσης των πελατών τους μπορούν να κατανοήσουν το πώς οι προσπάθειες τους για τη διατήρηση καλών σχέσεων, συμβάλουν στο επίπεδο των κερδών τους.

Οι διευθυντές των επιχειρήσεων δείχνουν να κατανοούν την έννοια της αφοσίωσης, αφού όπως προκύπτει από έρευνα της διαδικτυακής κοινότητας επιχειρηματιών

CRMGuru (www.customerthink.com), το 64% προσδιορίζει την αφοσίωση του πελάτη ως τη συμπεριφορά επαναλαμβανόμενης αγοράς, το 58% ως συμπεριφορά καταναλωτών που προτείνουν μια εταιρία σε φίλους και συναδέλφους, το 54% ως τη συναισθηματική δέσμευση του πελάτη προς την εταιρία και μόνο το 32% των ερωτηθέντων θεώρησε την αφοσίωση ως τη συμπεριφορά του πελάτη εκείνου που

ξοδεύει όλο και περισσότερο με την πάροδο του χρόνου.

Επιπροσθέτως, η αύξηση του επιπέδου αφοσίωσης μπορεί:

- Να μειώσει τα έξοδα μάρκετινγκ, καθώς και το κόστος συναλλαγής.
- Να αυξήσει τις συναλλαγές ανά πελάτη.
- Να βελτιώσει τη φήμη της εταιρίας.

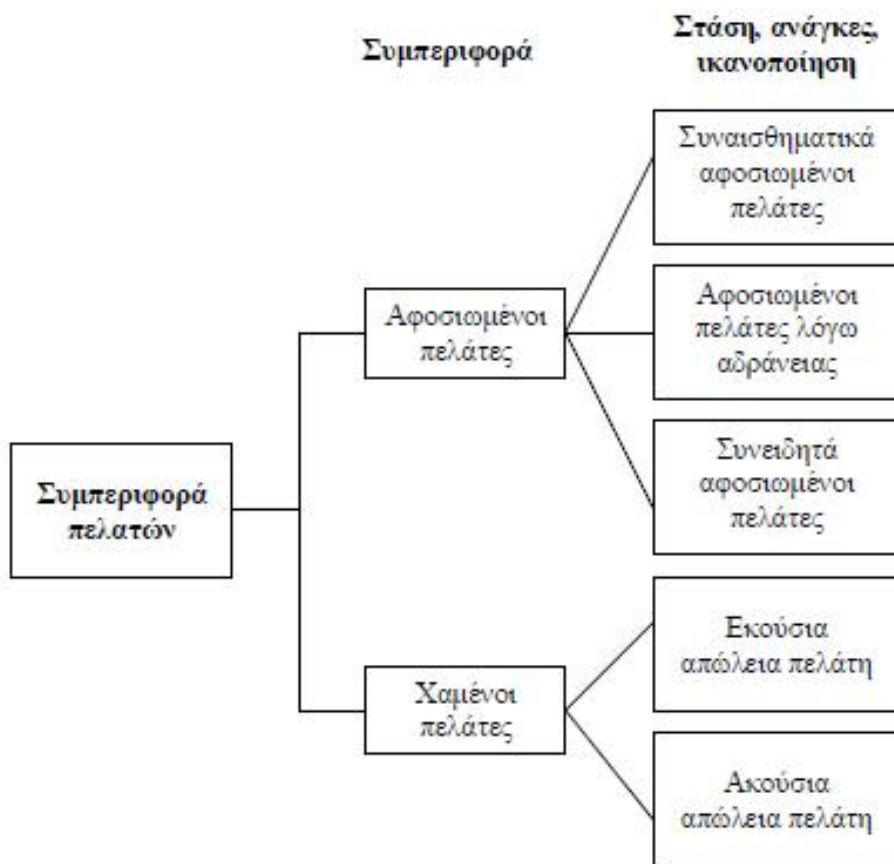
➤ Αξία διάρκειας ζωής του πελάτη (Customer Lifetime Value-CLV)

Η αξία διάρκειας ζωής του πελάτη είναι η ολική αξία των άμεσων και έμμεσων συνεισφορών στα γενικά έξοδα και έσοδα ενός πελάτη καθόλη τη διάρκεια ζωής του με την επιχείρηση (Hoekstra et al. 1999). Η αξία διάρκειας ζωής του πελάτη αντιπροσωπεύει τη συνολική προσδοκώμενη συνεισφορά του πελάτη στα κέρδη της εταιρίας και βασίζεται σε κάποια μέτρα: 1) στον προσδοκώμενο χρόνο συνεργασίας του πελάτη με την εταιρία, 2) τα ανά περίοδο έξοδα (π.χ. μηνιαίως) που πληρώνει ο πελάτης στην εταιρία, 3) τα ανά περίοδο έσοδα που κάνει η εταιρία προκειμένου να παράσχει στον πελάτη μια υπηρεσία. Η εταιρία, μέσα από την καλύτερη κατανόηση της αξίας διάρκειας ζωής του πελάτη μπορεί να μετρήσει τα τωρινά και τα μελλοντικά έσοδα από τους πελάτες, και να καλλιεργήσει τη διατήρηση των πελατών και την αφοσίωσή τους που θα οδηγήσουν σε υψηλότερη κερδοφορία (Jain et al. 2003). Γίνεται φανερό λοιπόν η σημαντικότητα της αξιολόγησης των πελατών μιας επιχείρησης υπό το πρίσμα της αξίας διάρκειας ζωής του πελάτη. Με το συνδυασμό πρόβλεψης απώλειας πελατών και ανάλυσης αξίας διάρκειας ζωής του πελάτη μπορούν να σχεδιαστούν οικονομικά αποδοτικές στρατηγικές διατήρησης πελατών.

1.2. Αιτίες απώλειας πελατών

1.2.1. Διαχωρισμός πελατών

Οι πελάτες κατηγοριοποιούνται με βάση το ύψος των επαναλαμβανόμενων αγορών και το βαθμό αφοσίωσης στο προϊόν ή στην υπηρεσία. Σύμφωνα με αυτά πρωταρχικά χωρίζονται σε αφοσιωμένους και μη και στη συνέχεια χωρίζονται σε περαιτέρω ομάδες με βάση τη στάση και συμπεριφορά τους.



Σχήμα 1: Απώλεια πελατών με βάση τη συμπεριφορά (Kim et al. 2004)

Αφοσιωμένοι πελάτες

Συναισθηματικά αφοσιωμένοι πελάτες (emotive customers): Είναι οι πιο αφοσιωμένοι πελάτες, οι οποίοι μέσα από μια μεγάλη περίοδο καλών αποδόσεων των προϊόντων/υπηρεσιών που έχουν επιλέξει να αγοράσουν, νιώθουν έντονα ότι οι επιλογές τους

και οι αγορές τους είναι οι καλύτερες και πιο έντονες για αυτούς, με αποτέλεσμα σπάνια να επαναξιολογούν τις αγοραστικές τους αποφάσεις και πολύ συχνά να ξοδεύουν παραπάνω. Οι συναισθηματικοί πελάτες έχουν χαμηλότερο επίπεδο απώλειας πελατών από άλλους.

Αφοσιωμένοι πελάτες λόγω αδράνειας (inertial customers): Αυτοί οι πελάτες σπάνια επαναξιολογούν τις αγορές τους και αυτό οφείλεται κυρίως στη συνήθεια και στον περιορισμό που νιώθουν λόγω του υψηλού κόστους αλλαγής εταιρίας. Κάνουν επαναλαμβανόμενες αγορές, αλλά η δέσμευσή τους για το προϊόν είναι μικρή.

Συνειδητά αφοσιωμένοι πελάτες (deliberators): Αυτό το είδος πελατών αποτελεί τη μεγαλύτερη αγοραστική ομάδα. Είναι πληροφορημένοι για αγοραστικές ευκαιρίες, επανεξετάζουν τις επιλογές τους βάση κριτηρίων όπως η τιμή, η ποσότητα και η απόδοση. Τα οφέλη από τον επηρεασμό αυτών των πελατών έχουν διπλάσια αξία σε σχέση με αυτά των δύο προηγούμενων ομάδων, καθώς αυτοί είναι οι πελάτες με τις μεγαλύτερες πιθανότητες απώλειας (Kim et al. 2004).

Μη Αφοσιωμένοι πελάτες

Εκούσια απώλεια πελάτη: Η εκούσια απώλεια πελατών χωρίζεται σε δύο κατηγορίες. Πρώτη είναι η τυχαία, η οποία είναι υπεύθυνη για ένα μικρό ποσοστό της εκούσιας απώλειας, κατά την οποία οι περιστάσεις είναι εκείνες που καθιστούν μη απαραίτητη την αγορά κάποιου προϊόντος ή υπηρεσίας (μετακόμιση, αλλαγή οικονομικής κατάστασης πελάτη, θάνατος). Δεύτερη κατηγορία είναι η εσκεμμένη απώλεια πελατών, όπου ο πελάτης συνειδητά αποφασίζει να αλλάξει εταιρία, διότι η ανταγωνίστρια εταιρία προσφέρει καλύτερα προϊόντα, χαμηλότερη τιμή και για άλλους παράγοντες όπως η ανεπαρκής κάλυψη, ή άσχημη εξυπηρέτηση (Kim et al. 2004). Έρευνα της συμβουλευτικής εταιρίας Booz·Allen (www.boozallen.com), υποστηρίζει ότι οι παράγοντες που σχετίζονται με την τιμή αποτελούν τη σημαντικότερη αιτία για την εκούσια απώλεια πελατών.

Ακούσια απώλεια πελάτη: Οι ακούσια χαμένοι πελάτες, είναι εκείνοι οι πελάτες των οποίων οι υπηρεσίες διακόπηκαν από την ίδια την εταιρία και έτσι είναι οι πιο εύκολο να αναγνωριστούν. Λόγοι για τους οποίους μια εταιρία θα μπορούσε να διακόψει τη σχέση της με ένα πελάτη, είναι η κατάχρηση μιας υπηρεσίας ή η μη κάλυψη των οικονομικών τους υποχρεώσεων. Η ανάλυση που πραγματοποιήθηκε στα πλαίσια της έρευνας της εταιρίας Booz Allen έδειξε ότι οι πελάτες που εμφάνισαν ακούσια απώλεια είχαν μεγαλύτερα επίπεδα μηνιαίας χρήσης σε σχέση με τους ενεργούς πελάτες, και ότι οι πελάτες που έχασαν την υπηρεσία λόγω αθέτησης πληρωμών αντιμετώπισαν αυτό το πρόβλημα επειδή δεν γνώριζαν πώς να ελέγχουν το επίπεδο της χρήσης τους, ώστε να την περιορίσουν πριν να είναι πολύ αργά.

1.2.2. Κίνδυνοι απώλειας πελατών

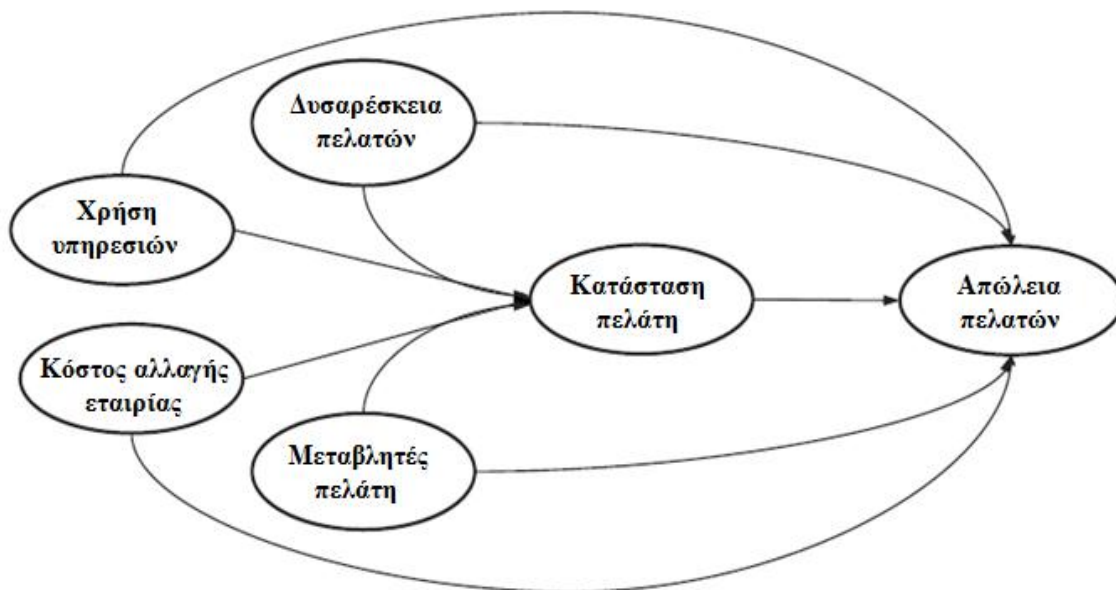
Οι A. Settle και P. Alreck (2000) ορίζουν τον κίνδυνο απώλειας πελατών ως:

- Οικονομικό: Για την αγορά ακριβών προϊόντων που σε εξάρτηση με την οικονομική κατάσταση των αγοραστών, μπορεί να αποτελέσει ανασταλτικό παράγοντα στη λήψη μιας αγοραστικής απόφασης του καταναλωτή.
- Λειτουργικό: Σχετίζεται με την ορθή λειτουργία του προϊόντος και την ανταπόκριση αυτών που ο πωλητής έχει υποσχεθεί στον καταναλωτή.

- Φυσικό: Αν το προϊόν προσφέρει ασφαλή χρήση και δεν θέτει σε κίνδυνο την ζωή και την ασφάλεια του καταναλωτή.
- Κοινωνικό-επαγγελματικό: Ο κοινωνικός και επαγγελματικός κίνδυνος που μπορεί να ελλοχεύει εναντίον στο κοινωνικό status των καταναλωτών. Πιστεύεται ότι οι καταναλωτές που περνούν απαρατήρητοι από το κοινωνικό τους περίγυρο προτιμούν να προβληθούν μέσα από γνωστά προϊόντα που τα χαρακτηρίζει prestige – image. Οι άνθρωποι που κατέχουν κοινωνικά και επαγγελματικά καλύτερη θέση, είναι δεδομένο ότι δεν προβάλλονται μόνο με αυτόν τον τρόπο και ο κίνδυνος που αντιμετωπίζουν είναι ασήμαντος.
- Ψυχολογικό: Σχετίζεται με μια εσφαλμένη επιρροή του πωλητή στον καταναλωτή η οποία οδηγεί συνήθως σε αγορά προϊόντος που τελικά δεν είναι χρήσιμο ή έχει προβλήματα.

1.2.3. Παράγοντες απώλειας πελατών

Παρακάτω περιγράφονται συγκεντρωτικά οι παράγοντες απώλειας πελατών και ο τρόπος επηρεασμού τους:



Σχήμα 2: Μοντέλο απεικόνισης παραγόντων απώλειας πελατών (Ahn et al. 2006)

Δυσαρέσκεια πελατών

Πληθώρα ερευνών έχει δείξει ότι η δυσαρέσκεια των πελατών σχετίζεται με την αύξηση της απώλειας πελατών (Keaveney 1995, Anderson et al. 1993, Boulding et al. 1993, Oliva et al. 1995, Oliver 1997, Reichheld et al. 1990).

Η δυσαρέσκεια με μία υπηρεσία ή ένα προϊόν είναι θετικά συσχετισμένη με τη συμπεριφορά παραπόνων των πελατών και κατά συνέπεια με την απώλεια πελατών (Bearden et al. 1983).

Κόστος αλλαγής εταιρίας

Το κόστος αλλαγής εταιρίας (switching cost) είναι ένας παράγοντας που δρα ως περιορισμός, εμποδίζοντας τους πελάτες από το να αλλάζουν ελεύθερα παρόχους υπηρεσιών.

Χρήση υπηρεσιών

Το επίπεδο χρήσης μιας υπηρεσίας, σε μηνιαία χρέωση, είναι ένας από τους πιο σημαντικούς παράγοντες απώλειας πελατών (Buckinx et al. 2005, Mozer et al. 2000). Οι Oliva et al. (1995) και Oliver (1997) υποστηρίζουν ότι όταν ένας πελάτης χρησιμοποιεί μια υπηρεσία συχνά, αναπτύσσει θετική στάση απέναντί της, και συνεπώς ένας πελάτης με πολλές θετικές εμπειρίες σε μία υπηρεσία δε θα είναι ευαίσθητος σε μεμονωμένες περιπτώσεις αποτυχίας και άρα δε θα έχει μεγάλη πιθανότητα απώλειας.

Μεταβλητές σχετικές με τον πελάτη

- **Ενασχόληση**

Οι πελάτες που ασχολούνται ενεργά με την επιλογή μιας υπηρεσίας ή μιας εταιρίας (involved consumers), έχει βρεθεί ότι εμφανίζουν μεγαλύτερα επίπεδα ικανοποίησης τα οποία και τείνουν να διατηρούνται (Richins et al. 1991). Αυτοί οι πελάτες κάνουν προαγοραστική έρευνα (Beatty et al. 1987) και δείχνουν μεγαλύτερη προσοχή στην επιλογή (Celsi et al. 1988).

- **Δημογραφικά στοιχεία**

Πελάτες με υψηλό εισόδημα και επίπεδο μόρφωσης μπορούν να αναπτύξουν σχετικά ακριβείς εκτιμήσεις του τι να περιμένουν από μια υπηρεσία, αφού για παράδειγμα οι πελάτες με μεγαλύτερο εισόδημα μπορεί να έχουν εμπειρία με πιο συχνή χρήση υπηρεσιών μεγαλύτερης ποικιλίας, γεγονός που τους προσδίδει εμπειρία.

- **Κατάσταση πελατών**

Ο παράγοντας αυτός δεν αποτελεί άμεσα αιτία απώλειας πελατών, αλλά θεωρείται ως ενδιάμεσο στοιχείο μεταξύ των παραγόντων απώλειας πελάτη και της ίδιας της απώλειας, επομένως η αλλαγή της κατάστασης ενός πελάτη μπορεί να αποτελέσει στοιχείο προειδοποίησης για μία εταιρία.

1.3. Διαχείριση πελατειακών σχέσεων (CRM)

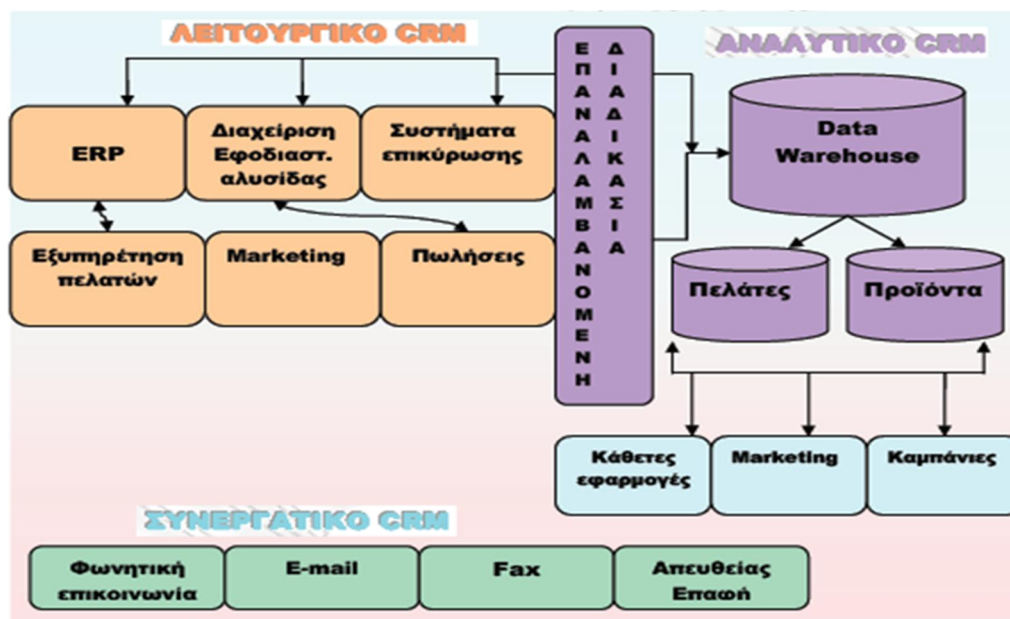
Οι σύγχρονες επιχειρήσεις την τελευταία δεκαετία καλύπτουν την ανάγκη για την επιβίωση σε ένα όλο και πιο ανταγωνιστικό περιβάλλον αγοράς και για την διατήρηση της κερδοφορίας τους, χρησιμοποιώντας Συστήματα Διαχείρισης Πελατών (Customer Relationship Management, CRM), τα οποία έχουν αντικαταστήσει τις παραδοσιακές στρατηγικές μαζικού μάρκετινγκ με επιλεκτικές πρακτικές οι οποίες βασίζονται στη μοντελοποίηση και ανάλυση πληροφοριών στο χρόνο σε σχέση με τα προφίλ των πελατών, δίνοντας τη δυνατότητα για διαχείριση απώλειας πελατών, διαχείριση εκστρατείας (campaign management), και εκτίμηση πιστωτικού κινδύνου (Credit Risk Assessment) (Coussement et al. 2008).

Ο όρος Διαχείριση Πελατειακών Σχέσεων παγκόσμια γνωστός ως Customer Relationship Management ή CRM, αναφέρεται στην επιχειρηματική στρατηγική, που στοχεύει στην μεγιστοποίηση των εσόδων και των κερδών και στην αύξηση της ικανοποίησης των πελατών. Οι τεχνολογίες που υποστηρίζουν την φιλοσοφία CRM συγκεντρώνουν και αποθηκεύουν δεδομένα για τους πελάτες, τους προμηθευτές, τους συνεργάτες και τις εσωτερικές διαδικασίες μιας επιχείρησης. (<http://epixeirein.gr/crm-epixeirisi/>).

1.3.1. Αρχιτεκτονική του CRM

Η δομή ενός CRM συστήματος μπορεί να συνοψιστεί σε τρία βασικά επίπεδα, τα οποία θα πρέπει να λειτουργούν μαζί ως μια ενιαία οντότητα (Rainer Alt, Thomas Puschmann, 2004):

- Λειτουργικό CRM (Operational CRM)
- Συνεργατικό CRM (Collaborative CRM)
- Αναλυτικό CRM (Analytical CRM)



Σχήμα 3: Το περιβάλλον του συστήματος CRM (META Group, Νοέμβριος 2000).

Λειτουργικό CRM (Operational CRM)

Αφορά στις επιχειρηματικές διαδικασίες που θα βελτιστοποιήσουν τη διαχείριση των σχέσεων με τους πελάτες (Chakravorti 2006).

Τα δεδομένα για τους πελάτες συλλέγονται από τα διάφορα σημεία επαφής με τον πελάτη, ύστερα αποθηκεύονται και οργανώνονται σε μια βάση δεδομένων που είναι διαθέσιμα ως ιστορικό πελατών για κάθε συγκεκριμένο πελάτη, σε όλους τους χρήστες μέσα σε μια επιχείρηση.

Ένα τυπικό λειτουργικό CRM περιλαμβάνει υποσυστήματα front-office που υποστηρίζουν τις εξής δραστηριότητες (Phillip Lauren):

- Να διαχειρίζονται ευκολότερα οι προωθητικές ενέργειες (campaign)
- Να αυτοματοποιούνται πολλές λειτουργίες marketing
- Αυτοματοποίηση των πωλήσεων και της παραγγελιοληψίας

Τα λειτουργικά CRM βοηθούν τις επιχειρήσεις να αλληλεπιδράσουν και να επικοινωνήσουν με τους πελάτες συνήθως με την μορφή τηλεφωνικών κέντρων, ιστοσελίδων, direct mails, direct sales και communities. (www.business&decision.com).

Συνεργατικό CRM (Collaborative CRM)

Αφορά στη συνεργασία και στην ενοποίηση των πόρων μέσα σε ένα οργανισμό και ανάμεσα στον οργανισμό, τους συνεργάτες και πελάτες του (Chakravorti et al. 2006).

Το συνεργατικό CRM στοχεύει στην χρησιμοποίηση των πληροφοριών που έχουν συγκεντρωθεί από όλα τα τμήματα της επιχείρησης ώστε να βελτιωθεί η ποιότητα των πελατειακών υπηρεσιών (Phillip Lauren). Παράλληλα, υποστηρίζει διαδικασίες που διεξάγονται σε επίπεδο back-office, οι οποίες επηρεάζουν τις δραστηριότητες των πελατών και την διατήρηση των πελατειακών σχέσεων.

Αναλυτικό CRM (Analytical CRM)

Αφορά εφαρμογή εξεζητημένων αναλυτικών εργαλείων στα δεδομένα των πελατών και των επαφών με αυτούς, ώστε να αποκαλυφθούν μελλοντικές τάσεις και αγοραστικά ή άλλα πρότυπα των πελατών (Chakravorti et al. 2006). Το Αναλυτικό CRM συνιστά την λογική συνέχεια του Λειτουργικού και του Συνεργατικού CRM.

Κατά το Αναλυτικό μέρος της CRM μια εταιρία προσπαθεί να συγκεντρώσει και να αναλύσει δεδομένα που έχουν να κάνουν με τις συναλλαγές της με τους πελάτες, έτσι ώστε να μπορέσει να αποκτήσει μεγαλύτερη κατανόηση της συμπεριφοράς του καταναλωτή, να αναγνωρίσει την πιο επικερδή μερίδα καταναλωτών και να βελτιώσει την αξία αυτών των συγκεκριμένων καταναλωτών.

Το Αναλυτικό CRM πραγματοποιεί (<http://epixeirein.blogspot.com>):

- Στοχευόμενες καμπάνιες marketing.
- Εξειδικευμένες καμπάνιες marketing, με σκοπό το cross-selling και το upselling.
- Ανάλυση της συμπεριφοράς των πελατών, ώστε να υποστηριχτεί η

διαδικασία λήψης αποφάσεων σχετικά με τα προϊόντα και τις προσφερόμενες υπηρεσίες.

- Προβλέψεις των μελλοντικών χρηματοροών.
- Ανάλυση κερδοφορίας (γενικότερα και ανά πελάτη)
- Το Αναλυτικό CRM γενικά κάνει συχνή χρήση του εργαλείου εξόρυξης δεδομένων (data mining).

1.3.2. Οφέλη του CRM

Ένα σύστημα CRM προσφέρει σημαντικά οφέλη στην εταιρία στην οποία εφαρμόζεται με κυριότερα την ανάπτυξη αφοσιωμένων πελατών, οι οποίοι παραμένουν περισσότερο χρόνο ως πελάτες της επιχείρησης, αγοράζουν περισσότερο ή αγοράζουν συχνότερα (Σαρμανιώτης, 2002) επιτυγχάνοντας έτσι και μεγαλύτερη αποδοτικότητα των πελατών (customer profitability) (Tae Hyup Roh, 2005) και την δημιουργία μακροχρόνιων διεπιχειρησιακών δεσμών με τους πελάτες. Ακόμη βοηθά στην ροή της πληροφορίας σε όλο το επιχειρησιακό οργανισμό που έρχεται σε επαφή με τους πελάτες, προσφέροντας έτσι εξαιρετικές υπηρεσίες από κάθε υπάλληλο στους πελάτες της.

Συγκεκριμένα, η ανάγκη για τη χρήση ενός συστήματος CRM γίνεται αντιληπτή από τις επιχειρήσεις αν αναλογιστούν το κόστος απώλειας πελατών. Υπάρχουν δύο συνισταμένες για αυτό το κόστος, η πρώτη αφορά το κόστος απόκτησης νέων πελατών και η δεύτερη αφορά τα διαφυγόντα κέρδη, όπως έδειξε έρευνα του Harvard Business Review αλλά και άλλες ακαδημαϊκές έρευνες και έρευνες που πραγματοποιήθηκαν σε πολλές εταιρίες τα τελευταία χρόνια κατέληξαν στα εξής :

- Κοστίζει έξι φορές περισσότερο να πουλήσεις σε ένα νέο πελάτη από το να πουλήσεις σε έναν υπάρχοντα.
- Ένας τυπικός δυσαρεστημένος πελάτης θα μιλήσει για την άσχημη εμπειρία του σε οκτώ έως δέκα άλλους ανθρώπους.
- Μια εταιρία μπορεί να αυξήσει τα κέρδη της έως και 85%, αυξάνοντας την ετήσια ικανότητα συγκράτησης πελατών μόνο κατά 5%.
- Οι πιθανότητες να πουλήσεις ένα προϊόν σε ένα νέο πελάτη είναι 15%, ενώ οι πιθανότητες να πουλήσεις το ίδιο προϊόν σε έναν υπάρχοντα πελάτη είναι 50%
- Εβδομήντα τοις εκατό των παραπονούμενων πελατών θα συνεργάζονταν ξανά με μια επιχείρηση, εάν αυτή διόρθωνε γρήγορα ένα πρόβλημα που είχε εμφανιστεί στο παρελθόν. (Κουρής, 2000)

1.3.3. Στόχοι του CRM

Προκειμένου να αυξηθεί ο αριθμός των ικανοποιημένων πελατών, επομένως και τα έσοδα από τις πωλήσεις οι επιχειρήσεις μέσω ενός συστήματος CRM στοχεύουν στην ποιοτική και προσωπική αλληλεπίδραση της επιχείρησης με τον πελάτη καθώς επίσης και η αξιοποίηση της γνώσης που αποκτάται μέσω της αλληλεπίδρασης αυτής.

Το CRM μπορεί να δώσει αξιόλογο ανταγωνιστικό προβάδισμα στις επιχειρήσεις που το χρειάζονται και αυτό συμβαίνει κυρίως:

- Λόγω της αυτοματοποίησης πολλών διαδικασιών, άρα και αυτόματης παροχής μέρους των υπηρεσιών της εταιρίας.

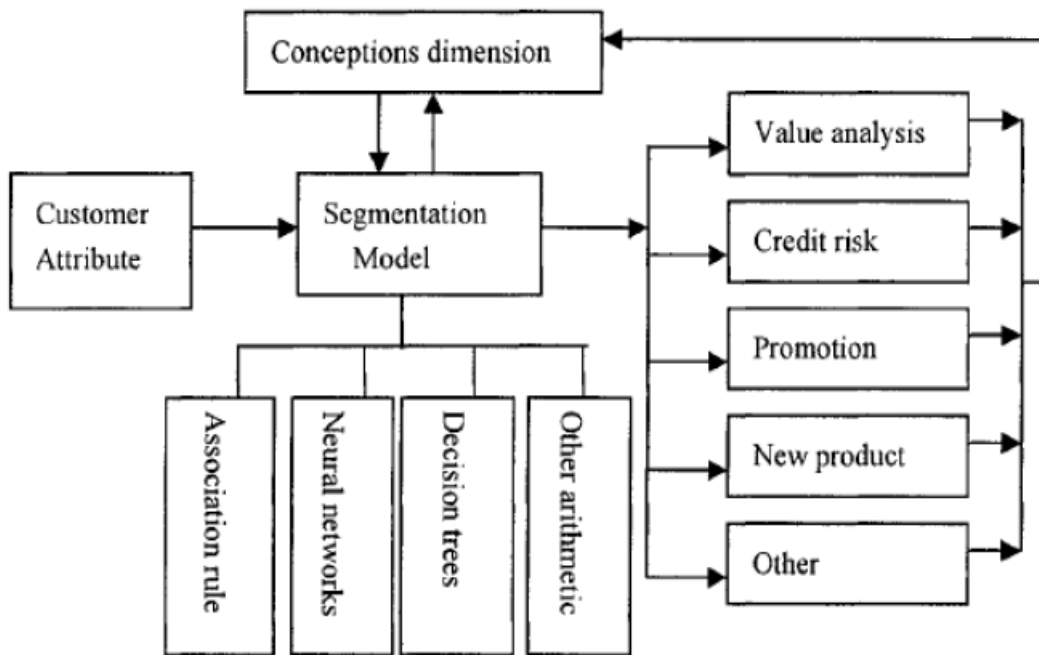
- Λόγω της αξιοποίησης των στατιστικών πληροφοριών που συγκεντρώνονται για τον σχεδιασμό καινούριων πακέτων υπηρεσιών ή την βελτίωση των ήδη υπαρχόντων.
- Λόγω της αποφυγής ανθρώπινων λαθών κατά τις διεργασίες ελέγχου, υπολογισμού κλπ.

Οι Kalakota και Robinson (2000) αναφέρουν τρεις βασικούς στόχους του CRM:

- Ενδυνάμωση της σχέσης «επιχείρηση - πελάτης» με σκοπό την αύξηση των εσόδων. Αυτό σημαίνει προσπάθεια από την πλευρά της επιχείρησης να κατανοήσει τις ανάγκες των πελατών και να προσπαθήσει να προσελκύσει και να διατηρήσει τους καλύτερους πελάτες της.
- Χρησιμοποίηση των πληροφοριών για άριστη εξυπηρέτηση. Αυτό σημαίνει ότι μέσω των πληροφοριών που έχει συγκεντρώσει η επιχείρηση για τους πελάτες είναι δυνατή η καλύτερη εξυπηρέτηση τους (άμεση και αποτελεσματική εξυπηρέτηση).
- Εισαγωγή καναλιών επικοινωνίας με τους πελάτες. Με την αύξηση των καναλιών επικοινωνίας, πολύ περισσότεροι υπάλληλοι εμπλέκονται στις διαδικασίες συναλλαγών. Ανεξάρτητα από το μέγεθος ή πολυπλοκότητα των καναλιών επικοινωνίας, οι επιχειρήσεις πρέπει να βελτιώνουν τις διαδικασίες και την συνέπεια τους στην διαχείριση των πωλήσεων.

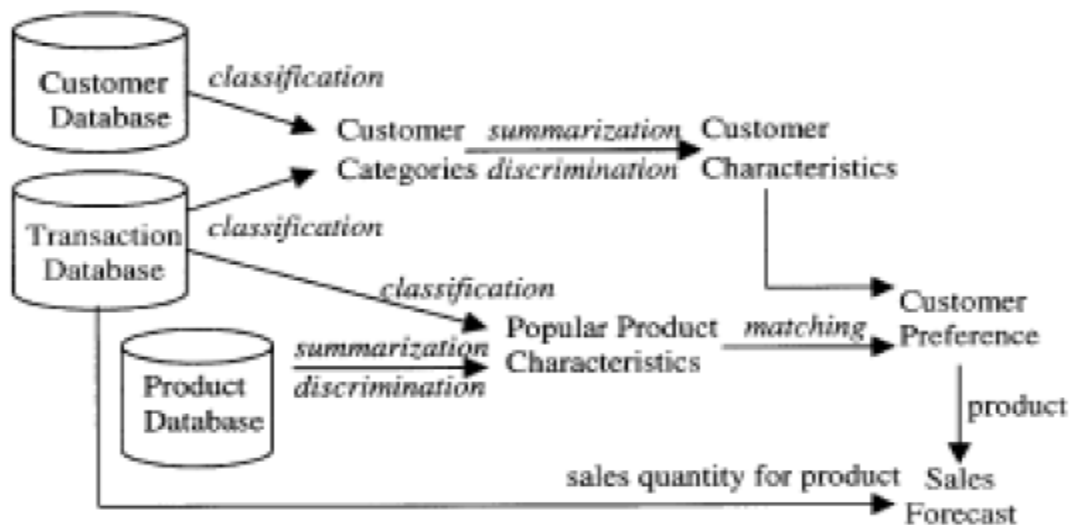
1.3.4. Εξόρυξη δεδομένων (Data Mining) και CRM

Τεχνικές όπως το Data Mining (εξόρυξης δεδομένων) χρησιμοποιούνται συχνά στο αναλυτικό κομμάτι του CRM για να μετατρέπουν όλα τα αδόμητα δεδομένα που έχουν να κάνουν με τις συναλλαγές με τους πελάτες, σε χρήσιμη, δομημένη και πολύτιμη γνώση που θα μπορούσε να ωφελήσει την λήψη αποφάσεων στο μάρκετινγκ και να προβλέψει τα αποτελέσματα του. Βασισμένοι σε τέτοιες τεχνικές εξόρυξης δεδομένων οι πελάτες μπορούν να διαχωριστούν σε ομογενείς ομάδες (segments), με βάση τα κοινά τους χαρακτηριστικά, όπως είναι οι κοινές συνήθειες, προτιμήσεις και άλλα. Πέραν του διαχωρισμού αυτού, οι πελάτες μπορούν επίσης να βαθμολογηθούν σε σχέση με την πιθανότητα που εμφανίζουν να συμπεριφερθούν με ένα συγκεκριμένο τρόπο, αγοράζοντας ένα συγκεκριμένο προϊόν ή να ανταποκριθούν σε μια συγκεκριμένη καμπάνια μάρκετινγκ. Ο διαχωρισμός και η κατανομή των πελατών ονομάζεται τμηματοποίηση των πελατών (customer segmentation), που με την βοήθεια της μια εταιρία μπορεί να προσεγγίσει πελάτες που έχουν επιλεγεί πολύ προσεκτικά, με αποτέλεσμα πετυχημένες εκστρατείες μάρκετινγκ και κρίνεται απαραίτητη προκειμένου να αντιμετωπιστούν οι συνεχώς μεταβαλλόμενες καταναλωτικές αγορές.



Σχήμα 4: Μοντέλο τμηματοποίησης πελατών με βάση την εξόρυξη δεδομένων
(Chen et al. 2006)

Μετά την τμηματοποίηση των πελατών μπορεί να δημιουργηθεί ένα μοντέλο συμπεριφοράς του πελάτη, το οποίο ονομάζεται προφίλ πελατών, που είναι μια βάση η οποία συγκεντρώνει πληροφορίες και περιγράφει κάθε πελάτη ξεχωριστά σύμφωνα με τα ιδιαίτερα αγοραστικά και προσωπικά γνωρίσματα, όπως είναι η ηλικία, δημογραφικά χαρακτηριστικά, το εισόδημα και ο τρόπος ζωής. Μέσω αυτής της βάσης η εταιρίες μπορούν να επικοινωνούν με τους υπάρχοντες πελάτες της και να εντοπίσουν τους δυνητικούς πελάτες με στόχο την διατήρηση τους.



Σχήμα 5: Σύστημα δημιουργίας προφίλ πελατών (Shaw et al. 2001)

Στις μέρες μας, το CRM θα ήταν αδύνατη χωρίς την βοήθεια τεχνικών εξόρυξης δεδομένων. Έτσι, οι ερευνητές προσπαθούν να βελτιώσουν το CRM ενισχύοντας τις τεχνικές του data mining, με αποτέλεσμα οι τεχνικές αυτές να έχουν εξελιχθεί σημαντικά. Οι πιο κοινοί τύποι πελατειακής τμηματοποίησης είναι η ανάλυση RFM (Recency, Frequency, Monetary) με τρεις δείκτες: πρόσφατη εντολή αγοράς, συχνότητα αγοράς και χρηματική αξία αγοράς, και αφορά τις αγοραστικές συνήθειες των πελατών. Χρησιμοποιούνται κυρίως για την βελτίωση της αποδοτικότητας του μάρκετινγκ ώστε να δημιουργήσει υπάρχοντες πελάτες (Hand et al, 2001).

- **Recency:** Αναφέρεται στον αριθμό των μηνών από την τελευταία αγορά. Θεωρείται το πιο ισχυρό από τα τρία χαρακτηριστικά για την πρόβλεψη αντιδράσεων για μια συνακόλουθη πρόσφορα. Και αυτό είναι λογικό. Είναι πιο πιθανό, κάποιος που έκανε μια αγορά από την επιχείρηση, να ξανααγοράσει από ότι κάποιος ο οποίος δεν έκανε καμία πρόσφατη αγορά.
- **Frequency:** Αναφέρεται στον αριθμό των αγορών. Μπορεί να αφορά όλες τις αγορές που πραγματοποιήθηκαν μέσα σε ένα συγκεκριμένο χρονικό περιθώριο ή να αναφέρεται σε όλες τις αγορές. Αποτελεί το δεύτερο σε δύναμη εργαλείο πρόβλεψης των επιχειρήσεων.
- **Monetary:** Αναφέρεται στην συνολική τρέχουσα ποσότητα. Παρόμοια με το frequency, μπορεί να αναφέρεται μέσα σε ένα συγκεκριμένο χρονικό διάστημα για προβλέψεις, όμως αν χρησιμοποιηθεί σε συνδυασμό, προσθέτει καινούριες διαστάσεις κατανόησης.

Αλλά και τεχνικές ταξινόμησης όπως η CHAID (Chi-squared Automatic Interaction Detection) που είναι επί της ουσίας μια τεχνική δέντρου αποφάσεων. Στην πράξη χρησιμοποιείται για να επιλέγουν επικερδείς ομάδες καταναλωτών. Τα συστήματα CHAID ανιχνεύουν την αλληλεπίδραση μεταξύ των μεταβλητών στο σύνολο στοιχείων.

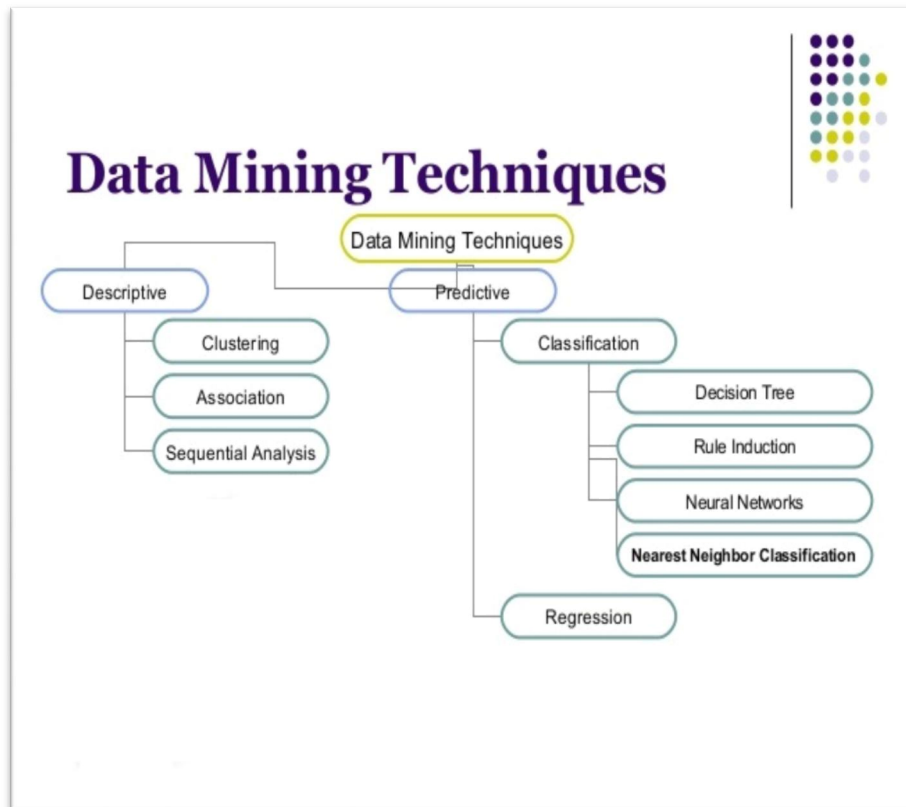
Οι Chaffey et al. (2003) λένε ότι οι τεχνικές της εξόρυξης δεδομένων στα συστήματα CRM δίνουν τις εξής δυνατότητες:

- Αυτοματοποίηση της διαδικασίας της πώλησης- οι πωλητές μπορούν να αυξήσουν το επίπεδο της αποτελεσματικότητάς τους αφού έχουν πλέον πιο ουσιαστική και στοχευόμενη πληροφόρηση σχετικά με τους πελάτες τους
- Καλύτερο μάντζμεντ της υπηρεσίας (Customer service management) – οι εταιρείες προσεγγίζουν καλύτερα τις απαιτήσεις των πελατών
- Ανάλυση δεδομένων – η εξόρυξη δεδομένων προσφέρει πολύ μεγάλες δυνατότητες ανάλυσης της συμπεριφοράς των αγοραστών και οι επιχειρήσεις μπορούν να προσωποποιήσουν το μείγμα μάρκετινγκ τους.
- Διαχείριση της διαδικασίας της πώλησης – αυτό συμβαίνει κυρίως σε περιβάλλοντα B2B (Business to Business)

Το data mining είναι η «ψυχή» του CRM, γιατί επιπρόσθετα παρέχει την δυνατότητα της ανάπτυξης κερδοφόρων μακροχρόνιων σχέσεων με τους πελάτες. Αυτό σημαίνει ότι με την ενίσχυση της τεχνολογίας και της εξόρυξης δεδομένων ενισχύεται η ποιότητα της σχέσης του καταναλωτή και της επιχείρησης (Chaffey et al. 2003).

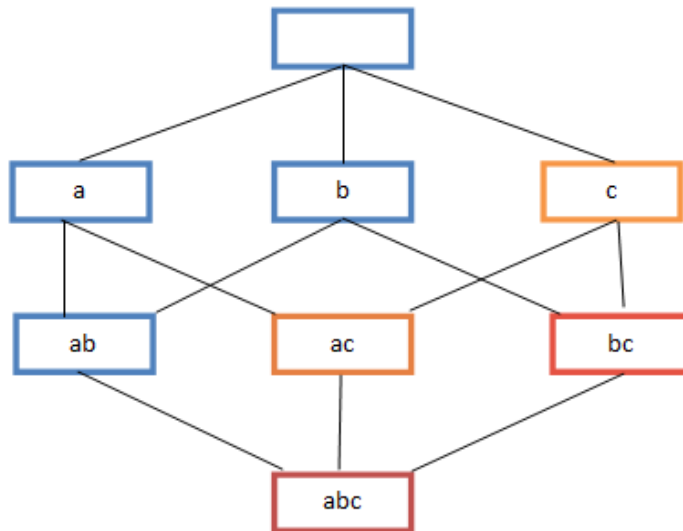
1.3.4.1. Κυριότερες τεχνικές εξόρυξης δεδομένων

Η εξόρυξη δεδομένων παράγει τα ακόλουθα είδη μοντέλων πληροφοριών:



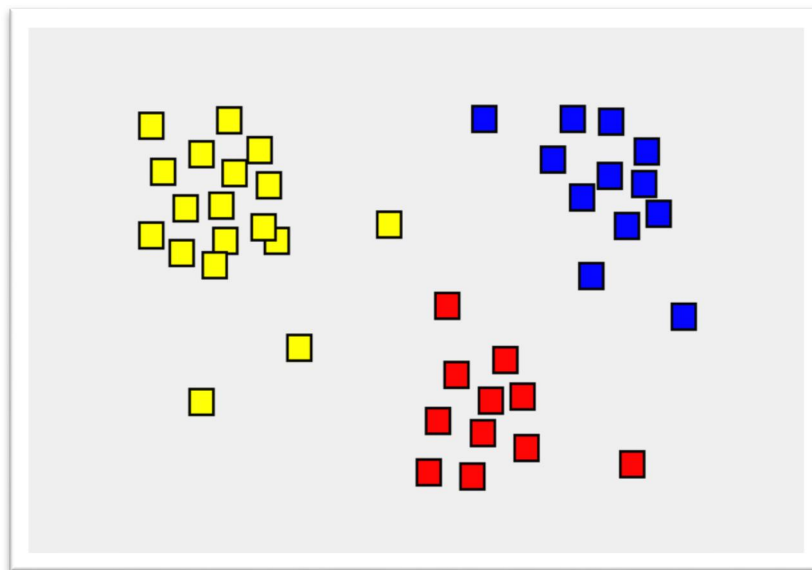
Σχήμα 6: Τεχνικές εξόρυξης δεδομένων στο CRM
(<http://www.slideshare.net/shyaaminibaludata-mining-techniques-for-crm>)

- **Συσχετίσεις (associations):** Η τεχνική της συσχέτισης προσπαθεί να βρει συσχετίσεις μεταξύ των μεταβλητών σε ένα σύνολο δεδομένων. Οι συσχετίσεις συμβαίνουν όταν οι μεταβλητές συνδέονται μεταξύ τους με ένα μεμονωμένο γεγονός. Η αγορά ενός προϊόντος, όταν αγοράζεται μαζί με ένα άλλο προϊόν αντιπροσωπεύει ένα κανόνα συσχέτισης. Χρησιμοποιώντας τον κανόνα συσχέτισης, οι επιχειρήσεις μπορούν να υπολογίζουν ποια προϊόντα αγοράζονται συνήθως μαζί και μπορούν να χρησιμοποιούν αυτή την πληροφορία για αγοραστικούς σκοπούς.



Σχήμα 7: Συσχετίσεις (http://en.wikipedia.org/wiki/Association_rule_learning)

- **Ακολουθίες (sequences):** Οι ακολουθίες σχηματίζονται από δεδομένα που σχετίζονται με την έννοια του χρόνου, όπως οι ακολουθίες των επισκέψεων σε ένα web site.
- **Κατηγοριοποίηση (classification):** Είναι η πιο διαδεδομένη δραστηριότητα εξόρυξης δεδομένων, η οποία είναι ικανή να προβλέψει ποιοι πελάτες είναι πιθανό να στραφούν σε ανταγωνιστές και μπορεί να βοηθήσει στη διατήρησή τους. Χρησιμοποιείται σε περιπτώσεις εταιριών που έχουν ένα μεγάλο όγκο ιστορικών δεδομένων.
- **Συσταδοποίηση (clustering):** Είναι η τμηματοποίηση ενός συνόλου δεδομένων, που μπορεί να εφαρμοστεί σε πολύπλοκα προβλήματα. Σκοπός είναι η δημιουργία συστάδων (ομάδων) αντικειμένων, με όσο το δυνατόν περισσότερα όμοια χαρακτηριστικά, που περιγράφουν την κάθε ομάδα και την ξεχωρίζουν από άλλες.



Σχήμα 8: Το αποτέλεσμα της ανάλυσης συστάδων εμφανίζεται με το χρωματισμό των τετραγώνων σε τρεις συστάδες (http://en.wikipedia.org/wiki/Cluster_analysis)

- **Πρόγνωση (forecasting):** Είναι μια διαφορετική μορφή πρόβλεψης, που εκτιμά την μελλοντική αξία των συνεχών μεταβλητών.
- **Παλινδρόμηση (regression):** Εισάγει ένα δεδομένο σε μιας πραγματικής αξίας μεταβλητή πρόβλεψης, ώστε να βρεθεί ζητούμενη συνάρτηση που θα κατηγοριοποιεί τα νέα δεδομένα.
- **Σύνοψη (summarize):** Παρέχει μια συμπαγή σύνοψη των δεδομένων.

1.3.5. Κινήσεις στρατηγικής CRM

Για να μπορέσει μια επιχείρηση να έχει τη μακροπρόθεσμη επίδοση που επιθυμεί θα πρέπει όχι απλά να χρησιμοποιεί ένα σύστημα CRM, αλλά να εκπονήσει μια στρατηγική CRM. Αυτή η στρατηγική ονομάζεται Στρατηγική Πελατειακών Σχέσεων (Customer Relationship Strategy, CRS) και περιλαμβάνει όλες τις λειτουργίες και τις δραστηριότητες της εταιρίας στην οποία θα εφαρμοστεί. Η στρατηγική CRM είναι απαραίτητο να ευθυγραμμίζεται πλήρως με την επιχειρησιακή στρατηγική. Αναφέρεται στη δημιουργία μιας ενιαίας εικόνας για τον πελάτη στην οποία είναι δυνατή η πρόσβαση από όλα τα τμήματα της επιχείρησης μέσα από πολλαπλά κανάλια επικοινωνίας (Greenberg, 2004). Η επιχείρηση χαράζει τη στρατηγική CRM και θέτει επιμέρους στόχους, η επίτευξη των οποίων θα οδηγήσει στη στρατηγική επιτυχία, η οποία επέρχεται με τη σωστή υλοποίησή του στρατηγικού σχεδιασμού.

Ένα από τα εργαλεία αξιολόγησης της απόδοσης του CRM, που δημιουργήθηκε από τους Kaplan και Norton το 1992, είναι η θεωρία της ισορροπημένης κάρτας επίδοσης (Balance Scorecard). Το μοντέλο του CRM Scorecard είναι χρήσιμο για την διαμόρφωση και την υλοποίηση της στρατηγικής CRM και βοηθά την επιχείρηση να καταλάβει που ήταν, που βρίσκεται και που θέλει να βρεθεί, με τη χρήση δεικτών και ενός δομημένου συστήματος επίδοσης της επιχείρησης.

2. ΜΕΘΟΔΟΙ ΤΑΞΙΝΟΜΗΣΗΣ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΑΠΩΛΕΙΑΣ ΠΕΛΑΤΩΝ

Στην ενότητα αυτή αναφέρονται οι τεχνικές και τα κριτήρια αξιολόγησης που είναι ικανές να αντιμετωπίσουν το πρόβλημα της πρόβλεψης απώλειας πελατών.

Οι τεχνικές που θα παρουσιαστούν είναι :

- Λογιστική Παλινδρόμηση (Logistic Regression)
- Δέντρα Απόφασης (Decision Trees)
- Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)
- Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines).

Και τα κριτήρια αξιολόγησης :

- Καμπύλη AUC
- Καμπύλη Top Decile Lift

2.1. Λογιστική Παλινδρόμηση

Η Λογιστική Παλινδρόμηση (Logistic Regression) είναι μια ευρέως διαδεδομένη τεχνική ταξινόμησης για την πρόβλεψη της πιθανότητας εμφάνισης ενός γεγονότος, όπου χρησιμοποιείται συχνά στην ανάλυση πελατών και υλοποιείται εντός των πλαισίων της εξόρυξης δεδομένων, αναπτύσσοντας μια μαθηματική σχέση η οποία προβλέπει την μελλοντική συμπεριφορά των πελατών. Είναι η παλαιότερη, γνωστή και απλή τεχνική, με σημαντική ωφέλεια σε σχέση με άλλες μεθόδους την ερμηνεία της. Παράγει συγκεκριμένες πληροφορίες σχετικά με το μέγεθος και την κατεύθυνση της επίδρασης ανεξάρτητων μεταβλητών. Επίσης, σχετικά με την ικανότητα πρόβλεψης και την στατιστική της ανθεκτικότητα, η λογιστική παλινδρόμηση μπορεί να ανταγωνιστεί πιο προχωρημένες τεχνικές εξόρυξης δεδομένων.

Ανήκει σε μια ομάδα μοντέλων που ονομάζονται Γενικευμένα Γραμμικά Μοντέλα (Generalized Linear Models, GLMs). Τα GLMs εφαρμόζουν παλινδρόμηση συνήθων ελάχιστων τετραγώνων (Ordinary Least Squares Regression) στις άλλες εξαρτώμενες μεταβλητές, όπως δίτιμα αποτελέσματα, χρησιμοποιώντας μια συνδετική συνάρτηση (link function).

Στο μοντέλο της λογιστικής παλινδρόμησης, η εξαρτημένη μεταβλητή είναι διχοτομική δηλαδή μπορεί να πάρει:

- την τιμή 1 με πιθανότητα
$$P(\mathbf{x}) = \frac{e^{-(\alpha+\beta x)}}{1+e^{-(\alpha+\beta x)}}$$
- και την τιμή 0 με πιθανότητα
$$Q(\mathbf{x})= 1-P(\mathbf{x}) = \frac{1}{1+e^{-(\alpha+\beta x)}}$$

Άρα η έξοδος του γραμμικού μοντέλου παλινδρόμησης με n εισόδους προκύπτει από τον τύπο:

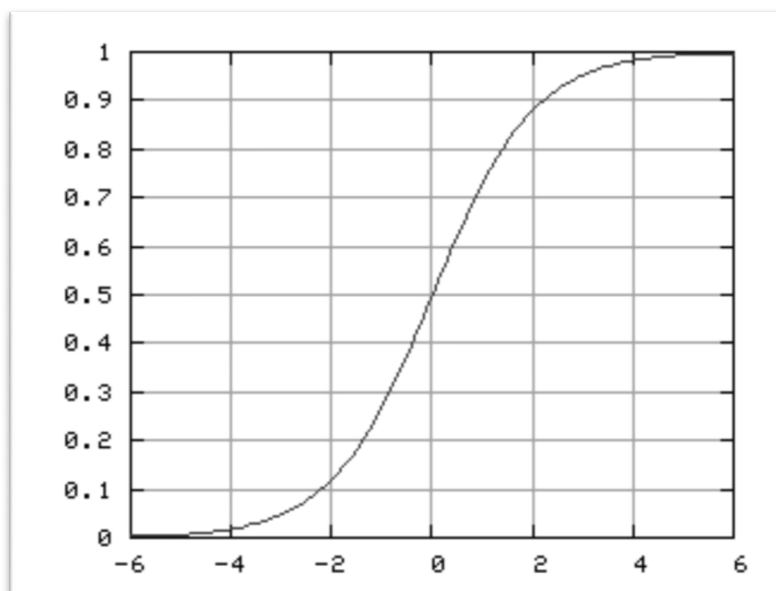
$$P(\mathbf{x}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Ο αντίστοιχος τύπος λογιστικής συνάρτησης στο μοντέλο λογιστικής παλινδρόμησης είναι:

$$Q(\mathbf{x}) = \frac{1}{1+e^{-(\alpha+\beta_1 x_1+\beta_2 x_2+\dots+\beta_n x_n)}} \quad \text{ή} \quad f(z) = \frac{1}{1+e^{-z}},$$

με είσοδο τη λογιστική (logit) μεταβλητή $z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$, (με σταθερό όρο α την τιμή του z) η οποία αντιπροσωπεύει την συνολική συνεισφορά του κάθε παράγοντα ρίσκου που χρησιμοποιήθηκε στο μοντέλο και έξοδο το $f(z)$, το οποίο αντιπροσωπεύει την πιθανότητα ενός αποτελέσματος, του συνόλου των παραγόντων ρίσκου (risk factors).

Τα δεδομένα μπορούν να προσαρμόζονται σε μια λογιστική καμπύλη (logistic curve)



Σχήμα 9: Παράδειγμα της λογιστικής συνάρτησης προσαρμοσμένη στη λογιστική καμπύλη (http://en.wikipedia.org/wiki/Logistic_regression).

Στην λογιστική παλινδρόμηση ένα σημαντικό στοιχείο είναι η συνάρτηση απώλειας (loss function), η οποία είναι ένα μέτρο προσαρμογής ανάμεσα στο μαθηματικό μοντέλο και στα δεδομένα. Η συνάρτηση που επιλέγεται είναι η μέγιστη λογαριθμική συνάρτηση πιθανοφάνειας (maximum loglikelihood function), όπου οι παράμετροι υπολογίζονται χρησιμοποιώντας τεχνικές μη γραμμικής βελτιστοποίησης, ώστε να πραγματοποιηθεί η κατάλληλη προσαρμογή στα δεδομένα.

Οι πελάτες μπορούν να βαθμολογηθούν σχετικά με την πιθανότητα να αγοράσουν ένα προϊόν, με τις πιθανότητες οι οποίες κυμαίνονται από το 0 έως το 1, τις οποίες θα δώσουν οι παρακάτω συναρτήσεις :

- $\pi_i = \frac{e^n}{1+e^n}$

- $n = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni}$

Όπου, το π_i συμβολίζει την εκ των υστέρων πιθανότητα αγοράς από έναν πελάτη i , το x_{ni} συμβολίζει τις ανεξάρτητες μεταβλητές για τον πελάτη i , το β_0 το σημείο τομής με τον κατακόρυφο άξονα, το β_n τις παραμέτρους που πρέπει να υπολογιστούν και το n τον αριθμό των ανεξάρτητων μεταβλητών.

Η μέθοδος της λογιστικής παλινδρόμησης είναι η πιο γνωστή τεχνική ταξινόμησης για την πρόβλεψη γεγονότων. Μερικά από τα πλεονεκτήματα της είναι:

- Μοντελοποίηση με την χρήση της, είναι γνωστή, απλή, εύκολη στην ερμηνεία σε σχέση με άλλα πιο σύγχρονα μοντέλα, όπως τα νευρωνικά δίκτυα.
- Είναι ικανή να γενικεύει τα γραμμικά μοντέλα, έτσι ώστε η εξαρτημένη μεταβλητή να ακολουθεί την εκθετική κατανομή.
- Η λογιστική παλινδρόμηση είναι ένας χρήσιμος τρόπος να περιγραφεί η σχέση μεταξύ μίας ή περισσότερων μεταβλητών κατά την πρόβλεψη απώλειας πελατών (πελατειακή σχέση, ηλικία, φύλλο κτλ) και ενός αποτελέσματος, εκφρασμένη ως πιθανότητα.
- Έχει αποδειχτεί σε σύγκριση με άλλες τεχνικές ταξινόμησης, σε μελέτες για την πρόβλεψη απώλειας πελατών, ότι παρέχει καλά και εύρωστα αποτελέσματα (Neslin et al. 2004)

2.2. Δέντρα Απόφασης

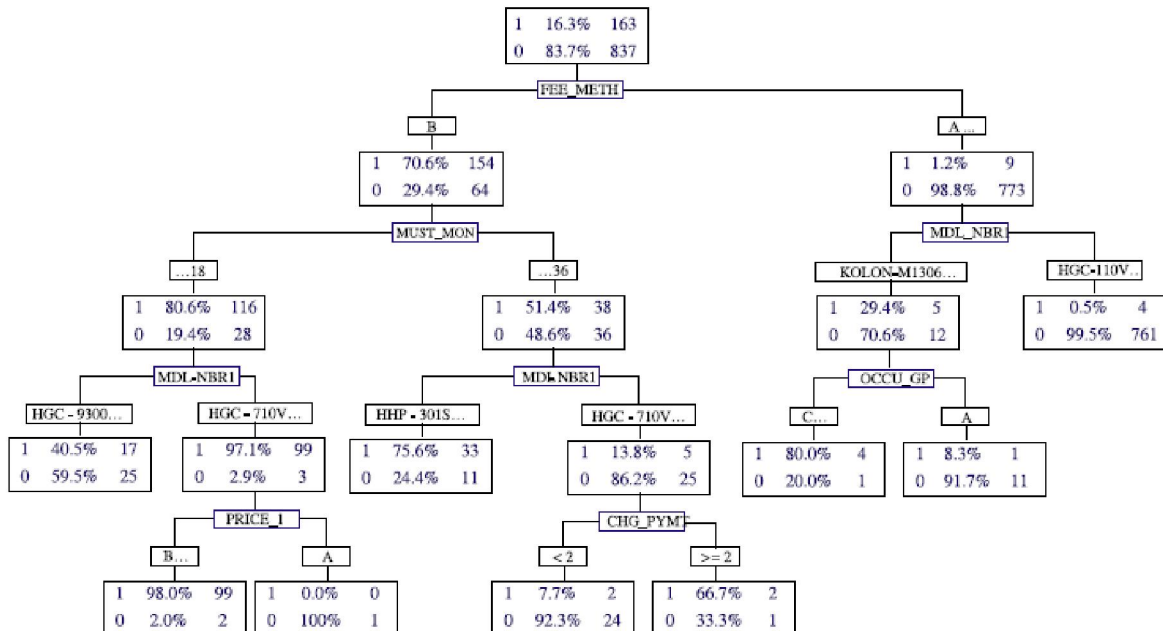
Τα δέντρα απόφασης (decision trees) είναι δομές σε σχήμα δέντρου που ταξινομούν τα αντικείμενα μιας βάσης δεδομένων στην πρόβλεψη μελλοντικών γεγονότων, όπως η πιθανότητα απώλειας πελατών, βάσει των τιμών των χαρακτηριστικών αυτών, μέσα από την εξαγωγή if_then κανόνων ταξινόμησης (Muata et al 2004 & Murphy, 1998). Αποδίδουν αποτελεσματικά στην ανάλυση μεγάλου όγκου δεδομένων, καθώς μπορούν να αποδώσουν σε μεγάλες βάσεις, με μικρή προετοιμασία δεδομένων, χωρίς να γίνει κανονικοποίηση δεδομένων, αφού το μέγεθος του δέντρου που δημιουργείται δεν εξαρτάται από το μέγεθος της βάσης.

Σε αυτές τις δομές, η ταξινόμηση ενός αρχείου ακολουθεί κάποια βήματα. Ένα αρχείο εισέρχεται στο δέντρο ξεκινώντας από τον κόμβο της ρίζας (αρχικός κόμβος) και έπειτα καθορίζεται από αυτό τον κόμβο, ποιοι εσωτερικοί κόμβοι θα ακολουθηθούν ως ότου το αρχείο καταλήξει σε έναν κόμβο φύλλων. Τα φύλλα αντιπροσωπεύουν τις ταξινομήσεις και τα κλαδιά τους συνδέσμους των χαρακτηριστικών που οδηγούν σε αυτές τις ταξινομήσεις.

Η ανάπτυξη τους αποτελείται από δύο φάσεις, την ανάπτυξη (tree building) και τη περικοπή (tree pruning) του δέντρου. Στην φάση της ανάπτυξης του δέντρου, το σύνολο των

δεδομένων εκμάθησης χωρίζεται πολλές φορές σύμφωνα με τις τιμές των χαρακτηριστικών τους, μέχρι όλα τα αντικείμενα σε ένα τμήμα του ανωτέρω συνόλου να ανήκουν στην ίδια κλάση, έχοντας την ίδια τιμή στην εξαρτημένη μεταβλητή.

Στην συνέχεια, κατά την φάση της περικοπής του δέντρου, περικόπτονται τα κλαδιά που περιέχουν το μεγαλύτερο ποσοστό λάθους, ώστε να μειωθεί η πολυπλοκότητά του και το δέντρο να έχει υψηλότερη ακρίβεια πρόβλεψης στην απώλεια πελατών (Au et al 2003).



Σχήμα 10: Το αποτέλεσμα του δέντρου απόφασης για την ανάλυση κατηγορίας πελατών με χαμηλής- αφοσίωσης (Kim et al. 2006).

Υπάρχουν διάφοροι μέθοδοι για την υλοποίηση των δέντρων απόφασης, όπως :

- **CHAID** : (CHi-squared Automatic Interaction Detection) που σημαίνει Τετραγωνισμένος Αυτόματος Ανιχνευτής Αλληλεπίδρασης. Είναι μια από τις παλαιότερες μεθόδους ταξινόμησης δέντρου. Δημιουργεί δέντρα απόφασης προσδιορίζοντας την βέλτιστη διάσπαση πολλαπλών επιπέδων, με την χρήση τεχνικών ελάχιστων τετραγώνων.
- **CART** : (Classification and Regression Tree) δηλαδή μέθοδος Ταξινόμησης και Παλινδρόμησης (C&R). Κατασκευάζει δέντρα απόφασης για να προβεί σε προβλέψεις ή να ταξινομήσει τις μελλοντικές παρατηρήσεις. Διαιρεί τις εγγραφές του συνόλου των δεδομένων βάσει των τιμών εξόδου, ώστε οι εγγραφές με την ίδια τιμή να συγκροτούν ένα τελικό κόμβο.
- **QUEST** : (Quick, Unbiased, Efficient Statistical Tree). Γρήγορα, αμερόληπτα, ικανά στατιστικά δέντρα. Είναι μέθοδος δυαδικής κατηγοριοποίησης, για την κατασκευή δέντρων απόφασης, αφού έχει δημιουργηθεί με στόχο να μειώσει το μεγάλο χρόνο επεξεργασίας. Χρησιμοποιεί μια σειρά κανόνων που βασίζονται σε test-σημαντικότητας significance για την αξιολόγηση των μεταβλητών πρόβλεψης.

Μερικά πλεονεκτήματα από τη χρήση δέντρων αποφάσεων κατηγοριοποίησης είναι :

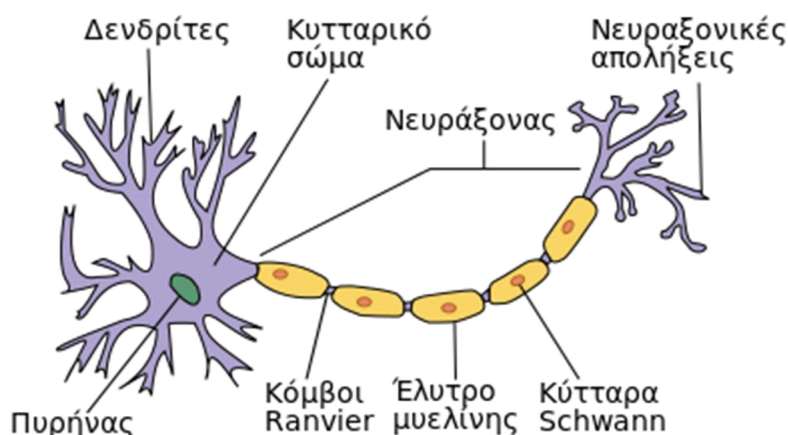
- Είναι αποτελεσματικά προβλεπτικά μοντέλα, απλά στην κατανόηση, την χρήση και την ερμηνεία που αποδίδουν καλά στην ανάλυση μεγάλου όγκου δεδομένων σε λίγο χρόνο.
- Μπορούν να λειτουργήσουν αρκετά καλά σε μεγάλες βάσεις δεδομένων λόγω του ότι το μέγεθος του δένδρου είναι ανεξάρτητο από το μέγεθος της βάσης.
- Η αξιοπιστία ενός μοντέλου δέντρου απόφασης μπορεί να ελεγχθεί με
- χρήση στατιστικών ελέγχων.
- Σε αντίθεση με άλλες τεχνικές που απαιτούν κανονικοποίηση δεδομένων ή δημιουργία ψευδομεταβλητών τα δέντρα απόφασης απαιτούν μικρή προετοιμασία δεδομένων.
- Όταν το δέντρο κατασκευαστεί, η ταξινόμηση νέων εγγραφών είναι πολύ γρήγορη.

Υπάρχουν όμως και κάποιοι περιορισμοί από τη χρήση των δέντρων απόφασης, όπως:

- Δεν μπορεί να χειριστεί περίπλοκες σχέσεις μεταξύ των γνωρισμάτων που αποτελούνται από συνεχείς τιμές.
- Μέσα από την εκπαίδευση δέντρων απόφασης δημιουργούνται σύνθετα δέντρα χωρίς δυνατότητες γενίκευσης και έτσι υπάρχει η πιθανότητα υπερ-προσαρμογής (overfitting).
- Υπάρχει πρόβλημα όταν λείπουν πολλά δεδομένα.

2.3. Τεχνητά Νευρωνικά Δίκτυα

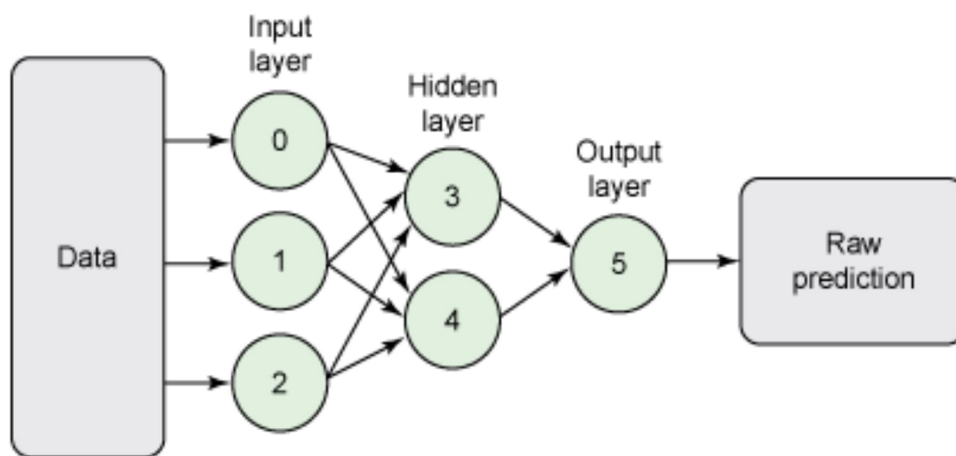
Μια άλλη διαδεδομένη μέθοδος που εφαρμόζεται στην πρόβλεψη και την ταξινόμηση είναι τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks -ANN), τα οποία αναπτύχθηκαν μέσα από έρευνα στην τεχνική νοημοσύνη και είναι βασισμένα στη δομή και τις συσχετίσεις των βιολογικών νευρωνικών δικτύων.



Σχήμα 11: Σχηματικό διάγραμμα ενός τυπικού βιολογικού νευρώνα (<http://el.wikipedia.org>)

Είναι πολύ χρήσιμα εργαλεία που έχουν την ικανότητα να εξάγουν συμπεράσματα από πολύπλοκα δεδομένα. Ξεπερνούν τα στατιστικά – οικονομικά μοντέλα, αφού μπορούν να παρέχουν μια πρόβλεψη με την πιθανότητά της σε αντίθεση με άλλες τεχνικές ταξινόμησης όπως τα δέντρα απόφασης (Fadalla et al. 2001). Έρευνες δείχνουν ότι τα νευρωνικά δίκτυα αποδίδουν καλύτερα από τα δέντρα απόφασης και τη λογιστική παλινδρόμηση σε προβλήματα πρόβλεψης απώλειας πελατών (Au et al. 2003).

Τα νευρωνικά δίκτυα είναι ένα υπολογιστικό μοντέλο που αποτελείται από ένα σύνολο νευρώνων, οι οποίοι είναι τα δομικά στοιχεία του νευρωνικού δικτύου στα οποία τελείται όλη η επεξεργασία της πληροφόρησης, όπου κάθε τέτοιο στοιχείο δέχεται δεδομένα από άλλους νευρώνες ή από το περιβάλλον, υπολογίζει με βάση τις εισόδους αυτών των δεδομένων και παράγει μια έξοδο. Ένα τεχνητό νευρωνικό δίκτυο αποτελείται από ένα σύνολο κόμβων συνδεδεμένων μεταξύ τους σε τρία αλληλένδετα επίπεδα (layers).

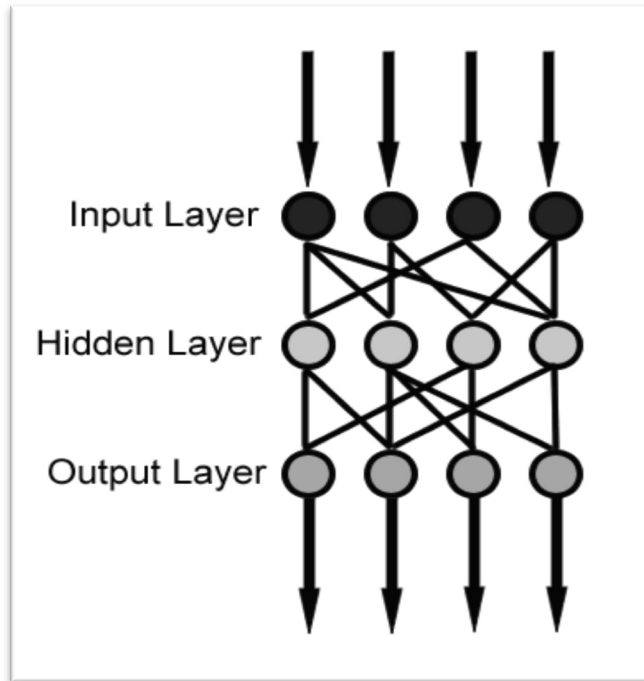


Σχήμα 12: Δομή νευρωνικού δικτύου με ένα κρυμμένο επίπεδο (<http://www.ibm.com/developerworks/library/ba-ind-PMML1/>).

Το πρώτο επίπεδο είναι το επίπεδο (input layer), όπου αποτελείται από την είσοδο των νευρώνων όπου και εισάγονται τα δεδομένα. Τα δεδομένα αυτά στέλνονται για ανάλυση στο/α ενδιάμεσο/α επίπεδο/α (hidden layer) και τέλος από εκεί στέλνεται η μεταβλητή εξόδου στο επίπεδο εξόδου (output layer) το οποίο και δίνει την πρόβλεψη (raw prediction).

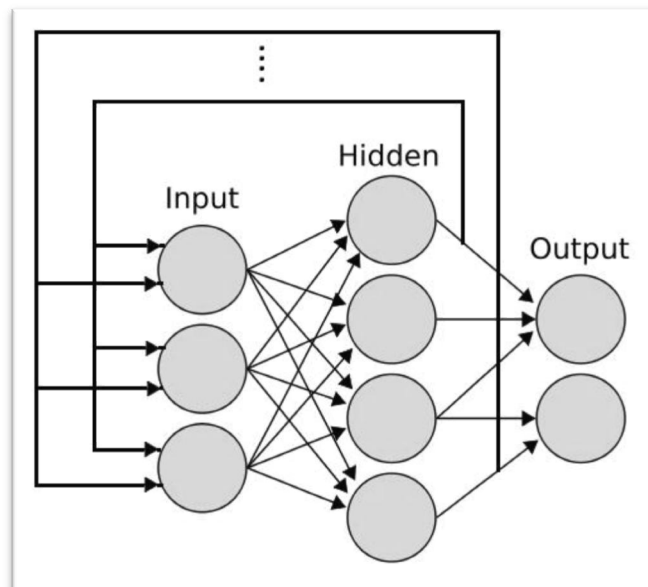
Τα τεχνικά νευρωνικά δίκτυα χωρίζονται σε δύο βασικές κατηγορίες:

- **Εμπρόσθιας τροφοδότησης Νευρωνικά Δίκτυα (Feed-forward):** Σε ένα νευρωνικό δίκτυο με προς τα εμπρός τροφοδοσία η πληροφορία ρέει μόνο προς μία κατεύθυνση, από το επίπεδο εισόδου, μέσω κάποιων κρυφών νευρώνων, προς το επίπεδο εξόδου. Οι νευρώνες εισόδου δεν κάνουν καμία επεξεργασία, παρά μόνο μεταφέρουν την πληροφορία στο επόμενο επίπεδο. Ενώ οι κρυφοί και οι εξόδου νευρώνες είναι υπολογιστικοί νευρώνες που ακολουθούν το μοντέλο του νευρώνα.



Σχήμα 13: Νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης
(http://en.wikipedia.org/wiki/Feedforward_neural_network).

- **Αναδραστικά Νευρωνικά Δίκτυα (Recurrent):** Είναι ένα επαναλαμβανόμενο νευρωνικό δίκτυο όπου οι συνδέσεις μεταξύ των μονάδων αποτελούν ένα κατευθυνόμενο κύκλο, όπου η πληροφορία επιστρέφει σε προηγούμενο επίπεδο που με την σειρά του ανατροφοδοτεί το ενδιάμεσο επίπεδο. Υπάρχει τουλάχιστον ένας κόμβος ανάδρασης. Η έξοδος κάθε νευρώνα ανατροφοδοτεί την είσοδο των άλλων νευρώνων ίδιου επιπέδου. Σε αντίθεση με τα δίκτυα feed-forward, τα αναδραστικά δίκτυα μπορούν να χρησιμοποιήσουν την εσωτερική μνήμη τους για να επεξεργαστούν αυθαίρετες ακολουθίες εισόδου.



Σχήμα 14: Αναδραστικό νευρωνικό δίκτυο
(http://en.wikibooks.org/wiki/Artificial_Neural_Networks/Print_Version)

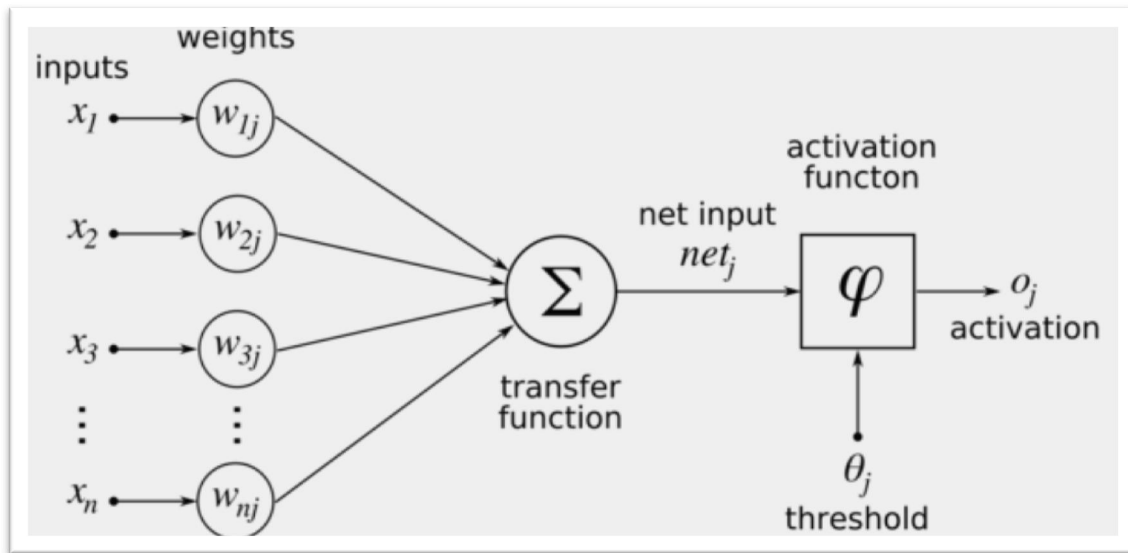
Με μαθηματικούς όρους ο νευρώνας μπορεί να περιγραφεί ως εξής (http://el.wikipedia.org/wiki/Νευρωνικό_δίκτυο) :

Οι υπολογιστικοί-κρυμμένοι νευρώνες πολλαπλασιάζουν κάθε είσοδό τους με το αντίστοιχο συναπτικό βάρος και υπολογίζουν το ολικό άθροισμα των γινομένων. Το άθροισμα αυτό τροφοδοτείται ως όρισμα στη συνάρτηση ενεργοποίησης, την οποία υλοποιεί εσωτερικά κάθε κόμβος. Η τιμή που λαμβάνει η συνάρτηση για το εν λόγω όρισμα είναι και η έξοδος του νευρώνα για τις τρέχουσες εισόδους και βάρη.

- Όπου x_{ki} είναι η i -οστή είσοδος του k νευρώνα, w_{ki} το i -οστό συναπτικό βάρος του k νευρώνα και φ η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου, τότε η έξοδος y_k του k νευρώνα δίνεται από την εξίσωση:

$$y_k = \varphi\left(\sum_{i=0}^N x_{ki}w_{ki}\right)$$

Στον k -οστό νευρώνα υπάρχει ένα συναπτικό βάρος w_{k0} με ιδιαίτερη σημασία, το οποίο καλείται πόλωση ή κατώφλι (bias,threshold). Η τιμή της εισόδου του πάντα είναι η μονάδα, $x_{k0}=1$. Εάν το συνολικό άθροισμα από τις υπόλοιπες εισόδους του νευρώνα είναι μεγαλύτερο από την τιμή αυτή, τότε ο νευρώνας ενεργοποιείται. Εάν είναι μικρότερο, τότε ο νευρώνας παραμένει ανενεργός.



Σχήμα 15: Νευρώνας (http://en.wikibooks.org/wiki/Artificial_Neural_Networks/Print_Version)

Οι βασικές συναρτήσεις που χρησιμοποιούνται είναι:

- Η σιγμοειδής συνάρτηση (sigmoid function):

$$\psi(x) = \frac{1}{1 + e^{-a(x)}}$$

- Η βηματική συνάρτηση (threshold function):

$$\psi(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- Η κατά τμήματα γραμμική συνάρτηση (piecewise linear function):

$$\psi(x) = \begin{cases} 1, & x \geq \frac{1}{2} \\ x, & -\frac{1}{2} < x < \frac{1}{2} \\ 0, & x < 0 \text{ ή } x \leq -\frac{1}{2} \end{cases}$$

Η χρήση των τεχνητών νευρωνικών δικτύων προσφέρει χρήσιμες δυνατότητες, όπως:

- Είναι ευλύγιστα, εύκολα, η απόδοση τους μπορεί να αυτοματοποιηθεί ελαχιστοποιώντας την ανθρώπινη παρουσία και μπορούν να χρησιμοποιηθούν τόσο για πρόβλεψη όσο και για κατηγοριοποίηση.
- Μπορούν να χρησιμοποιηθούν για την εκτέλεση μη γραμμικών στατιστικών μοντέλων, καθώς εκτελεί εργασίες που ένα γραμμικό πρόγραμμα δεν μπορεί, ικανά να μοντελοποιούν ιδιαίτερα σύνθετες συναρτήσεις.
- Είναι ικανά να μοντελοποιήσουν εξαιρετικά πολύπλοκες λειτουργίες.
- Τα νευρωνικά δίκτυα μπορούν να ενημερωθούν με νέα δεδομένα, άρα είναι ιδανικά για δυναμικά περιβάλλοντα, καθώς μπορούν να συντηρηθούν εύκολα.

Από την άλλη πλευρά επειδή τα τεχνητά νευρωνικά δίκτυα αποτελούν ένα σχετικά πρόσφατο μοντέλο υπολογισμού υπάρχουν μειονεκτήματα, όπως:

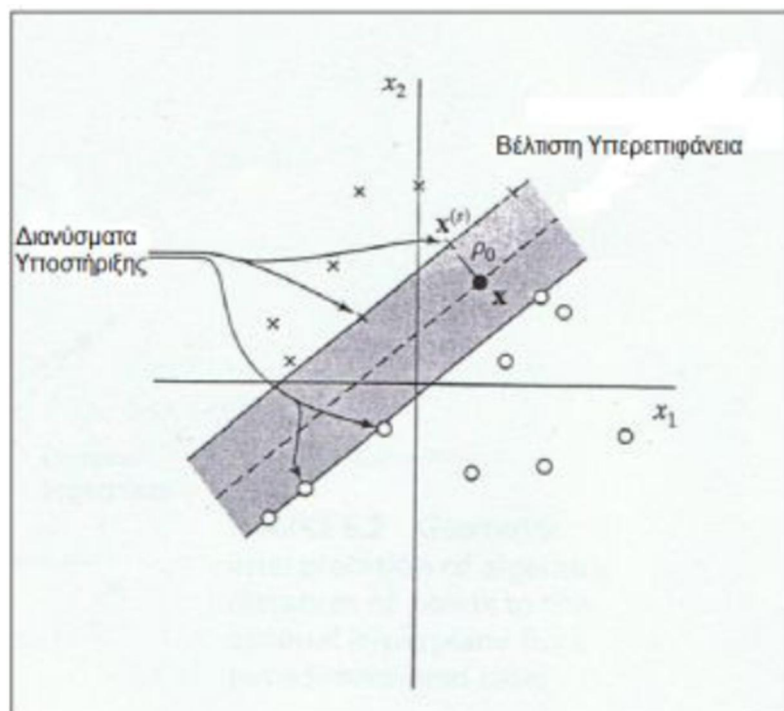
- Ο υπολογισμός των βαρών του δικτύου απαιτεί μεγάλα ποσά δεδομένων, και αυτό συνεπάγεται μεγάλο υπολογιστικό κόστος.
- Τα νευρωνικά δίκτυα εξαρτώνται σημαντικά από την ποσότητα και την ποιότητα των διαθέσιμων δεδομένων.
- Δεν έχουν κλασικές στατιστικές ιδιότητες, όπως τα διαστήματα εμπιστοσύνης και ο έλεγχος υπόθεσης.
- Η επιλογή των ενδιάμεσων κόμβων και των παραμέτρων εκπαίδευσης είναι εμπειρική. Η εκπαίδευση μπορεί να είναι δύσκολη ή αδύνατη.

- Δεν υπάρχουν σαφείς κανόνες για να επιλεγεί ο πιο κατάλληλος αλγόριθμος τεχνητού νευρωνικού δικτύου, για οποιαδήποτε εφαρμογή.

2.4. Μηχανές Διανυσμάτων Υποστήριξης.

Οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines, SVM) είναι μια πρόσφατη προσέγγιση εκπαίδευσης των νευρωνικών δικτύων στο πρόβλημα της ταξινόμησης και χρησιμοποιούνται για την πρόβλεψη μελλοντικών δεδομένων για δυαδική κατηγοριοποίηση. Τα συστήματα ταξινόμησης που βασίζονται στις μηχανές διανυσμάτων υποστήριξης αποτελούν σήμερα μια από τις δημοφιλέστερες τεχνικές στο χώρο της κατηγοριοποίησης. Τα μοντέλα αυτά ανάλογα με το είδος του προβλήματος μπορεί να είναι γραμμικά (linear) SVMs, αν τα δεδομένα είναι γραμμικά διαχωρισμένα αλλά και μη γραμμικά (non-linear) στην αντίθετη περίπτωση.

Σκοπός τους είναι η εύρεση ενός υπερεπιπέδου Π που να διαχωρίζει δύο κατηγορίες αντικειμένων με μεγιστοποίηση ενός διαχωριστικού περιθωρίου (margin) μεταξύ τους, σε ένα πολυδιάστατο χώρο χαρακτηριστικών.



Σχήμα 16: Γεωμετρική κατασκευή βέλτιστης υπερεπιφάνειας (υπερεπιπέδου)
(Γεωργίου Ε. , 2009)

Θεωρώντας ένα σύνολο n διανυσμάτων εκπαίδευσης, διάστασης $l+1$ έστω $\mathbf{x} = \{ \vec{x}_1, \vec{x}_2, \dots, \vec{x}_n \}$ όπου $\vec{x}_i = \langle \mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_{l-1} \rangle$ με τα \mathbf{a} να αποτελούν τα l χαρακτηριστικά του χώρου \mathbf{S} και με $\mathbf{y}_i \in \{-1, 1\}$ την κλάση στην οποία ανήκει το i -οστό στιγμιότυπο. Το υπερεπίπεδο αυτό διαχωρίζει κατά βέλτιστο τρόπο τα διανύσματα εκπαίδευσης με την εξίσωση:

$$\vec{w} \cdot \vec{x} + b = 0$$

Όπου \vec{w} το κάθετο διάνυσμα προς το υπερεπίπεδο και $\frac{b}{\|\vec{w}\|}$ τη κάθετη απόσταση από το υπερεπίπεδο προς την αρχή. Βάσει αυτών το διάνυσμα εκπαίδευσης \vec{x}_i θα ισχύουν τα εξής:

$$\vec{w} \cdot \vec{x}_i + b \geq 1, \quad \text{αν } y_i = 1$$

$$\vec{w} \cdot \vec{x}_i + b \leq -1, \quad \text{αν } y_i = -1$$

Στην πιο απλή μορφή, η συνάρτηση $f(\mathbf{x})$ ορίζεται ως :

$$f(\mathbf{x}) = \text{sgn}(\mathbf{x}w - \gamma)$$

όπου το υπερεπίπεδο ορίζεται $\mathbf{x}w = \gamma$, με w το κάθετο υπερεπίπεδο διάνυσμα και γ ένα σταθερό όρο.

Η βέλτιστη συνάρτηση απόφασης f είναι αυτή που μεγιστοποιεί το περιθώριο ($\frac{2}{\|\mathbf{w}_2\|}$) του υπερεπιπέδου ταξινόμησης ανάμεσα σε δύο κλάσεις, τα όρια των οποίων τείνουν ή είναι να είναι γραμμικά.

Έχοντας υπολογίσει τα διανύσματα υποστήριξης κατά την εκπαίδευση του αλγορίθμου, η μεγιστοποίηση του περιθωρίου μπορεί να γίνει με τη λύση του προβλήματος τετραγωνικού προγραμματισμού:

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + \mathbf{C} \mathbf{e}^T \mathbf{y} \\ \text{D}(\mathbf{X}\mathbf{w} - \mathbf{e}\mathbf{y}) + \mathbf{y} & \geq \mathbf{e} \\ \mathbf{y} & \geq \mathbf{0}, \mathbf{w}, \mathbf{y} \in \mathbf{R} \end{aligned}$$

όπου \mathbf{D} είναι ο πίνακας $n \times n$, το \mathbf{X} είναι ένας πίνακας $n \times m$ με τα δεδομένα εκπαίδευσης, το \mathbf{y} ένα διάνυσμα $n \times 1$ με θετικές μεταβλητές απόκλισης, το \mathbf{C} είναι μια παράμετρος για τα σφάλματα ταξινόμησης και το \mathbf{e} είναι ένα διάνυσμα με μονάδες.

Όταν τα όρια δεν είναι γραμμικά, χρησιμοποιούνται μη γραμμικές μηχανές διανυσμάτων υποστήριξης. Τα δεδομένα του προβλήματος μεταφέρονται σε ένα χώρο υψηλότερων διαστάσεων \mathbf{H} μέσα από μια θετικά ορισμένη συνάρτηση πυρήνα (kernel function) για το σχηματισμό του προβλήματος σε αντίστοιχο γραμμικό, που τα δεδομένα θα είναι γραμμικά διαχωρισμένα. Η συνάρτηση πυρήνα ϕ που μεταφέρει τα δεδομένα στο χώρο \mathbf{H} είναι η :

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \phi^T(\mathbf{x}_j)$$

Τα μοντέλα των μηχανών διανυσμάτων υποστήριξης εκφράζονται τόσο σε γραμμική όσο και μη γραμμική μορφή τα οποία όμως απαιτούν μεγάλο υπολογιστικό φορτίο σε περιπτώσεις μεγάλου όγκου δεδομένων. Για να αντιμετωπιστεί αυτό το πρόβλημα κατασκευάστηκε νέο μοντέλο, οι προσθετικές μηχανές διανυσμάτων υποστήριξης (Additive Support Vector Machines, ASVM) οι οποίες έχουν στόχο να συνδυάζουν την απλότητα και διαφάνεια ερμηνείας των γραμμικών μοντέλων με τη δυνατότητα γενίκευσης των μη γραμμικών μοντέλων και εφαρμόζονται σε προβλήματα όπου οι τεχνικές ταξινόμησης πρέπει να είναι κατανοητές και ακριβείς. Μπορούν να αναπτυχθούν από το συνδυασμό τμημάτων γραμμικών

συναρτήσεων, όπου κάθε μια συνάρτηση αναπτύσσεται χρησιμοποιώντας απλές μηχανές διανυσμάτων υποστήριξης. Ένα τέτοιο μοντέλο έχει τη μορφή:

$$f(x) = \sum_{j=1}^n f(x_j) - \gamma$$

όπου f_i είναι μια συνάρτηση του χαρακτηριστικού x_j και γ μία σταθερά.

Μερικά από τα πλεονεκτήματα της χρήσης των μηχανών διανυσμάτων υποστήριξης είναι τα εξής:

- Μπορούν να παράγουν καλά αποτελέσματα γενίκευσης σε προβλήματα κατηγοριοποίησης, ενώ ξεπερνούν το πρόβλημα υπερπροσαρμογής (overfitting) στα δεδομένα, καθώς εκπαιδεύονται με ένα ειδικό γραμμικό όριο απόφασης, το υπερεπίπεδο μέγιστου περιθωρίου. Αυτό το χαρακτηριστικό είναι μοναδικό στις μηχανές διανυσμάτων υποστήριξης.
- Βασίζονται σε απλές και ξεκάθαρες ιδέες από την θεωρία στατιστικής μάθησης, όπου σε πληθώρα εφαρμογών έχουν καλύτερη επίδοση σε σχέση με άλλες μεθόδους.
- Έχουν χαμηλό υπολογιστικό κόστος, ακόμη και στην περίπτωση μη γραμμικότητας.
- Η μέθοδος δεν απαιτεί γνώση της στατιστικής κατανομής των δεδομένων.

Όμως όσον αφορά τους χρόνους εκπαίδευσης και ελέγχου του μοντέλου, αποδεικνύονται αυξημένοι, ιδιαίτερα όταν η διάσταση του χώρου είναι μεγάλη ή όταν η συνάρτηση δεν είναι γραμμική.

2.5. Κριτήρια αξιολόγησης μοντέλων

Για να εκτιμηθεί η αποδοτικότητα και η ακρίβεια των μοντέλων ταξινόμησης και συνεπώς η ικανότητα τους για την πρόβλεψη απώλειας πελατών, χρησιμοποιούνται κριτήρια αξιολόγησης μοντέλων, τα κυριότερα των οποίων είναι το εμβαδόν κάτω από την καμπύλη AUC (area under curve) ή AUROC (area under receiver operating characteristic curve) και η καμπύλη top-decile lift, τα οποία έχουν καθιερωθεί ως μέτρα αξιολόγησης στην βιβλιογραφία διατήρησης πελατών.

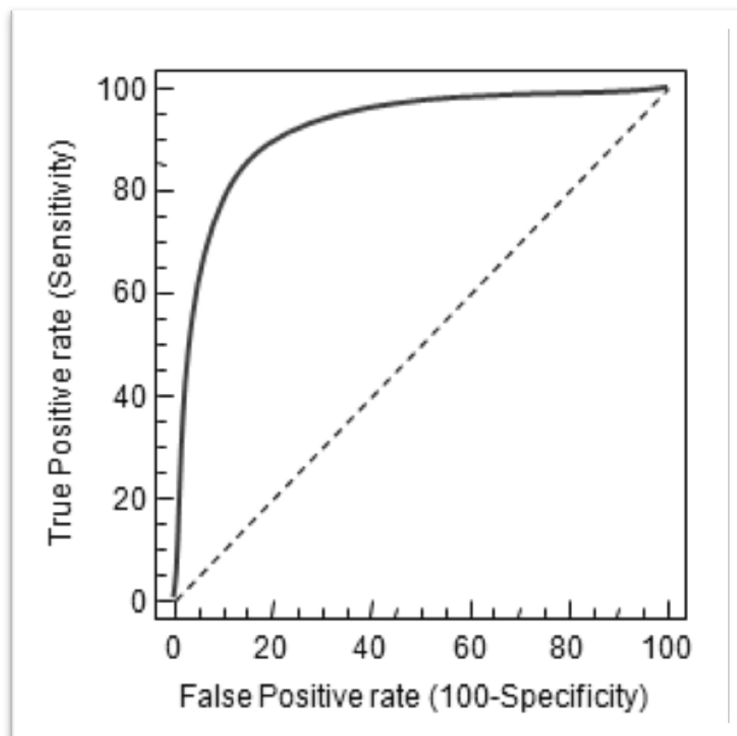
Καμπύλη AUC

Η καμπύλη AUC είναι ένα επαρκές αντικειμενικό εργαλείο μέτρησης της απόδοσης, ιδανικό για συγκρίσεις μοντέλων ταξινόμησης. Η αποτελεσματικότητα του μοντέλου πρόβλεψης εξετάζεται βάσει του εμβαδού που υπάρχει κάτω από την καμπύλη AUC:

$$AUC = \int_0^1 \frac{TP}{P} d \frac{FP}{N} = \frac{1}{P*N} \int_0^N TP dFP$$

όπου $TP+FN=P$ και $TN+FP=N$.

Αυτό το κριτήριο συνοψίζει την αποδοτικότητα ενός ταξινομητή που αναπαρίσταται με την καμπύλη λειτουργικών χαρακτηριστικών ROC curve, η οποία για κάθε δυνατή οριακή τιμή (threshold value), είναι η γραφική παράσταση δύο αξόνων, της ευαισθησίας ως προς την σαφήνεια, όπου ο οριζόντιος άξονας δείχνει το ποσοστό των πελατών που μένουν στην εταιρία, αλλά το μοντέλο πρόβλεψης έχει δείξει ότι θα την εγκαταλείψουν και ο κάθετος άξονας το ποσοστό πελατών που θα εγκαταλείψουν την εταιρία και το μοντέλο πρόβλεψης έχει θεωρήσει την ίδια κατάσταση. Μπορεί να λάβει τιμές μεταξύ του 0.5 (τυχαία πρόβλεψη- διακεκομμένη γραμμή) και του 1 (τέλεια πρόβλεψη- συνεχής καμπύλη), όπου οι μεγαλύτερες τιμές αντιπροσωπεύουν και καλύτερη προβλεπτική ικανότητα του αλγορίθμου. Επίσης η AUC είναι στενά συνδεδεμένη με το δείκτη Gini ως εξής: $Gini=2*AUC-1$, ο οποίος είναι διπλάσιος από το χώρο ανάμεσα στην διαγώνιο και στην καμπύλη ROC.(Alker, 1965).

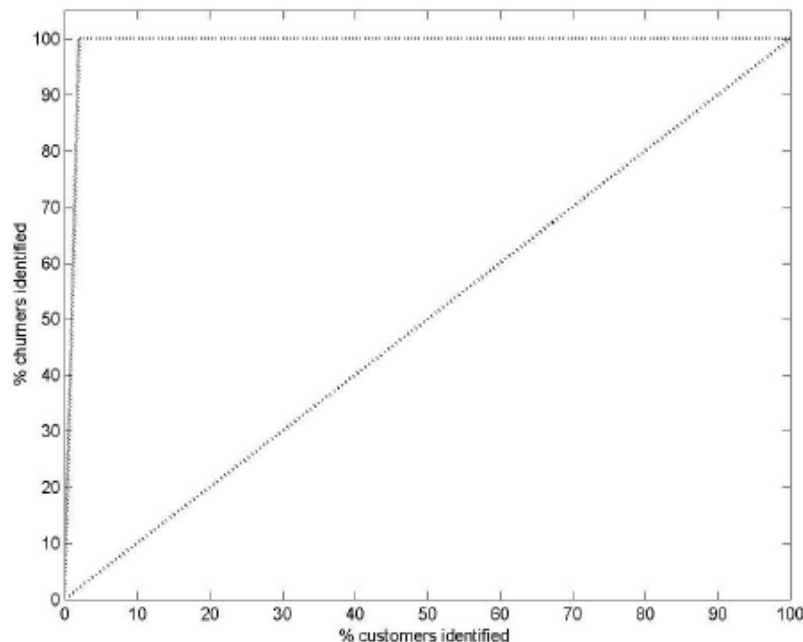


Σχήμα 17: Καμπύλη AUC (Neslin et al. 2002)

Καμπύλη Top Decile Lift

Πολλές μελέτες τονίζουν την σημασία του top decile lift στην αξιολόγηση και στην σύγκριση της απόδοσης των μοντέλων πρόβλεψης απώλειας πελατών. Αναφέρεται στον λόγο μεταξύ του ποσοστού των πελατών που πραγματικά εγκατέλειψαν μια εταιρία στο 10% των υψηλότερων προβλέψεων απώλειας πελατών, με το ποσοστό των πελατών που πραγματικά έπαψαν να είναι πελάτες σε όλο το δείγμα. Έτσι, στοχεύει στο να εντοπίσει τους πιο επικίνδυνους σχετικά με το ρίσκο διαφυγής πελάτες και εστιάζει στην ικανότητα των μοντέλων πρόβλεψης να αναγνωρίζουν τους πελάτες εκείνους οι οποίοι είναι πιθανότερο να φύγουν από την εταιρία στο μέλλον. Με αυτό τον τρόπο οι εταιρίες μπορούν να ξεκινήσουν

μια καμπάνια για να διατηρήσουν τους πελάτες αυτούς, αλλά και αυτούς με ρίσκο αποχώρησης. Όσο υψηλότερη είναι η κλίση της καμπύλης αυτής, τόσο πιο ακριβές είναι το μοντέλο που χρησιμοποιήθηκε.



Σχήμα 18: Καμπύλη Top Decile Lift (Mutanen, 2006)

Στο σχήμα 17 απεικονίζεται ο διαχωρισμός των κατηγοριών των πελατών που εγκαταλείπουν μία εταιρία από εκείνους που παραμένουν αφοσιωμένοι σε αυτή. Η καμπύλη πάνω αριστερά δείχνει τους πελάτες που πρόκειται να φύγουν, οι οποίοι εντοπίζονται από το μοντέλο πρόβλεψης. Ενώ στη μεσαία καμπύλη αναπαρίσταται μία κατάσταση όπου κανένας διαχωρισμός μεταξύ των πελατών δεν έχει γίνει, πράγμα που συμβαίνει όταν οι πιθανότητες απώλειας πελατών είναι τυχαίες.

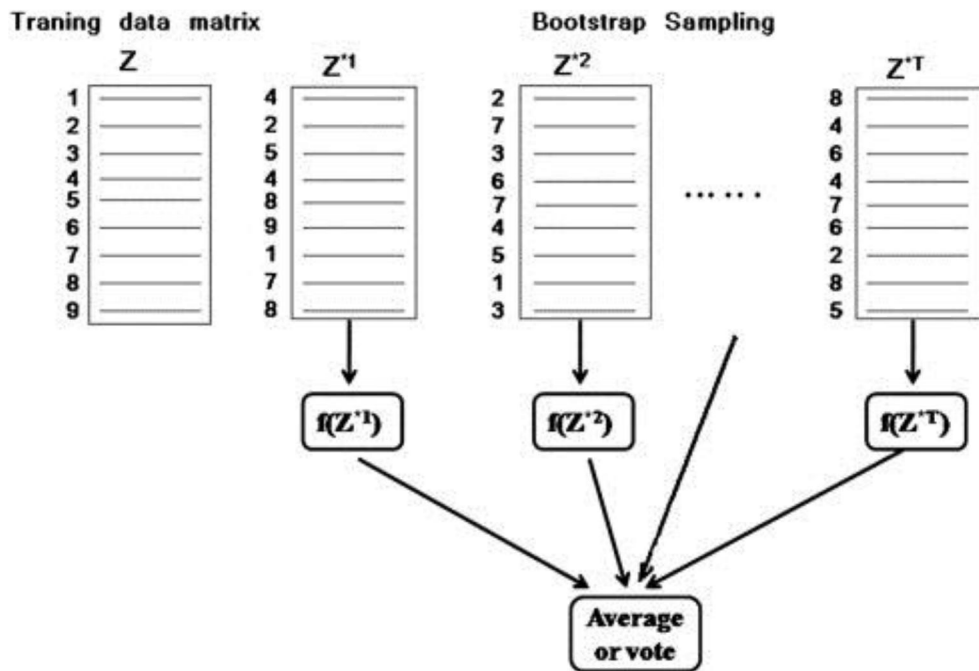
2.6. Ταξινομητές Συνόλων

Για την βελτίωση της ακρίβειας της πρόβλεψης που δίνουν οι ταξινομητές που αναλύθηκαν, χρησιμοποιούνται μαζί με αυτούς, ταξινομητές συνόλων (ensemble classifiers). Οι οποίοι έχουν στόχο τον συνδυασμό αδύναμων ταξινομητών σε ένα ισχυρό ταξινομητή. Δύο από τις πιο δημοφιλείς μεθόδους ταξινομητών συνόλων είναι η Ομαδοποίηση - Bagging και η Ενίσχυση - Boosting.

Ομαδοποίηση (Bagging)

Στο Bagging (Bootstrap AGGregatING) το αρχικό σύνολο δεδομένων δειγματοληπτείται ώστε να δημιουργηθεί ένας μεγάλος αριθμός μικρότερων συνόλων δεδομένων. Στόχος του

είναι να βελτιώσει την ακρίβεια των μοντέλων αδύναμης μάθησης σε μοντέλα ισχυρής μάθησης. Το σύνολο των μικρότερων δεδομένων κατασκευάζεται από ταξινομητές που είναι Bootstrap (μέθοδος που χρησιμοποιείται για την αναπαραγωγή των δειγμάτων εκπαίδευσης στην περίπτωση όπου το σύνολο εκπαίδευσης είναι μικρό) αντίγραφα του συνόλου εκπαίδευσης. Ύστερα οι ταξινομητές συνδυάζονται ώστε, για κάθε διάνυσμα εισόδου ο κάθε ταξινομητής προβλέπει την μεταβλητή εξόδου και τελικά η τιμή που εμφανίζεται περισσότερες φορές επιλέγεται ως η μεταβλητή απόκρισης για το συγκεκριμένο διάνυσμα.



Σχήμα 19: Σχηματική απεικόνιση της Bagging (<http://ars.els-cdn.com/content/image>).

Ενίσχυση (Boosting)

Το Boosting είναι μια παραλλαγή του Bagging, με διαφορά ότι εισάγει βάρη κατά την δημιουργία των συνόλων εκπαίδευσης. Είναι μια αποτελεσματική μέθοδος παραγωγής ενός ισχυρού και ακριβούς ταξινομητή αξιολογώντας τα λάθη των προηγούμενων ταξινομητών που δημιουργήθηκαν, συνδυάζοντας πολλούς διαδοχικούς ανακριβείς ταξινομητές. Περιλαμβάνει μια επαναληπτική διαδικασία στην οποία προσαρμόζεται η κατανομή των δειγμάτων ώστε οι ταξινομητές βάσης να εστιάζουν στα δείγματα που ταξινομούνται δύσκολα. Κατά την πρόβλεψη χρησιμοποιούνται τα αποτελέσματα όλων των ταξινομητών με βάρος ψηφοφορίας ανάλογο με την ποιότητα τους.

2.7. Άλλες μέθοδοι

Σύμφωνα με την βιβλιογραφία υπάρχουν και άλλες μέθοδοι ταξινόμησης που δεν θα αναλυθούν σε αυτή την εργασία, όπως οι Γενετικοί Αλγόριθμοι και οι αλυσίδες Markov.

Οι μελέτες που έχουν γίνει για το πρόβλημα της πρόβλεψης απώλειας πελατών χρησιμοποιώντας Γενετικούς Αλγόριθμους είναι λίγες, γεγονός που υποδηλώνει ότι η έρευνα με σκοπό να βελτιωθεί η ακρίβειά για τις τεχνολογίες αυτές δεν έχει εξαντληθεί ακόμη. Οι Γενετικοί Αλγόριθμοι είναι μαθηματικές διαδικασίες που χρησιμοποιούν τη διαδικασία της γενετικής κληρονομιάς. Η χρήση τους προσφέρει υψηλή ακρίβεια, είναι πολύ εύκολο να αναπτυχθούν παρέχοντας εύρωστες προβλέψεις και μπορούν να χρησιμοποιηθούν για μεγάλο όγκο δεδομένων αποτελεσματικά. Όμως, έχουν μεγαλύτερο υπολογιστικό φορτίο καθώς επίσης για τα πολύ μεγάλα σύνολα στοιχείων, είναι απαραίτητη η δειγματοληψία, οι οποία οδηγεί σε διαφορετικά αποτελέσματα (Hadden et al. 2007).

Μια ακόμα μέθοδος για την πρόβλεψη απώλειας πελατών αποτελούν οι αλυσίδες Markov, όπου είναι μια στοχαστική διαδικασία που χρησιμοποιείται για την ανάλυση πολύπλοκων δυναμικών συστημάτων (Migueis et al. 2004), οι οποίες πρόσφατα μπήκαν στο προσκήνιο της πρόβλεψης απώλειας πελατών και οι οποίες δίνουν σωστές και με συνάφεια πληροφορίες. Είναι μια πολύ ελκυστική μέθοδος για να αντιμετωπιστούν απαιτητικά υπολογιστικά προβλήματα.

3. Βιβλιογραφική ανασκόπηση

Μέσα από την βιβλιογραφική ανασκόπηση των μεθόδων ταξινόμησης στην πρόβλεψη απώλειας πελατών που αναλύθηκαν, εντοπίστηκαν μελέτες για την εφαρμογή των μεθόδων αυτών.

Η Λογιστική Παλινδρόμηση, τα Μοντέλα Δέντρων Αποφάσεων και τα Νευρωνικά Δίκτυα υλοποιούνται με το στατιστικό πακέτο SPSS, ενώ οι Μηχανές Διανυσμάτων Υποστήριξης υλοποιούνται στο περιβάλλον MATLAB.

3.1. Βιβλιογραφική ανασκόπηση εφαρμογών των μεθόδων ταξινόμησης.

- Σε έρευνα που διεξήχθη από τους Hwang et al (2004) για την περίπτωση της ασύρματης βιομηχανίας τηλεπικοινωνιών, στην οποία συμπεριελήφθησαν ένα μοντέλο λογιστικής παλινδρόμησης, ένα δέντρο απόφασης και ένα νευρωνικό δίκτυο, έδειξαν ότι η λογιστική παλινδρόμηση αποδίδει καλύτερα σε τέτοιες βάσεις δεδομένων, καθώς επίσης το δέντρο απόφασης έδειξε ελαφρώς καλύτερη ακρίβεια σε σχέση με τις άλλες τεχνικές, όμως αναφέρουν πως αυτή η τεχνική δεν αποδίδει το ίδιο και σε άλλες περιπτώσεις.
- Οι Μηχανές Διανυσμάτων Υποστήριξης είναι μια νέα μέθοδος που μελετάται για την χρήση της στην απώλεια πελατών. Σε μελέτη των Coussement et al. (2008) συμβάλλει στην βιβλιογραφία για την διερεύνηση της αποτελεσματικότητας των Μηχανών Διανυσμάτων Υποστήριξης στην πρόβλεψη απώλειας πελατών σε συνδρομητικές υπηρεσίες. Χρησιμοποιώντας ένα επαρκές μεγέθος δείγμα σε αντίθεση με προηγούμενες μελέτες όπου οι Μηχανές Διανυσμάτων Υποστήριξης εφαρμόζονται σε ένα πολύ μικρό δείγμα. Σε αυτή την μελέτη χρησιμοποιήθηκαν δεδομένα εφημερίδας βελγικού εκδοτικού οίκου, στον οποίο οι συνδρομητές πρέπει να καταβάλουν ένα σταθερό ποσό ανάλογα με την διάρκεια εγγραφής και τη δεδομένη προωθητική προσφορά, η εταιρία δεν επιτρέπει να λήγει η συνδρομή πριν από την ημερομηνία λήξης. Σε αυτή την μελέτη έδειξαν ότι οι Μηχανές Διανυσμάτων Υποστήριξης είναι σε θέση να προβλέψουν την απώλεια πελατών στις συνδρομητικής υπηρεσίες οι οποίες δείχνουν καλή απόδοση γενίκευσης, ξεπερνούν την παραδοσιακή τεχνική μοντελοποίησης της λογιστικής παλινδρόμησης.
- Οι Mozer et al. (2000) ερεύνησαν με λογιστική παλινδρόμηση, δέντρα απόφασης και νευρωνικά δίκτυα, την βάση δεδομένων 47.000 οικιακών συνδρομητών μιας αμερικανικής εταιρίας ασύρματης τηλεφωνίας με σκοπό τον εντοπισμό των δυνητικών πελατών και την πρόβλεψη απώλειας τους, αλλά και τον προσδιορισμό των κινήτρων τα οποία θα βελτιώσουν την διατήρηση των πελατών και θα μεγιστοποιήσουν την κερδοφορία της εταιρίας. Η βάση αυτή περιελάμβανε πληροφορίες σχετικά με τη χρήση, την τιμολόγηση, την πίστωση και το ιστορικό παραπόνων. Συγκρίνοντας τις μεθόδους που χρησιμοποιήθηκαν με βάση την καμπύλη lift, έδειξαν ότι τα νευρωνικά δίκτυα απέδωσαν καλύτερα από τις άλλες μεθόδους.

- Οι Ng et al (2001) δείχνουν ότι για την διατήρηση πελατών σε μεγάλη εμπορική εταιρία, η διαδικασία πρέπει να είναι πολύ ακριβής για αυτό επέλεξαν το δέντρο απόφασης ως μέθοδο ταξινόμησης επειδή ισχυρίζονται ότι έχουν αποδεδειγμένα καλές επιδόσεις και γιατί δημιουργούν αυτόματα τους κανόνες ταξινόμησης. Το Σχήμα 19 δείχνει πως οι κανόνες ταξινόμησης μπορούν να εφαρμοστούν στα δεδομένα για να εκτελέσουν τις εργασίες λήψης αποφάσεων.

Jobless	Bought	Sex	...	Age	Savings	Granted
No	Car	Male	...	38	\$150K	Yes
Yes	Jewel	Female	...	26	\$60K	Yes
Yes	Stereo	Male	...	20	\$10K	No

Σχήμα 20: Παράδειγμα των κανόνων ταξινόμησης (Ng et al. 2001)

3.2. Βιβλιογραφική ανασκόπηση μελέτης: **The Relevant Length of Customer Event History for Churn Prediction: How long is long enough? Working Paper of Faculty of Economics and Business Administration, Department of Marketing, Ghent University**

Στην μελέτη των Ballings και Van den Poel το 2012 με τίτλο *The Relevant Length of Customer Event History for Churn Prediction: How long is long enough?*, του βελγικού Πανεπιστημίου Ghent, το βασικό ερώτημα που πραγματεύονται είναι το: ποια θα πρέπει να είναι η χρονική διάρκεια του πελατειακού ιστορικού προκειμένου να δημιουργηθεί ένα μοντέλο πρόβλεψης απώλειας πελατών; Χρησιμοποιώντας Λογιστική Παλινδρόμηση (Logistic Regression), μοντέλα Δέντρων Αποφάσεων (Classification Trees) και την τεχνική bagging σε συνδυασμό με δέντρα αποφάσεων, η παρούσα μελέτη αναλύει την βελτίωση της απόδοσης του μοντέλου πρόβλεψης απώλειας παρατείνοντας το πελατειακό ιστορικό από 1 στα 16 χρόνια.

3.2.1. Δεδομένα

Μετά την ανάλυση ολόκληρης της βάσης δεδομένων των πελατών μιας εφημερίδας, με δεδομένα που ξεκινούν από το σημείο t και τελειώνουν στις 01/03/2009, όπου το t παίρνει τις τιμές από 03/01/2009 έως 03/01/1994 και με δεδομένα για τον υπολογισμό της αποχώρησης από την περίοδο μεταξύ 01/08/2010 έως 02/03/2011, πήραν τα εξής δεδομένα.

Όσοι πελάτες ήταν ενεργοί μετά την χρονική στιγμή 01/03/2009, την εξαρτημένη συμπεριφορά, η περίοδος αυτή ονομάζεται περίοδος συγκράτησης.) συμπεριλήφθησαν στην ανάλυση (n=129,892) από τους οποίους το 75% χρησιμοποιήθηκε για τον υπολογισμό και το 25% για την επικύρωση. Οι ρυθμοί αποχώρησης των πελατών ήταν αντίστοιχα 11,15% και 11,47%. Η τιμή που πληρώνουν οι πελάτες εξαρτιόταν από την διάρκεια της συνδρομής στην εφημερίδα και τυχόν προσφορές προώθησης. Η εφημερίδα στέλνει στον πελάτη ειδοποίηση ότι πλησιάζουν στο τέλος της συνδρομής τους και τους ρωτά εάν επιθυμούν να την ανανεώσουν, μαζί με οδηγίες για την ανανέωση. Οι πελάτες δεν μπορούν να ακυρώσουν την συνδρομή τους και τους παρέχεται μια δωρεάν περίοδος τεσσάρων εβδομάδων από την στιγμή που η συνδρομή λήξει για να την ανανεώσουν. Οπότε στην συγκεκριμένη περίπτωση η πρόβλεψη έχει να κάνει με το αν ο πελάτης θα ανανεώσει την συνδρομή του ή όχι εντός της περιόδου των τεσσάρων εβδομάδων από την λήξη της συνδρομής.

3.2.2. Τεχνικές

Στην βιβλιογραφία απώλειας πελατών οι πιο διαδεδομένες τεχνικές ανάλυσης είναι η Λογιστική Παλινδρόμηση και τα μοντέλα Δέντρων Αποφάσεων, τα οποία ακαδημαϊκοί και επαγγελματίες χρησιμοποιούν κατά 68% (Kamakura et al. 2006). Η αποδοτικότητα των δύο αυτών τεχνικών είναι σχεδόν όμοια και εξαρτάται από πολλαπλούς παράγοντες όπως η κανονικότητα των δεδομένων, ο αριθμός των κατηγορικών μεταβλητών και το μέγεθος του δείγματος εκπαίδευσης (Perlich et al. 2004). Έχει δειχτεί ότι τα Δέντρα Απόφασης μπορούν να ωφεληθούν σημαντικά από μεθόδους συνόλων. Με την τεχνική bagging αυξάνεται σημαντικά η ικανότητα πρόβλεψης των Δέντρων Απόφασης, ενώ δεν συμβαίνει το ίδιο και με την Λογιστική Παλινδρόμηση λόγω ότι το Logit είναι λιγότερο ευαίσθητο σχετικά με τα Δέντρα, οπότε ο υπολογισμός μέσω όρων προβλέψεων των διαφόρων δειγμάτων bootstrap δεν φέρει σημαντικά αποτελέσματα (Simonoff 2004 και Risselada 2010).

Συνοψίζοντας, η Λογιστική Παλινδρόμηση και τα Δέντρα Απόφασης χρησιμοποιούνται ευρέως από ακαδημαϊκούς και από επαγγελματίες, ενώ το bagging είναι μια απλή μέθοδος να αυξήσουμε την αποδοτικότητα της πρόβλεψης των τεχνικών ταξινόμησης. Στην παρούσα μελέτη χρησιμοποιήθηκε η μέθοδος CHAID για την δημιουργία των Δέντρων, αφού μετά από μια προκαταρκτική έρευνα βγήκε το αποτέλεσμα ότι η CHAID υπερτερούσε της CART. Για να αποφύγουν το overfitting στην CHAID δοκίμασαν πολλαπλά την χρήση σημείων ελαχίστου μεγέθους για την διχοτόμηση και στην Λογιστική Παλινδρόμηση χρησιμοποίησαν μια κλιμακωτή επιλογή μεταβλητών.

Η τεχνική bagging συνίσταται στην δημιουργία υπολογιστικών μοντέλων βασισμένα σε έναν αριθμό bootstrap δειγμάτων του αρχικού δείγματος. Στην μελέτη χρησιμοποιήθηκαν δέκα bootstrap δείγματα για τον υπολογισμό δέκα μοντέλων, γιατί το δείγμα ήταν αρκετά μεγάλο. Αν και περισσότερα bootstrap δείγματα μπορούσαν να προσδώσουν μεγαλύτερη αποδοτικότητα, η ανάλυσή έδειξε ότι ακόμα και με δέκα bootstrap δείγματα υπήρξε μεγάλη αύξηση στην ικανότητα πρόβλεψης του μοντέλου σε σχέση με ένα Δέντρο μόνο καθώς το ίδιο είναι χρήσιμο και στις εταιρίες που έχουν πολλούς πελάτες και δεν έχουν χρόνο να υπολογίσουν έναν μεγάλο αριθμό μοντέλων.

3.2.3. Μεταβλητές

Σε ένα μοντέλο βασισμένο σε μια ανεξάρτητη περίοδο του ενός έτους, 113 μεταβλητές εξετάστηκαν για την στατιστική τους σημαντικότητα. Κάθε φορά που προστίθενται τα δεδομένα του επόμενου έτους, οι μεταβλητές υπολογίζονται εκ νέου και εκείνες από τα προηγούμενα έτη φυλάσσονται. Έτσι, όσο μεγαλύτερη είναι η διάρκεια της ανεξάρτητης περιόδου, τόσο πιο πολλές μεταβλητές εξετάζονται στην ανάλυση. Οι μεταβλητές εκείνες που μένουν σταθερές στον χρόνο (όπως είναι το φύλο και η πελατειακή σχέση) και εκείνες που σχετίζονται απόλυτα με τον χρόνο (όπως η ηλικία) φυσικά δεν υπολογίζονται εκ νέου.

Έτσι, 118 μεταβλητές προστίθενται με κάθε επιπλέον χρονιά, φτιάχνοντας 1733 μεταβλητές συνολικά προς εξέταση όταν η διάρκεια της ανεξάρτητης περιόδου θα φτάσει τα 16 χρόνια. Ο αριθμός των μεταβλητών που επιλέγουν τα διάφορα μοντέλα και ο οποίος συμπεριλαμβάνεται στα τελικά μοντέλα, κυμαίνεται μεταξύ 31 με 54.

Τα χαρακτηριστικά των πελατών και η πελατειακή σχέση περιλαμβάνονται στις αναλύσεις. Όπως έδειξαν έρευνες υπάρχουν τρεις μεταβλητές των πελατειακών σχέσεων που καθιστούν τους πιο σημαντικούς παράγοντες πρόβλεψης καταναλωτικής συμπεριφοράς: η συχνότητα (συχνότητα και η διάρκεια της σχέσης), η πρόσφατα (συχνότητα αγορών) και η χρηματική αξία της αγοράς (Van den Poel et al. 2003). Εκτός από τις μεταβλητές αυτές, η διάρκεια της πελατειακής σχέσης (LOR- Length Of Relationship) φαίνεται ότι αποτελεί έναν επίσης σημαντικό παράγοντα (Van den Poel et al. 2005)

Στον Πίνακα 1 υπάρχει μια σύνοψη των μεταβλητών που συμπεριελήφθησαν στην παρούσα μελέτη μαζί με μια σύνοψη της σχετικής βιβλιογραφίας.

Variable	Variable Type	Study
<ul style="list-style-type: none"> • Language • Gender • Age, Whether the age is known • Relationship type: person, company, advertiser • Whether information is available: telephone number, fax, email, bank account number 	<ul style="list-style-type: none"> • Customer characteristics • Socio demographics 	Larivière & Van den Poel, 2005
<ul style="list-style-type: none"> • Frequency: Number of subscriptions • Number of newspapers in last subscription • Sum number of newspapers across subscriptions • Number of changes in newspaper editions: sum, mean • Subscription type: Renewal or not • Newspaper edition • Payment method: direct debit • Changes in payment method: sum, mean • Price single newspaper: • -Last formula: Net • -All formulas: Sum net, mean gross • Price product formula (Monetary value): • -Last formula: Net, Gross • -All formulas: Sum net, mean gross, sum gross, mean gross • Length of last subscription • Which month and season the end of subscription falls in • Length of relationship with interruptions • Length of relationship without interruptions • Mean length of subscription • Length of last subscription divided by mean length of subscription • Recency: elapsed time since start date last subscription • Elapsed time since last change in edition • Prior churn: not having a subscription in between subscriptions • -Y/es/no: there is a start date 1, 10, 20, >30 day(s) than the expiry date of the previous subscription (both for last and all subscriptions) 	<ul style="list-style-type: none"> • Relationship characteristics • Frequency related • Monetary value related • Recency and duration related 	<ul style="list-style-type: none"> • Cullinan, 1977; Bauer, 1988; Van den Poel, 2003; Morrison, 1966; Lawrence, 1980 • Ganesan 1994; Levin & Zahavi 1996; Cullinan 1977 • Van den Poel, 2003; Cullinan, 1977; Van den Poel, 2003; Buckinx & Van den Poel, 2005; Simpson, 1987

-Count	Combinations	This study
Duration: sum, mean		
• Sum of length last 50%(20%) of subscriptions divided by sum of length first 50%(20%) of subscriptions		
• Monetary value of last 50%(20%) of all subscriptions divided by monetary value of first 50%(20%) of all subscriptions		
• Interactions: Recency x Frequency x Monetary x LOR, Recency x		
Frequency x Monetary, Recency x Frequency x LOR, Recency x		
Monetary x LOR, Frequency x Monetary x LOR, Recency x Frequency,		
Recency x Monetary, Recency x LOR, Monetary x LOR, Frequency x		
Monetary, Frequency x LOR,		
Divisions: Monetary / LOR, Frequency / LOR, Monetary / Frequency,		
LOR / Frequency, Recency / length of last subscription, Recency / LOR		

3.2.4. Απόδοση Μοντέλου

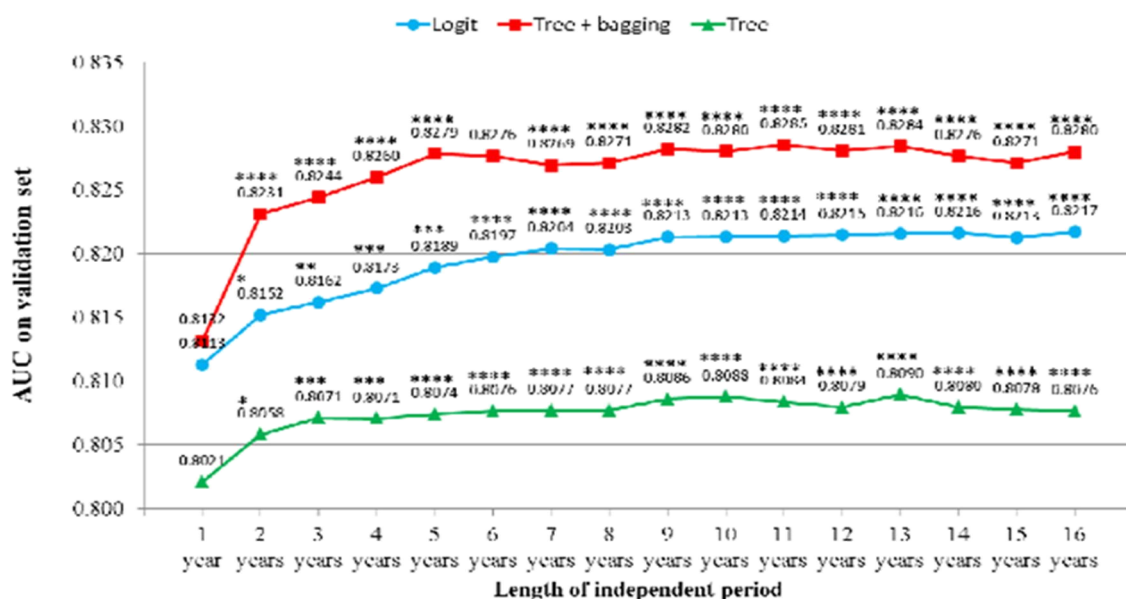
Για να εκτιμήσουν την αποδοτικότητα του μοντέλου ταξινόμησης, οι συγγραφείς χρησιμοποίησαν το εμβαδόν κάτω από την καμπύλη AUC. Η χρήση της καμπύλης AUC έχει εκτιμηθεί ως καλό αντικειμενικό κριτήριο για την αποδοτικότητα της ταξινόμησης από πολλούς ερευνητές. Η AUC παίρνει τιμές από 0.5 αν η πρόβλεψη είναι τυχαία, έως 1, αν η πρόβλεψη είναι τέλεια (Baecke et al. 2011).

Χρησιμοποίησαν την AUC έναντι του Σωστά Ταξινομημένου Ποσοστού (Percentage of Correctly Classified, PCC) επειδή η AUC σε σχέση με το PCC δεν είναι τόσο ευαίσθητη στη διαχωριστική τιμή των εκ των υστέρων πιθανοτήτων. Έτσι, η AUC αποτελεί ένα πιο επαρκές εργαλείο μέτρησης απόδοσης (Baesens et al. 2002).

3.2.5. Αποτελέσματα

Το Σχήμα 13 δείχνει την ικανότητα πρόβλεψης σε σχέση με την AUC, στα διάφορα σημεία της ανεξάρτητης περιοχής. Αναφέρεται και η στατιστική σημαντικότητα για τις διαφορές μεταξύ των διαφόρων χρονικών σημείων της ανεξάρτητης περιόδου και κάθε έτους.

Μια πρώτη παρατήρηση είναι ότι η Λογιστική Παλινδρόμηση υπερτερεί της ενός Δέντρου ταξινόμησης. Το αποτέλεσμα αυτό συμφωνεί και με τη βιβλιογραφία ότι και η Λογιστική Παλινδρόμηση και τα Δέντρα Απόφασης εξαρτώνται πολύ από τα χαρακτηριστικά του δείγματος, όπως την κανονικότητα των δεδομένων και τις κατηγορικές μεταβλητές (Perlich et al. 2004). Έτσι, εξαρτάται από το δείγμα αν η Λογιστική Παλινδρόμηση ή τα Δέντρα Απόφασης αποδώσουν καλύτερα. Μια δεύτερη παρατήρηση είναι η τεχνική bagging μαζί με Δέντρα Απόφασης υπερτερεί της Λογιστικής Παλινδρόμησης, κάτι που επίσης συμφωνεί με την βιβλιογραφία: ο συνδυασμός bagging και Δέντρων συχνά βελτιώνει σε μεγάλο βαθμό την ταξινόμηση και έχει δειχθεί ότι υπερτερεί της Λογιστικής Παλινδρόμησης (Perlich et al. 2004). Έτσι, συνοπτικά, η επίδοση των ταξινομητών που χρησιμοποιήθηκε είναι σύμφωνη με τη βιβλιογραφία.



* $p < .10$, ** $p < .05$, *** $p < .01$, **** $p < .001$

Σχήμα 21: Πρόβλεψη απόδοσης σε διάφορα μήκη της ανεξάρτητης περιόδου

Είναι ενδιαφέρον ότι και οι τρεις ταξινομητές δείχνουν μια λογαριθμική αύξηση στην απόδοσή τους όταν η διάρκεια της ανεξάρτητης περιόδου αυξάνεται. Μετά το πέμπτο έτος αυτή στην πρόβλεψη φαίνεται ότι σταθεροποιείται στους δύο από τους τρεις ταξινομητές (Δέντρα και Δέντρα μαζί με bagging). Επιπρόσθετα, η σημαντικότητα της διαφοράς με τη μικρότερη διάρκεια του ενός έτους δεν φέρει βελτίωση. Έτσι, αν η εταιρία που έλαβε μέρος στην συγκεκριμένη μελέτη πάρει τα πέντε χρόνια ως διαχωριστική τιμή, θα μπορούσε να πετάξει περίπου το 69% των δεδομένων που φυλάει και συνεπώς να κάνει σημαντική οικονομία στα δεδομένα που αποθηκεύει, προετοιμάζει και αναλύει, με ελάχιστο κόστος στην προβλεψιμότητα των μοντέλων. Άλλες διαχωριστικές θα μπορούν να διαχωριστούν επίσης. Για παράδειγμα, για την ανεξάρτητη περίοδο των 9 χρόνων, η αποδοτικότητα πρόβλεψης της τρίτης τεχνικής, της Λογιστικής Παλινδρόμησης, επίσης σταθεροποιείται. Ακόμα και έτσι μία εταιρία μπορεί να κάνει 44% οικονομία στις βάσεις δεδομένων.

3.2.6. Συμπεράσματα

Χρησιμοποιήθηκαν Λογιστική Παλινδρόμηση, Δέντρα Απόφασης και Δέντρα Απόφασης μαζί με την τεχνική bagging προκειμένου να μελετηθεί η σχέση μεταξύ της διάρκειας της ανεξάρτητης περιόδου και την αποδοτικότητα ταξινόμησης. Από την παρούσα μελέτη βγαίνει το συμπέρασμα ότι η διάρκεια της ανεξάρτητης περιόδου συνδέεται λογαριθμικά με την αποδοτικότητα της ταξινόμησης.

Οι πρακτικές συνέπειες των ευρημάτων είναι ότι η εταιρίες μπορούν να απολαμβάνουν πολύ πιο αποτελεσματικά οφέλη κατά την διαδικασία μοντελοποίησης. Πιο συγκεκριμένα, τα αποτελέσματα έδειξαν ότι η εταιρία στην μελέτη αυτή θα μπορούσε να μειώσει τα δεδομένα που χρησιμοποιεί κατά 69%, χωρίς να χάσουν τα μοντέλα της σε προβλεψιμότητα (στην παρούσα μελέτη το κόστος σε προβλεψιμότητα κυμαινόταν μεταξύ 0.0028 και 0.0001, ανάλογα με το σύστημα ταξινόμησης που χρησιμοποιήθηκε) μειώνοντας την διάρκεια της ανεξάρτητης περιόδου από τα 16 χρόνια στα 5. Τα αποτελέσματα έδειξαν ότι μετά το 5^ο έτος, η επίδοση πρόβλεψης αυξήθηκε μόνο οριακά, δηλαδή σε πρακτικό επίπεδο οι αναλυτές μπορούν να μειώσουν σημαντικά το μέγεθος των δεδομένων που αποθηκεύουν, προετοιμάζουν και αναλύουν. Αυτό είναι πολύ σημαντικό σε περιπτώσεις μεγάλων βάσεων δεδομένων όπου η υπολογιστική πολυπλοκότητα είναι σημαντική.

3.2.7. Περιορισμοί και Μελλοντικές Κατευθύνσεις

Αν και οι Ballings και Van den Poel νιώθουν ασφαλείς για τα αποτελέσματα που παρουσίασαν στην μελέτη, δεν είναι βέβαιο το κατά πόσο αυτά μπορούν να γενικευτούν σε ένα μεγαλύτερο φάσμα συνδρομητικών υπηρεσιών. Αν τα αποτελέσματα αυτά επικυρωθούν με άλλες μελέτες πάνω σε πελατειακές βάσεις δεδομένων σε άλλους τομείς. Αυτό θα

αποδείκνυε αν η λογαριθμική σχέση μεταξύ της διάρκειας της ανεξάρτητης περιόδου και της προβλεψιμότητας ισχύει μόνο για τον τομέα του Τύπου ή αν μπορεί να γενικευτεί και αλλού. Μια δεύτερη ερευνητική προοπτική θα ήταν η χρήση και άλλων αναλυτικών τεχνικών. Αν και αυτές που χρησιμοποιήσαν σε αυτή την μελέτη είναι εκείνες που χρησιμοποιούνται συχνότερα και από ακαδημαϊκούς και από επαγγελματίες, θα ήταν χρήσιμο να δούμε αν τα ίδια αποτελέσματα προκύπτουν χρησιμοποιώντας και πιο εξελιγμένες τεχνικές, όπως τα Νευρωνικά Δίκτυα και οι Μηχανές Διανυσμάτων Υποστήριξης.

4. ΣΥΜΠΕΡΑΣΜΑΤΑ – ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ

4.1. Συμπεράσματα

Στον τομέα της πρόβλεψης πελατειακής απώλειας, η απόδοση της ταξινόμησης παίζει σημαντικό ρόλο στην διατήρηση των πελατών και της επιχειρησιακής αποδοτικότητας. Έτσι, για να παραμείνουν ανταγωνιστικές πολλές εταιρίες προσπαθούν με την βοήθεια της τεχνολογίας να βελτιώσουν τη στρατηγική του μάρκετινγκ του, με την μετάβαση από το μαζικό (mass marketing) στο επιλεκτικό μάρκετινγκ (one-to-one marketing). Ως αποτέλεσμα, η βάση δεδομένων μιας εταιρίας αποτελεί ολοένα και πιο σημαντικό στοιχείο στην υποστήριξη των αποφάσεων μάρκετινγκ σε θέματα όπως η διαχείριση πελατειακών σχέσεων. Την ίδια στιγμή ακαδημαϊκοί ερευνητές προσπαθούν συνεχώς να βελτιώσουν τα μοντέλα CRM γενικά και την ανάλυση προγνώσεων ειδικότερα. Μια βελτίωση ενός μοντέλου πρόβλεψης CRM μπορεί να επιτευχθεί και εστιάζοντας σε τεχνικές εξόρυξης δεδομένων, αλλά και μέσω της μείωσης του όγκου των δεδομένων της βάσης αυτής, πάνω στην οποία θα τρέξουν οι τεχνικές εξόρυξης.

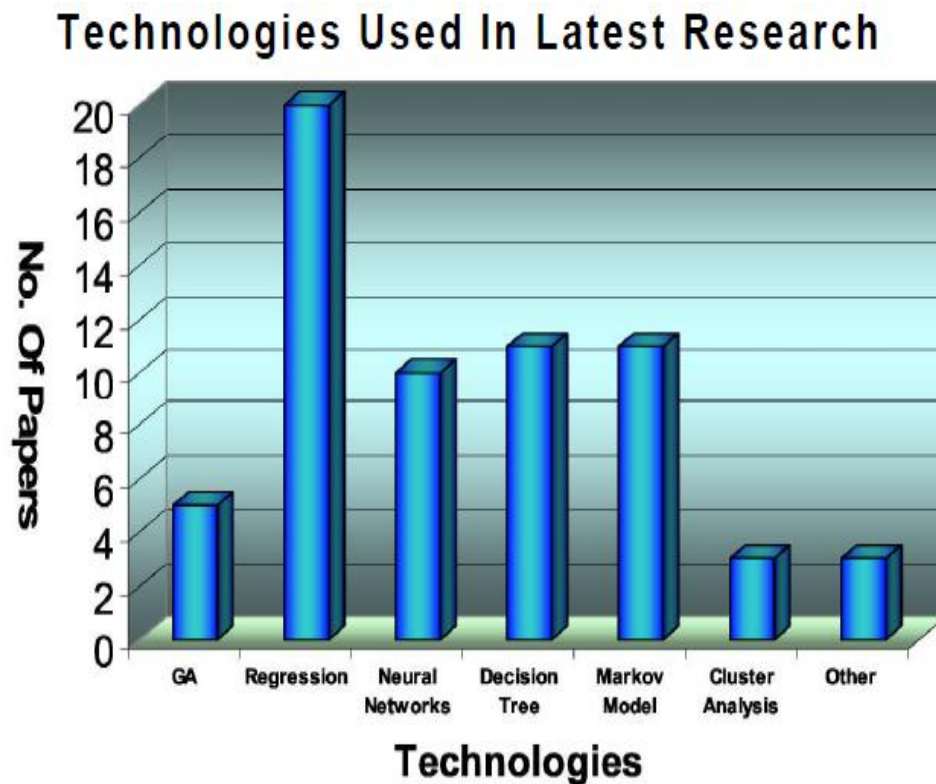
Υποστηρίζεται πως το μέγεθος των δεδομένων που αντλούνται από το πελατειακό ιστορικό τα οποία αποθηκεύονται, προετοιμάζονται και αναλύονται, μπορούν να μειωθούν σημαντικά, αν μειωθεί πρώτα η χρονική διάρκεια του πελατειακού ιστορικού ενεργειών.

Η εξόρυξη δεδομένων η οποία και αποτελεί το μέσο κλειδί ανάμεσα σε μια επιτυχημένη επιχείρηση και την κατανόηση των αναγκών των πελατών της, βρίσκεται σε αρχικό στάδιο, υπάρχει ήδη ένα ευρύ φάσμα εταιριών όπως είναι οι τηλεπικοινωνίες και οι τράπεζες, που χρησιμοποιούν εργαλεία και τεχνικές εξόρυξης δεδομένων έτσι ώστε να επωφεληθούν από το ιστορικό ενεργειών των πελατών τους.

Η τάση που εμφανίζουν οι πελάτες να διακόπτουν τις σχέσεις και τις συναλλαγές τους με μια εταιρία, τονίζει ακόμα περισσότερο τη σημαντικότητα της πρόβλεψης της μελλοντικής συμπεριφοράς των πελατών μέσω της προβλεπτικής μοντελοποίησης και συνεπώς επιτρέπει τις εταιρίες να οργανώσουν το μάρκετινγκ και τις στρατηγικές τους. Στην παρούσα εργασία έγινε ανάλυση του φαινομένου αυτού, της απώλειας πελατών από θεωρητικής βάσης παρουσιάζοντας τις αιτίες απώλειας πελατών, καθώς επίσης και την διαχείριση της απώλειας πελατών και του πληροφοριακού συστήματος διαχείρισης πελατειακών σχέσεων. Το πρόβλημα της διαχείρισης της απώλειας πελατών και της πρόβλεψής της με την καλύτερη ακρίβεια θα συνεχίσει να απασχολεί την έρευνα σε αυτό τον τομέα.

Ακόμη, όπως δείχνει η βιβλιογραφική ανασκόπηση η επιλογή της τεχνικής μοντελοποίησης επηρεάζει σημαντικά την αποδοτικότητα των προσπαθειών πρόβλεψης πελατειακής απώλειας και τονίζεται η σημασία της σύγκρισης μεταξύ των διαφόρων αλγορίθμων για την επιλογή του καλύτερου. Αυτός ο ισχυρισμός αποτυπώνεται στην πλειάδα των μεθόδων που χρησιμοποιούνται στον τομέα της πρόβλεψης και διατήρησης πελατών, όπου οι πιο διαδεδομένες τεχνικές ανάλυσης είναι η Λογιστική Παλινδρόμηση (Logistic Regression) και τα Μοντέλα Δέντρων Αποφάσεων (Classification Trees), η αποδοτικότητα των δύο τεχνικών είναι σχεδόν όμοια και εξαρτάται από πολλαπλούς παράγοντες, όπως η ικανότητα των δεδομένων, ο αριθμός των κατηγορικών μεταβλητών, το μέγεθος του δείγματος και την σηματοθορυβική σχέση (Perlich et al. 2004). Καθώς και οι πιο προχωρημένες τεχνικές που έχουν χρησιμοποιηθεί στην βιβλιογραφία του μάρκετινγκ, τα Νευρωνικά Δίκτυα (Neural Networks) και οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines), όπου η χρήση τους παραμένει ακόμα περιορισμένη αφού αυξάνουν την πολυπλοκότητα (Risselada et al. 2010).

Σε μελέτες για την ανάλυση απώλειας πελατών, την πρώτη θέση κατέχουν η ανάλυση με Λογιστική Παλινδρόμηση ενώ ακολουθούν τα Δέντρα Απόφασης και οι αλυσίδες Markov και ύστερα τα νευρωνικά δίκτυα.



Σχήμα 22: Ανάλυση των μελετών απώλειας πελατών σχετικά με τις τεχνολογιών που χρησιμοποίησαν (Hadden, 2007)

Η επιτυχία μίας μεθόδου και η απόδοση της, εξαρτώνται από το είδος της επιχείρησης που εφαρμόζεται, την ποιότητα και τον όγκο των δεδομένων και των παραμέτρων αλλά και τον χρόνο που χρειάζεται ένα μοντέλο για την φάση της εκπαίδευσης. Άρα δεν μπορεί να υπάρξει ένα συμπέρασμα για το ποια μέθοδος υπερτερεί των υπολοίπων στην πρόβλεψη απώλειας πελατών.

Όσο αφορά την ακρίβεια και την αποδοτικότητα πρόβλεψης των μοντέλων ταξινόμησης που αναλύθηκαν στην εργασία και συνεπώς την ικανότητα τους για την πρόβλεψη, χρησιμοποιούνται πρώτον τα κριτήρια αξιολόγησης μοντέλων, τα κυριότερα των οποίων είναι όπως έχουν καθιερωθεί στην βιβλιογραφία διατήρησης πελατών, το εμβαδό κάτω από την καμπύλη AUC και η καμπύλη top - decile lift και δεύτερον οι ταξινομητές συνόλου που χρησιμοποιούνται μαζί με τα μοντέλα ταξινόμησης, οι οποίοι όπως και παρουσιάστηκαν στοχεύουν στο συνδυασμό αδύναμων μοντέλων σε ένα ισχυρό, αυτοί είναι η Ομαδοποίηση (Bagging) και η Ενίσχυση (Boosting).

4.2. Μελλοντικές Κατευθύνσεις

Σκοπός της παρούσας πτυχιακής εργασίας ήταν η μελέτη του προβλήματος της διαχείρισης απώλειας πελατών, ώστε να διαπιστωθεί πως η εφαρμογή των μεθόδων ταξινόμησης μπορεί να οδηγήσει στην πρόβλεψη απώλειας πελατών, το οποίο πρόβλημα θα συνεχίσει να απασχολεί με την περαιτέρω έρευνα ακαδημαϊκούς και επαγγελματίες.

Την επέκταση των ερευνών του προβλήματος της πρόβλεψης καθορίζει η μελλοντική επέκταση της ανάλυσης και πρόβλεψης απώλειας πελατών σε περισσότερους και διαφορετικούς τομείς εταιριών και βιομηχανιών, όπου τα δεδομένα των πελατειακών βάσεων και οι παράγοντες που επηρεάζουν την απώλεια πελατών διαφοροποιούνται σημαντικά.

Έτσι, οι επενδύσεις πληροφορικής στην διαχείριση πελατών είναι μια κύρια προτεραιότητα επένδυσης των εταιριών στο μελλοντικό υπερανταγωνιστικό περιβάλλον τους. Όμως, για να επιτύχουν την καλύτερη ακρίβεια προβλέψεων θα πρέπει σε κάθε τμήμα να εφαρμοστούν οι κατάλληλες επιχειρησιακές διαδικασίες για τον κάθε πελάτη μέσω της πελατοκεντρικής στρατηγικής σε συνδυασμό με την σωστή τεχνολογία η οποία θα επεξεργαστεί τα ιστορικά στοιχεία ενεργειών των πελατών.

5. ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική Βιβλιογραφία

- Γεωργίου Ε. (2009) Χρήση Μηχανών Διανυσμάτων Υποστήριξης για την εκτίμηση ακινήτων, Πανεπιστήμιο Κύπρου.
- Δημήτριος Β. Κοσμάτος (2011) “CRM Διαχείριση Πελατειακών Σχέσεων Αρχές και Τεχνολογίες” 2η Έκδοση, Κλειδάριθμος.
- Θεοδωρίδης Γ., Πελέκης Ν. (2011) Εξόρυξη Γνώσης από Δεδομένα –Συσταδοποίηση, Ομάδα Διαχείρισης Δεδομένων Πανεπιστήμιο Πειραιώς.
- Κραδασάκης, Γ. (2006), Επαναπροσδιορίζοντας το CRM μία ολοκληρωμένη θεώρηση.
- Κουρής Γ., “Μια νέα ταυτότητα για τους πελατοκεντρικούς” ,XRAM, Δεκέμβριος 2000, σελ. 9-13
- Οικονόμου Π. , Καρώνη Χ. (2010) , Στατιστικά Μοντέλα Παλινδρόμησης, Εκδόσεις Συμεών Αθήνα 2010.
- Χαραμή Γ. , Ανάλυση και Σχεδιασμός Πληροφοριακών Συστημάτων, Θεσσαλονίκη 2002.

Ξένη Βιβλιογραφία

- Athanassopoulos, A.D. (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. *Journal of Business Research*, 47(3), 191–207.
- A. P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recognit.* 7(1997) 1145-1159.
- Ahn, J. H., S. P. Han and Y. S. Lee (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy* 30, 552–568.
- Alker, H. R. (1965). *Mathematics and politics*. New York: The Macmillan Company.
- Allison, P. D. (1999). *Logistic regression using the SAS system: Theory and application*, Cary, NC: SAS Institute.
- Anderson, E. and M. Sullivan (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science* (12), 125-143.
- Au, W. H., C. K. Chan and X. Yao (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on evolutionary computation* 7(6), 532-545.
- Baecke, P., & Van den Poel, D. (2011). Data Augmentation by Predicting Spending Pleasure Using Commercially Available External Data. *Journal of Intelligent Information Systems*. 36(3), 367-383.
- Baecke, P., & Van den Poel, D. (2010). Improving purchasing behavior predictions by data augmentation with situational variables. *International Journal of Information Technology & Decision Making*, 9(6), 853-872.
- Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., & Dedene, G. (2002). Bayesian neural network learning for repeat purchase modelling in direct marketing. *European Journal of Operational Research*, 138(1), 191–211.

- Ballings, M., & Van den Poel, D. (2012). The Relevant Length of Customer Event History for Churn Prediction: How long is long enough? Faculty of Economics and Business Administration, Department of Marketing, Ghent University, 2012/804
- Beatty, S. E. and S. M. Smith (1987). External search effort: An investigation across several product categories? *Journal of Consumer Research* (1), 83-95.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Belmont: Wadsworth
- Buckinx, W., & Van den Poel, D. (2005). Customer base analysis: Partial defection of behaviorally-loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1), 252–268.
- Burez, J., & Van den Poel, D. (2007). CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2), 277-288.
- Burez, J. and Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications* 36(3), 4626–4636
- Celsi, R. L. and J. C. Olson (1988). The role of involvement in attention and comprehension processes. *Journal of Consumer Research* (15), 210-224.
- Chandar, M., A. Laha and P. Krishna (2006). Modeling churn behavior of bank customers using predictive data mining techniques. National conference on soft computing techniques for engineering applications.
- Chen, Y., Zhang, G., Hu, D., and Wang, S. (2006) Customer Segmentation in Customer Relationship Management Based on Data Mining. In International Federation for Information Processing (IFIP), Knowledge Enterprise: Intelligent Strategies In Product Design, Manufacturing and Management eds. K Wang, Kovacs G., Wonzy M., Fang M., (Boston: Springer). Vol. 207. Pp.288-293.
- Colgate, M. and P. J. Danaher (2000). Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution. *Journal of the Academy of Marketing Science* 28(3), 375-387.
- Coussement, K. and D. Van den Poel (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications* 34(1), 313– 327.
- D. Van den Poel, and W. Buckinx, Predicting online-purchasing behavior, *Eur. J. Oper. Res.* 166(2005) 557–575.
- De Bock, K.W., & Van den Poel, D. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*, 38(10), 12293-12301.
- Dick, A. S., and K. Basu (1994). Customer loyalty: Toward an integrated conceptual framework. *Journal of the Academy of Marketing Science* 22(2), 99–113.
- Fadalla, A., C. H. Lin (2001). An analysis of the applications of neural networks in finance. *Interfaces* 31(4), 112-122.
- F. F. Reichheld, and W. E. Sasser, Zero defections: Quality comes to services, *Harv. Bus. Rev.* 68(1990) 105–112.
- Hwang, H., Taesoo, J. and Euiho, S. (2004) An Ltv Model and Customer Segmentation Based on Customer Value: A Case Study on the Wireless Telecommunications Industry, *Expert systems with applications*, 26, 181-188.
- Hung, S. Y., D. C. Yen, and H. Y. Wang (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3), 515–524.

- Zan Philipp Heyn (2010), CRM – optimize your company: Benefits and downsides of implementing CRM systems. University of Bedfordshire. course: Applied Management Project / Master Thesis.
- Kalakota,R.- Robinson,M. (2000), “e-Business 2.0, 2nd Edition”, Edited by Tapscott D.,Addison – Wesley.
- Keaveney, S. M. (1995). Customer switching behavior in service industries: An exploratory study. *Journal of Marketing*, 59(2), 71–82.
- Kim, M. K, M. C. Park, and D. H. Jeong, (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. *Telecommunications Policy* 28(2), 145–159.
- Kim, S. Y., T. S. Jung, E. H. Suh and H. S. Hwang (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications* 31(1), 101–107.
- Kim, H. S. and C. H. Yoon (2004). Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, 28(9/10), 751–765.
- Larivière, B., & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29(2),472-484.
- Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(2), 276-286.
- Miguéis, V. L.; den Poel, D. V.; Camanho, A. S. & e Cunha, J. F. (2012), 'Predicting partial customer churn using Markov for discrimination for modeling first purchase sequences.', *Adv. Data Analysis and Classification* 6 (4) , 337-353 .
- Mozer, M. C., R. Wolniewicz, D. B. Grimes, E. Johnson and H. Kaushansky (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunication industry. *IEEE Transactions on Neural Networks* 11(3), 690–696.
- Muata, K. and O. Bryson (2004). Evaluation of Decision Trees: A MultiCriteria Approach. *Computers and Operational Research* 31(11), 1933-1945.
- Neslin, S. A., S. Gupta, W. Kamakura, J. Lu and C. H. Mason (2006). Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research* 43(2), 204–211.
- Ng, K. and Liu, H. (2001) Customer Retention Via Data Mining, *Issues on the application of data mining*, 14, 569-590.
- Oliva, T. A., R. L. Oliver, and W. O. Bearden. 1995. The relationships among consumer satisfaction, involvement, and product performance: A catastrophe theory application? *Behavioral Science* (40), 104-132.
- Oliver, R. L. (1997). *Satisfaction: A behavioral perspective on the consumer*. Boston: Irwin McGraw-Hill.
- Payne, A., & Frow, P. (2005). A strategic framework for customer relationship management. *Journal of Marketing*, 69(4), 167-176.
- Perlich, C., Provost, F., & Simonoff, J. S. (2004). Tree Induction vs. Logistic Regression: A Learning-curve Analysis. *Journal of Machine Learning Research*, 4(2), 211–255.
- P. Baecke, and D. Van den Poel, Data augmentation by predicting spending pleasure using commercially available external data, *J. Intell. Inf. Syst.* (2010) (forthcoming).
- Reinartz, W.J., & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, 67(1), 77-99.

- Richins, M. L. and P. H. Bloch. (1991). Post-purchase product satisfaction: incorporating the effects of involvement and time? *Journal of Business Research* (23), 145-158.
- Risselada, H., Verhoef P.C., Bijmolt, T.H.A. (2010). Staying Power of Churn Prediction Models. *Journal of Interactive Marketing*, 24(3), 198-208.
- Rygieski, C., Wang, J., Yen, D. (2002), Data mining techniques for customer relationship management, *Technology in Society* 24 (2002) 483–502.
- Sarmaniotis, C., Tilikidou, I. and Adamson, I.(2002). The Measurement Instrument of Ecologically Conscious Consumer Behaviour. *MEDIT*, vol.1, no. 4. pp. 46-53.
- Shaw, M. J., Subramaniam, C., Tan, G. W., and Welge, M. E. (2001) Knowledge Management and Data Mining for Marketing. *Decision Support Systems*, Vol 31, pp 127-137
- Siber, R. (1997). Combating the churn phenomenon. *Telecommunications*, 31(10), 77–81.
- Tae Hyup Roh, Cheol Kyung Ahn, Ingoo Han (2005). The priority factor model for customer relationship management system success. *Expert Syst. Appl.* Vol. 28 No. 4 Pg. 641-654
- Van den Poel D. (2003). Predicting mail-order repeat buying: Which variables matter? *Tijdschrift voor Economie en Management*, 48(3), 371–403.
- Van Wezel, M., & Potharst, R. (2007). Improved customer choice predictions using ensemble methods. *European Journal of Operational Research*, 181(1), 436-452.
- Winer, R. S. (2001). A framework for customer relationship management. *California Management Review*, 43(4), 89-108.
- Xie, Y., X. Li, E.W. T. Ngai and W. Ying (2009). Customer churn prediction using improved balanced random forests. *Expert systems with applications* 36(3), 5445–5449.

Ηλεκτρονικές πηγές

<http://ars.els-cdn.com/content/image>

www.business&decision.com

www.boozallen.com

www.crm2day.com

www.customerthink.com

<http://www.crm.ugent.be>

www.datamining.gr

<http://epixeirei.gr/crm-epixeirisi>

<http://epixeirein.blogspot.com>

http://en.Wikipedia.org/wiki/Decision_tree

http://en.wikipedia.org/wiki/Logistic_regression

http://en.wikipedia.org/wiki/Neural_network

<http://www.ibm.com/developerworks/library/ba-ind-PMML1/>

www.meta-group.com

www.statsoft.com/textbook/neural-networks/

[http:// www.slideshare.net/shyaaminibaludata-mining-techniques-for-crm](http://www.slideshare.net/shyaaminibaludata-mining-techniques-for-crm)

www.plant-management.gr