



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

**Υπολογιστικές Διαδικασίες στο Σχεδιασμό
Στρατηγικών Άμεσου Μάρκετινγκ**

*Διατριβή που υπεβλήθη για τη μερική ικανοποίηση των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης*

Υπό
ΚΑΡΑΜΠΑΤΖΑΚΗ ΔΑΚΗ

Επίβλεψη
ΔΟΥΜΠΟΣ ΜΙΧΑΛΗΣ, Επίκουρος Καθηγητής

Χανιά 2010

Στον κυρ Παντελή

ΠΡΟΛΟΓΟΣ

Η παρούσα εργασία αποτελεί τη Διπλωματική μου εργασία στα πλαίσια των μεταπτυχιακών μου σπουδών στον τομέα Οργάνωσης και Διοίκησης, του τμήματος Μηχανικών Παραγωγής και Διοίκησης του Πολυτεχνείου Κρήτης. Η εκπόνηση της πραγματοποιήθηκε υπό την επίβλεψη του Επικουρου Καθηγητή κ. Δούμπου Μιχάλη.

Η επιτυχής ολοκλήρωση αυτής της εργασίας οφείλεται σε πολύ μεγάλο βαθμό στην επίβλεψη και καθοδήγηση από τον καθηγητή μου, κ. Δούμπο Μιχάλη, ο οποίος με τις καταλυτικές του συμβουλές και παρεμβάσεις, μου έλυνε οποιαδήποτε απορία εμφανιζόταν. Η άριστη επικοινωνία και το άριστο κλίμα συνεργασίας που δημιουργούσε σε κάθε μας συνάντηση, μου έδινε την ώθηση να συνεχίσω αγνοώντας τις δυσκολίες που θα εμφανίζονταν. Ένα ευχαριστώ πιστεύω πως δεν θα αρκούσε.

Επίσης, ένα ευχαριστώ δεν θα αρκούσε στην οικογένεια μου, για την εμπιστοσύνη που μου δείχνει όλα αυτά τα χρόνια και με στηρίζει σε κάθε μου προσπάθεια δίνοντας μου θάρρος και κουράγιο να συνεχίσω.

Τέλος, η εργασία αυτή αφιερώνεται στο “δάσκαλο” των προβλέψεων και αναλύσεων, κυρ Παντελή, ως ένδειξη εκτίμησης και σεβασμού, για τα καθημερινά του μαθήματα προς εμάς τους νεώτερους.

ΠΕΡΙΛΗΨΗ

Το άμεσο μάρκετινγκ δεν αποτελεί απλά μια άλλη τακτική μάρκετινγκ αλλά μια αποτελεσματική στρατηγική που αλληλεπιδρά μεταξύ πελάτη και επιχείρησης. Ο έντονος ανταγωνισμός μεταξύ των επιχειρήσεων για τη διατήρηση και αύξηση του μεριδίου τους στην αγορά έφερε στην επιφάνεια την ανάγκη για εφαρμογή προγραμμάτων μάρκετινγκ που θα επικεντρώνονταν περισσότερο στον πελάτη. Το κλασικό μάρκετινγκ περνάει μηνύματα μαζικά στους πελάτες, καθώς όμως η αποτελεσματικότητα του μειωνόταν οι πωλητές άρχισαν να στοχεύουν σε μικρότερα τμήματα αγορών. Το άμεσο μάρκετινγκ, σε αντιδιαστολή με το κλασικό μάρκετινγκ, είναι πελατοκεντρικό καθώς βασίζεται στην άμεση και εξατομικευμένη επικοινωνία, στοχεύοντας στην ανάπτυξη μιας σχέσης με τους πελάτες και μετά την πώληση. Στόχος του άμεσου μάρκετινγκ, είναι η αύξηση που ποσοστού ανταπόκρισης συγκεκριμένων πελατών.

Πολλοί άνθρωποι λαμβάνουν σε μηνιαία βάση ένα μεγάλο όγκο αυτόκλητων επιστολών, δέχονται τηλεφωνικές κλήσεις που στοχεύουν στην αγορά προϊόντων και ξεφυλλίζουν καταλόγους προϊόντων ελάχιστα εκ των οποίων τους ενδιαφέρουν πραγματικά. Ένα σημαντικό ζήτημα που αντιμετωπίζουν οι εταιρίες άμεσου μάρκετινγκ είναι πως θα καταφέρουν να στέλνουν διαφημιστική αλληλογραφία με μεγαλύτερη αποτελεσματικότητα προκειμένου να αυξηθούν τα ποσοστά ανταπόκρισης και να μειωθούν τα έξοδα αποστολής. Εάν η εταιρία είχε μια καλύτερη εκτίμηση για το ποιοι μπορεί να είναι οι πιθανοί πελάτες της, θα γνώριζε με μεγαλύτερη ακρίβεια που να στείλει την αλληλογραφία. Το πρόβλημα επικεντρώνεται στο τρόπο με τον οποίο θα κατορθώσουν οι εταιρίες να προβλέψουν το ενδιαφέρον των πελατών για ένα προϊόν (π.χ. οι τράπεζες, για ένα δάνειο ή οι ασφαλιστικές εταιρείες, για μία ασφάλεια ζωής).

Για την επίλυση του προβλήματος αυτού έχουν προταθεί και μελετηθεί διάφορες μεθοδολογικές προσεγγίσεις από το χώρο της στατιστικής, της επιχειρησιακής έρευνας και της τεχνητής νοημοσύνης. Στην παρούσα εργασία, εφαρμόζονται η Λογιστική Παλινδρόμηση, τα Τεχνητά Νευρωνικά Δίκτυα, τα Δέντρα Αποφάσεων, οι Μηχανές Διανυσμάτων Υποστήριξης, καθώς επίσης ένα συνδυασμένο μοντέλο και υβριδικά μοντέλα, για τον εντοπισμό υποψήφιων πελατών στα πλαίσια μιας στρατηγικής άμεσου μάρκετινγκ για μια ασφαλιστική επιχείρηση. Στόχος είναι η συγκριτική αξιολόγηση των παραπάνω υπολογιστικών διαδικασιών. Τέλος, δίνονται προτάσεις για πιθανές επεκτάσεις και εφαρμογές άλλων τεχνικών για την επίλυση του προβλήματος του εντοπισμού πιθανών πελατών.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ

Άμεσο Μάρκετινγκ, Μάρκετινγκ με τη χρήση Βάσεων δεδομένων, Εξόρυξη Δεδομένων, Λογιστική Παλινδρόμηση, Τεχνητά Νευρωνικά Δίκτυα, Δέντρα Αποφάσεων, Μηχανές Διανυσμάτων Υποστήριξης, Συνδυασμένες Τεχνικές, Υβριδικά Μοντέλα.

ABSTRACT

Direct marketing constitutes not only another approach to marketing but an effective and interactive strategy among a customer and a company. The intense competition between marketing companies forces them to preserve and improve their market share using marketing programs based on customers needs. Mass marketing uses mass media to broadcast messages to customers. However mass marketing has become less effective and marketers need to focus their campaigns in smallest segments of market. Direct marketing compared with mass marketing, is customer-oriented, based on direct and personal communication in order to develop after sales relationship with customers. The main objective of direct marketing is the improvement of the response rate for the selected customers.

Many people receive monthly a huge bulk of unsolicited mails, receive calls for a product promotion and looking at catalogues of products that in fact they don't need. The major issue in the direct marketing mail industry is that mistargeted and unwanted mails wastes financial resources and results in low response rates. If the company had a better prediction to mail "smarter" to prospective customers, it would know with more accuracy where to send it. The problem is in which way companies can predict the response of potential customers for a specific product (e.g. a bank for a loan, an insurance company for a life insurance).

Various techniques from statistics, operation research and artificial intelligence proposed and used for the solution of this problem. In this thesis logistic regression, artificial neural networks, decision trees, support vector machines, hybrid and ensemble models are used to predict potential customers in direct marketing strategy for an insurance company. The objective is the evaluation of the above computational methods. Finally, suggestions for possible enhancements or usage of other techniques on this topic are given.

KEYWORDS

Direct Marketing, Database Marketing, Data Mining, Logistic Regression, Artificial Neural Networks, Decision Trees-CART, Support Vector Machines, Bagging, Hybrid Models.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ ΣΤΟ ΑΜΕΣΟ ΜΑΡΚΕΤΙΝΓΚ	1
1.1 ΕΙΣΑΓΩΓΗ.....	1
1.2 ΟΡΙΣΜΟΣ ΤΟΥ ΑΜΕΣΟΥ ΜΑΡΚΕΤΙΝΓΚ.....	1
1.3 ΤΑ ΠΡΟΣΘΕΤΑ ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΤΟΥ ΑΜΕΣΟΥ ΜΑΡΚΕΤΙΝΓΚ.....	3
1.4 ΟΙ ΜΕΤΑΒΛΗΤΕΣ ΑΠΟΦΑΣΗΣ ΤΟΥ ΑΜΕΣΟΥ ΜΑΡΚΕΤΙΝΓΚ.....	5
1.5 ΤΑ ΜΕΣΑ ΚΑΙ ΟΙ ΜΟΡΦΕΣ ΤΟΥ ΑΜΕΣΟΥ ΜΑΡΚΕΤΙΝΓΚ.....	6
1.6 ΤΟ ΑΜΕΣΟ ΜΑΡΚΕΤΙΝΓΚ ΣΕ ΣΥΝΕΡΓΑΣΙΑ ΜΕ ΤΟ ΜΑΡΚΕΤΙΝΓΚ ΜΕ ΤΗ ΧΡΗΣΗ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ	7
1.7 ΑΝΑΛΥΣΗ ΠΡΟΦΙΛ ΠΕΛΑΤΩΝ.....	9
 ΚΕΦΑΛΑΙΟ 2: ΥΠΟΛΟΓΙΣΤΙΚΑ ΜΟΝΤΕΛΑ ΣΤΟ ΑΜΕΣΟ ΜΑΡΚΕΤΙΝΓΚ	13
2.1 ΕΙΣΑΓΩΓΗ.....	13
2.2 ΜΕΘΟΔΟΙ ΓΙΑ ΤΗΝ ΕΠΙΛΟΓΗ ΣΤΟΧΩΝ	13
2.2.1 Βασικές στατιστικές τεχνικές	14
2.2.2 Προηγμένες στατιστικές τεχνικές	15
2.2.3 Τεχνικές μηχανικής μάθησης.....	16
2.2.4 Συνδυασμένες και υβριδικές τεχνικές	18
2.3 ΜΕΘΟΔΟΙ ΓΙΑ ΟΜΑΔΟΠΟΙΗΣΗ ΠΕΛΑΤΩΝ & ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ ...	20
2.4 ΜΕΘΟΔΟΙ ΓΙΑ CROSS-SELLING & UP-SELLING	21
2.5 ΜΕΘΟΔΟΙ ΓΙΑ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΕΠΙΧΕΙΡΗΜΑΤΙΚΩΝ ΑΠΑΙΤΗΣΕΩΝ	22
2.6 ΑΝΑΛΥΣΗ ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΜΟΝΤΕΛΩΝ	23
2.6.1 Λογιστική Παλινδρόμηση	23
2.6.2 Δέντρα Αποφάσεων.....	24
2.6.3 Τεχνητά Νευρωνικά Δίκτυα	25
2.6.4 Μηχανές Διανυσμάτων Υποστήριξης	26
2.6.5 Συνδυασμένες τεχνικές.....	28

ΚΕΦΑΛΑΙΟ 3: ΕΦΑΡΜΟΓΗ ΣΤΟ ΑΜΕΣΟ ΜΑΡΚΕΤΙΝΓΚ	29
3.1 ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ	29
3.2 ΕΠΙΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΜΕΤΑΒΛΗΤΩΝ	30
3.2.1 Επιλογή μεταβλητών	30
3.2.2 Επεξεργασία μεταβλητών	30
3.2.3 Υπολογιστικά μοντέλα και αξιολόγηση	31
3.3 ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ.....	33
3.3.1 Πίνακες ταξινόμησης	33
3.3.2 Καμπύλη ROC	36
3.3.3 Επίπεδο σημαντικότητας μεταβλητών	38
3.4 BAGGING ΚΑΙ ΥΒΡΙΔΙΚΑ ΜΟΝΤΕΛΑ	41
3.4.1 Πίνακες ταξινόμησης και καμπύλες ROC	41
3.4.2 Επίπεδο σημαντικότητας μεταβλητών	42
ΚΕΦΑΛΑΙΟ 4: ΣΥΜΠΕΡΑΣΜΑΤΑ & ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ	45
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	49
ΠΑΡΑΡΤΗΜΑ.....	55

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ ΣΤΟ ΑΜΕΣΟ ΜΑΡΚΕΤΙΝΓΚ

1.1 ΕΙΣΑΓΩΓΗ

Οι περισσότερες εταιρίες που πωλούν προϊόντα και υπηρεσίες χρειάζεται να διαφημίζουν και να προωθούν τα προϊόντα και τις υπηρεσίες τους. Ορισμένα χαρακτηριστικά παραδείγματα είναι οι τράπεζες, οι ασφαλιστικές εταιρίες και τα εμπορικά καταστήματα. Υπάρχουν δύο προσεγγίσεις για τη διαφήμιση και την προώθηση προϊόντων: το μαζικό μάρκετινγκ (*mass marketing*) και το άμεσο μάρκετινγκ (*direct marketing*).

Το μαζικό μάρκετινγκ, το οποίο χρησιμοποιεί τα Μ.Μ.Ε, όπως την τηλεόραση, το ραδιόφωνο και τις εφημερίδες, προβάλλει μηνύματα/διαφημίσεις στο κοινό χωρίς διάκριση. Όμως, στη σημερινή εποχή, όπου τα προϊόντα είναι αμέτρητα και η αγορά εξαιρετικά ανταγωνιστική, το μαζικό μάρκετινγκ δεν επιτυγχάνει πάντα τα επιθυμητά αποτελέσματα.

Η δεύτερη προσέγγιση για την προώθηση προϊόντων είναι το άμεσο μάρκετινγκ. Σε αντίθεση με τη γενικευμένη, χωρίς διάκριση, προώθηση προϊόντων το άμεσο μάρκετινγκ μελετά τα χαρακτηριστικά και τις ανάγκες των πελατών και επιλέγει συγκεκριμένους πελάτες για την προώθηση των προϊόντων. Στόχος του άμεσου μάρκετινγκ, είναι η αύξηση που ποσοστού ανταπόκρισης συγκεκριμένων πελατών. Το άμεσο μάρκετινγκ, σε αντιδιαστολή με το κλασικό μάρκετινγκ, είναι πελατοκεντρικό καθώς βασίζεται στην άμεση και εξατομικευμένη επικοινωνία, στοχεύοντας στην ανάπτυξη μιας σχέσης με τους πελάτες και μετά την πώληση (Ling και Li, 1998).

Η παρούσα εργασία μελετά το πρόβλημα του εντοπισμού υποψήφιων πελατών στα πλαίσια μιας στρατηγικής άμεσου μάρκετινγκ για μια ασφαλιστική εταιρία. Απώτερος στόχος της, είναι η συγκριτική αξιολόγηση διάφορων υπολογιστικών διαδικασιών από το χώρο της στατιστικής, της επιχειρησιακής έρευνας και της τεχνητής νοημοσύνης.

1.2 ΟΡΙΣΜΟΣ ΤΟΥ ΑΜΕΣΟΥ ΜΑΡΚΕΤΙΝΓΚ

Υπάρχουν διάφοροι ορισμοί δημοσιευμένοι για το άμεσο μάρκετινγκ. Κανείς όμως, από τους υπάρχοντες ορισμούς δεν έχει κατορθώσει να συμπεριλάβει όλες τις πτυχές του σύγχρονου άμεσου μάρκετινγκ. Ορισμένοι από αυτούς δίνονται παρακάτω:

“Η παράδοση ενός μηνύματος ή μιας πρότασης μάρκετινγκ σε ένα συγκεκριμένο πελάτη ή έναν υποψήφιο πελάτη, με μία ευνοϊκή προσφορά για τον πελάτη, μέσω ταχυδρομείου, τηλεφώνου, ηλεκτρονικού ταχυδρομείου και άλλων μέσων χωρίς τη μεσολάβηση ενδιάμεσου πρόσωπου ή έμμεσων μέσων” – M. Roddy, 2002

*“...επικοινωνίες όπου τα δεδομένα χρησιμοποιούνται συστηματικά για να επιτύχουν τους ποσοτικά (quantifiable) αντικειμενικούς στόχους του μάρκετινγκ και η άμεση επαφή γίνεται ή προσκαλείται, μεταξύ μιας επιχείρησης και των πελατών της ή των υποψήφιων πελατών της” – Οργανισμός Άμεσου Μάρκετινγκ
(Direct Marketing Association)*

Ένας ακόμη ορισμός από την DMA, είναι ο ακόλουθος:

“Το άμεσο μάρκετινγκ είναι ένα αμφίδρομο (interactive) σύστημα μάρκετινγκ που χρησιμοποιεί ένα ή περισσότερα διαφημιστικά μέσα, προκειμένου να πραγματοποιήσει μια μετρούμενη ανταπόκριση (response) ή συναλλαγή (transaction) σε οποιαδήποτε σημείο” – Οργανισμός Άμεσου Μάρκετινγκ, DMA

Από τους παραπάνω ορισμούς, γίνεται σαφές πως το άμεσο μάρκετινγκ αντιπροσωπεύει μία νέα προσέγγιση στο χώρο του μάρκετινγκ, καθώς ταξινομεί τους πελάτες, έτσι ώστε η εξατομικευμένη διαφήμιση και οι προωθητικές ενέργειες να μπορούν να στοχεύσουν σε συγκεκριμένες κατηγορίες πελατών (Bose και Chen, 2009). Επικεντρώνεται στην ανάπτυξη σχέσεων με τους πελάτες, αποσκοπώντας σε νέες πωλήσεις με την πάροδο του χρόνου. Αυτό πραγματοποιείται με τη συλλογή και την επεξεργασία των δεδομένων των πελατών για να παραχθούν νέες στρατηγικές μάρκετινγκ, οι οποίες θα είναι εξυπνότερες, περισσότερο συγκεκριμένες και εξατομικευμένες από τις παραδοσιακές στρατηγικές μάρκετινγκ (Roberts και Berger, 1999).

Στον Πίνακα 1.1 παρουσιάζονται οι διαφορές ανάμεσα στο κλασικό-μαζικό μάρκετινγκ και το άμεσο μάρκετινγκ. Οι βασικότερες διαφορές που διαφοροποιούν το άμεσο μάρκετινγκ από το μαζικό είναι, πως το άμεσο μάρκετινγκ:

- διευκολύνει τη στοχευμένη/συγκεκριμένη επικοινωνία με τον πελάτη
- επιτρέπει στην επικοινωνία αυτή να είναι προσωπική
- ενθαρρύνει τον υποψήφιο πελάτη να λάβει άμεση, συγκεκριμένη δράση
- κατορθώνει να κάνει λιγότερο ορατές τις στρατηγικές μάρκετινγκ στους ανταγωνιστές
- είναι μετρήσιμο, παρέχοντας τη δυνατότητα υπολογισμού των δαπανών μάρκετινγκ, κάτι το οποίο δεν υφίσταται στο κλασικό μάρκετινγκ

Οι παραπάνω πρόσθετες ικανότητες είναι σε μεγάλο μέρος υπεύθυνες για τη συνεχόμενη επέκταση του άμεσου μάρκετινγκ.

Μαζικό Μάρκετινγκ	Άμεσο Μάρκετινγκ
Προσεγγίζει ένα μαζικό κοινό μέσω των Μ.Μ.Ε	Επικοινωνεί άμεσα με τον πελάτη ή τον υποψήφιο πελάτη
Η επικοινωνία είναι απρόσωπη	Μπορεί να εξατομικεύει την επικοινωνία απευθυνόμενο με όνομα/επώνυμο/τίτλο
Η επικοινωνία είναι μονόπλευρη, από τον διαφημιστή στον πελάτη	Η επικοινωνία είναι αμφίδρομη
Οι προωθητικές ενέργειες είναι ιδιαίτερα ορατές	Οι προωθητικές ενέργειες (ειδικά οι δοκιμαστικές) είναι σχετικά “αόρατες”
Το ποσό των προωθητικών ενεργειών ελέγχεται από το μέγεθος του προϋπολογισμού (<i>budget</i>)	Το μέγεθος του προϋπολογισμού μπορεί να καθοριστεί από την επιτυχία των δοκιμαστικών/προωθητικών ενεργειών
Το επιθυμητό αποτέλεσμα είτε καθυστερεί είτε είναι ασαφές	Συγκεκριμένο αποτέλεσμα ζητείται πάντα από την έρευνα ή την αγορά
Ελλιπή ή δείγματα δεδομένων, για λόγους λήψης αποφάσεων λαμβάνονται από έρευνες μάρκετινγκ	Αναλυτική/Περιεκτική βάση δεδομένων οδηγεί τα προγράμματα μάρκετινγκ
Τα δεδομένα πρέπει να συλλέγονται ξεχωριστά από τη διαδικασία πωλήσεων	Τα δεδομένα παράγονται ως αναπόσπαστο κομμάτι της διαδικασίας πωλήσεων
Η ανάλυση διεξάγεται σε τμηματικό επίπεδο	Η ανάλυση διεξάγεται σε ατομικό/σταθερό επίπεδο
Χρησιμοποιεί μεταβλητές, όπως η πρόθεση αγοράς για την μέτρηση της αποτελεσματικότητας	Μετρήσιμο και επομένως ιδιαίτερα ελέγξιμο

Πίνακας 1.1: Βασικές διαφορές μεταξύ άμεσου και μαζικού μάρκετινγκ.
(Πηγή: *Direct Marketing Management, Roberts και Berger, 1999*)

1.3 ΤΑ ΠΡΟΣΘΕΤΑ ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΤΟΥ ΑΜΕΣΟΥ ΜΑΡΚΕΤΙΝΓΚ

Σύμφωνα με τους Roberts και Berger (1999), το γεγονός ότι το άμεσο μάρκετινγκ λειτουργεί ιδιαίτερα καλά, οφείλεται στο ότι η επικοινωνία των άμεσων πωλητών (*direct marketers*) είναι στοχευμένη σε συγκεκριμένα άτομα, και όχι στις τμηματικές μαζικές αγορές μέσω των Μ.Μ.Ε.

1. Ακριβής στόχευση (*precision targeting*)

Μέσω της χρήσης των προσεκτικά επιλεγμένων καταλόγων αλληλογραφίας και των πληροφοριών που περιλαμβάνονται στις βάσεις δεδομένων των πελατών, ένας άμεσος πωλητής μπορεί να κατευθύνει τις επικοινωνίες σε ένα συγκεκριμένο πελάτη ή ένα συγκεκριμένο εμπορικό πελάτη, που έχει αρχικά προσδιοριστεί ως βιώσιμη προοπτική. Για παράδειγμα, ένας πελάτης συνηθίζει να στέλνει ευχετήριες κάρτες στην περίοδο των γιορτών σε διάφορους φίλους του (κάτι το οποίο οι άμεσοι πωλητές το γνωρίζουν, γιατί έχουν σταλεί σε διαφορετικές διευθύνσεις) άρα ο συγκεκριμένος πελάτης θα μπορούσε προσεγγιστεί ούτως ώστε να πράξει το ίδιο και το επόμενο έτος. Θα ήταν χρήσιμο λοιπόν, να υπενθυμίσουν στον πελάτη πού και τι έστειλε το

προηγούμενο έτος. Την πληροφορία αυτή την περιέχουν οι βάσεις δεδομένων του άμεσου μάρκετινγκ.

2. Εξατομίκευση (personalization)

Ένας συγκεκριμένος πελάτης θα πρέπει να προσφωνείται με το όνομά του, ενώ ένας εμπορικός πελάτης με το όνομα και τη θέση που κατέχει. Παρόλα αυτά, η δυνατότητα της προσωποποίησης εκτείνεται πέραν της απλής χρήσης των ονομάτων. Πληροφορίες από τις βάσεις δεδομένων θα μπορούσαν να χρησιμοποιηθούν για την επιλογή μιας κατάλληλης επίκλησης προς τους πελάτες (για παράδειγμα, “*Η καλλιέργεια του φυτού Α ενδείκνυται για την περιοχή σας κατά τους καλοκαιρινούς μήνες*”) ή για να προτρέψουν τον πελάτη, ο οποίος είχε αγοράσει παλαιότερα, να πραγματοποιήσει μια νέα αγορά (για παράδειγμα, “*Ως ένα ευχαριστώ προς τους πιστούς μας πελάτες, σας προσφέρουμε μια μέρα εκπτώσεων πριν ανοίξουμε τις πόρτες μας για το ευρύ κοινό*”).

3. Πρόσκληση για άμεση δράση (call for immediate action)

Οι άμεσοι πωλητές χρησιμοποιούν προσκλήσεις για συγκεκριμένη και άμεση δράση των πελατών, είτε για την αγορά ενός προϊόντος είτε για να μάθουν περισσότερες πληροφορίες όσον αφορά το προϊόν. Η πρόσκληση για άμεση δράση, λειτουργεί σχεδόν πάντα ενάντια στην τάση της λογικής να χαθεί αυτή η προοπτική και οι πελάτες ενδίδουν. Ορισμένες παραινήσεις που συχνά χρησιμοποιούνται είναι, “*Προσφορά περιορισμένης χρονικής διάρκειας*”, “*Η προσφορά δεν θα επαναληφθεί*” και “*Απαντήστε αμέσως*”.

4. “Αόρατες” στρατηγικές (invisible strategies)

Οι στρατηγικές και οι τακτικές των άμεσων πωλητών είναι λιγότερο ορατές προς τους ανταγωνιστές σε σχέση με τις στρατηγικές που εφαρμόζονται στα Μ.Μ.Ε. Οι εταιρίες χρησιμοποιούν τακτικές για να είναι ενήμερες για τις κινήσεις των ανταγωνιστών, όπως να περιλαμβάνονται στους παραλήπτες καταλόγων. Όμως, αυτό δεν αποδεικνύεται και τόσο αποτελεσματικό, γιατί η στιγμή που ο ανταγωνιστής θα λάβει τον κατάλογο είναι η ίδια που τον λαμβάνει και ο πελάτης, άρα μπορεί να είναι πάρα πολύ αργά για να απαντήσει αποτελεσματικά. Επίσης, η έλλειψη πληροφοριών όσον αφορά τις στρατηγικές είναι ιδιαίτερα σημαντική κατά τις δοκιμές νέων στρατηγικών άμεσου μάρκετινγκ. Μια παραδοσιακή τεχνική αποστολής δοκιμαστικού διαφημιστικού υλικού, η οποία μπορεί να κατορθώσει να περάσει απαρατήρητη από τους ανταγωνιστές, πραγματοποιείται με την απόκρυψη/μη αποστολή του διαφημιστικού υλικού σε εκείνους τους ταχυδρομικούς κώδικες, οι οποίοι βρίσκονται στην ευρύτερη περιοχή της έδρας ενός ανταγωνιστή, υποθέτοντας ότι οι οικογένειες των στελεχών είναι πιθανόν να βρίσκονται σε αυτούς τους ταχυδρομικούς κώδικες.

5. Μετρήσιμο (Measurability)

Το σημαντικότερο πλεονέκτημα που κατευθύνει το άμεσο μάρκετινγκ πέρα από τις άλλες τεχνικές μάρκετινγκ είναι αδιαμφισβήτητη η δυνατότητά του να ακολουθηθεί και επομένως να μετρήσει την αποτελεσματικότητα, συγκεκριμένων ενεργειών μάρκετινγκ. Γνωρίζοντας επακριβώς τι δουλίζει και τι όχι, ο πωλητής μπορεί να κατανέμει τα διαθέσιμα κεφάλαια περισσότερο αποτελεσματικά.

Λαμβάνοντας υπόψη όλες τις παραπάνω πρόσθετες ικανότητες, οι άμεσοι πωλητές χρειάζεται να εξετάσουν το σύνολο των μεταβλητών απόφασης, που θα συνδυαστούν για ένα προσεγμένο και αποτελεσματικό πρόγραμμα προώθησης.

1.4 ΟΙ ΜΕΤΑΒΛΗΤΕΣ ΑΠΟΦΑΣΗΣ ΤΟΥ ΑΜΕΣΟΥ ΜΑΡΚΕΤΙΝΓΚ

Στο κλασικό μάρκετινγκ οι μεταβλητές που χρησιμοποιεί ο πωλητής με τέτοιον τρόπο ώστε να επιτύχει την ανταπόκριση που θέλει, είναι γνωστές ως τα 4Ps, από τα αρχικά των αγγλικών λέξεων, “*Product*” (Προϊόν), “*Price*” (Τιμή), “*Place*” (Διανομή), “*Promotion*” (Προώθηση). Αντίστοιχες μεταβλητές με τα 4Ps του κλασικού μάρκετινγκ μπορούν να προσδιοριστούν και για το άμεσο μάρκετινγκ. Μόνο που στο άμεσο μάρκετινγκ προστίθεται και μια πέμπτη μεταβλητή, η εξυπηρέτηση πελατών, ως απαραίτητο στοιχείο της διαδικασίας του μάρκετινγκ (Roberts και Berger, 1999).

Οι μεταβλητές απόφασης που διαμορφώνουν το βασικό πλαίσιο εργασίας γύρω από τα προγράμματα προώθησης άμεσου μάρκετινγκ είναι:

- Προσφορά–*Offer*
- Δημιουργικότητα–*Creative*
- Μέσα Επικοινωνίας–*Media*
- Συγχρονισμός–*Timing/Sequencing*
- Εξυπηρέτηση πελατών–*Customer service*

1. Προσφορά

Η προσφορά είναι μια πλήρης πρόταση που γίνεται από τον πωλητή σε έναν υποψήφιο πελάτη. Περιλαμβάνει το προϊόν ή την υπηρεσία, την τιμή στην οποία προσφέρεται, οποιουσδήποτε διακανονισμούς στην τιμή και άλλα στοιχεία που προσδιορίζουν τη στρατηγική που ακολουθείται για την προώθηση του προϊόντος.

2. Δημιουργικότητα

Το κομμάτι της δημιουργικότητας περιλαμβάνει την πλατφόρμα του σχεδιασμού, τα γραφικά στοιχεία και όλες τις τεχνικές που αφορούν την προσωποποίηση του μηνύματος.

3. Μέσα Επικοινωνίας

Τα μέσα επικοινωνίας τα οποία είναι διαθέσιμα στο άμεσο μάρκετινγκ περιλαμβάνουν όλα αυτά που υπάρχουν στο κλασικό μάρκετινγκ και επιπλέον το κλασικό ταχυδρομείο, το τηλέφωνο, καθώς και τους νέους τρόπους ψηφιακής επικοινωνίας όπως, το διαδίκτυο ή το ηλεκτρονικό ταχυδρομείο.

4. Συγχρονισμός

Αφορά τη χρονική στιγμή που θα προβληθεί το μήνυμα στους πελάτες και ως ένα βαθμό είναι όμοιο με το κλασικό μάρκετινγκ. Πιο συγκεκριμένα οι πωλητές προσπαθούν να πετύχουν τον πελάτη σε εκείνη τη χρονική περίοδο που θα είναι πιο πιθανό να δεχτεί θετικά το προϊόν που του προβάλλεται. Παράλληλα, περιλαμβάνει τη συχνότητα με την οποία θα πρέπει να διενεργούνται ενέργειες άμεσου μάρκετινγκ δηλαδή, το πόσο τακτικά θα έρχονται οι πωλητές σε επικοινωνία με τους πελάτες στη διάρκεια του χρόνου. Αυτό είναι πολύ σημαντικό, ώστε οι ενέργειες άμεσου μάρκετινγκ να μην γίνονται ενοχλητικές για τον πελάτη, πράγμα που μπορεί να οδηγήσει στην αποστροφή του από την εταιρία.

5. Εξυπηρέτηση πελατών

Αν και αναγνωρίστηκε τα τελευταία χρόνια δεν μπορεί να υποτιμηθεί και αποτελεί την πέμπτη βασική μεταβλητή του άμεσου μάρκετινγκ. Οι τηλεφωνικοί αριθμοί χωρίς χρέωση, η δοκιμαστική περίοδος χρήσης ενός προϊόντος/υπηρεσίας, η αποδοχή πιστωτικών καρτών και οι άτοκες δόσεις για την εξόφληση ενός προϊόντος αποτελούν βασικές τεχνικές για να αντιμετωπιστούν οι αντιρρήσεις των πελατών. Ακόμη, πιο σημαντικό είναι το επίπεδο της εξυπηρέτησης που παρέχεται, η ταχύτητα και η ακρίβεια στην εξυπηρέτηση, καθώς και ο σωστός χειρισμός των παραπόνων δυσαρεστημένων πελατών.

Οι επιχειρήσεις οι οποίες ξεχωρίζουν για την τελειότητα στην εξυπηρέτηση πελατών, αναζητούν συνεχώς νέους τρόπους για να βελτιώσουν περαιτέρω τις υπηρεσίες που προσφέρουν στους πελάτες τους. Για παράδειγμα, η Federal Express, γνωστή και ως FedEx, προσφέρει μέσω της ιστοσελίδας της, τη δυνατότητα στον πελάτη να σχεδιάσει και να παρακολουθεί την αποστολή των δεμάτων του.

1.5 ΤΑ ΜΕΣΑ ΚΑΙ ΟΙ ΜΟΡΦΕΣ ΤΟΥ ΑΜΕΣΟΥ ΜΑΡΚΕΤΙΝΓΚ

Γενικότερα οποιοδήποτε μέσο που μπορεί να χρησιμοποιηθεί για να παραδώσει ένα μήνυμα σε έναν πελάτη μπορεί να χρησιμοποιηθεί στο άμεσο μάρκετινγκ. Ο Πίνακας 1.2 παρουσιάζει την κατανομή των διαφημιστικών δαπανών στα μέσα επικοινωνίας, του άμεσου μάρκετινγκ σε σύγκριση με τις συνολικές δαπάνες διαφήμισης για το έτος 2008 στις ΗΠΑ, σύμφωνα με τα στοιχεία του Οργανισμού Άμεσου Μάρκετινγκ, DMA.

Η πρώτη εκδοχή του άμεσου μάρκετινγκ ήταν το ταχυδρομικό μάρκετινγκ, δηλαδή η εξατομικευμένη επικοινωνία μέσω του κλασικού ταχυδρομείου. Η πρώτη αυτή εκδοχή κρατά, όπως φαίνεται και από τον Πίνακα 1.2, ακόμη και σήμερα, την πρώτη θέση στους τρόπους επικοινωνίας του άμεσου μάρκετινγκ. Επίσης, εκτός από το τηλεφωνικό μάρκετινγκ και κυρίως λόγω της τεχνολογικής ανάπτυξης, το ηλεκτρονικό ταχυδρομείο και άλλοι νέοι τρόποι ψηφιακής επικοινωνίας κατέχουν σημαντικό μερίδιο των διαφημιστικών δαπανών.

2008	Διαφημιστικές δαπάνες DM	Συνολικές διαφημιστικές δαπάνες	Ποσοστό % DM στις συνολικές δαπάνες
Ταχυδρομείο	56.5	61.7	91.5%
Τηλέφωνο	42.5	86.6	49.1%
Εφημερίδες	13.0	38.8	33.5%
Τηλεόραση	22.8	75.9	30.1%
Περιοδικά	8.9	18.2	49.0%
Ραδιόφωνο	4.6	18.7	24.6%
Νέες Τεχνολογίες	28.6	39.4	72.5%
Σύνολο	176.9	339.3	52.1%

1.DM: Άμεσο Μάρκετινγκ.

2.Ποσά σε δισεκατομμύρια δολάρια.

Πίνακας 1.2: Διαφημιστικές δαπάνες άμεσου μάρκετινγκ σε σύγκριση με τις συνολικές δαπάνες για τις ΗΠΑ για το έτος 2008.

(Πηγή: *Direct Marketing Association*, www.marketingcharts.com)

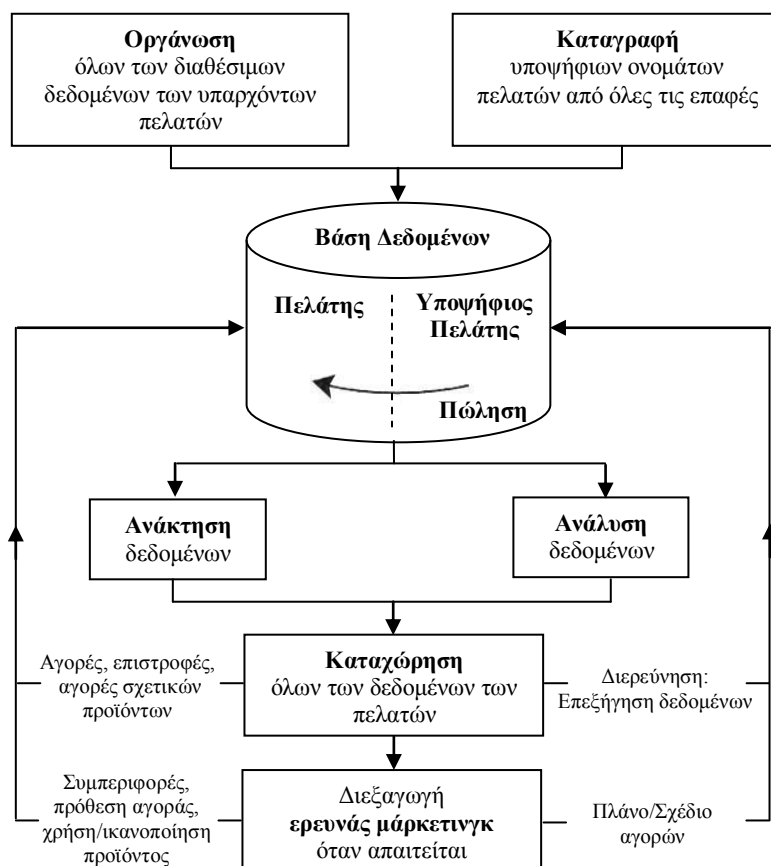
Είναι σημαντικό να αναφερθεί ότι, αν και πολλά από τα παραπάνω μέσα επικοινωνίας χρησιμοποιούνται και από το κλασικό και από το άμεσο μάρκετινγκ, όμως στο άμεσο μάρκετινγκ η χρησιμοποίηση και η αξιολόγηση της αποτελεσματικότητας των μέσων επικοινωνίας είναι διαφορετική.

1.6 ΤΟ ΑΜΕΣΟ ΜΑΡΚΕΤΙΝΓΚ ΣΕ ΣΥΝΕΡΓΑΣΙΑ ΜΕ ΤΟ ΜΑΡΚΕΤΙΝΓΚ ΜΕ ΤΗ ΧΡΗΣΗ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

Το μάρκετινγκ με τη χρήση βάσεων δεδομένων (*Data Base Marketing-DBM*) είναι μια μορφή του άμεσου μάρκετινγκ χρησιμοποιώντας βάσεις δεδομένων των πελατών ή των πιθανών πελατών, για να παράγει τις εξατομικευμένες επικοινωνίες με σκοπό την προώθηση ενός προϊόντος ή μιας υπηρεσίας. Στηρίζεται στην τεχνολογία της συλλογής, επεξεργασίας και χρήσης μεγάλου όγκου πληροφοριών για πελάτες, πραγματικούς ή πιθανούς.

Η διαφορά ανάμεσα στο άμεσο μάρκετινγκ και το DBM βρίσκεται στην ιδιαίτερη προσοχή που δίδεται στην ανάλυση των δεδομένων. Το DBM δίνει έμφαση στη χρήση στατιστικών και άλλων τεχνικών για την ανάπτυξη μοντέλων συμπεριφοράς των πελατών, τα οποία στη συνέχεια χρησιμοποιούνται για την επιλογή νέων πελατών. Η διαδικασία που ακολουθείται στο DBM παρουσιάζεται στο Σχήμα 1.1.

Μια αεροπορική εταιρία, για παράδειγμα, σχεδιάζει μια βάση δεδομένων με σκοπό να παρακολουθήσει τα μέλη του προγράμματος τακτικών επιβατών, χρησιμοποιώντας τη διαδικασία που παρουσιάζεται στο Σχήμα 1.1. Η συγκεκριμένη βάση δεδομένων με την πάροδο του χρόνου αποκτά ένα σημαντικό όγκο δεδομένων όσον αφορά το κάθε μέλος του προγράμματος, που περιλαμβάνει την απόσταση των μιλίων κάθε ταξιδιού, την ενοικίαση αυτοκινήτου ή κράτηση δωματίου, καθώς ακόμη και τα μίλια που απαιτούνται ώστε το μέλος να λάβει ορισμένα προνόμια σύμφωνα με την πολιτική του προγράμματος.



Σχήμα 1.1: Η διαδικασία του DBM.

(Πηγή: *Direct Marketing Management, Roberts και Berger, 1999*)

Η αεροπορική εταιρία κατορθώνει να συγκεντρώσει όλες αυτές τις σημαντικές πληροφορίες από τις συναλλαγές των μελών της (*επαφές πελατών*). Λαμβάνει έτσι, πολύτιμες πληροφορίες για το πόσο συχνά πετούν τα μέλη, σε ποιους προορισμούς, την διάρκεια του ταξιδιού, καθώς και τον αριθμό και τον τύπο των εισιτηρίων που αγοράζουν. Ακόμη, κατορθώνει να λαμβάνει πληροφορίες για την ενοικίαση αυτοκινήτου ή κράτηση δωματίου, παρέχοντας ειδικά προνόμια στα μέλη που το πράττουν, με αποτέλεσμα ένα ουσιαστικό ποσοστό να ανταποκριθεί συμβάλλοντας θετικά στην έρευνα μάρκετινγκ.

Η διαχείριση των δεδομένων αυτών (*ανάλυση δεδομένων*) μπορεί να πραγματοποιηθεί με διαφορετικούς τρόπους με σκοπό την εξαγωγή χρήσιμων πληροφοριών. Μπορούν να αναλυθούν τα σχέδια του ταξιδιού κάθε μέλους ξεχωριστά. Για παράδειγμα, γνωρίζοντας ότι κάποιο μέλος ταξιδεύει συχνά από την πόλη Α στην πόλη Β, μπορεί να επικεντρωθεί στο μέλος αυτό (*ανάκτηση δεδομένων*) πληροφορώντας για ειδικές τιμές που υπάρχουν σε ορισμένες πτήσεις από την πόλη Α στην πόλη Β ή προτείνοντας διαμονή σε ένα ξενοδοχείο στην πόλη Β. (Η εταιρία γνωρίζει από την καρτέλα του πελάτη ή από τον ταχυδρομικό του κώδικα ότι διαμένει στην πόλη Α, επομένως ίσως να τον ενδιέφερε η πρόταση διαμονής σε ένα ξενοδοχείο στην πόλη Β).

Επειδή, τα μέλη είναι περισσότερο προσιτά η εταιρία μπορεί να επιθυμεί να συλλέξει δεδομένα (*έρευνα μάρκετινγκ*) από τις συμπεριφορές των μελών, την

ικανοποίησή τους από τα δρομολόγια, καθώς επίσης και από το επίπεδο ικανοποίησης από την εξυπηρέτηση που δέχονται. Η πιθανότητα να απαντήσει ένα μέλος σε ένα ερωτηματολόγιο ή μια αίτηση είναι πολύ μεγαλύτερη από το να επιλεγεί τυχαία ένας πελάτης της εταιρίας. Η εταιρία κατορθώνει έτσι να συλλέξει πληροφορίες οι οποίες μπορούν να χρησιμοποιηθούν άμεσα για την ανάπτυξη υπηρεσιών και προωθητικών ενεργειών που θα ήταν ελκυστικές για το μέλος και άλλους πελάτες με παρόμοια συμπεριφορά.

Από το παραπάνω χαρακτηριστικό παράδειγμα της διαδικασίας που ακολουθείται στο DBM εξάγεται το συμπέρασμα, ότι η αεροπορική εταιρία ή η εκάστοτε εταιρία μπορεί να διαχειριστεί τις πληροφορίες που η ίδια συλλέγει για να κατανοήσει με μεγαλύτερη ακρίβεια τη συμπεριφορά των πελατών της. Είναι επίσης σημαντικό να τονιστεί, ότι οι βάσεις δεδομένων είναι δυναμικές και όχι στατικές. Στοιχεία προστίθενται συνεχώς σε αυτές, είτε ως αποτέλεσμα της αλληλεπίδρασης με τους πελάτες, είτε από τις έρευνες μάρκετινγκ, είτε με τον εμπλουτισμό στοιχείων από εξωτερικές πηγές.

Η χρησιμοποίηση όλων των πληροφοριών που περιλαμβάνονται στις βάσεις δεδομένων παρέχουν στις εταιρίες τη δυνατότητα να σχεδιάσουν και να αναπτύξουν τα προγράμματα άμεσου μάρκετινγκ, τα οποία θα απευθύνονται σε συγκεκριμένους πελάτες. Γίνεται σαφές, ότι ο συνδυασμός άμεσου μάρκετινγκ και DBM σε μια εταιρία, αυξάνει την αποτελεσματικότητα κάθε ενέργειας προώθησης και μειώνει το κόστος άσκοπων δαπανών της (Roberts και Berger, 1999).

1.7 ΑΝΑΛΥΣΗ ΠΡΟΦΙΛ ΠΕΛΑΤΩΝ

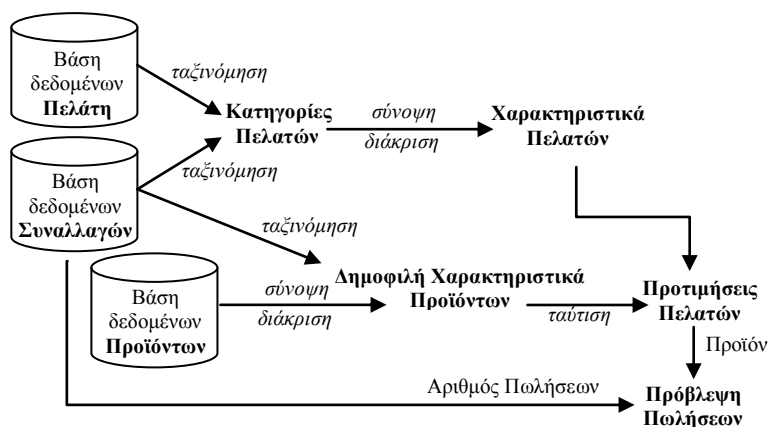
Ο εντοπισμός των αναγκών και των κινήτρων των πελατών αποτελεί βασική προϋπόθεση για κάθε πετυχημένο πρόγραμμα άμεσου μάρκετινγκ. Η διαμόρφωση του αγοραστικού προφίλ των πελατών, κατανοώντας δηλαδή τις συμπεριφορές τους, μπορεί να προσφέρει στην εταιρία το πλεονέκτημα να προσαρμόσει τα προϊόντα καλύτερα στους πελάτες αποκτώντας σημαντικό προβάδισμα έναντι του ανταγωνισμού (Thomas et al., 2007).

Στο σημερινό σύνθετο και συνεχώς μεταβαλλόμενο, όσον αφορά τις προτιμήσεις των πελατών, περιβάλλον οι αποφάσεις μάρκετινγκ που λαμβάνονται σύμφωνα με την ενημέρωση από τη γνώση (*knowledge*) για κάθε συγκεκριμένο πελάτη αποδεικνύονται κρίσιμης σημασίας. Στην πραγματικότητα, κάθε πελάτης θέλει να εξυπηρετηθεί σύμφωνα με τις ιδιαίτερες και μοναδικές του ανάγκες (Shaw et al., 2001).

Συμφωνά με τους Shaw et al. (2001), τα εργαλεία εξόρυξης δεδομένων (*data mining*) παρέχουν στους σημερινούς πωλητές (*marketers*) ακριβώς το σωστό είδος της γνώσης για να κατορθώσουν να λάβουν τις κατάλληλες αποφάσεις μάρκετινγκ. Ένας σημαντικός παράγοντας για την εφαρμογή των εργαλείων εξόρυξης δεδομένων είναι η ανάλυση του προφίλ των πελατών.

Μια χρήσιμη γνώση όσον αφορά τον πελάτη είναι το προφίλ του, το οποίο χρησιμοποιείται για να ληφθούν διάφορες σημαντικές αποφάσεις μάρκετινγκ. Το προφίλ του πελάτη απεικονίζει τα χαρακτηριστικά και τις ανάγκες ενός πελάτη, βάσει

των οποίων ο πωλητής αποφασίζει τις σωστές στρατηγικές και τακτικές για την ικανοποίηση των αναγκών του πελάτη. Το Σχήμα 1.2 παρουσιάζει ένα σύστημα ανάλυσης του προφίλ πελάτη που χρησιμοποιεί εργαλεία εξόρυξης δεδομένων.



Σχήμα 1.2: Σύστημα ανάλυσης προφίλ πελάτη.
(Πηγή: Shaw et al.,2001)

Μαθαίνοντας το προφίλ του πελάτη, ο πωλητής ενδιαφέρεται για τις δημογραφικές λεπτομέρειες του πελάτη, καθώς και για τα χαρακτηριστικά των συναλλαγών που πραγματοποιεί. Στη συνέχεια αναφέρονται ορισμένα χαρακτηριστικά που μπορούν να συμβάλλουν στην κατανόηση και ανάλυση του προφίλ πελατών.

1. Συχνότητα αγορών

Γνωρίζοντας τη συχνότητα των αγορών ή των επισκέψεων σε ένα κατάστημα, ο πωλητής μπορεί να χτίσει στοχευμένες προωθητικές ενέργειες, όπως τα “προγράμματα συχνών αγοραστών”.

2. Ποσό χρημάτων για τις αγορές

Η πληροφορία του ποσού που ξοδεύει ο πελάτης σε μια τυπική συναλλαγή παρέχει στον πωλητή τη δυνατότητα να αφιερώσει του κατάλληλους πόρους στον πελάτη που ξοδεύει περισσότερο.

3. Χρόνος τελευταίας αγοράς

Ο πωλητής μπορεί να ερευνήσει τους λόγους για τους οποίους ο πελάτης ή μια ομάδα πελατών δεν έχει πραγματοποιήσει κάποια συναλλαγή για ένα μεγάλο χρονικό διάστημα και να λάβει τα κατάλληλα μέτρα. Πολλές φορές, αυτό οφείλεται στο γεγονός ότι είτε ο πελάτης έχει στρέψει το ενδιαφέρον του ή έχει μετακινηθεί κάπου αλλού.

4. Προσδιορισμός τυπικών ομάδων πελατών

Τα χαρακτηριστικά κάθε ομάδας μπορούν να ληφθούν από την κατηγορία όπου ανήκει κάθε ομάδα ή την περιγραφή συγκεκριμένης ενέργειας. Για παράδειγμα, στο προφίλ ενός πελάτη εμφανίζεται ότι έχει αγοράσει ένα καινούργιο σπίτι, αυτό μπορεί να οδηγήσει τον πωλητή να του κάνει μια προσφορά για την επίπλωση. Η γνώση του πελάτη και η στοχευμένη προσφορά προς αυτόν αποδίδει περισσότερο από ένα γενικό μήνυμα.

5. Υπολογισμός της αξίας διατήρησης υπαρχόντων πελατών

Για να κατορθώσει η εταιρία να υπολογίσει το κόστος που απαιτεί η διατήρηση ενός πελάτη, πρέπει να γνωρίζει πόσα έσοδα ή καλύτερα κέρδη προσδοκεί από αυτόν στο μέλλον (*customer LifeTime Values-LTV*). Αυτά μπορούν να υπολογιστούν χρησιμοποιώντας πληροφορίες από το προφίλ των πελατών σε συνδυασμό με το προϊόν και τα στατιστικά στοιχεία των προωθητικών ενεργειών.

6. Υποψήφιοι πελάτες

Το προφίλ πελατών και ειδικότερα τα πρότυπα (*patterns*) αγορών τους προσφέρουν στοιχεία για υποψήφιους πελάτες. Για παράδειγμα, το πρότυπο ότι η αγορά παιχνιδιών για τις ηλικίες 3-5 ακολουθείται από την αγορά ενός ποδηλάτου μέσα σε διάστημα 6 μηνών σε ποσοστό 90% από πελάτες με υψηλά εισοδήματα, ανακαλύφθηκε από εργαλεία εξόρυξης δεδομένων (Shaw et al.,2001). Γνωρίζοντας το παραπάνω πρότυπο ο πωλητής μπορεί να προσδιορίσει ενδεχόμενους πελάτες βασιζόμενος στα στοιχεία των αγορών παιχνιδιών και να προσαρμόσει ένα κατάλογο προσφορών, αυξάνοντας έτσι την προοπτική προσέλκυσης νέων πελατών.

7. Επιτυχία/Αποτυχία προγραμμάτων μάρκετινγκ

Οι βάσεις δεδομένων προσφέρουν ακριβής πληροφορίες για τα αποτελέσματα των προγραμμάτων μάρκετινγκ. Ο πωλητής μπορεί να χρησιμοποιήσει τα πρότυπα αγορών τα οποία ανακαλύφθηκαν από τις βάσεις δεδομένων και σχετικά προγράμματα μάρκετινγκ για να μετρήσει τα βραχυπρόθεσμα και μακροπρόθεσμα αποτελέσματα των προγραμμάτων.

Τέλος, αξίζει να αναφερθεί πως υπάρχουν δύο τύποι δεδομένων για τα χαρακτηριστικά των πελατών που χρησιμοποιούνται στο άμεσο μάρκετινγκ.

Ο πρώτος τύπος περιλαμβάνει δημογραφικά, γεωγραφικά, και χαρακτηριστικά του τρόπου ζωής και της κοινωνικής κατάστασης των πελατών. Τα δημογραφικά χαρακτηριστικά περιλαμβάνουν την ηλικία, το φύλο κτλ. Τα χαρακτηριστικά του τρόπου ζωής περιλαμβάνουν συνήθειες των πελατών, την ανάγνωση συγκεκριμένης εφημερίδας κτλ. Τέλος, τα γεωγραφικά περιλαμβάνουν την τοποθεσία κατοικίας, του γραφείου ή της εργασίας των πελατών.

Ο δεύτερος τύπος δεδομένων περιλαμβάνει την αμφίδρομη επικοινωνία των πελατών με τους πωλητές. Ο συνδυασμός των τριών πρώτων μεταβλητών της ανάλυσης του προφίλ των πελατών, που παρουσιάστηκαν παραπάνω, αναφέρεται στη βιβλιογραφία ως μεταβλητές RFM, από τα αρχικά των λέξεων Recency-Frequency-Monetary. Οι εταιρίες θεωρούν ως καλύτερα τμήματα της αγοράς, με αυξημένη την πιθανότητα οι πελάτες να ξανά αγοράσουν στο μέλλον, αυτά στα οποία οι πελάτες έχουν πραγματοποιήσει πρόσφατα συναλλαγές, αγοράζουν με μεγάλη συχνότητα και ξοδεύουν τα περισσότερα χρήματα (Bose και Chen,2009).

ΚΕΦΑΛΑΙΟ 2

ΥΠΟΛΟΓΙΣΤΙΚΑ ΜΟΝΤΕΛΑ ΣΤΟ ΑΜΕΣΟ ΜΑΡΚΕΤΙΝΓΚ

2.1 ΕΙΣΑΓΩΓΗ

Για την επίλυση προβλημάτων άμεσου μάρκετινγκ έχουν προταθεί και χρησιμοποιηθεί διάφορες μεθοδολογίες από τους ερευνητές. Αναλόγως με το προς επίλυση πρόβλημα άμεσου μάρκετινγκ οι μεθοδολογίες χωρίζονται σε τέσσερις κατηγορίες: Τις μεθόδους για την επιλογή στόχων άμεσου μάρκετινγκ, τις μεθόδους για ομαδοποίηση πελατών και αναγνώριση προτύπων, τις μεθόδους για cross-selling και up-selling και τις μεθόδους για βελτιστοποίηση των επιχειρηματικών απαιτήσεων. Διάφορες τεχνικές από το χώρο της στατιστικής, της επιχειρησιακής έρευνας και της τεχνητής νοημοσύνης χρησιμοποιούνται για την επίλυση των προβλημάτων που ανήκουν στις παραπάνω κατηγορίες.

2.2 ΜΕΘΟΔΟΙ ΓΙΑ ΤΗΝ ΕΠΙΛΟΓΗ ΣΤΟΧΩΝ

Στους επόμενους πίνακες αναφέρονται τεχνικές από το χώρο της στατιστικής οι οποίες διαχωρίζονται σε τεχνικές παραδοσιακής στατιστικής (Πίνακας 2.1) αλλά και τεχνικές προηγμένης στατιστικής (Πίνακας 2.2). Αυτές περιλαμβάνουν τη γραμμική παλινδρόμηση (*Linear Regression-LR*), το λογιστικό/κανονικό υπόδειγμα πιθανότητας (*Logit/Probit*), το Tobit, το διωνυμικό μοντέλο (*Beta-binomial*), το Gamma-Poisson μοντέλο και τη διακριτική ανάλυση (*Discriminant Analysis-DA*). Στον Πίνακα 2.2 περιλαμβάνονται τεχνικές προηγμένης στατιστικής που αποτελούνται από δύο στάδια μοντέλων, για παράδειγμα Beta+Gamma, ή εναλλακτικής μορφής ανάλυση απροσδιόριστης κατάταξης (*latent class analysis-LCA*). Τέλος, οι τεχνικές που αναφέρονται στον Πίνακα 2.3, αποτελούν τεχνικές εξόρυξης δεδομένων βασισμένες στη μηχανική μάθηση (*machine learning*). Ορισμένες από αυτές είναι τα τεχνητά νευρωνικά δίκτυα (*Artificial Neural Networks-ANN*), οι γενετικοί αλγόριθμοι (*Genetic Algorithms-GA*), ο γενετικός προγραμματισμός (*Genetic Programming-GP*), τα δέντρα αποφάσεων (*Decision Trees-DT*) και οι μηχανές διανυσμάτων υποστήριξης (*Support Vector Machines-SVM*), (Bose και Chen,2009).

2.2.1 Βασικές στατιστικές τεχνικές

Τα μοντέλα παλινδρόμησης είναι οι περισσότερο χρησιμοποιούμενες τεχνικές. Είναι απλά αλλά έχουν περιορισμένη αναλυτική ικανότητα. Η γραμμική παλινδρόμηση, το λογιστικό/κανονικό και το tobit μοντέλο αποτελούν μοντέλα παλινδρόμησης. Τα μοντέλα γραμμικής παλινδρόμησης μπορούν να παράγουν συνεχόμενα αποτελέσματα όπως, το ενδιαφέρον απάντησης (Levin & Zahavi,1998) και το κατά εκτίμηση διαθέσιμο ποσό αγορών ενός πελάτη (Malthouse,1999). Ένα κατώτατο όριο τίθεται συνήθως εκ των προτέρων όταν επιλέγονται στοχευμένοι πελάτες, βασισμένο στα αποτελέσματα που δίνονται από ένα μοντέλο γραμμική παλινδρόμησης. Αν το αποτέλεσμα ενός πελάτη είναι μεγαλύτερο από το κατώτατο όριο, τότε αυτοί επιλέγονται. Το ποσοστό των πελατών που επιλέγεται αναφέρεται ως το ποσοστό απάντησης των μοντέλων γραμμικής παλινδρόμησης.

Τα λογιστικά, κανονικά και tobit μοντέλα διαφέρουν από τα μοντέλα γραμμικής παλινδρόμησης, διότι μπορούν να αντιμετωπίσουν άμεσα τη διακριτή απάντηση (*discrete response*) με τη χρησιμοποίηση απροσδιόριστων (*latent*) μεταβλητών. Τα λογιστικά και κανονικά μοντέλα μπορούν να παράγουν μία δυαδική (*binary*) επιλογή αποτελέσματος (Bult et al.,1997; Bult & Wansbeek,1995; Hansotia & Wang,1997; Levin & Zahavi,1998; Van den Poel & Buckinx,2005) ή κατηγορικές (*categorical*) τιμές αν υπάρχουν περισσότερες από δύο επιλογές (Heilman et al.,2003). Οι Bodapati και Gupta (2004), διατύπωσαν τον τρόπο με τον οποίο τα λογιστικά/κανονικά μοντέλα προβλέπουν διακριτοποιημένη απόκριση ως άμεση προσέγγιση, το οποίο χρησιμοποιείται επίσης και από τα μοντέλα γραμμικής παλινδρόμησης έμμεσης προσέγγισης. Σύμφωνα με τους συγγραφείς, η άμεση προσέγγιση επιτυγχάνει καλύτερη πρόγνωση για τα μεγάλα δείγματα, γιατί η απόκλιση (*bias*) στην άμεση προσέγγιση είναι μικρότερη. Οι απροσδιόριστοι μεταβλητές χρησιμοποιούνται στα tobit μοντέλα για να περικόψουν τιμές μεγαλύτερες ή μικρότερες από το κατώτατο όριο. Για παράδειγμα, στο άμεσο μάρκετινγκ οι πελάτες παράγουν εισοδήματα ή μεγαλύτερα του μηδέν ή καθόλου. Έτσι οι αρνητικές τιμές εισοδήματος που παράγονται από τα μοντέλα δεν έχουν καμία σημασία. Τα tobit μοντέλα μετατρέπουν τις αρνητικές τιμές σε μηδενικές και κρατούν μόνο τις θετικές. Επίσης, συνήθως χρησιμοποιούνται συνεχόμενες τιμές, όπως το ποσό χρημάτων που ξοδεύει ο πελάτης σε μία συναλλαγή (Hansotia & Wang,1997; Levin & Zahavi,1998).

Τέλος, σύμφωνα με τους Bult και Wittink, (1996) και άλλοι τύποι στατιστικών τεχνικών χρησιμοποιούνται στο άμεσο μάρκετινγκ, όπως η διακριτική ανάλυση και σε ορισμένες περιπτώσεις δίνουν καλύτερα αποτελέσματα από τα μοντέλα παλινδρόμησης και τα διωνυμικά μοντέλα (Rao&Steckel,1995). Παρόλα αυτά, διαφορετικές στατιστικές τεχνικές απαιτούν διαφορετικές υποθέσεις. Η παραβίαση αυτών των υποθέσεων μπορεί να προκαλέσει ανακριβείς εκτιμήσεις των παραμέτρων των μοντέλων με συνέπεια υπερβολική ανακρίβεια στην πρόβλεψη.

2.2.2 Προηγμένες στατιστικές τεχνικές

Στο παρελθόν έχουν χρησιμοποιηθεί ορισμένα προηγμένα στατιστικά μοντέλα. Τα μοντέλα αυτά συνδυάζουν δύο απλά στατιστικά μοντέλα με σκοπό να μειώσουν τις αδυναμίες κάθε μοντέλου, ενώνοντας τα δυνατά τους σημεία σε δύο επίπεδα. Στο πρώτο επίπεδο προσπαθούν να μοντελοποιήσουν την πιθανότητα απάντησης και στο επόμενο επίπεδο το χρηματικό ποσό που θα ξοδέψει ο πελάτης σε μια δραστηριότητα άμεσου μάρκετινγκ. Το αποτέλεσμα που παράγεται από το μοντέλο των δύο επιπέδων είναι το προβλεπόμενο/αναμενόμενο ποσό που μπορεί ο πελάτης να διαθέσει. Οι Colombo και Jiang (1999) χρησιμοποίησαν ένα beta μοντέλο κατανομής για να αντιμετωπίσουν την πιθανότητα απάντησης των πελατών και ένα μοντέλο gamma κατανομής για το ποσό των χρημάτων που διαθέτουν οι πελάτες για αγορά. Οι Levin και Zahavi (1998) και ο Van der Sheer (1998), χρησιμοποίησαν το λογιστικό μοντέλο για να υπολογίσουν την πιθανότητα απάντησης και το μοντέλο γραμμικής παλινδρόμησης για να υπολογίσουν τη χρηματική αξία. Οι Baumgartner και Hruschka (2005), χρησιμοποίησαν ένα κανονικό μοντέλο για να διαχωρίσουν τους πελάτες που είχαν πραγματοποιήσει μία αγορά και κράτησαν το προϊόν από αυτούς που το επέστρεψαν πίσω. Η ανάλυση αυτή ακολουθήθηκε με μοντέλα μη-γραμμικής παλινδρόμησης για να υπολογιστεί το ποσό χρημάτων των αγορών και η επιστροφή των προϊόντων από τους πελάτες.

Μια περισσότερο πολύπλοκη τεχνική, όπως η εναλλακτικής μορφής ανάλυση απροσδιόριστης κατάταξης (LCA) χρησιμοποιείται όταν υπάρχουν τμήματα (*segments*) αγοράς μεταξύ των πελατών. Η LCA εκτελεί ταυτόχρονα την τμηματοποίηση και δημιουργεί μοντέλα επιλογής μέσα στα τμήματα. Το πλεονέκτημά της είναι ότι μπορεί να παράγει τμήματα της αγοράς και μπορεί ταυτόχρονα να μοντελοποιήσει την απάντηση των πελατών. Οι Desarbo και Ramaswamy (1994), χρησιμοποίησαν την παραπάνω ανάλυση για τη δημιουργία ενός λογιστικού μοντέλου σε κάθε τμήμα για την επίλυση προβλημάτων δυαδικής απάντησης. Επίσης, οι Wedel et al. (1993) με την LCA δημιούργησαν ένα μοντέλο παλινδρόμησης poisson σε κάθε τμήμα για την εκτίμηση της ποσότητας των προϊόντων που ένας πελάτης θα μπορούσε να αγοράσει. Τέλος, η LCA μπορεί να χρησιμοποιηθεί για να καθορίσει την απάντηση των πελατών, καθώς επίσης και το χρόνο απάντησης. Οι Gonul et al. (2000) συμπεριέλαβαν το λογιστικό μοντέλο στην LCA και μελέτησαν το αποτέλεσμα/αντίκτυπο του συγχρονισμού των προτάσεων άμεσου μάρκετινγκ στις απαντήσεις των πελατών και οι Bitran και Mondschein (1996), το κανονικό μοντέλο για να υπολογίσουν την LTV των πελατών σε διάφορες προτάσεις άμεσου μάρκετινγκ.

Τεχνικές	Τύπος Απάντησης	Τύπος Αποτελέσματος	Ερευνητές
Γραμμική Παλινδρόμηση	Ενδιαφέρον Απάντησης Εισόδημα	Συνεχής	Levin & Zahavi (1998), Berger & Magliozzi (1992) Malthouse (1999)
Λογιστικό/Κανονικό Υπόδειγμα Πιθανότητας	Δυναδική	Δυναδικός	Bodapati & Gupta (2004), Bult et al. (1997), Bult & Wansbeek (1995), Hansotia & Wang (1997), Levin & Zahavi (1998), Van den Poel & Buckinx (2005) Levin & Zahavi (1998)
	Αριθμός προϊόντων	Ακέραιος	Heilman et al. (2003)
Tobit	Κατηγορική Εισόδημα	Ακέραιος	Hansotia & Wang (1997), Levin & Zahavi (1998)
Beta/Gamma	Εισόδημα	Συνεχής	Rao & Steckel (1995)
Διακριτική Ανάλυση	Δυναδική	Συνεχής	Bult (1993), Bult & Wittink (1996)

Πίνακας 2.1: Βασικές στατιστικές τεχνικές.
(Πηγή: Bose και Chen,2009)

Τεχνικές	Τύπος Απάντησης	Τύπος Αποτελέσματος	Ερευνητές
Two stage Beta + Gamma	Δυναδική και Εισόδημα	Συνεχής	Colombo & Jiang (1999)
Two stage Logit + Linear	Δυναδική και Εισόδημα	Συνεχής	Levin&Zahavi (1998), V.Sheer (1998)
Two stage Probit + Non-linear	Δυναδική και Εισόδημα	Συνεχής	Baumgartner & Hruschka (2005)
Latent class model Logit	Εισόδημα	Συνεχής	Gonul et al. (2000)
	Δυναδική	Συνεχής	Desarbo & Ramaswamy (1994)
Latent class model Poisson	Αριθμός προϊόντων	Ακέραιο	Wedel et al. (1993)
Latent class model Probit	Δυναδική	Δυναδικό	Bitran & Mondschein (1996)

Πίνακας 2.2: Προηγμένες στατιστικές τεχνικές.
(Πηγή: Bose και Chen,2009)

2.2.3 Τεχνικές μηχανικής μάθησης

Η τελευταία κατηγορία ποσοτικών μοντέλων βασίζεται στη μηχανική μάθηση. Σύμφωνα με τον Mitchell (1997): “Η μηχανική μάθηση είναι μια διαδικασία κατά την οποία τα προγράμματα H/Y μπορούν να μάθουν να βελτιώνουν την απόδοση τους από την εμπειρία να κάνουν ορισμένες εργασίες”. Ο Kantardzic (2003) διαφοροποίησε τις προσεγγίσεις βασισμένες στη στατιστική και τη μηχανική μάθηση με δύο τρόπους. Πρώτον, τα στατιστικά μοντέλα δίνουν έμφαση στα μαθηματικά και τις συναρτήσεις. Δεύτερον, οι προηγμένες στατιστικές μέθοδοι είναι σχεδόν εξ ολοκλήρου οδηγούμενες από προκαθορισμένη δομή (*postulated structure*), ή μια προσέγγιση στη δομή, η οποία θα μπορούσε να οδηγήσει στα δεδομένα. Αντιθέτως, η μηχανική μάθηση τείνει να εμβαθύνει στους αλγόριθμους που συνήθως χρησιμοποιούνται, υπό τη μορφή της διαδικασίας της “μάθησης” εξάγοντας κανόνες από τα δεδομένα. Οι διάφορες τεχνικές

μηχανικής μάθησης που βρίσκουν εφαρμογή στο άμεσο μάρκετινγκ παρατίθενται στον Πίνακα 2.3.

Από τις τεχνικές που παρουσιάζονται στον Πίνακα 2.3 τα τεχνητά νευρωνικά δίκτυα (ANN) αποδίδουν ικανοποιητικά στη μάθηση μη-γραμμικών σχέσεων μεταξύ της εισόδου και της εξόδου. Τα δέντρα αποφάσεων (DT) είναι μία τεχνική ταξινόμησης βασισμένη σε κανόνες και οι μηχανές διανυσμάτων υποστήριξης (SVM) μπορούν να εκτελέσουν είτε γραμμικές είτε μη-γραμμικές ταξινομήσεις (Burgess,1998). Η δημοφιλέστερη εφαρμογή των παραπάνω τεχνικών στο άμεσο μάρκετινγκ είναι η δυαδική ταξινόμηση των πελατών (για παράδειγμα, ο πελάτης που αποδέχεται ένα προϊόν παίρνει την τιμή 1 ενώ την τιμή 0 αντίθετα). Επίσης, τα ANN και τα DT έχουν χρησιμοποιηθεί για ταξινομήσεις με παραπάνω από δύο κατηγορίες. Για παράδειγμα, οι Heilman et al. (2003) χρησιμοποίησαν ένα ANN για την ταξινόμηση των πελατών σε τρεις κατηγορίες με σκοπό να προσδιοριστούν οι μάρκες που οι πελάτες ήταν πιστοί και οι Haughton και Oulabi (1997) την τεχνική DT για να παράγουν, με τη μορφή δέντρου, τμηματοποιήσεις των πελατών στο άμεσο μάρκετινγκ.

Τα ANN είναι η περισσότερο διαδεδομένη τεχνική εξόρυξης δεδομένων που χρησιμοποιείται στο άμεσο μάρκετινγκ. Εκτός από την επιλογή των πελατών χρησιμοποιείται επίσης και για άλλους σκοπούς. Οι Shin & Sohn (2004) ενσωμάτωσαν στην περιβάλλουσα ανάλυση δεδομένων (*data envelopment analysis*) τεχνητά νευρωνικά δίκτυα για να αξιολογήσουν την αποδοτικότητα κάθε πελάτη. Επίσης, οι Kaefer et al. (2005) χρησιμοποίησαν ANN για να ταξινομήσουν τους πελάτες με τέτοιο τρόπο ώστε ο συγχρονισμός του άμεσου μάρκετινγκ να βελτιώνεται.

Τα DT έχουν ισχυρή ικανότητα να αντιμετωπίζουν κατηγορικές μεταβλητές εισόδου. Κατά την επεξεργασία των κατηγορικών δεδομένων με περισσότερα από δύο επίπεδα τιμών τα ANN, τα SVM και οι στατιστικές τεχνικές δημιουργούν ψευδομεταβλητές (*dummy*) για κάθε επίπεδο των τιμών των σχετικών μεταβλητών εισόδου, προσθέτοντας έτσι υπολογιστικό φόρτο στην επεξεργασία των μοντέλων αυτών. Αντιθέτως, τα DT μπορούν να παράγουν κανόνες χρησιμοποιώντας τα κατηγορικά δεδομένα απευθείας δίχως τη δημιουργία ψευδομεταβλητών. Παρόλα αυτά, τα DT δεν μπορούν να χρησιμοποιήσουν συνεχείς μεταβλητές απευθείας και χρειάζεται η διακριτοποίησή τους.

Η τέταρτη τεχνική μηχανικής μάθησης, η οποία εφαρμόζεται στο άμεσο μάρκετινγκ, είναι αρκετά διαφορετική από τις παραπάνω τεχνικές. Αναφέρεται στην οικογένεια των προσεγγίσεων που προσομοιώνουν τις αρχές της φυσικής εξέλιξης και περιλαμβάνει τους γενετικούς αλγόριθμους (GA), τον γενετικό προγραμματισμό (GP) και τον εξελικτικό προγραμματισμό (EP). Οι τεχνικές αυτές είναι βασισμένες στη διαδικασία της αναζήτησης που προσομοιώνει τη φυσική επιλογή και εξέλιξη περιλαμβάνοντας τα ακόλουθα βήματα: επιλογή-*selection*, διασταύρωση-*crossover* και μετάλλαξη-*mutation* (Goldberg,1989). Οι GA, GP, EP είναι κατάλληλες τεχνικές για προβλήματα βελτιστοποίησης, όπως η επιλογή των στόχων για αποστολή προτάσεων άμεσου μάρκετινγκ, ικανοποιώντας ταυτόχρονα τις επιχειρηματικές απαιτήσεις. Οι Kwon και Moon (2001) πρότειναν ένα μοντέλο GP για το πρόβλημα της επιλογής στόχων για το ηλεκτρονικό ταχυδρομείο βασισμένο στις ανάγκες του άμεσου μάρκετινγκ. Επίσης, ο Bhattacharyya (1999&2000) πρότεινε ένα GA και ένα GP

αντίστοιχα, για την αναζήτηση των καλύτερων λύσεων, που μεγιστοποιούν τα συνολικά ποσοστά απάντησης και το κέρδος σε διάφορα “βάθη” αποστολής στο άμεσο μάρκετινγκ. Το “βάθος” αποστολής αναφέρεται στους πελάτες με την υψηλότερη πιθανότητα απάντησης σε σχέση με τη διαδικασία αποστολής προτάσεων. Για παράδειγμα, το βάθος αποστολής 10%, σημαίνει ότι το 10% των πλέον πιθανών πελατών χρησιμοποιήθηκε για τη διαδικασία αποστολής. Στην ανάλυση βάθους αποστολής οι μεμονωμένοι πελάτες ταξινομούνται με φθίνουσα σειρά, σύμφωνα με ορισμένα κριτήρια, όπως η πιθανότητα απάντησης ή το εισόδημα τους.

Τεχνικές	Τύπος Απάντησης	Τύπος Αποτελέσματος	Ερευνητές
ANN	Πιθανότητα απάντησης	Συνεχής	Kim & Street (2004), Kim et al. (2005), Shin & Sohn (2004)
ANN	Δυαδική	Δυαδικό	Ha et al. (2005), Kaefer et al. (2005), Viaene et al. (2001a), Zahavi & Levin (1997)
ANN	Κατηγορική	Ακέραιο	Heilman et al. (2003)
Bayesian ANN	Δυαδική	Δυαδικό	Baesens et al. (2002)
CHAID/CART	Δυαδική	Δυαδικό	Haughton & Oulabi (1997)
DT and Naive Bayes	Πιθανότητα απάντησης	Συνεχής	Ling & Li (1998)
DT	Δυαδική	Δυαδικό	Buckinx et al. (2004)
LS-SVM	Δυαδική	Δυαδικό	Viaene et al. (2001b)
SVM-SVR	Πιθανότητα απάντησης	Συνεχής	Drucker et al. (1997), Muller et al. (1997), Kim et al. (2008)
GP	Δυαδική	Δυαδικό	Kwon & Moon (2001)
GA	Εισόδημα	Συνεχής	Bhattacharyya (1999)
GA and GP	Δυαδική και Εισόδημα	Δυαδικό και Συνεχής	Bhattacharyya (2000)
Hybrid $f(\text{ANN} + \text{Logit} + \text{RFM})$	Πιθανότητα απάντησης	Συνεχής	Suh et al. (1999)
Hybrid $f(\text{ANN} + \text{DT} + \text{Logit})$	Πιθανότητα απάντησης	Συνεχής	Suh et al. (2004)
Hybrid $f(\text{BBN} + \text{GP})$	Πιθανότητα απάντησης	Συνεχής	Cui et al. (2006)

Πίνακας 2.3: Τεχνικές μηχανικής μάθησης.
(Πηγή: Bose και Chen, 2009)

2.2.4 Συνδυασμένες και υβριδικές τεχνικές

Μια τεχνική ανάλυσης δεδομένων μπορεί συχνά να παράγει περισσότερες από μια πιθανές λύσεις, εξαιτίας των διαφορετικών κατανομών των δεδομένων εκπαίδευσης. Μια προσέγγιση που το αντιμετωπίζει αυτό, είναι η αξιολόγηση της απόδοσης κάθε μοντέλου και η επιλογή του καλύτερου εξ αυτών. Μια άλλη προσέγγιση είναι η χρησιμοποίηση του συνόλου όλων των πιθανών μοντέλων. Η συγκεκριμένη προσέγγιση προσπαθεί να συνδυάσει τα αποτελέσματα όλων των πιθανών μοντέλων βασισμένη σε ορισμένες αρχές.

Υπάρχουν δύο βασικές προσεγγίσεις συνδυασμού μοντέλων που βρίσκουν εφαρμογή στο άμεσο μάρκετινγκ: το bagging και το boosting. Το bagging αντιγράφει τα δείγματα εκπαίδευσης με δειγματοληψία με επανατοποθέτηση και συνδυασμό των

αποτελεσμάτων όλων των λύσεων χρησιμοποιώντας υπολογισμό του μέσου όρου ή ψηφοφορία με πλειοψηφία. Το boosting ορίζει βάρη στα χαρακτηριστικά του συνόλου δεδομένων. Όσο υψηλότερο βάρος έχει ένα χαρακτηριστικό τόσο μεγαλύτερη επιρροή θα έχει στη διαδικασία εκμάθησης. Τα βάρη ρυθμίζονται έπειτα από κάθε επανάληψη. Το bagging και το boosting, συνήθως παρουσιάζουν καλές επιδόσεις στο δείγμα εκπαίδευσης επειδή συνδυάζουν πολλαπλές λύσεις που ενδεχομένως χωριστά να παρουσιάσουν ορισμένα μειονεκτήματα. Οι Ha et al. (2005) χρησιμοποίησαν ένα bagging ANN για την πρόβλεψη αν ένας πελάτης θα ανταποκριθεί σε μία πρόταση άμεσου μάρκετινγκ. Επίσης, για το ίδιο πρόβλημα οι Ling και Li (1998) χρησιμοποίησαν ένα bagging DT. Τέλος, οι Kim και Street (2004) πρότειναν μία προσέγγιση που συνδυάζει τις παραπάνω δυο προσεγγίσεις.

Οι συνδυασμένες τεχνικές που περιγράφηκαν παραπάνω χρησιμοποιούν μια συγκεκριμένη τεχνική. Εκτός όμως, από το συνδυασμό των αποτελεσμάτων της ίδιας τεχνικής, διάφορες τεχνικές μπορούν επίσης να συνδυαστούν χρησιμοποιώντας μια υβριδική προσέγγιση με σκοπό την ελαχιστοποίηση των μειονεκτημάτων των ξεχωριστών τεχνικών. Οι Suh et al. (1999) ανέλυσαν το συντελεστή συσχέτισης μεταξύ των προβλεπόμενων πιθανοτήτων απόκρισης που εξάχθηκαν από ένα ANN, ένα λογιστικό μοντέλο και μια στοχευμένη προσέγγιση που χρησιμοποιούσε τις RFM μεταβλητές των πελατών για τμηματοποίηση. Διαπιστώθηκε ότι ο συντελεστής συσχέτισης, μεταξύ των αποτελεσμάτων του ANN και των RFM καθώς και τα αποτελέσματα των RFM με το λογιστικό μοντέλο ήταν χαμηλός, ενώ ο συντελεστής συσχέτισης μεταξύ των αποτελεσμάτων ANN και λογιστικού ήταν υψηλός. Στη συνέχεια διαπιστώθηκε ότι ο συνδυασμός ANN και RFM σε ένα ενιαίο μοντέλο επιτυγχάνει βελτιωμένη ακρίβεια ταξινόμησης από το συνδυασμό λογιστικού-RFM.

Οι Zahavi και Levin (1997) πρότειναν τη μέθοδο Best Double-Scoring (BDS) για το συνδυασμό των αποτελεσμάτων ενός ANN και ενός λογιστικού μοντέλου για κάθε πελάτη στο άμεσο μάρκετινγκ, χρησιμοποιώντας συναρτήσεις της μορφής $\text{Min}[A,B]$, $\text{Max}[A,B]$, $A+B$ κτλ. Στη συνάρτηση αυτή, τα A και B αντιπροσωπεύουν τα αποτελέσματα (scores) που λαμβάνονται από τους πελάτες από δυο διαφορετικές τεχνικές. Οι Suh et al. (1999) προσάρμοσαν και τροποποίησαν το BDS για την τμηματοποίηση των καταλόγων πελατών στο άμεσο μάρκετινγκ. Εκτός από τα μοντέλα ANN και το λογιστικό, οι συγγραφείς χρησιμοποίησαν στα υβριδικά μοντέλα και μια προσέγγιση βασισμένη στις RFM μεταβλητές. Το BDS τροποποιήθηκε έτσι ώστε να κατορθώσει να συνδυάσει τρεις διαφορετικές τεχνικές τα ANN, το λογιστικό μοντέλο και την προσέγγιση RFM. Οι Lin και McClean (2001) πρότειναν ένα υβριδικό μοντέλο που χρησιμοποιούσε την ακρίβεια ως βάρος για κάθε δυαδική ταξινόμηση για τον υπολογισμό ενός συνδυαζόμενου αποτελέσματος για την πρόβλεψη της πτώχευσης μιας επιχείρησης. Οι Suh et al. (2004) προσάρμοσαν το παραπάνω μοντέλο για άμεσο μάρκετινγκ πραγματικού χρόνου και συγκεκριμένα για το διαδικτυακό μάρκετινγκ (Direct Web Marketing). Οι τεχνικές που χρησιμοποιήθηκαν ήταν τα ANN, το λογιστικό μοντέλο και τα DT. Οι Cui et al. (2006) ανέπτυξαν ένα άλλο είδος υβριδικού μοντέλου. Οι τεχνικές που χρησιμοποιήθηκαν στην έρευνά τους περιελάμβαναν Bayesian Belief Networks (BBN) και GP. Αρχικά, κατασκευαζόταν μια προσανατολισμένη γραφική αναπαράσταση χωρίς κυκλώματα που συνέδεε τις

μεταβλητές για να εξετάσει αιτιώδεις σχέσεις μεταξύ τους. Έπειτα, το μοντέλο αξιολογήθηκε χρησιμοποιώντας το κριτήριο ελάχιστης περιγραφής μήκους (*Min. Description Length-MDL*), (Rissanen,1978). Το MDL χρησιμοποιείται ως μέτρο ικανότητας για να βοηθήσει το GP να βρει τη βέλτιστη τοπολογία BBN. Η εργασία των Cui et al. (2006) ήταν διαφορετική από τις υπόλοιπες υβριδικές προσεγγίσεις, διότι αντί να αξιολογεί τους πελάτες ξεχωριστά χρησιμοποιώντας μια τεχνική και συνδυάζοντας τα αποτελέσματα, χρησιμοποιήθηκε GP για την κατασκευή του καλύτερου BBN μοντέλου με επαναληπτικό τρόπο.

2.3 ΜΕΘΟΔΟΙ ΓΙΑ ΟΜΑΔΟΠΟΙΗΣΗ ΠΕΛΑΤΩΝ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ

Η ομαδοποίηση πελατών αποσκοπεί στην ταξινόμηση των πελατών σε διαφορετικές ομάδες (*clusters*). Οι πελάτες που βρίσκονται στην ίδια ομάδα επιλέγονται να έχουν περισσότερα κοινά χαρακτηριστικά μεταξύ τους σε σχέση με τους πελάτες άλλων ομάδων.

Διάφορες προσεγγίσεις έχουν αναπτυχθεί για την αναγνώριση του προφίλ πελατών. Ένας τύπος προσέγγισης είναι η επιλογή μιας συγκεκριμένης αποκοπής (*cutoff*) για τις ανεξάρτητες μεταβλητές έτσι ώστε το αποτέλεσμα της τμηματοποίησης να έχει υψηλότερο όφελος στο κέρδος. Ένα χαρακτηριστικό παράδειγμα της προσέγγισης αυτής αποτελεί το RFM προφίλ. Οι Jonker et al. (2004) και οι Bitran και Mondschein (1996) ταξινόμησαν τους πελάτες χρησιμοποιώντας τις μεταβλητές RFM. Οι πελάτες αρχικά, χωρίζονταν σε ομάδες βασισμένοι στη *R* τιμή τους. Μέσα σε αυτές τις ομάδες οι πελάτες υποδιαιρέθηκαν σύμφωνα με τη *F* τιμή και έπειτα υποδιαιρέθηκαν σύμφωνα με τη *M* τιμή τους. Ο δεύτερος τύπος προσέγγισης του προφίλ εφαρμόζει στατιστικά μοντέλα, όπως η εναλλακτικής μορφής ανάλυση απροσδιόριστης κατάταξης. Η περαιτέρω ανάλυση των πελατών πραγματοποιείται με την κατασκευή μοντέλων παλινδρόμησης τα οποία χρησιμοποιούν το τμήμα των πελατών ως εξαρτημένη μεταβλητή και τα χαρακτηριστικά των πελατών ως ανεξάρτητη (Desarbo & Ramaswamy,1994). Τέλος, ένας τρίτος τύπος προσέγγισης του προφίλ περιλαμβάνει τη χρήση τεχνικών εξόρυξης δεδομένων χωρίς επίβλεψη, όπως την ομαδοποίηση K-means και τα αυτό-οργανούμενα δίκτυα όπως τα Self Organizing Map (SOM). Τα SOM αποτελούνται από ένα σύνολο τεχνητών νευρώνων, των οποίων τα βάρη προσαρμόζονται για να ταιριάζουν στα διανύσματα εισαγωγής στο δείγμα εκπαίδευσης. Οι Weng και Liu (2004) χρησιμοποίησαν μια προσέγγιση ομαδοποίησης δυο επιπέδων η οποία ενσωμάτωνε τα SOM και την ομαδοποίηση K-means. Η τεχνική K-means ομαδοποιεί τους πελάτες σε ένα συγκεκριμένο αριθμό ομάδων, ενώ τα SOM έχουν τη δυνατότητα να αποφασίσουν τον αριθμό των ομάδων αυτόματα. Τα αποτελέσματα της έρευνας τους έδειξαν ότι η ομαδοποίηση δύο επιπέδων παρουσιάζει μεγαλύτερη συγκέντρωση πελατών (που μετρούνται σύμφωνα με το συντελεστή της διασποράς των ομάδων) σε μια ομάδα σε σύγκριση με την ομαδοποίηση SOM.

Η αναγνώριση προτύπων των πελατών χρησιμοποιείται για να ανακαλύψει συσχετισμούς των γεγονότων που μπορούν να αντιπροσωπευθούν ως πιθανολογικοί κανόνες μέσα σε κάθε ομάδα πελατών. Τεχνικές βασισμένες στους κανόνες συσχέτισης

συχνά χρησιμοποιούνται για να βρουν τα πρότυπα συμπεριφοράς των πελατών. Οι κανόνες συσχέτισης προτάθηκαν αρχικά από τους Agrawal και Srikant (1994) και χρησιμοποιήθηκαν για να ανακαλύψουν τις συσχετίσεις των γεγονότων που θα μπορούσαν να αντιπροσωπευθούν ως πιθανολογικοί κανόνες. Στο άμεσο μάρκετινγκ τα γεγονότα θα μπορούσαν να είναι τα χαρακτηριστικά των πελατών ή οι αγορές σχετικών προϊόντων (Changchien et al.,2004; Chen et al.,2005). Με την εξέλιξη της τεχνολογίας των H/Y και του διαδικτύου δόθηκε ιδιαίτερη έμφαση στη συμπεριφορά των πελατών μέσω του διαδικτύου και του ηλεκτρονικού εμπορίου, η οποία μπορεί επίσης να αναλυθεί με κανόνες συσχέτισης σύμφωνα με τους Liao και Chen (2004) και Suh et al., (2004). Στη συγκεκριμένη προσέγγιση χρησιμοποιούνται κανόνες συσχέτισης για να ανακαλύψουν πρότυπα του πελάτη, κατά τη χρήση μιας ιστοσελίδας ηλεκτρονικού εμπορίου καταγράφοντας διάφορα στοιχεία, όπως τις σελίδες που επισκέπτεται ο πελάτης, τον αριθμό επισκέψεων, την IP διεύθυνση, καθώς επίσης και άλλες ενέργειες σχετικές με τις αγορές των πελατών μέσω διαδικτύου. Όλα τα παραπάνω χρησιμοποιούνται ως είσοδοι σε ένα υβριδικό μοντέλο το οποίο περιέχει τρία είδη τεχνικών εξόρυξης δεδομένων (DT, NN, και LR) με σκοπό την πρόβλεψη της πιθανότητας αγοράς. Το υβριδικό μοντέλο καθορίζει τους συσχετισμούς μεταξύ της συμπεριφοράς των πελατών στο διαδίκτυο και τις συμπεριφορές αγορών. Η εξακρίβωση των προτύπων αυτών βοηθά τους πωλητές στον προσδιορισμό σημαντικών κατηγοριών προϊόντων και στοχευμένων πελατών.

2.4 ΜΕΘΟΔΟΙ ΓΙΑ CROSS-SELLING ΚΑΙ UP-SELLING

Οι πωλητές πρέπει να ερευνούν συνεχώς τις μεταβαλλόμενες ανάγκες των πελατών και να αποφασίσουν έπειτα ποια προϊόντα ή υπηρεσίες πρέπει να προσφερθούν. Οι δύο περισσότερο χρησιμοποιημένες στρατηγικές μάρκετινγκ είναι το cross-selling και το up-selling. Το cross-selling είναι η πρακτική της πρότασης παρόμοιων προϊόντων ή υπηρεσιών σε ένα πελάτη ο οποίος έχει πραγματοποιήσει μια αγορά. Για παράδειγμα, σε ένα ηλεκτρονικό βιβλιοπωλείο, προτείνονται παρόμοια βιβλία με κάποιο που έχει επιλέξει ένας πελάτης. Το up-selling είναι η πρακτική της πρότασης προϊόντων ή υπηρεσιών (με υψηλότερη τιμή, περισσότερα χαρακτηριστικά και καλύτερη ποιότητα) σε ένα πελάτη ο οποίος σκέφτεται να πραγματοποιήσει μια αγορά. Για παράδειγμα, η πρόταση αγοράς νέου H/Y τελευταίας τεχνολογίας.

Υπάρχουν κυρίως δύο προσεγγίσεις συστάσεων: αυτές που βασίζονται στο περιεχόμενο (*content*) (Lang,1995) και το συνεργατικό φιλτράρισμα (*collaborative filtering*) (Shardanand & Maes,1995). Η προσέγγιση που βασίζεται στο περιεχόμενο μετρά το βαθμό ομοιότητας μεταξύ των υποψήφιων πελατών και των στοιχείων που περιλαμβάνονται στο προφίλ του πελάτη. Η προσέγγιση του συνεργατικού φιλτραρίσματος προσπαθεί να βρει ομοιότητες μεταξύ των προτιμήσεων των πελατών. Στα πλαίσια του άμεσου μάρκετινγκ το στοιχείο είναι το προϊόν ή τα χαρακτηριστικά του προϊόντος. Το προφίλ των πελατών και οι προτιμήσεις βρίσκονται με την ανάλυση των αρχείων συναλλαγής των πελατών. Τα προϊόντα που παρουσιάζουν παρόμοια συμπεριφορά με το προφίλ ενός πελάτη, συνήθως αγοράζονται από άλλο πελάτη με παρόμοιο προφίλ. Επομένως, η προσέγγιση συστάσεων απαιτεί τη γνώση του προφίλ

του πελάτη, αλλά επίσης και το προφίλ των προϊόντων. Το προφίλ των προϊόντων είναι παρόμοιο με το προφίλ των πελατών. Τα προϊόντα περιγράφονται από στοιχεία τα οποία αντιπροσωπεύουν τα διαφορετικά χαρακτηριστικά και έπειτα ομαδοποιούνται. Οι τεχνικές ομαδοποίησης που χρησιμοποιούνται είναι οι ίδιες για την ομαδοποίηση των πελατών και περιγράφηκαν παραπάνω.

Τεχνικές	Σκοπός έρευνας	Ερευνητές
SVM	Είδος προϊόντων που πρέπει να προσφέρονται σε συγκεκριμένους πελάτες	Cheung et al. (2003)
Ανάλυση συνεργατικού φιλτραρίσματος	Ανάπτυξη συστήματος συστάσεων προϊόντων	Min & Han (2005), Li et al. (2005)
Υβριδικό μοντέλο	Βελτίωση διαδικασίας ανάλυσης προφίλ πελάτη και προϊόντος	Changchien et al., (2004), Weng & Liu, (2004)

Πίνακας 2.4: Έρευνες για cross-selling και up-selling.

2.5 ΜΕΘΟΔΟΙ ΓΙΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΕΣ ΑΠΑΙΤΗΣΕΙΣ

Ο μαθηματικός προγραμματισμός και οι εξελικτικοί αλγόριθμοι συνήθως χρησιμοποιούνται για βελτιστοποίηση στο άμεσο μάρκετινγκ. Συνήθως λαμβάνονται υπόψη παράμετροι κέρδους ή κόστους για την εκτίμηση των σχετικών με το μοντέλο συντελεστών. Στον Πίνακα 2.5 παρουσιάζονται ορισμένες σημαντικές έρευνες.

Τεχνικές	Σκοπός έρευνας	Ερευνητές
Παράγοντες κέρδους και κόστους στο λογιστικό μοντέλο	Μεγιστοποίηση του κέρδους	Bult & Wansbeek (1995), Van der Sheer (1998)
Δυναμική παλινδρόμηση-Weighted Maximum-Likelihood	Οικονομικά του βάθους αποστολής	Prinzie & Van den Poel (2005)
Στοχαστικά μοντέλα-Αλυσίδες Markov	Μεγιστοποίηση του κέρδους	Bitran & Mondschein (1996), Gonul & Shi (1998)

Πίνακας 2.5: Έρευνες για βελτιστοποίηση επιχειρηματικών απαιτήσεων.

Ο μαθηματικός προγραμματισμός αποτελείται από μια αντικειμενική συνάρτηση και από ένα σύνολο περιορισμών. Τα αποτελέσματα από τα μοντέλα επιλογής, όπως αυτά που αναφέρθηκαν παραπάνω, χρησιμοποιούνται ως είσοδοι στα μοντέλα μαθηματικού προγραμματισμού ως πιθανές λύσεις (Cohen, 2004). Χρησιμοποιώντας ως εισαγωγή τις πιθανές λύσεις ο μαθηματικός προγραμματισμός επιλέγει την καλύτερη λύση που βελτιστοποιεί την αντικειμενική συνάρτηση ικανοποιώντας ταυτόχρονα όλους τους περιορισμούς. Ο Cohen, (2004) χρησιμοποίησε μοντέλα μαθηματικού προγραμματισμού για τη λήψη απόφασης όσον αφορά ποιο είδος προϊόντων πρέπει να σταλεί σε συγκεκριμένο είδος πελατών. Ο στόχος ήταν η μεγιστοποίηση του κέρδους και οι περιορισμοί, μεταξύ άλλων περιελάμβαναν το διαθέσιμο αριθμό προϊόντων και τα όρια στα διαθέσιμα κεφάλαια.

Οι εξελικτικοί αλγόριθμοι ενσωματώνουν την επιλογή στόχων στη διαδικασία βελτιστοποίησης μέσω GA και παρόμοιων τεχνικών. Οι επιχειρηματικοί στόχοι και οι

περιορισμοί ορίζονται ως λειτουργικές ικανότητες στη διαδικασία. Ο Bhattacharyya (2000) χρησιμοποίησε τεχνικές GA και GP για την αναζήτηση βέλτιστων λύσεων κατά μήκος του συνόρου Pareto (*Pareto-frontier*) (αποτελούμενο από ένα σύνολο λύσεων) στο άμεσο μάρκετινγκ. Η λύση αυτή αντιπροσώπευε το μέγιστο συνολικό ποσοστό απόκρισης και το κέρδος στο αντίστοιχο “βάθος” αποστολής. Οι Kim και Street (2004) πρότειναν ένα συνδυασμό GA/ANN για την επιλογή συγκεκριμένων πελατών επιτυγχάνοντας ταυτόχρονα δύο κύριους στόχους για κάθε συγκεκριμένο πελάτη: την ερμηνεία των μοντέλων και την ακρίβεια πρόβλεψης. Τα ANN χρησιμοποιούνται για να βαθμολογήσουν (*score*) τους πελάτες και ο GA για να αναζητήσει τους πιθανούς συνδυασμούς των χαρακτηριστικών εισόδου στο ANN που θα μπορούσαν να βελτιστοποιήσουν πολλαπλούς στόχους. Οι Kim et al. (2005) πρότειναν μια παρόμοια με την προηγούμενη προσέγγιση με τη μόνη διαφορά ότι χρησιμοποίησαν Evolutionary Local Selection Algorithm (*ELSA*), μια βελτίωση του κλασικού GA. Ο ELSA λειτουργεί καλύτερα από τον κλασικό GA, διότι διερευνεί ολόκληρο το σύνολο Pareto, παρέχοντας στους πωλητές περισσότερες πληροφορίες για τις παραχωρήσεις (*trade-offs*) μεταξύ διαφόρων αντικειμενικών στόχων.

Ο μαθηματικός προγραμματισμός είναι περισσότερο ευέλικτος για τη βελτιστοποίηση στις εφαρμογές άμεσου μάρκετινγκ, διότι μπορεί να ενσωματώσει επιχειρηματικούς στόχους και περιορισμούς. Οι εξελικτικοί αλγόριθμοι μπορούν να βελτιστοποιήσουν πολλούς αντικειμενικούς στόχους, αλλά καμία έρευνα μέχρι στιγμής δεν έχει αποδείξει την ικανότητα να ενσωματώνουν διάφορους -σχετικούς με την επιχειρηματικότητα- περιορισμούς. Ο ορισμός των περιορισμών ως αντικειμενικές συναρτήσεις θα μπορούσε να είναι μια πιθανή λύση, αλλά το πρόβλημα εμφανίζεται όταν υπάρχουν πολλοί περιορισμοί.

2.6 ΑΝΑΛΥΣΗ ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΜΟΝΤΕΛΩΝ

Από τα υπολογιστικά μοντέλα που αναφέρθηκαν παραπάνω, στην παρούσα εργασία χρησιμοποιήθηκαν, από το χώρο της στατιστικής, η λογιστική παλινδρόμηση (*Binary-Logistic Regression*) και από το χώρο της μηχανικής μάθησης, τα δέντρα αποφάσεων, τα τεχνητά νευρωνικά δίκτυα και οι μηχανές διανυσμάτων υποστήριξης. Τέλος, χρησιμοποιήθηκαν συνδυασμένες (*bagging*) και υβριδικές τεχνικές. Στις υποενότητες που ακολουθούν παρουσιάζονται τα βασικά χαρακτηριστικά κάθε μοντέλου.

2.6.1 Λογιστική Παλινδρόμηση

Η λογιστική παλινδρόμηση, όπως αναφέρθηκε παραπάνω ανήκει στην κατηγορία των στατιστικών μεθόδων. Είναι μια μέθοδος πολυπαραγοντικής στατιστικής ανάλυσης (*multivariate statistical analysis*) που χρησιμοποιεί ένα σύνολο ανεξαρτήτων μεταβλητών για τη διερεύνηση της κίνησης μιας κατηγορικής εξαρτημένης μεταβλητής. Η διαφορά με τη γραμμική παλινδρόμηση έγκειται στο γεγονός ότι η εξαρτημένη μεταβλητή δεν παίρνει συνεχείς τιμές, π.χ εκφράζει ένα ποσοτικό μέγεθος, αλλά είναι δυαδική μεταβλητή, για παράδειγμα λαμβάνει την τιμή 1, αν ένας πελάτης αποδεχθεί ένα προϊόν και την τιμή 0, αν αρνηθεί. Έτσι για παράδειγμα, θα μπορούσε να χρησιμοποιηθεί γραμμική παλινδρόμηση για την πρόβλεψη των πωλήσεων με βάση τα

χαρακτηριστικά των προϊόντων και η λογιστική για τον εντοπισμό των πελατών που είναι πιθανότερο να ανταποκριθούν σε στοχευμένες ενέργειες μάρκετινγκ. Εκτός από την πρόβλεψη ένα μοντέλο λογιστικής παλινδρόμησης δίνει τη δυνατότητα να εκτιμηθεί η επίδραση κάθε ανεξάρτητης μεταβλητής στη διαμόρφωση των τιμών της εξαρτημένης μεταβλητής.

Η πιο διαδεδομένη στη βιβλιογραφία έκφραση της λογιστικής παλινδρόμησης είναι:

$$\ln(ODDS) = a + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (2.1)$$

$$ODDS = \left(\frac{p}{1-p} \right) \quad (2.2)$$

Το δεξί μέλος της εξίσωσης (2.1) δημιουργείται από ένα γραμμικό συνδυασμό των ανεξάρτητων μεταβλητών που συμμετέχουν στο μοντέλο παλινδρόμησης. Το αριστερό μέλος περιέχει τις τιμές της εξαρτημένης μεταβλητής με τη μορφή του λογαρίθμου των *ODDS*, δηλαδή του λογαρίθμου της σχέσης. Το *ODDS* εναλλακτικά ονομάζεται *Logit* και ο όρος $\hat{p}$ εκφράζει την πιθανότητα του συμβάντος του γεγονότος. Οι συντελεστές των ανεξάρτητων μεταβλητών στην εξίσωση παλινδρόμησης εκτιμούνται βάσει της μεθόδου Μεγίστης Πιθανοφάνειας (*maximum likelihood*). Σύμφωνα με τη μέθοδο αυτή η τιμή των συντελεστών των ανεξάρτητων μεταβλητών είναι αυτή που κάνει τις παρατηρηθείσες τιμές της εξαρτημένης μεταβλητής πιο πιθανές, βάσει του συνόλου των ανεξαρτήτων μεταβλητών (Wuensch, 2009).

2.6.2 Δέντρα Αποφάσεων

Τα δέντρα αποφάσεων είναι διαισθητικές (*intuitive*) μέθοδοι ταξινόμησης. Ακολουθούν μια σειρά ερωτήσεων ή κανόνων αποφάσεων, όπου η επόμενη ερώτηση εξαρτάται από την απάντηση στην τρέχουσα ερώτηση. Συνήθως είναι χρήσιμα για κατηγορικά δεδομένα, καθώς οι κανόνες δεν απαιτούν καμία μετρίσιμη έννοια. Για την ανάπτυξη ενός συνόλου κανόνων αποφάσεων έχουν αναπτυχθεί διάφοροι αλγόριθμοι, όπως ο ID3, ο C4.5 και τα CART ή CHAID. Μια δημοφιλής προσέγγιση για τη διαμορφωση δέντρων αποφάσεων βασίζεται στη θεωρητική έννοια της πληροφορίας της εντροπίας (Crone et al., 2006).

Ο αλγόριθμος ξεκινώντας από ένα σύνολο παραδειγμάτων τα οποία διαμορφώνουν το δείγμα εκπαίδευσης, οδηγεί στην ανάπτυξη ενός συνόλου κανόνων αποφάσεων για την ταξινόμηση. Κάθε κόμβος του δέντρου περιλαμβάνει το κριτήριο αξιολόγησης το οποίο ελέγχεται βάσει των συνθηκών που καθορίζουν τα κλαδιά του δέντρου. Το κριτήριο το οποίο παρουσιάζει τη μικρότερη εντροπία επιλέγεται ως εκείνο το οποίο παρέχει την υψηλότερη νέα πληροφορία στην ταξινόμηση. Τέλος, τα φύλλα του δέντρου υποδεικνύουν την κατηγορία στην οποία πρέπει να ενταχθεί ένας πελάτης. Η ανάπτυξη ενός δέντρου πραγματοποιείται μέσω επαναληπτικής διαδικασίας.

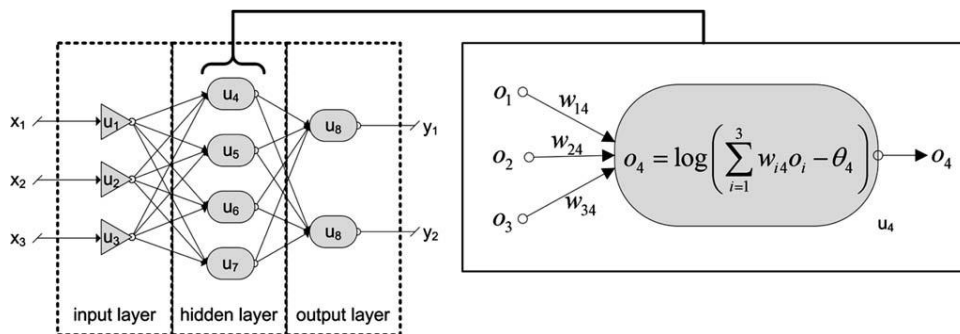
Λόγω της επαναλαμβανόμενης διαδικασίας, τα δέντρα αποφάσεων τείνουν να οδηγήσουν στην ανάπτυξη ενός ιδιαίτερα περίπλοκου και εξειδικευμένου δέντρου κάθε φύλλο του οποίου θα αφορά μόνο ένα πελάτη. Το φαινόμενο αυτό είναι αποτέλεσμα της μεγάλης προσαρμογής του αναπτυσσόμενου δέντρου στα δεδομένα του δείγματος εκπαίδευσης (*overfitting*). Η αντιμετώπιση του προβλήματος αυτού πραγματοποιείται

σε ένα δεύτερο στάδιο του αλγορίθμου, μέσω της περικοπής του αναπτυσσόμενου δέντρου (*pruning*) ώστε αυτό να καταστεί περισσότερο γενικό (Δούμπος και Ζοπουνίδης,2001).

2.6.3 Τεχνητά Νευρωνικά Δίκτυα

Τα ANN αναπτύχθηκαν από ερευνητές του χώρου της τεχνητής νοημοσύνης ως μια καινοτόμος μεθοδολογία μοντελοποίησης πολύπλοκων προβλημάτων. Κύρια μεθοδολογική βάση των ANN αποτελεί η προσπάθεια εξομίωσης του τρόπου λειτουργίας του ανθρώπινου εγκεφάλου κατά την επεξεργασία των σημάτων/μηνυμάτων που δέχεται από το εξωτερικό περιβάλλον. Η περισσότερο διαδεδομένη μορφή ANN είναι τα πολύ-επίπεδα (*multilayer*) perceptrons (*MLP*), (Δούμπος και Ζοπουνίδης,2001).

Η αρχιτεκτονική ενός MLP περιλαμβάνει ένα δίκτυο παράλληλων μονάδων επεξεργασίας οι οποίες είναι οργανωμένες σε μια σειρά επιπέδων (*layers*), όπως παρουσιάζεται στο Σχήμα 2.1.



Σχήμα 2.1: Αρχιτεκτονική MLP, επεξεργασία πληροφοριών μέσα σε έναν κόμβο χρησιμοποιώντας ένα σταθμισμένο ορό ως συνάρτηση εισαγωγής, τη λογιστική συνάρτηση ως συνάρτηση μετασχηματισμού και μια συνάρτηση εξόδου.
(Πηγή: Crone et al.,2006)

Επίπεδο εισόδου (*input layer*)

Ένα επίπεδο εισόδου αποτελείται από μια σειρά κόμβων u_j (επιμέρους μονάδες επεξεργασίας), ένα για κάθε είσοδο του ANN. Κάθε σύνδεση μεταξύ δύο κόμβων του ANN έχει ένα βάρος w_{ij} (*weight*), το οποίο αναπαριστά την ισχύ της σύνδεσης. Ο καθορισμός των βαρών αυτών (εκμάθηση του δικτύου) επιτυγχάνεται μέσω διαδικασιών βελτιστοποίησης, με σκοπό την ελαχιστοποίηση των αποκλίσεων μεταξύ των αποτελεσμάτων του δικτύου από το πραγματικό αποτέλεσμα (προκαθορισμένη ταξινόμηση) το οποίο καθορίζεται για ένα δείγμα εκπαίδευσης. Ως μέτρο των αποκλίσεων συνήθως χρησιμοποιείται το άθροισμα των τετραγώνων των σφαλμάτων, κατά αναλογία με την στατιστική παλινδρόμηση. Βάσει των βαρών των συνδέσεων, η είσοδος στον κάθε κόμβο του δικτύου υπολογίζεται ως ο σταθμισμένος μέσος των εξόδων όλων των κόμβων με τους οποίους συνδέεται (Δούμπος και Ζοπουνίδης,2001-Crone et al,2006).

Ενδιάμεσα επίπεδα (hidden layers)

Όλοι οι κόμβοι διαδοχικών επιπέδων συνδέονται πλήρως μεταξύ τους, ενώ είναι επίσης δυνατή η ανάπτυξη ANN στα οποία υπάρχουν συνδέσεις ακόμα και μεταξύ κόμβων που δεν ανήκουν σε διαδοχικά επίπεδα. Κάθε ενδιάμεσο επίπεδο j υπολογίζει ένα σταθμισμένο γραμμικό συνδυασμό $\mathbf{w}^T \mathbf{o}$ για κάθε διάνυσμα εισόδου $\mathbf{o}$, σταθμίζοντας κάθε εισόδο $\mathbf{o}_i$ του κόμβου i στο επίπεδο με τον ανάστροφο πίνακα $\mathbf{w}^T$ των προς εκπαίδευση βαρών w_{ij} περιλαμβάνοντας μια σταθερά εκπαίδευσης θ_j . Ο γραμμικός συνδυασμός μετασχηματίζεται με τη βοήθεια λειτουργιών μιας οριακής, αύξουσας, μη γραμμικής ενεργοποίησης σε κάθε κόμβο για να διαμορφώσει τη διαφορετική συμπεριφορά του δικτύου.

Επίπεδο εξόδου (output layer)

Ένα επίπεδο εξόδου το οποίο αποτελείται από ένα ή περισσότερους κόμβους, ανάλογα με τη μορφή που έχει το αποτέλεσμα του ANN. Στην περίπτωση προβλημάτων ταξινόμησης το πλήθος των κόμβων του επιπέδου εξόδου καθορίζεται ανάλογα με το πλήθος των κατηγοριών. Έτσι στην περίπτωση δύο κατηγοριών, το δίκτυο μπορεί να έχει μόνο ένα κόμβο εξόδου $y_i = \{0;1\}$ ή χρησιμοποιώντας πολλαπλές ταξινομήσεις $y_i = \{0,1;1,0\}$. Επιπλέον, η επιλογή της συνάρτησης εξόδου επιτρέπει στην πρόβλεψη της δυαδικής ταξινόμησης, καθώς επίσης και στον υπολογισμό της πιθανότητας για την ταξινόμηση κάθε περίπτωσης πελατών.

2.6.4 Μηχανές Διανυσμάτων Υποστήριξης

Οι μηχανές διανυσμάτων υποστήριξης μπορούν να χαρακτηριστούν ως αλγόριθμοι επιβλεπόμενης μάθησης ικανές να λύσουν γραμμικά και μη-γραμμικά δυαδικά προβλήματα ταξινόμησης. Η λογική τους στηρίζεται στον προσδιορισμό ενός υπερεπίπεδου χρησιμοποιώντας κατάλληλη συνάρτηση απόφασης που διαχωρίζει το χώρο των δεδομένων σε κλάσεις (Vapnik,1995).

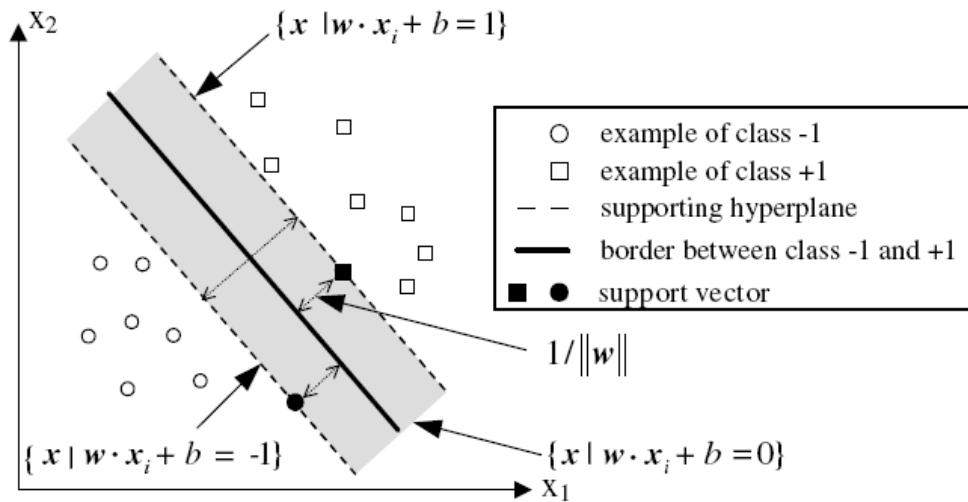
Δοθέντος ενός δείγματος εκπαίδευσης με m πρότυπα $\{(\mathbf{x}_i, y_i)\}_{i=1}^m$, όπου $\mathbf{x}_i$ το διάνυσμα εισόδου που χρησιμοποιείται για την εκπαίδευση και $y_i \in \{-1, +1\}$ η αντίστοιχη έξοδος. Στόχος της ανάλυσης είναι η κατασκευή μιας συνάρτησης $f(x)$ η οποία να διαχωρίζει τις θετικές από τις αρνητικές περιπτώσεις. Στην απλούστερη περίπτωση η $f(x)$ ορίζεται από ένα υπερεπίπεδο $w\mathbf{x} = b$ ως εξής:

$$f(x) = \text{sgn}(w\mathbf{x} - b) \quad (2.3)$$

όπου, w είναι ένα διάνυσμα των συντελεστών των x χαρακτηριστικών και b μια σταθερά. Αποδεικνύεται ότι το βέλτιστο υπερεπίπεδο (και συνεπώς η βέλτιστη συνάρτηση) είναι εκείνο με το μέγιστο περιθώριο διαχωρισμού ανάμεσα στις δύο κλάσεις (Vapnik,1995). Προκειμένου να πραγματοποιηθεί το μέγιστο περιθώριο διαχωρισμού θα πρέπει να ελαχιστοποιηθεί το $\|\mathbf{w}\|$ με τον περιορισμό ότι τα πρότυπα εκπαίδευσης κάθε κλάσης βρίσκονται στις αντίθετες πλευρές της διαχωριστικής επιφάνειας, όπως φαίνεται στο Σχήμα 2.2. Δηλαδή να ισχύει:

$$x_i w - b \geq 1, y = +1 \quad (2.4)$$

$$x_i w - b \leq -1, y = -1 \quad (2.5)$$



Σχήμα 2.2: Γραμμικός διαχωρισμός δύο κλάσεων -1,+1 με SVM.
(Πηγή: Crone et al.,2006)

Οι δύο παραπάνω ανισότητες (2.4)–(2.5) μπορούν να συνδυαστούν στην ακόλουθη:

$$y_i((\mathbf{w} \cdot \mathbf{x}_i) + b) \geq 1, \quad i = 1, \dots, m. \quad (2.6)$$

Τα σημεία που ικανοποιούν την ισότητα της (2.6), βρίσκονται πάνω στα όρια του περιθωρίου και ονομάζονται διανύσματα υποστήριξης (*Support Vectors*). Τα σημεία αυτά ταξινομούνται δυσκολότερα και αποτελούν τα κρίσιμότερα στοιχεία για την ταξινόμηση του δείγματος εκπαίδευσης. Επιπλέον, εάν τα SV μετακινηθούν θα αλλάξουν τη θέση του διαχωριστικού υπερεπιπέδου.

Για να ληφθούν υπόψη οι ταξινομήσεις, όπου ο περιορισμός (2.6) δεν ικανοποιείται, εισάγονται μεταβλητές απόκλισης ξ_i (*slack variables*). Αρκεί επομένως να επιλυθεί το εξής πρόβλημα ελαχιστοποίησης τετραγωνικού προγραμματισμού:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i$$

$$\text{Υπ: } y_i((\mathbf{w} \cdot \mathbf{x}_i) + b) \geq 1 - \xi_i, \quad i = 1, \dots, m. \quad (2.7)$$

όπου, C είναι μια σταθερά οριζόμενη από το χρήστη, η οποία δηλώνει την παραχώρηση ανάμεσα στους δύο συγκρουόμενους στόχους (μεγιστοποίηση του διαχωριστικού περιθωρίου και ελαχιστοποίηση των σφαλμάτων εκπαίδευσης). Τα βέλτιστα $\mathbf{w}^*$ και b^* που προκύπτουν από την επίλυση της (2.7) αντικαθιστούνται στη σχέση (2.3), (Crone et al,2006).

Τέλος, αξίζει να αναφερθεί ότι για τα μη-γραμμικά προβλήματα η έννοια του βέλτιστου υπερεπιπέδου διαφέρει αφού στη συγκεκριμένη περίπτωση δεν γίνεται απευθείας διαχωρισμός των κλάσεων. Στην περίπτωση αυτή το βέλτιστο υπερεπίπεδο είναι αυτό που ελαχιστοποιεί την πιθανότητα λανθασμένης ταξινόμησης κατά μέσο όρο για ολόκληρο το σύνολο των δεδομένων εκπαίδευσης.

2.6.5 Συνδυασμένες τεχνικές

Βασική ιδέα των συνδυασμένων τεχνικών είναι ο συνδυασμός πολλαπλών μοντέλων βάσει ορισμένων αρχών-κανόνων, αντί της χρήσης του μοντέλου που έχει την καλύτερη απόδοση, όπως συμβαίνει στις παραδοσιακές τεχνικές.

Μια συνδυασμένη τεχνική είναι και το bagging, το όνομα της οποίας προέρχεται από την ονομασία της διαδικασίας του αλγορίθμου, “**bootstrap aggregating**”. Η διαδικασία του bagging ξεκινά με το bootstrapping του δείγματος εκπαίδευσης. Δοθέντος ενός δείγματος εκπαίδευσης με N πρότυπα (π.χ. εγγραφές πελατών) κατασκευάζεται ένα δείγμα bootstrap, από τυχαία επιλογή με ομοιόμορφη κατανομή και επανατοποθέτηση, από το αρχικό δείγμα εκπαίδευσης. Γενικότερα, η παραπάνω διαδικασία πραγματοποιείται N φορές και λόγω της επανατοποθέτησης καθένα από τα N πρότυπα στο αρχικό δείγμα μπορεί να βρεθεί στο νέο από καμία έως N φορές. Έχει αποδειχθεί ότι ένα δείγμα bootstrap περιέχει περίπου το 63,2% του αρχικού δείγματος (Breiman, 1996). Με αυτή τη διαδικασία προκύπτει το δείγμα εκπαίδευσης. Το νέο αυτό δείγμα λέγεται bootstrap αντίγραφο του αρχικού. Αν πρέπει να συνδυαστούν L μοντέλα, τότε πραγματοποιείται η ίδια διαδικασία L φορές. Το αποτέλεσμα της καθεμιάς είναι το δείγμα εκπαίδευσης του αντίστοιχου μοντέλου. Προφανώς, τα δείγματα αυτά μπορούν να είναι από ίδια μέχρι ριζικά διαφορετικά μεταξύ τους. Όλα πάντως προέρχονται από το ίδιο αρχικό δείγμα οπότε είναι μέρη του ίδιου προβλήματος.

Στην συνέχεια, κάθε δείγμα bootstrap χρησιμοποιείται για την εκπαίδευση κάθε μοντέλου. Η εκπαίδευση κάθε μοντέλου γίνεται με τον τρόπο που αναλογεί στον τύπο του, π.χ. με backpropagation για MLP.

Ο τρόπος συνδυασμού των αποτελεσμάτων εξαρτάται από τη μορφή των εξόδων. Αν οι έξοδοι είναι συνεχείς σε κάποιο διάστημα, π.χ. $[0,1]$, τότε το bagging παίρνει το μέσο όρο των εξόδων από κάθε μοντέλο. Αν αντίθετα, οι έξοδοι είναι κατατάξεις κλάσεων (*labels*), η τελική απόφαση για την τάξη λαμβάνεται με ψήφο πλειοψηφίας, δηλαδή η κλάση που προτείνει τελικά το bagging θα είναι αυτή που προτείνουν τα περισσότερα μοντέλα. Ο συνδυασμός των αποτελεσμάτων των μοντέλων για να ληφθεί η τελική απόφαση της μεθόδου λέγεται aggregating. Έτσι προκύπτει η ονομασία *bootstrap aggregating* από την οποία παραπάνω αναφέρθηκε ότι προέρχεται το όνομα bagging (Bauer και Kohavi, 1999).

Βήμα 1. Παραγωγή δειγμάτων bootstrap

$\{B_i\}, i = 1, 2, \dots, L$: Σύνολο αντιγράφων *bootstrap* του αρχικού δείγματος B

$\{f_i\}, i = 1, 2, \dots, L$: Σύνολο μοντέλων για συνδυασμό

Βήμα 2. Εκπαίδευση L μοντέλων

$f_i(x, y), i = 1, 2, \dots, L, (x, y) \in B_i$: Εκπαίδευση κάθε L μοντέλου βάσει το δείγμα *bootstrap*

Βήμα 3. Aggregating L αποτελεσμάτων

$f_{bag}(x) = \sum_{i=1}^L \alpha_i f_i(x)$: Συνολικό αποτέλεσμα βασισμένο σε βάρη α_i ή με πλειοψηφία

Σχήμα 2.3: Βήματα μεθόδου bagging.
(Πηγή: Ha et al., 2005)

ΚΕΦΑΛΑΙΟ 3

ΕΦΑΡΜΟΓΗ ΣΤΟ ΑΜΕΣΟ ΜΑΡΚΕΤΙΝΓΚ

3.1 ΠΕΡΙΓΡΑΦΗ ΔΕΔΟΜΕΝΩΝ

Το πρόβλημα που μελετάται στην παρούσα εργασία αφορά τον εντοπισμό υπονήφιων πελατών στα πλαίσια μιας στρατηγικής άμεσου μάρκετινγκ για μια ασφαλιστική εταιρία. Το ασφαλιστικό προϊόν, για το οποίο πραγματοποιήθηκε η έρευνα αφορά ασφαλιστήριο για τροχόσπιτα. Τα δεδομένα που χρησιμοποιήθηκαν είναι από το “CoIL Challenge” και βασίζονται σε πραγματικό επιχειρηματικό πρόβλημα. Τα δεδομένα είναι διαθέσιμα στο διαδίκτυο: <http://www.liacs.nl/~putten/library/cc2000/Tic2000.zip>.

Το αρχείο των δεδομένων περιλαμβάνει 9.822 εγγραφές, κάθε μια από τις οποίες αντιπροσωπεύει ένα πελάτη. Οι πρώτες 5.822 εγγραφές πελατών αποτελούν το δείγμα εκπαίδευσης (*train sample*) και το ποσοστό απάντησης (*response rate*) ήταν 5,97%, δηλαδή 348 πελάτες αποδέχτηκαν το προϊόν (κατηγορία 1) και 5474 πελάτες δεν το αποδέχτηκαν (κατηγορία 0). Οι υπόλοιπες 4.000 εγγραφές αποτελούν το δείγμα ελέγχου (*test sample*), στο οποίο 238 πελάτες ανήκουν στην κατηγορία 1 και 3762 στην κατηγορία 0. Η εγγραφή κάθε πελάτη αποτελείται από 86 μεταβλητές, οι οποίες περιλαμβάνουν κοινωνικό-δημογραφικά στοιχεία για τους πελάτες (μεταβλητές 1-43), τα οποία έχουν συλλεχθεί από τον ταχυδρομικό κώδικα των πελατών, στοιχεία για τη χρήση ασφαλιστικών προϊόντων (μεταβλητές 44-85) και τέλος τη δυαδική μεταβλητή απόφασης 0/1 (μεταβλητή 86). Η σχέση μεταξύ των μεταβλητών εισαγωγής και της μεταβλητής απόφασης είναι πολύ ασθενής (*weak*). Ακόμη, πρέπει να αναφερθεί πως δεν υπάρχουν ελλιπή δεδομένα για καμία εγγραφή και όλες οι συνεχείς μεταβλητές έχουν μετασχηματιστεί σε διακριτές, το πολύ σε δώδεκα τιμές.

Όλοι οι πελάτες που διαμένουν σε περιοχές με τον ίδιο ταχυδρομικό κώδικα έχουν ίδια κοινωνικό-δημογραφικά χαρακτηριστικά. Δηλαδή, τα κοινωνικό-δημογραφικά δεδομένα κάθε εγγραφής πελάτη συνδέονται με τον ταχυδρομικό κώδικα και όχι με κάθε μεμονωμένο πελάτη. Για παράδειγμα, η τιμή 5 για τη μεταβλητή “Ιδιοκτήτης Σπιτιού” σημαίνει ότι, το 50-62% των ανθρώπων που ζουν στην ίδια περιοχή ταχυδρομικού κώδικα, όπως αυτός ο πελάτης είναι ιδιοκτήτες ενός σπιτιού. Δεδομένου ότι οι μεταβλητές αυτές συνδέονται με μία ενιαία κρυφή μεταβλητή, την περιοχή διαμονής, μπορεί να είναι ιδιαίτερα συσχετισμένες μεταξύ τους.

Στο Παράρτημα της εργασίας παρουσιάζεται αναλυτικός πίνακας με την περιγραφή κάθε μεταβλητής (Πίνακας Π1), (Van der Putten και Someren, 2004).

3.2 ΕΠΙΛΟΓΗ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΜΕΤΑΒΛΗΤΩΝ

3.2.1 Επιλογή μεταβλητών

Από τις 85 μεταβλητές (αναφέρθηκε παραπάνω ότι η μεταβλητή 86 αποτελεί τη δυαδική μεταβλητή απόφασης) που περιέχει κάθε εγγραφή πελάτη, αποφασίστηκε ορισμένες εξ αυτών να τεθούν εκτός της έρευνας που πραγματοποιείται στην παρούσα εργασία. Στον Πίνακα 3.1 παρουσιάζονται τα αποτελέσματα από τον έλεγχο Pearson χ^2 και δίδεται το επίπεδο σημαντικότητας κάθε μεταβλητής, το οποίο σχετίζεται με τη διακριτική ικανότητα κάθε μεταβλητής στο διαχωρισμό των πελατών σε δύο κατηγορίες, αυτούς που αποδέχονται το ασφαλιστικό προϊόν (κατηγορία 1) και αυτούς που δεν το αποδέχονται (κατηγορία 0).

Το κριτήριο σύμφωνα με το οποίο αποφασίζεται αν μια μεταβλητή τίθεται εκτός έρευνας, είναι η τιμή του επιπέδου σημαντικότητας κάθε μεταβλητής (Asymp. Sig). Μεταβλητές για τις οποίες η τιμή του επιπέδου σημαντικότητας είναι μεγαλύτερη του 0,5 θεωρείται ότι δεν συνεισφέρουν στο τελικό αποτέλεσμα και τίθενται εκτός έρευνας. Σύμφωνα με τον παραπάνω κανόνα από τις αρχικές 85 μεταβλητές τέθηκαν εκτός 25 μεταβλητές (παρουσιάζονται με γκρι χρώμα στον Πίνακα 3.1). Στις 25 μεταβλητές που εξαιρέθηκαν, υπάρχουν 7 μεταβλητές των οποίων η τιμή του επιπέδου σημαντικότητας είναι μικρότερη ή βρίσκεται πολύ κοντά στο 0,5 (μεταβλητές 20-51-60-63-72-77-81). Παρόλα αυτά, εξαιρέθηκαν διότι από την κατανομή των απαντήσεων παρατηρήθηκε ότι δεν υπήρχε σε κάθε βαθμίδα ικανοποιητικός αριθμός απαντήσεων και οι απαντήσεις συγκεντρώνονταν σε μια μόνο βαθμίδα. Για το λόγο αυτό, κρίθηκε απαραίτητο, αν και η τιμή που επιπέδου σημαντικότητας ήταν μικρότερη του 0,5 οι μεταβλητές αυτές να τεθούν εκτός έρευνας.

3.2.2 Επεξεργασία μεταβλητών

Οι 60 μεταβλητές που παρέμειναν για κάθε εγγραφή πελάτη επεξεργάστηκαν περαιτέρω σύμφωνα με τη συχνότητα εμφάνισης των απαντήσεων σε κάθε βαθμίδα κάθε μεταβλητής. Συγκεκριμένα, χρησιμοποιήθηκε η αναλογία των πελατών που ανήκουν στην κατηγορία 0 προς το συνολικό αριθμό απαντήσεων σε κάθε βαθμίδα, για να εξεταστεί εάν διαφορετικές βαθμίδες κάθε μεταβλητής παρέχουν διαφορετική πληροφορία ως προς την ταξινόμηση των πελατών. Αντίστοιχα, θα μπορούσαν να χρησιμοποιηθούν οι αναλογίες 1/σύνολο, 0/1 ή 1/0. Οι βαθμίδες της μεταβλητής που έχουν ίδια αναλογία, παρέχουν την ίδια ακριβώς πληροφορία στην πρόβλεψη και επομένως μπορούν να εξεταστούν ως μια ενιαία κατάσταση. Το στοιχείο αυτό εξετάστηκε σε συνδυασμό με το πλήθος των πελατών σε κάθε βαθμίδα των μεταβλητών. Η ανάλυση αυτή βοηθά στην “απλοποίηση” της κωδικοποίησης των μεταβλητών μέσω της αναμόρφωσης της ποιοτικής κλίμακας μέτρησης τους ώστε αυτή να αποκτήσει μεγαλύτερη πληροφοριακή ισχύ.

Τα ποσοστά κάθε βαθμίδας των 60 μεταβλητών παρουσιάζονται στον Πίνακα Π2 του Παραρτήματος. Ακόμη, στον Πίνακα Π3 παρουσιάζονται οι ενοποιημένες βαθμίδες για όσες μεταβλητές πραγματοποιήθηκε η διαδικασία που παραπάνω περιγράφηκε.

Pearson χ^2			Asymp. Sig. (2-sided)		
Μεταβλητή			Μεταβλητή		
1	124,81	0,00	44	57,47	0,00
2	3,46	0,90	45	3,91	0,69
3	9,32	0,05	46	2,85	0,42
4	3,29	0,66	47	194,68	0,00
5	88,66	0,00	48	0,83	0,84
6	9,12	0,43	49	23,68	0,00
7	22,4	0,01	50	0,57	0,90
8	11,74	0,04	51	4,03	0,55
9	19,4	0,02	52	3,91	0,42
10	30,90	0,00	53	1,34	0,85
11	7,83	0,35	54	12,87	0,02
12	27,79	0,00	55	10,34	0,32
13	17,95	0,04	56	0,83	0,99
14	6,71	0,67	57	14,44	0,00
15	12,26	0,20	58	7,96	0,09
16	47,37	0,00	59	140,39	0,00
17	18,70	0,03	60	6,96	0,03
18	57,46	0,00	61	80,94	0,00
19	39,78	0,00	62	4,79	0,03
20	4,46	0,48	63	5,59	0,47
21	20,17	0,02	64	29,36	0,00
22	25,16	0,00	65	48,30	0,00
23	32,1	0,00	66	0,07	0,97
24	23,52	0,01	67	2,64	0,10
25	52,14	0,00	68	136,75	0,00
26	20,96	0,01	69	0,58	0,97
27	4,22	0,90	70	0,86	0,84
28	34,53	0,00	71	0,57	0,90
29	26,12	0,00	72	2,24	0,52
30	49,44	0,00	73	1,82	0,77
31	48,99	0,00	74	1,34	0,85
32	36,99	0,00	75	11,94	0,00
33	1,52	0,98	76	20,66	0,00
34	41,92	0,00	77	0,42	0,52
35	26,41	0,00	78	6,55	0,01
36	25,07	0,00	79	7,96	0,02
37	52,23	0,00	80	33,53	0,00
38	6,18	0,72	81	3,99	0,05
39	38,91	0,00	82	66,37	0,00
40	27,61	0,00	83	10,86	0,01
41	3,82	0,80	84	2,35	0,31
42	73,79	0,00	85	27,80	0,00
43	82,89	0,00			

Πίνακας 3.1: Αποτελέσματα ελέγχου Pearson χ^2 .

3.2.3 Υπολογιστικά μοντέλα και αξιολόγηση

Στα πλαίσια της παρούσας εργασίας χρησιμοποιήθηκαν τα εξής υπολογιστικά μοντέλα:

- **BINLOG60:** Μοντέλο λογιστικής παλινδρόμησης με βηματική επιλογή μεταβλητών, χρησιμοποιώντας το σύνολο των μεταβλητών που έχουν επεξεργαστεί σύμφωνα με τη διαδικασία που περιγράφηκε στην υποενότητα 3.2.2.

- **BINLOG27:** Μοντέλο λογιστικής παλινδρόμησης με βηματική επιλογή μεταβλητών, χρησιμοποιώντας τις 27 μεταβλητές που επιλέχθηκαν στο μοντέλο από την πρώτη εφαρμογή της λογιστικής παλινδρόμησης (BINLOG60).
- **CRT60:** Δέντρα αποφάσεων, με το σύνολο των μεταβλητών χωρίς την επεξεργασία της υποενότητας 3.2.2.
- **CRT21:** Δέντρα αποφάσεων, με τις 21 μεταβλητές που επιλέχθηκαν στο μοντέλο από την εφαρμογή της BINLOG27.
- **MLP21:** Τεχνητά νευρωνικά δίκτυα, με τις 21 μεταβλητές που επιλέχθηκαν στο μοντέλο από την εφαρμογή της BINLOG27.
- **SVM21:** Μηχανές διανυσμάτων υποστήριξης, με τις 21 μεταβλητές που επιλέχθηκαν στο μοντέλο από την εφαρμογή της BINLOG27.

Επίσης, χρησιμοποιήθηκε η συνδυασμένη τεχνική bagging και κατασκευάστηκαν τα παρακάτω υβριδικά μοντέλα:

- **Ensemble:** Bagging, με τις 21 μεταβλητές που επιλέχθηκαν στο μοντέλο από την εφαρμογή της BINLOG27.
- **Hybrid:** $f(CRT60 + BINLOG27)$.
- **Hybrid:** $f(CRT21 + BINLOG27)$.
- **Hybrid:** $f(MLP21 + BINLOG27)$.
- **Hybrid:** $f(SVM21 + BINLOG27)$.

Στο πρώτο μοντέλο της λογιστικής παλινδρόμησης, συμπεριελήφθη το σύνολο των μεταβλητών σύμφωνα με την επεξεργασία που περιγράφηκε στην υποενότητα 3.2.2 και παρουσιάζεται στον Πίνακα Π3 του Παραρτήματος. Στο δεύτερο μοντέλο της λογιστικής παλινδρόμησης χρησιμοποιήθηκαν οι 27 μεταβλητές που επιλέχθηκαν στο πρώτο μοντέλο της λογιστικής παλινδρόμησης (BINLOG60). Οι 27 μεταβλητές επεξεργάστηκαν περαιτέρω σύμφωνα με τα αποτελέσματα της σημαντικότητας τους και των διαστημάτων εμπιστοσύνης ($\Delta.E$) των συντελεστών των μεταβλητών στο μοντέλο. Η λογική της ενοποίησης των ψευδομεταβλητών (βαθμίδων) κάθε μεταβλητής είναι παρόμοια με αυτή που περιγράφηκε στην υποενότητα 3.2.2, όσον αφορά την επεξεργασία του συνόλου των 60 μεταβλητών. Οι ψευδομεταβλητές όπου τα $\Delta.E$ ταυτίζονται ή καλύπτει το ένα το άλλο, σημαίνει ότι παρέχουν την ίδια ακριβώς πληροφορία στην πρόβλεψη και επομένως μπορούν να εξεταστούν ως μια ενιαία κατάσταση. Στον Πίνακα Π4 παρουσιάζονται οι συντελεστές των μεταβλητών στο μοντέλο (B), οι τιμές των $\Delta.E$, καθώς επίσης και η τελική ενοποίηση των ψευδομεταβλητών που αντιστοιχούν σε επιμέρους βαθμίδες κάθε μεταβλητής. Αυτή η επιπλέον επεξεργασία των δεδομένων επεκτείνει τους μετασχηματισμούς που αναλύθηκαν στην υποενότητα 3.2.2, λαμβάνοντας έτσι υπόψη τις αλληλεπιδράσεις των μεταβλητών σε ένα πολυμεταβλητό μοντέλο πρόβλεψης.

Στα μοντέλα των δέντρων αποφάσεων χρησιμοποιήθηκαν αρχικά, οι 60 μεταβλητές χωρίς να υποστούν καμία επεξεργασία και στη συνέχεια οι 21 επεξεργασμένες μεταβλητές που επιλέχθηκαν από το μοντέλο BINLOG27. Ο λόγος, για τον οποίο αρχικά χρησιμοποιήθηκαν οι μεταβλητές χωρίς καμία επεξεργασία είναι γιατί τα δέντρα ταξινόμησης, εκτός από την ανάπτυξη ενός μοντέλου πρόβλεψης μπορούν να χρησιμοποιηθούν και ως μια διαδικασία επιλογής μεταβλητών. Στην περίπτωση αυτή, το συγκριτικό τους πλεονέκτημα έναντι της λογιστικής παλινδρόμησης είναι ότι δεν

υποθέτουν μια γραμμική σχέση μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων μεταβλητών. Παρά όμως το πλεονέκτημα αυτό, στην παρούσα ανάλυση δεν διαπιστώθηκε κάποια βελτίωση από τη χρήση των μεταβλητών που επιλέγονται μέσω των δέντρων ταξινόμησης στην ανάπτυξη μη γραμμικών μοντέλων πρόβλεψης. Έτσι, όλα τα υπόλοιπα μη γραμμικά μοντέλα βασίστηκαν στις μεταβλητές που προέκυψαν μέσω της λογιστικής παλινδρόμησης. Συγκεκριμένα, οι 21 μεταβλητές της λογιστικής παλινδρόμησης χρησιμοποιήθηκαν και στα νευρωνικά δίκτυα (MLP21), καθώς και στις μηχανές διανυσμάτων υποστήριξης (SVM21). Όσον αφορά το SVM21 μοντέλο, χρησιμοποιήθηκε το γραμμικό, με τιμή κόστους $C=10$ και η συνάρτηση Radial Basis (RBF), με τιμή κόστους $C=100$ και Gamma $g=0,1$. Τέλος, στη συνδυασμένη τεχνική, συνδυαστήκαν τα αποτελέσματα από 50 δέντρα αποφάσεων και χρησιμοποιήθηκαν οι 21 μεταβλητές της BINLOG27 με το τελικό αποτέλεσμα να λαμβάνεται με ψήφο πλειοψηφίας (*bagging*). Στα 4 υβριδικά μοντέλα προστέθηκε σαν εξαρτημένη μεταβλητή και η πιθανότητα πρόβλεψης που εξήχθη από τη BINLOG27. Η επιλογή να εισαχθεί ως μία ακόμη εξαρτημένη μεταβλητή στα υβριδικά μοντέλα, η πιθανότητα πρόβλεψης που εξήχθη από τη BINLOG27, οφείλεται στο γεγονός ότι η BINLOG27 επιτυγχάνει καλύτερα αποτελέσματα πρόβλεψης σε σχέση με τα υπόλοιπα μοντέλα.

Για την εφαρμογή των παραπάνω υπολογιστικών μοντέλων, χρησιμοποιήθηκε το λογισμικό SPSS για τα μοντέλα της λογιστικής παλινδρόμησης, των δέντρων ταξινόμησης και των τεχνητών νευρωνικών δικτύων και το λογισμικό MATLAB για τα μοντέλα των μηχανών διανυσμάτων υποστήριξης και της συνδυασμένης τεχνικής *bagging*.

Στο άμεσο μάρκετινγκ υπάρχουν διάφορα κριτήρια για την αξιολόγηση των υπολογιστικών μοντέλων. Στην παρούσα εργασία χρησιμοποιείται ο πίνακας ταξινόμησης (*confusion matrix*) και η καμπύλη ROC. Τέλος, παρουσιάζεται η σημαντικότητα των μεταβλητών στην πρόβλεψη, για όσα υπολογιστικά μοντέλα παρέχεται η πληροφορία αυτή.

3.3 ΠΑΡΟΥΣΙΑΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

3.3.1 Πίνακες ταξινόμησης

Οι πίνακες ταξινόμησης κατασκευάζονται συνήθως για την αξιολόγηση της ακρίβειας των υπολογιστικών μοντέλων. Ένας πίνακας ταξινόμησης έχει τη μορφή του Πίνακα 3.2. Στον Πίνακα 3.3 παρουσιάζονται οι ποσότητες που υπολογίζονται από τον πίνακα ταξινόμησης.

		Πρόβλεψη		
		0	1	SUM
Πραγματικές	0	<i>A</i>	<i>B</i>	<i>A+B</i>
	1	<i>C</i>	<i>D</i>	<i>C+D</i>

Πίνακας 3.2: Πίνακας ταξινόμησης.

Ποσότητα	Αναλογία
Ακρίβεια (AC)	$(A+D)/(A+B+C+D)$
Ανάκληση ή True positive (TP)	$D/(C+D)$
False positive (FP)	$B/(A+B)$
True negative (TN)	$A/(A+B)$
False negative (FN)	$C/(C+D)$
Σαφήνεια/Precision (P)	$D/(B+D)$

Πίνακας 3.3: Αναλογίες πίνακα ταξινόμησης.

Αρχικά, πρέπει να αναφερθεί ότι το *A* αποτελεί τον αριθμό των σωστά ταξινομημένων πελατών, οι οποίοι δεν αποδέχθηκαν το προϊόν. Το *B* αποτελεί τον αριθμό των λάθος ταξινομημένων πελατών, οι οποίοι δεν αποδέχθηκαν το προϊόν αλλά ταξινομήθηκαν ως πελάτες που το αποδέχθηκαν. Αντίστοιχα, το *D* αποτελεί τον αριθμό των σωστά ταξινομημένων πελατών, οι οποίοι αποδέχθηκαν το προϊόν και το *C* τον αριθμό των λάθος ταξινομημένων πελατών, οι οποίοι αποδέχθηκαν το προϊόν αλλά ταξινομήθηκαν ως πελάτες που δεν το αποδέχθηκαν.

Μια ποσότητα για τη μέτρηση της ακρίβειας είναι η αναλογία του αθροίσματος των σωστών ταξινομήσεων ($A+D$) με το σύνολο ($A+B+C+D$). Η αναλογία αυτή είναι γνωστή ως PCC (*Percentage Correctly Classified instances*). Το μειονέκτημα που παρουσιάζει η αναλογία PCC και δεν είναι επαρκής για ασφαλή συμπεράσματα, δημιουργείται όταν ο αριθμός του *A* είναι αρκετά μεγαλύτερος από τον αριθμό του *D*. Για παράδειγμα, η ακρίβεια PCC ενός μοντέλου μπορεί να είναι 81%, όμως η ακρίβεια της πρόβλεψης των πελατών που ανήκουν στην κατηγορία 1 να είναι μόλις 15%. Αυτό αποδεικνύει ότι το παραπάνω μοντέλο είναι πολύ καλύτερο στην πρόβλεψη των πελατών που ανήκουν στην κατηγορία 0 (Gonul et al.,2000). Παρόλα αυτά, μόνο ο υπολογισμός της ακρίβειας PCC δεν μπορεί να το αποδείξει αυτό.

Για την αντιμετώπιση του παραπάνω προβλήματος, συνήθως υπολογίζονται δύο ακόμη ποσότητες. Η αναλογία ανάκλησης (*recall*) και η σαφήνεια (*precision*). Η αναλογία ανάκλησης είναι γνωστή στο άμεσο μάρκετινγκ, ως ποσοστό απάντησης και υπολογίζεται ως: $D/(C+D)$. Δηλαδή, η αναλογία ανάκλησης υπολογίζει το ποσοστό των σωστών ταξινομήσεων, που ανήκουν στην κατηγορία 1 (*D*), προς το σύνολο των πραγματικών ταξινομήσεων της κατηγορίας 1 ($C+D$). Αντίστοιχα, η σαφήνεια υπολογίζει το ποσοστό των σωστών ταξινομήσεων, που ανήκουν στην κατηγορία 1 (*D*), προς το σύνολο των προβλεπόμενων ταξινομήσεων της κατηγορίας 1 ($B+D$). Ο γεωμετρικός μέσος όρος (*geometric mean*) αποτελεί ακόμη μια ποσότητα για τη μέτρηση της ακρίβειας χρησιμοποιώντας τις αναλογίες ανάκλησης και σαφήνειας σύμφωνα με τη σχέση (Kubat et al.,1998):

$$gmean = \sqrt{recall \cdot precision} \quad (3.1)$$

Στους πίνακες που ακολουθούν παρουσιάζονται τα αποτελέσματα για το δείγμα ελέγχου των 4000 πελατών. Στον Πίνακα 3.4 παρουσιάζονται οι πίνακες ταξινόμησης για κάθε υπολογιστικό μοντέλο σε ποσοστά και στον Πίνακα 3.5 τα αποτελέσματα των ποσοτήτων PCC, recall, precision και g-mean.

Μοντέλο	Τιμές	Πρόβλεψη	
		0	1
BINLOG60	Πραγματικές	0	0,734
		1	0,450
BINLOG27	Πραγματικές	0	0,717
		1	0,412
CRT60	Πραγματικές	0	0,607
		1	0,328
CRT21	Πραγματικές	0	0,676
		1	0,399
MLP21	Πραγματικές	0	0,725
		1	0,441
SVM21-linear	Πραγματικές	0	0,717
		1	0,408
SVM21-RBF	Πραγματικές	0	0,934
		1	0,807

Πίνακας 3.4: Πίνακες ταξινόμησης υπολογιστικών μοντέλων.

Μέθοδος	SVM21					
Αναλογία	BINLOG60	BINLOG27	CRT60	CRT21	MLP21	linear
PCC	72,27%	70,92%	61,05%	67,15%	71,48%	71,00%
Recall	55,04%	58,82%	67,23%	60,08%	55,88%	59,24%
Precision	11,57%	11,62%	9,76%	10,50%	11,38%	11,71%
g-mean	25,24%	26,14%	25,61%	25,12%	25,21%	26,34%

Πίνακας 3.5: Αποτελέσματα υπολογιστικών μοντέλων.

Όπως, παρουσιάζεται στον Πίνακα 3.5 το μεγαλύτερο ποσοστό PCC επιτυγχάνεται από το μοντέλο SVM21-RBF, με ποσοστό 89,00%, όμως στο μοντέλο αυτό παρουσιάζεται το μειονέκτημα που εμφανίζει η αναλογία PCC, όπως αναφέρθηκε παραπάνω. Ενώ, η ακρίβεια PCC του μοντέλου SVM21-RBF είναι 89,00%, η ακρίβεια της πρόβλεψης των πελατών που ανήκουν στην κατηγορία 1 (*recall*) είναι μόλις 19,33%. Το μοντέλο παρουσιάζει καλύτερη συμπεριφορά στην πρόβλεψη των πελατών που ανήκουν στην κατηγορία 0 ($3514/3762=93,41\%$). Ακολουθούν τα μοντέλα BINLOG60 και MLP21 με ποσοστά 72,27% και 71,48% και τα μοντέλα BINLOG27 και SVM21-linear με 70,92% και 71,00%, αντίστοιχα. Τέλος, τα μοντέλα των δέντρων αποφάσεων εμφανίζουν τις μικρότερες ακρίβειες με 67,15% το CRT21 και 61,05% το CRT60.

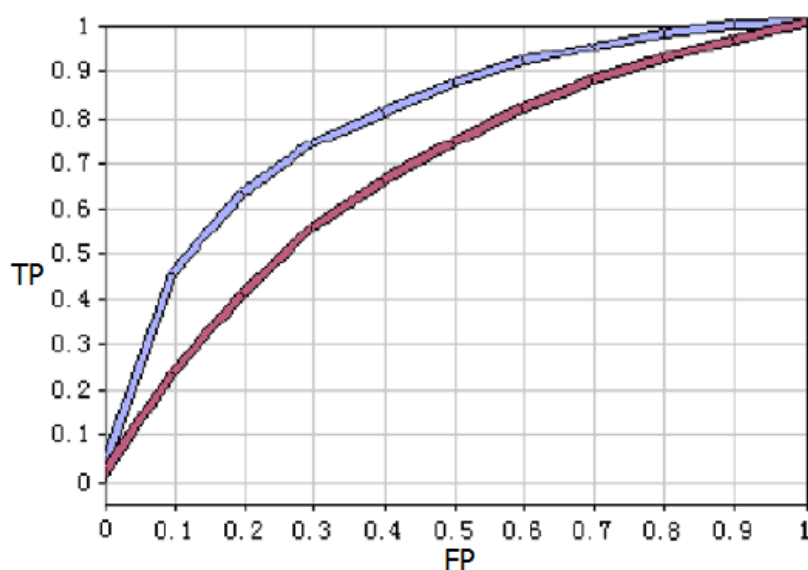
Ο υπολογισμός των αναλογιών ανάκλησης και σαφήνειας, αποδεικνύει την αδυναμία της ακρίβειας PCC. Ειδικότερα, όσον αφορά το μοντέλο SVM21-RBF η αναλογία ανάκλησης είναι μόλις 19,33% και η σαφήνεια 15,65%, η τιμή της οποίας είναι η μεγαλύτερη για όλα τα μοντέλα και οφείλεται στη σωστή ταξινόμηση (93,41%) των πελατών που ανήκουν στην κατηγορία 0. Τα μοντέλα των δέντρων αποφάσεων παρουσιάζουν τις υψηλότερες τιμές όσον αφορά την αναλογία ανάκλησης με ποσοστά 67,23% το μοντέλο CRT21 και 60,08% το CRT60. Ακολουθούν τα μοντέλα SVM21-linear και BINLOG27 με κοντινές τιμές, 59,24% και 58,82%, αντίστοιχα. Λίγο χαμηλότερα βρίσκονται τα μοντέλα MLP21 (55,88%) και BINLOG60 (55,04%).

Η αναλογία σαφήνειας, παρουσιάζει παρόμοια συμπεριφορά για όλα τα μοντέλα εκτός, από το SVM21-RBF, για το λόγο που παραπάνω αναφέρθηκε. Οι τιμές για τα υπόλοιπα έξι μοντέλα κυμαίνονται από 9,76% έως 11,71% με τα μοντέλα των δέντρων αποφάσεων να παρουσιάζουν τιμές κάτω από 11,00%.

Τέλος, η τιμή του ποσοστού του γεωμετρικού μέσου, *g-mean*, παρουσιάζει πολύ κοντινές τιμές για όλα τα μοντέλα, εκτός από το μοντέλο SVM21-RBF που βρίσκεται αρκετά χαμηλότερα. Υψηλότερη τιμή εμφανίζει το μοντέλο SVM21-linear με 26,34% και οι τιμές για τα υπόλοιπα μοντέλα δεν πέφτουν κάτω από 25,12%. Ενώ, το μοντέλο SVM21-RBF βρίσκεται στο τέλος με 17,39%.

3.3.2 Καμπύλη ROC

Η καμπύλη ROC (*Receiver Operating Characteristics*) χρησιμοποιείται συνήθως μαζί με τους πίνακες ταξινόμησης για την αξιολόγηση των υπολογιστικών μοντέλων. Ένα παράδειγμα καμπύλης ROC παρουσιάζεται στο Σχήμα 3.1. Η καμπύλη ROC σχεδιάζεται συνήθως ως δισδιάστατη γραφική παράσταση, έχοντας στον Y-άξονα το ποσοστό των πελατών, οι οποίοι αποδέχθηκαν το προϊόν και ταξινομήθηκαν σωστά (*True positives-TP*), ενώ στον X-άξονα το ποσοστό των πελατών, οι οποίοι δεν αποδέχθηκαν το προϊόν αλλά ταξινομήθηκαν ως πελάτες που το αποδέχθηκαν (*False positives-FP*). Ως *TP* ορίζεται η αναλογία του $D/(C+D)$, η οποία εκτός από *recall* αναφέρεται και ως ευαισθησία (*sensitivity*). Ενώ, ως *FP* ορίζεται η αναλογία του $B/(A+B)$ και αναφέρεται ως $1-\text{specificity}$. Το σημείο (0,1) αντιπροσωπεύει την άριστη ταξινόμηση, ενώ το (1,0) την τελειώς ανακριβής ταξινόμηση. Όσο υψηλότερη είναι μια καμπύλη ROC τόσο καλύτερο είναι ένα υπολογιστικό μοντέλο. Σε ορισμένες περιπτώσεις η καμπύλη ROC σχεδιάζεται έχοντας στον Y-άξονα την ποσότητα $1-\text{ευαισθησία}$, για να μελετήσει τις παραχωρήσεις μεταξύ *TP* και *FP*. Η μείωση του *TP* προκαλεί αύξηση στο *FP* και αντίστροφα.



Σχήμα 3.1: Τυπική μορφή καμπύλης ROC.

(Πηγή: Bose και Chen, 2009)

Ένα μέτρο αξιολόγησης της προβλεπτικής ικανότητας ενός μοντέλου, το οποίο προκύπτει από την καμπύλη ROC είναι το εμβαδόν της περιοχής κάτω από την καμπύλη (*Area Under ROC-AUROC*). Μεγαλύτερες τιμές του εμβαδού AUROC αντιπροσωπεύουν καλύτερο μοντέλο. Το εμβαδόν αυτό ισούται με τη μονάδα για ένα τέλειο προβλεπτικό σύστημα, ενώ τιμές από 0.9 έως 1 αντιστοιχούν σε εξαιρετικές προβλέψεις, από 0.8 έως 0.9 σε πολύ καλές προβλέψεις, από 0.7 έως 0.8 σε καλές, από 0.6 έως 0.7 σε μέτρια προβλεπτική ικανότητα, από 0.5 έως 0.6 σε τυχαίες προβλέψεις και τιμές μικρότερες από 0.5 αντιστοιχούν σε ένα ανεπαρκές σύστημα πρόβλεψης (Swets,1988).

Τα χαρακτηριστικά μιας καμπύλης ROC μπορούν να συνοψιστούν στα εξής (Yu και Cho,2006) :

- Μια καμπύλη ROC ή ένα σημείο (*FP,TP*) είναι ανεξάρτητα από τα σφάλματα ταξινόμησης και από τον τρόπο κατανομής των πελατών ανεξαρτήτως του μεγέθους των κατηγοριών (*class distribution*).
- Μια γραφική παράσταση ROC εμπεριέχει όλες τις πληροφορίες που περιλαμβάνονται στον πίνακα ταξινόμησης, δεδομένου ότι το *FN* είναι συμπλήρωμα του *TP* και το *TN* συμπλήρωμα του *FP*.
- Μια καμπύλη ROC παρέχει ένα οπτικό εργαλείο για την εξέταση των παραχωρήσεων, μεταξύ της ικανότητας ενός μοντέλου να προσδιορίσει σωστά τις ταξινομήσεις της κατηγορίας 1 και τον αριθμό των ταξινομήσεων της κατηγορίας 0 που είναι λάθος ταξινομημένες.

Στον Πίνακα 3.6 παρουσιάζονται τα αποτελέσματα του εμβαδού AUROC για κάθε ένα από τα υπολογιστικά μοντέλα.

Μέθοδος	BINLOG60	BINLOG27	CRT60	CRT21	MLP21	SVM21	
Τιμή						linear	RBF
AUROC	0,702	0,711	0,698	0,694	0,709	0,712	0,654

Πίνακας 3.6: Αποτελέσματα AUROC υπολογιστικών μοντέλων.

Από τον Πίνακα 3.6 παρατηρείται ότι μεγαλύτερες τιμές AUROC εμφανίζουν τα μοντέλα SVM21-linear και BINLOG27 με τιμές 0,712 και 0,711, αντίστοιχα. Ακολουθεί το μοντέλο MLP21 με 0,709 και το μοντέλο BINLOG60 με 0,702. Τα μοντέλα αυτά θα μπορούσαν να χαρακτηριστούν ως ικανοποιητικά μοντέλα. Οι τιμές των μοντέλων των δέντρων αποφάσεων, CRT60 και CRT21, βρίσκονται λίγο χαμηλότερα με 0,698 και 0,694, αντίστοιχα. Στο τέλος βρίσκεται το μοντέλο SVM21-RBF με τιμή 0,654, επιβεβαιώνοντας ότι πρόκειται για ένα όχι τόσο ικανοποιητικό μοντέλο κάτι το οποίο φάνηκε και από τις τιμές των αναλογιών που υπολογίστηκαν στην παραπάνω υποενότητα.

Σύμφωνα με τις τιμές των αναλογιών που υπολογίστηκαν παραπάνω και τις τιμές AUROC, τα επτά υπολογιστικά μοντέλα θα μπορούσαν να χωριστούν ως εξής:

1^ο Group: Τα μοντέλα, SVM21-linear και BINLOG27, των οποίων η συμπεριφορά είναι σχεδόν παρόμοια. Με ικανοποιητικές τιμές ακρίβειας PCC, όπως επίσης και των

υπόλοιπων τριών αναλογιών. Επίσης, ικανοποιητικές κρίνονται και οι τιμές AUROC και για τα δύο μοντέλα. Τα αποτελέσματα ουσιαστικά αποδεικνύουν ότι και τα δύο μοντέλα παρουσιάζουν ικανοποιητική προσαρμογή και στις δύο κατηγορίες πρόβλεψης.

2^ο Group: Τα μοντέλα MLP21 και BINLOG60, με τις υψηλότερες τιμές της ακρίβειας PCC. Σε συνδυασμό όμως, με τα αποτελέσματα τις τιμές AUROC και όχι τόσο των υπόλοιπων τριών αναλογιών, εμφανίζονται να βρίσκονται αισθητά χαμηλότερα από το 1^ο Group.

3^ο Group: Τα δύο μοντέλα των δέντρων αποφάσεων, CRT60 και CRT21, τα οποία εμφανίζουν τα μεγαλύτερα ποσοστά, όσον αφορά την αναλογία ανάκλησης, με τις τιμές όμως των υπόλοιπων αναλογιών να βρίσκονται σε μέτρια επίπεδα, όπως επίσης και της τιμής AUROC.

Στο τέλος βρίσκεται το μοντέλο SVM21-RBF, το οποίο εμφανίζει την μεγαλύτερη τιμή της ακρίβειας PCC, όμως οι τιμές των υπόλοιπων αναλογιών και AUROC δεν εμφανίζουν ικανοποιητικά αποτελέσματα.

Οι τιμές του AUROC σε μεγάλο βαθμό αποδεικνύουν και συμπληρώνουν όσα παραπάνω αναφέρθηκαν σύμφωνα με τις τιμές των αναλογιών που υπολογίστηκαν βάσει των αποτελεσμάτων στους πίνακες ταξινόμησης. Αποδεικνύεται, πως τα δύο αυτά κριτήρια σε συνεργασία μπορούν σε μεγάλο βαθμό να αξιολογήσουν τα αποτελέσματα της πρόβλεψης ενός μοντέλου ταξινόμησης.

3.3.3 Επίπεδο σημαντικότητας μεταβλητών

Σε ορισμένες τεχνικές δίδεται η δυνατότητα του προσδιορισμού της τιμής του επιπέδου σημαντικότητας των μεταβλητών που περιλαμβάνονται στο μοντέλο. Παρέχεται ουσιαστικά η δυνατότητα της επίδρασης κάθε ανεξάρτητης μεταβλητής, που περιλαμβάνεται στο μοντέλο, στη διαμόρφωση των τιμών της εξαρτημένης μεταβλητής ή του αποτελέσματος. Η πληροφορία αυτή δεν παρέχεται για τεχνικές, όπως οι μηχανές διανυσμάτων υποστήριξης.

Οι παρακάτω πίνακες παρουσιάζουν τις σημαντικότερες μεταβλητές για κάθε υπολογιστικό μοντέλο, εκτός από τα μοντέλα SVM21-linear και SVM21-RBF.

BINLOG60	BINLOG27
newv47	newv47BR
newv59	newv59BR
newv23	newv1BR
newv61	newv61BR
newv7	newv7BR
newv1	newv8BR
newv16	newv23BR
newv37	newv34BR
newv34	newv16BR
newv32	newv18BR

Πίνακας 3.7: Οι 10 σημαντικότερες μεταβλητές στα μοντέλα BINLOG60 και BINLOG27 ($p < 0.01$).

MLP21	Σημαντικότητα	Κανονικοποιημένη Σημαντικότητα
newv61BR	0,113	100,0%
newv59BR	0,073	65,2%
newv47BR	0,071	62,7%
newv58BR	0,065	57,6%
newv1BR	0,061	54,2%
newv7BR	0,060	53,0%
newv16BR	0,051	45,1%
newv5BR	0,047	42,1%
newv34BR	0,046	40,8%
newv23BR	0,045	39,6%

Πίνακας 3.8: Σημαντικότητα μεταβλητών μοντέλου MLP21.

CRT60	Επίπεδο Δέντρου	CRT21	Επίπεδο Δέντρου
V47	1 ^ο	newv47BR	1 ^ο
V59	2 ^ο	newv59BR	2 ^ο
V37	2 ^ο	newv1BR	2 ^ο
V23	3 ^ο	newv5BR	3 ^ο
V42	3 ^ο	newv54BR	3 ^ο
V18	3 ^ο		

Πίνακας 3.9: Σημαντικότητα μεταβλητών μοντέλων CRT60 και CRT27.

Σύμφωνα με τους παραπάνω πίνακες εξάγονται ορισμένα χρήσιμα συμπεράσματα όσον αφορά τις μεταβλητές που περιλαμβάνονται σε κάθε υπολογιστικό μοντέλο. Συγκεκριμένα, όσον αφορά τα μοντέλα της λογιστικής παλινδρόμησης στις πρώτες θέσεις σημαντικότητας παρατηρούνται σχεδόν οι ίδιες μεταβλητές με εναλλαγές των θέσεων μεταξύ τους, με τις μεταβλητές 47 και 59 να παραμένουν στην πρώτη και δεύτερη θέση, αντίστοιχα. Πρέπει επίσης να αναφερθεί, ότι και για τα δύο μοντέλα το πρόσημο του συντελεστή των μεταβλητών 59, 1 και 34 είναι αρνητικό. Δηλαδή, οι συγκεκριμένες ανεξάρτητες μεταβλητές σχετίζονται αρνητικά με το τελικό αποτέλεσμα (εξαρτημένη μεταβλητή). Στο μοντέλο BINLOG27, εκτός από τις τρεις παραπάνω μεταβλητές, αρνητικό πρόσημο εμφανίζει και η μεταβλητή 18. Οι υπόλοιπες μεταβλητές και για τα δύο μοντέλα εμφανίζουν θετικά πρόσημα συντελεστών, άρα σχετίζονται θετικά με τη μεταβλητή απόφασης. Τέλος, αξίζει να τονιστεί, ότι το μοντέλο BINLOG27 εφαρμόστηκε με τις 27 μεταβλητές που συμπεριελήφθησαν στο μοντέλο BINLOG60 και ενοποιήθηκαν σύμφωνα με τα κριτήρια που παρουσιάστηκαν στην υποενότητα 3.2.3.

Σχεδόν παρόμοια συμπεριφορά με τα μοντέλα της λογιστικής παλινδρόμησης, όσον αφορά την εκτίμηση της σημαντικότητας των μεταβλητών που συμπεριλαμβάνονται σε ένα μοντέλο, παρουσιάζει και το μοντέλο τεχνητών νευρωνικών δικτύων. Στο συγκεκριμένο μοντέλο η μεταβλητή 61 εμφανίζεται με την υψηλότερη σημαντικότητα και ακολουθούν οι 59 και 47. Οι μεταβλητές που παρατηρούνται παρακάτω είναι σε πολύ μεγάλο βαθμό (πέντε στον αριθμό) οι ίδιες με το μοντέλο BINLOG27. Υπενθυμίζεται ότι και τα δύο μοντέλα χρησιμοποίησαν το ίδιο

σύνολο μεταβλητών. Παρατηρείται λοιπόν, πως τα δύο παραπάνω μοντέλα συμπεριφέρονται παρόμοια στην εκτίμηση της σημαντικότητας των μεταβλητών για την πρόβλεψη.

Στον Πίνακα 3.9 παρουσιάζονται τα επίπεδα σημαντικότητας των μοντέλων των δέντρων ταξινόμησης. Η μεταβλητή 47 και στα δύο αυτά μοντέλα κατέχει την υψηλότερη σημαντικότητα και ακολουθεί η 59. Μεγαλύτερη ομοιότητα όσον αφορά τα επίπεδα σημαντικότητας των μεταβλητών, σε σχέση με τα παραπάνω μοντέλα παρουσιάζει το μοντέλο CRT21. Δηλαδή, το μοντέλο στο οποίο εφαρμόστηκε το ίδιο σύνολο μεταβλητών με το MLP21. Ενώ στο μοντέλο CRT60, στο οποίο εφαρμόστηκε το σύνολο των 60 μεταβλητών, δίχως να έχουν υποστεί καμία επεξεργασία, με εξαίρεση τις δύο πρώτες θέσεις εμφανίζονται διαφορετικές μεταβλητές. Παρόλα αυτά, στα πρώτα δύο επίπεδα σημαντικότητας εμφανίζονται οι ίδιες μεταβλητές, ενώ διαφορά υπάρχει στα κατώτερα άρα και λιγότερο σημαντικά.

Τέλος, όπως αναφέρθηκε στην αρχή του κεφαλαίου το ασφαλιστικό προϊόν, για το οποίο διεξάγεται η έρευνα, είναι η αποδοχή ή όχι ενός ασφαλιστήριου τροχόσπιτου. Σύμφωνα με τον Πίνακα Π1 του Παραρτήματος, ο οποίος περιλαμβάνει την επεξήγηση του συνόλου των μεταβλητών, παρατηρείται πως οι μεταβλητές που εκτιμώνται από τα μοντέλα με την υψηλότερη σημαντικότητα αναφέρονται σε:

- Εισφορά ασφαλιστηρίου αυτοκινήτου: Μεταβλητή 47
- Εισφορά ασφαλιστηρίου φωτιάς: Μεταβλητή 59
- Εισφορά ασφαλιστηρίου βάρκας: Μεταβλητή 61
- Υποκατηγορία πελατών: Μεταβλητή 1

Ο συνδυασμός των παραπάνω μεταβλητών παρατηρείται να καθορίζει σε μεγάλο βαθμό την πρόβλεψη που πραγματοποιείται από τα υπολογιστικά μοντέλα.

3.4 BAGGING ΚΑΙ ΥΒΡΙΔΙΚΑ ΜΟΝΤΕΛΑ

3.4.1 Πίνακες ταξινόμησης και καμπύλες ROC

Στους επόμενους τρεις πίνακες παρουσιάζονται τα αποτελέσματα των πινάκων ταξινόμησης, των τιμών των αναλογιών που υπολογίστηκαν βάσει των πινάκων ταξινόμησης και το εμβαδόν της περιοχής κάτω από την καμπύλη ROC, για ένα συνδυασμένο μοντέλο (bagging) και τέσσερα υβριδικά μοντέλα.

Μοντέλο	Τιμές	Πρόβλεψη	
		0	1
Bagging	Πραγματικές	0	0,779
		1	0,466
CRT60+ BINLOG27	Πραγματικές	0	0,724
		1	0,412
CRT21+ BINLOG27	Πραγματικές	0	0,666
		1	0,349
MLP21+ BINLOG27	Πραγματικές	0	0,710
		1	0,378
SVM21-linear+ BINLOG27	Πραγματικές	0	0,724
		1	0,412
SVM21-RBF+ BINLOG27	Πραγματικές	0	0,935
		1	0,807

Πίνακας 3.10: Πίνακες ταξινόμησης μοντέλου bagging και υβριδικών μοντέλων.

Μέθοδος	CRT60+	CRT21+	MLP21+	SVM21+BL27		Bagging
Αναλογία	BL27	BL27	BL27	linear	RBF	
PCC	69,58%	66,50%	70,47%	71,60%	89,05%	76,48%
Recall	55,88%	65,13%	62,18%	58,82%	19,33%	53,36%
Precision	10,68%	10,98%	11,95%	11,88%	15,75%	13,27%
g-mean	24,43%	26,74%	27,25%	26,44%	17,45%	26,61%

Πίνακας 3.11: Αποτελέσματα υβριδικών μοντέλων και bagging.

Μέθοδος	CRT60+	CRT21+	MLP21+	SVM21+BL27		Bagging
Τιμή	BL27	BL27	BL27	linear	RBF	
AUROC	0,678	0,708	0,707	0,704	0,653	0,726

Πίνακας 3.12: Αποτελέσματα AUROC υβριδικών μοντέλων και bagging.

Από τους παραπάνω πίνακες γίνεται κατανοητό, πως το συνδυασμένο μοντέλο bagging (συνδυασμός 50 δέντρων αποφάσεων) επιτυγχάνει καλύτερα αποτελέσματα σε σύγκριση με αυτά που επιτυγχάνει ένα μεμονωμένο μοντέλο. Ειδικότερα, το ποσοστό PCC βρίσκεται σε ικανοποιητικά επίπεδα, καθώς επίσης και οι τιμές των υπόλοιπων αναλογιών είναι ικανοποιητικές και υψηλότερες από τις τιμές των υπόλοιπων μοντέλων. Τα αποτελέσματα για την ικανοποιητική πρόβλεψη του συνδυασμένου μοντέλου επιβεβαιώνονται και από την τιμή AUROC. Η τιμή AUROC είναι 0,726 και

είναι η μεγαλύτερη που επιτυγχάνεται, σε σύγκριση με τα αποτελέσματα των μεμονωμένων μοντέλων, που παρουσιάστηκαν στον Πίνακα 3.6 και τα αποτελέσματα των υβριδικών μοντέλων του Πίνακα 3.12.

Όσον αφορά τα υβριδικά μοντέλα, κανένα δεν καταφέρνει να επιτύχει καλύτερα αποτελέσματα από το 1^ο γκρουπ των μεμονωμένων μοντέλων, όπως επίσης και από το συνδυασμένο μοντέλο bagging. Στην περίπτωση όμως, που κάθε υβριδικό μοντέλο συγκριθεί με το αντίστοιχο μεμονωμένο μοντέλο μπορούν να εξαχθούν ορισμένα χρήσιμα συμπεράσματα. Συγκεκριμένα, όσον αφορά το υβριδικό μοντέλο CRT60+BL27 οι τιμές των αναλογιών εκτός, από την PCC, παρουσιάζονται σε χαμηλότερα επίπεδα. Αυτό αντικατοπτρίζεται και στην τιμή AUROC, η οποία βρίσκεται χαμηλότερα με τιμή 0,678 από 0,698. Αντίθετα, στο υβριδικό μοντέλο CRT21+BL27 εμφανίζεται ικανοποιητική βελτίωση της τιμής AUROC, από 0,694 σε 0,708. Το γεγονός ότι το υβριδικό μοντέλο CRT21+BL27 επιτυγχάνει καλύτερα αποτελέσματα από το CRT60+BL27, ίσως να οφείλεται στην επεξεργασία του συνόλου των 27 μεταβλητών και ο συνδυασμός τους με την εκτιμώμενη πιθανότητα που εξήχθη από το μοντέλο λογιστικής παλινδρόμησης BINLOG27, σε αντίθεση με τις 60 μεταβλητές του μοντέλου CRT60, οι οποίες χωρίς καμία επεξεργασία συνδυάστηκαν με την εκτιμώμενη πιθανότητα του BINLOG27.

Σχεδόν παρόμοια αποτελέσματα επιτυγχάνονται για τα υβριδικά μοντέλα MLP21+BL27 και SVM21-RBF+BL27 σε σύγκριση με τα αντίστοιχα μεμονωμένα μοντέλα. Η τιμή AUROC μειώνεται για το υβριδικό μοντέλο MLP21+BL27 από 0,709 σε 0,707, ενώ παρατηρούνται καλύτερα ποσοστά στις αναλογίες ακρίβειας. Το υβριδικό μοντέλο SVM21-RBF+BL27 παρουσιάζει ακριβώς παρόμοια συμπεριφορά, όπως το SVM21-RBF. Τέλος, το υβριδικό μοντέλο SVM21-linear+BL27 επιτυγχάνει χαμηλότερα αποτελέσματα σε σύγκριση με το SVM21-linear.

3.4.2 Επίπεδο σημαντικότητας μεταβλητών

Οι πίνακες που ακολουθούν παρουσιάζουν τις τιμές του επιπέδου σημαντικότητας των μεταβλητών που περιλαμβάνονται στα παραπάνω υβριδικά μοντέλα πρόβλεψης και στο συνδυασμένο μοντέλο bagging.

Bagging	Σημαντικότητα
newv47BR	1,215269
newv54BR	0,657679
newv23BR	0,399909
newv59BR	0,362404
newv8BR	0,327303
newv5BR	0,240227
newv31BR	0,183149
newv18BR	0,151298
newv16BR	0,100971
newv32BR	0,063417

Πίνακας 3.13: Σημαντικότητα μεταβλητών μοντέλου bagging.

MLP21+BL27	Σημαντικότητα	Κανονικοποιημένη Σημαντικότητα
Πιθανότητα πρόβλεψης 27_BRvar	0,303	100,0%
newv1BR	0,063	20,8%
newv42BR	0,047	15,5%
newv6BR	0,046	15,0%
newv16BR	0,043	14,2%
newv8BR	0,043	14,1%
newv34BR	0,042	14,0%
newv7BR	0,040	13,3%
newv47BR	0,040	13,0%
newv22BR	0,036	11,7%

Πίνακας 3.14: Σημαντικότητα μεταβλητών υβριδικού μοντέλου MLP21+BL27.

CRT60+BL27	Επίπεδο Δέντρου	CRT21+BL27	Επίπεδο Δέντρου
Πιθανότητα πρόβλεψης 27_BRvar	1 ^ο	Πιθανότητα πρόβλεψης 27_BRvar	1 ^ο
V42	2 ^ο	newv1BR	2 ^ο
V22	2 ^ο		
V15	3 ^ο		
V31	3 ^ο		

Πίνακας 3.15: Σημαντικότητα μεταβλητών υβριδικών μοντέλων CRT60+BL27 και CRT27+BL27.

Το εκτιμώμενο επίπεδο σημαντικότητας, όσον αφορά το συνδυασμένο μοντέλο bagging, λειτουργεί μεταβάλλοντας την τιμή μιας μεταβλητής, με τυχαία επιλογή, παρατηρώντας τη μεταβολή που επιφέρει η μεταβολή αυτή στο τελικό αποτέλεσμα. Η κατάταξη των μεταβλητών με την υψηλότερη σημαντικότητα παρουσιάζεται στον Πίνακα 3.13. Σύμφωνα με τον Πίνακα 3.13, το συνδυασμένο μοντέλο θεωρεί την μεταβλητή 47, ως την μεταβλητή με την υψηλότερη σημαντικότητα. Όπως αναφέρθηκε και παραπάνω η μεταβλητή 47, αναφέρεται στην εισφορά του ασφαλιστηρίου αυτοκινήτου, συμβάλει καθοριστικά στο τελικό αποτέλεσμα όλων των μεμονωμένων μοντέλων. Η συγκεκριμένη μεταβλητή εμφανίζεται πρώτη σε σημαντικότητα και στο μοντέλο των δέντρων αποφάσεων, CRT21. Το γεγονός αυτό, σε μεγάλο βαθμό οφείλεται στο ότι το συνδυασμένο μοντέλο συνδυάζει τα αποτελέσματα από 50 δέντρα αποφάσεων για το ίδιο σύνολο μεταβλητών.

Σε σύγκριση με τα υπόλοιπα υπολογιστικά μοντέλα, που και σε αυτά χρησιμοποιήθηκε το ίδιο σύνολο μεταβλητών, η σημαντική διαφορά παρουσιάζεται στην απουσία της μεταβλητής 61 (εισφορά ασφαλιστηρίου βάρκας) από τα επίπεδα υψηλής σημαντικότητας. Ενώ, αρκετές από τις υπόλοιπες μεταβλητές εμφανίζονται να κατέχουν καθοριστικό ρόλο και στα μεμονωμένα υπολογιστικά μοντέλα, όπως και στο συνδυασμένο.

Τέλος, όσον αφορά τα υβριδικά μοντέλα υψηλότερη σημαντικότητα εμφανίζεται να κατέχει η πιθανότητα πρόβλεψης που υπολογίζεται από το μοντέλο BINLOG27 και προστέθηκε στο σύνολο των ανεξάρτητων μεταβλητών κάθε μοντέλου. Αυτό τονίζει τον τρόπο “λειτουργίας” και “συμπεριφοράς” των υβριδικών μοντέλων. Δηλαδή, η

πιθανότητα πρόβλεψης που υπολογίζεται από κάποιο άλλο μοντέλο πρόβλεψης, εκτιμάται ως η μεταβλητή με τη μεγαλύτερη σημαντικότητα από το μοντέλο που επεξεργάζεται το σύνολο των μεταβλητών, στο οποίο έχει προστεθεί η πιθανότητα ως ανεξάρτητη μεταβλητή. Οι υπόλοιπες μεταβλητές σε μεγάλο βαθμό είναι οι ίδιες που εμφανίζονται στα μεμονωμένα μοντέλα.

ΚΕΦΑΛΑΙΟ 4

ΣΥΜΠΕΡΑΣΜΑΤΑ &

ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ

Στην εργασία αυτή παρουσιάστηκε το πρόβλημα του εντοπισμού υπονήφιων πελατών στα πλαίσια μιας στρατηγικής άμεσου μάρκετινγκ για ένα ασφαλιστικό προϊόν.

Όπως αναφέρθηκε, το άμεσο μάρκετινγκ δεν αποτελεί απλά μια άλλη τακτική μάρκετινγκ αλλά μια αποτελεσματική στρατηγική που αλληλεπιδρά μεταξύ πελάτη και επιχείρησης. Σε αντιδιαστολή με το κλασικό μάρκετινγκ, είναι πελατοκεντρικό, καθώς βασίζεται στην άμεση και εξατομικευμένη επικοινωνία, στοχεύοντας στην ανάπτυξη μιας σχέσης με τους πελάτες και μετά την πώληση. Στόχος του άμεσου μάρκετινγκ, είναι η αύξηση που ποσοστού ανταπόκρισης συγκεκριμένων πελατών.

Το πρόβλημα για τις εταιρίες άμεσου μάρκετινγκ επικεντρώνεται στο τρόπο με τον οποίο θα κατορθώσουν οι εταιρίες να προβλέψουν το ενδιαφέρον των πελατών για ένα προϊόν ή μία υπηρεσία, αυξάνοντας το ποσοστό ανταπόκρισης και μειώνοντας το κόστος των προγραμμάτων μάρκετινγκ.

Για την επίλυση του προβλήματος αυτού έχουν προταθεί και μελετηθεί διάφορες μεθοδολογικές προσεγγίσεις από το χώρο της στατιστικής, της επιχειρησιακής έρευνας και της τεχνητής νοημοσύνης. Στην παρούσα εργασία, εφαρμόστηκαν η λογιστική παλινδρόμηση, τα τεχνητά νευρωνικά δίκτυα, τα δέντρα αποφάσεων, οι μηχανές διανυσμάτων υποστήριξης, καθώς επίσης ένα συνδυασμένο μοντέλο και υβριδικά μοντέλα. Τα κύρια συμπεράσματα ήταν τα ακόλουθα:

1. Τα μοντέλα των μηχανών διανυσμάτων υποστήριξης (SVM21-γραμμικό) και της λογιστική παλινδρόμησης (BINLOG27), τα οποία εφαρμόστηκαν με το ίδιο σύνολο των επεξεργασμένων μεταβλητών, σύμφωνα με τα αποτελέσματα των πινάκων ταξινόμησης και του εμβαδού της περιοχής κάτω από την καμπύλη ROC, επιτυγχάνουν τα καλύτερα αποτελέσματα πρόβλεψης και κρίνονται ως ικανοποιητικά μοντέλα. Αισθητά χαμηλότερα σε απόδοση βρίσκονται τα μοντέλα των τεχνητών νευρωνικών δικτύων (MLP21) και λογιστικής παλινδρόμησης (BINLOG60, με το σύνολο των 60 μεταβλητών).
2. Τα μοντέλα των δέντρων αποφάσεων επιτυγχάνουν μέτρια αποτελέσματα ανεξαρτήτως του συνόλου μεταβλητών που εφαρμόστηκαν. Αυτό οφείλεται στο γεγονός, ότι οι εκτιμήσεις των πιθανοτήτων πρόβλεψης που εξάγονται από τα μοντέλα δέντρων αποφάσεων παρουσιάζουν 15-20 διαφορετικές τιμές πιθανοτήτων, σε σχέση με τις τιμές των πιθανοτήτων των άλλων μοντέλων που

παρουσιάζουν ασυνέχεια, δίχως να επαναλαμβάνονται οι ίδιες τιμές. Το γεγονός αυτό αντικατοπτρίζεται στον υπολογισμό του εμβαδού της περιοχής κάτω από την καμπύλη ROC, που λαμβάνει ως πληροφορία τις εκτιμώμενες πιθανότητες που εξάγει κάποιο μοντέλο. Επίσης, ίσως να οφείλεται και στον τύπο των δεδομένων, καθώς τα δέντρα αποφάσεων εμφανίζουν ισχυρή ικανότητα με κατηγορικά δεδομένα.

3. Η χρήση ενός συνδυασμένου μοντέλου (Bagging), το οποίο συνδυάζει τα αποτελέσματα 50 δέντρων αποφάσεων, στην παρούσα εργασία επιτυγχάνει καλύτερα αποτελέσματα από όλα τα μοντέλα που εφαρμόστηκαν.
4. Συγκρίνοντας τα αποτελέσματα των υβριδικών μοντέλων, τα οποία συνδυάζουν δύο ή περισσότερα διαφορετικά μοντέλα, με τα αντίστοιχα των μεμονωμένων μοντέλων δεν παρουσιάζεται ιδιαίτερη βελτίωση. Στα περισσότερα μοντέλα παρουσιάζεται παρόμοια συμπεριφορά ίσως και αισθητά χαμηλότερα, εκτός από το υβριδικό μοντέλο των δέντρων αποφάσεων (CRT21+BINLOG27), στο οποίο παρουσιάζεται ικανοποιητική βελτίωση σε σύγκριση με το μεμονωμένο μοντέλο δέντρου αποφάσεων (CRT21). Σε καμία έρευνα δεν έχει εμφανιστεί πως ο συνδυασμός συγκεκριμένων μοντέλων κατορθώνει να επιτύχει πάντα καλύτερα αποτελέσματα από τα μεμονωμένα, αλλά αυτό εξαρτάται σε μεγάλο βαθμό από τη φύση των δεδομένων και τον τύπο των μοντέλων που συνδυάζονται.
5. Οι μεταβλητές με το υψηλότερο εκτιμώμενο επίπεδο σημαντικότητας, σε μεγάλο βαθμό, είναι παρόμοιες στα περισσότερα μοντέλα που εφαρμόστηκαν. Δεν παύουν όμως να εμφανίζονται διαφορές σε ορισμένα μοντέλα, σύμφωνα με τον τρόπο λειτουργίας και τους κανόνες που ακολουθούν, καθώς επίσης και το χώρο που αντιπροσωπεύουν, εκτιμώντας διαφορετική-ες μεταβλητή-ες στα υψηλότερα επίπεδα σημαντικότητας μεταβλητών.

Ορισμένες εφικτές μελλοντικές επεκτάσεις της παρούσας εργασίας είναι οι ακόλουθες:

1. Όπως παρουσιάστηκε στην παρούσα εργασία η επεξεργασία των μεταβλητών επιφέρει διαφορετικά αποτελέσματα στην πρόβλεψη. Η διαδικασία αυτή είναι ιδιαίτερα σημαντική στο χώρο του άμεσου μάρκετινγκ, καθώς τις περισσότερες φορές υπάρχουν ελλιπή δεδομένα ή σε αυτά εμφανίζεται “θόρυβος”. Σε μία μελλοντική έρευνα θα μπορούσε να εξεταστεί σε ποίο βαθμό διαφορετικές μέθοδοι επεξεργασίας/προετοιμασίας των δεδομένων, επηρεάζουν το αποτελέσματα των υπολογιστικών μοντέλων.
2. Εκτός από την επεξεργασία των μεταβλητών, θα μπορούσε επίσης να μελετηθεί, η μεταβολή που επιφέρει στο τελικό αποτέλεσμα της πρόβλεψης η επιλογή συγκεκριμένων μεταβλητών.
3. Θα μπορούσε επίσης, να εφαρμοστούν και άλλες τεχνικές από τους χώρους της στατιστικής, της επιχειρησιακής έρευνας, καθώς και της τεχνητής νοημοσύνης με στόχο την αξιολόγηση και σύγκριση των αποτελεσμάτων περισσότερων τεχνικών. Έντονο ενδιαφέρον στους ερευνητές παρουσιάζουν οι τεχνικές των προσεγγιστικών συνόλων (*Rough Sets*), (Jensen και Shen, 2004) και των Bayesian Belief Networks (Baesens et al., 2002), οι οποίες μέχρι σήμερα

χρησιμοποιούνται σπάνια στο άμεσο μάρκετινγκ. Το ενδιαφέρον αυτό, οφείλεται στο γεγονός ότι οι παραπάνω δύο τεχνικές μπορούν καθορίζουν σχέσεις εξάρτησης μεταξύ των μεταβλητών είτε να προσδιορίζουν σημαντικούς κανόνες μεταξύ τους, επιτυγχάνοντας υψηλή ακρίβεια και παρέχοντας συγχρόνως σαφή ερμηνεία.

4. Τέλος, η χρήση άλλων μεθόδων ανάπτυξης συνδυασμένων μοντέλων (π.χ boosting) και επέκταση σε άλλες μεθόδους (π.χ συνδυασμός νευρωνικών δικτύων ή μηχανών διανυσμάτων υποστήριξης).

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Δούμπος Μ., Ζοπουνίδης Κ. (2001). *Πολυκριτήριες τεχνικές ταξινόμησης: Θεωρία και εφαρμογές*, Κλειδάριθμος, Αθήνα.
- Agrawal, R., Srikant, R. (1994). "Fast algorithms for mining association rules. In: Proceedings of the Twentieth International Conference on Very Large Databases", Santiago, Chile, pp. 487–499.
- Baesens, B., Viaene, S., Van den Poel, D., Vanthienen, J., Dedene, G., (2002). "Bayesian neural network learning for repeat purchase modeling in direct marketing", *European Journal of Operational Research*, 138 (1), 191–211.
- Bauer E., Kohavi R. (1999). "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting and Variants", *Machine Learning*, 36, 105–139.
- Baumgartner, B., Hruschka, H., (2005). "Allocation of catalogs to collective customers based on semiparametric response models". *European Journal of Operational Research*, 162 (3), 839–849.
- Berger, P., T. Magliozzi. (1992). "The effect of sample size and proportion of buyers in the sample on the performance of list segmentation equations generated by regression analysis". *J. Direct Marketing*, 6(1) 13–22.
- Bhattacharyya, S., (1999). "Direct marketing performance modeling using generic algorithms". *INFORMS Journal on Computing*, 11 (3), 248–257.
- Bhattacharyya, S., (2000). "Evolutionary algorithms in data mining: Multi-objective performance modeling for direct marketing". In: Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining, Boston, MA, USA, pp. 465–473.
- Bitran, G.R., Mondschein, S.V., (1996). "Mailing decisions in the catalog sales industry". *Management Science*, 42 (9), 1364–1381.
- Bodapati, A., Gupta, S., (2004). "A direct approach to predicting discretized response in target marketing". *Journal of Marketing Research*, 41 (1), 73–85.
- Bose, I., Chen, X. (2009). "Quantitative models for direct marketing: A review from systems perspective", *European Journal of Operational Research*, 195, 1–16.
- Breiman L. (1996). "Bagging predictors", *Machine Learning*, 24(2), 123–140.
- Bult, J.R., Van der Scheer, H., Wansbeek, T., (1997). "Interaction between target and mailing characteristics in direct marketing, with an application to healthcare fund raising". *International Journal of Research in Marketing*, 14 (4), 301–308.
- Bult, J.R., Wansbeek, T., (1995). "Optimal selection for direct mail". *Marketing Science*, 14 (4), 378–393.

- Bult, J.R., Wittink, D.R., (1996). "Estimating and validating asymmetric heterogeneous loss functions applied to health care fund raising". *International Journal of Research in Marketing*, 13 (3), 215–226.
- Burges, C.J.C. (1998). "A tutorial on support vector machines for pattern recognition", *Data Mining and Knowledge Discovery*, 2(2), 121–167.
- Changchien, S.W., Lee, C.-F., Hsu, Y.-J., (2004). "On-line personalized sales promotion in electronic commerce". *Expert Systems with Applications*, 27 (1), 35–52.
- Chen, M.-C., Chiu, A.-L., Chang, H.-H., (2005). "Mining changes in customer behavior in retail marketing". *Expert Systems with Applications*, 28 (4), 773–781.
- Cheung, K.-W., Kwok, J.T., Law, M.H., Tsui, K.-C. (2003), "Mining customer product ratings for personalized marketing", *Decision Support Systems*, 35(2), 231–243.
- Cohen, M.-D., (2004). "Exploiting response models – optimizing cross-sell and up-sell opportunities in banking". *Information Systems*, 29 (4), 327–341.
- Colombo, R., Jiang, W., (1999). "A stochastic RFM model". *Journal of Interactive Marketing*, 13 (3), 2–12.
- Crone S. F., Lessmann S., Stahlbock R. (2006). "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing", *European Journal of Operational Research*, 173, 781–800.
- Cui, G., Wong, M.L., Lui, H.-K. (2006). "Machine learning for direct marketing response models: Bayesian networks with evolutionary programming", *Management Science*, 52(4), 597–612.
- Desarbo, W.S., Ramaswamy, W., (1994). "CRISP: Customer response-based interactive segmentation procedures for response modeling in direct marketing". *Journal of Direct Marketing*, 8 (3), 7–20.
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V.(1997). "Support vector regression machines". In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.). *Advances in neural information processing system* (9, pp. 155–161). Cambridge, MA: MIT Press.
- Goldberg, D.E., (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing, Boston.
- Gonul, F., Kim, B.D., Shi, M.Z. (2000). "Mailing smarter to catalog customers", *Journal of Interactive Marketing*, 14(2), 2–16.
- Gonul, F., Shi, M.Z., (1998). "Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models". *Management Science*, 44 (9), 1249–1262.
- Ha, K., Cho, S., Maclachlan, D. (2005). "Response models based on bagging neural networks", *Journal of Interactive Marketing*, 19 (1), 17–30.

- Hand, D.J., Henley, W.E. (1997). "Statistical classification methods in consumer credit scoring: A review", *Journal of the Royal Statistical Society, Series A* 160 (3), 523–541.
- Hansotia, B.J., Wang, P., (1997). "Analytical challenges in customer acquisition". *Journal of Direct Marketing*, 11 (2), 7–19.
- Haughton, D., Oulabi, S. (1997). "Direct marketing modeling with CART and CHAID", *Journal of Direct Marketing*, 11(4), 42–52.
- Heilman, C.M., Kaefer, F., Ramenofsky, S.D., (2003). "Determining the appropriate amount of data for classifying consumers for direct marketing purposes". *Journal of Interactive Marketing*, 14 (3), 5–28.
- Jensen, R., Shen, Q., (2004). "Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches". *IEEE Transactions on Knowledge and Data Engineering*, 16 (12), 1457–1471.
- Jonker, J.-J., Piersma, N., Van den Poel, D., (2004). "Joint optimization of customer segmentation and marketing policy to maximize long-term profitability". *Expert Systems with Applications*, 27 (2), 159–168.
- Kaefer, F., Heilman, C.M., Ramenofsky, S.D., (2005). "A neural network application to consumer classification to improve the timing of direct marketing activities". *Computer & Operations Research*, 32 (10), 2595–2615.
- Kantardzic, M., (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. Wiley-Interscience, Hoboken.
- Kim D., Lee H-J., Cho, S. (2008), "Response modeling with support vector regression", *Expert Systems with Applications*, 34, 1102–1108.
- Kim, Y., Street, W.N., (2004). "An intelligent system for customer targeting: A data mining approach". *Decision Support Systems*, 37 (2), 215–228.
- Kim, Y., Street, W.N., Russell, G.J., Menczer, F., (2005). "Customer targeting: A neural network approach guided by genetic algorithms". *Management Science*, 51 (2), 264–276.
- Kubat, M., Holte, R. Matwin, S. (1997). "Learning when negative examples abound", Proceedings of the 9th European conference on machine learning, ECML097, Prague.
- Kwon, Y.-K., Moon, B.-R., (2001). "Personalized email marketing with a genetic programming circuit model". In: Proceedings of the Genetic and Evolutionary Computation Conference, San Francisco, CA, USA, pp. 1352–1358.
- Lang, K., (1995). "Newsweeder: Learning to filter Netnews". In: Proceedings of the Twelfth International Conference on Machine Learning, San Francisco, CA, USA, pp. 331–339.
- Levin, N., Zahavi, J., (1998). "Continuous predictive modeling – a comparative analysis". *Journal of Interactive Marketing*, 12 (2), 5–22.

- Liao, S.-H., Chen, Y.-J., (2004). "Mining customer knowledge for electronic catalog marketing". *Expert Systems with Applications*, 27 (4), 521–532.
- Li, Y., Lu, L., Li, X.F., (2005). "A hybrid collaborative filtering method for multiple interests and multiple-content recommendation in E-commerce". *Expert Systems with Applications*, 28 (1), 67–77.
- Lin, F.Y., McClean, S., (2001). "A data mining approach to the prediction of corporate failure". *Knowledge-based Systems*, 14 (3), 189–195.
- Ling, C.X., Li, C.-H., (1998). "Data mining for direct marketing: Problems and solutions". In: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, pp. 73–79.
- Malhotra, N. K., Peterson M., Bardi S. (1999). "Marketing research: A state-of-the-art review and directions for the twenty-first century", *J. Acad. Marketing Science*, 27(2), 160–183.
- Malhouse, E.C., (1999). "Ridge regression and direct marketing scoring models". *Journal of Interactive Marketing*, 13 (4), 10–23.
- Min, S.-H., Han, I., (2005). "Detection of the customer time-variant pattern for improving recommender systems". *Expert Systems with Applications*, 28 (2), 189–199.
- Mitchell, T.M., (1997). *Machine Learning*. McGraw-Hill, New York.
- Muller, K.-R., Smola, A., Ratsch, G., Scholkopf, B., Kohlmorgen, J., & Vapnik, V. (1997). "Predicting time series with support vector machines". *Lecture Notes in Computer Science*, 1327, 999–1004.
- Prinzie, A., Van den Poel, D., (2005). "Constrained optimization of data-mining problems to improve model performance: A direct-marketing application". *Expert Systems with Applications*, 29 (3), 630–640.
- Rao, V.R., Steckel, J.H., (1995). "Selecting, evaluating, and updating prospects in direct mail marketing". *Journal of Direct Marketing*, 9 (2), 20–31.
- Rissanen, J., (1978). "Modeling by shortest data description". *Automatica*, 14 (5), 465–471.
- Roberts, M. L. and Berger, P. D. (1999). *Direct marketing management-2nd ed.*, Prentice Hall, New Jersey.
- Roddy, M., (2002). "Direct Marketing: A Step-by-step Guide to Effective Planning and Targeting". Longdon, Kogan Page.
- Shardanand, U., Maes, P., (1995). "Social information filtering: Algorithms for automating "Word of Mouth" ". In: *Proceedings of the Conference on Human Factors in Computing Systems*, New York, NY, USA, pp. 210–217.
- Shaw, M.J., Subramaniam, C., Tan, G.-W., Welge, M.E., (2001). "Knowledge management and data mining for marketing". *Decision Support Systems*, 31 (1), 127–137.

- Shin, H.W., Sohn, S.Y., (2004). "Multi-attribute scoring method for mobile telecommunication subscribers". *Expert Systems with Applications*, 26 (3), 363–368.
- Smith, K.A., Gupta, J.N.D. (2000). "Neural Network in Business: Techniques and Applications for the Operations Research", *Computers & Operations Research*, 27(11–12), 1023–1044.
- Suh, E., Lim, S., Hwang, H., Kim, S., (2004). "A prediction model for the purchase probability of anonymous customers to support real time Web marketing: A case study". *Expert Systems with Applications*, 27 (2), 245–255.
- Suh, E., Noh, K.C., Suh, C.K., (1999). "Customer list segmentation using the combined response model". *Expert Systems with Applications*, 17 (2), 89–97.
- Swets JA: (1988), "Measuring the accuracy of diagnostic systems", *Science*, PubMed Abstract OpenURL, 240:1285-1293.
- Thomas, A. R., Lewison D. M., Hauser W. J., and Foley L. M. (2007). *Direct marketing in action : cutting-edge strategies for finding and keeping the best customers*, Praeger Publishers, Westport.
- Van den Poel, D., Buckinx, W., (2005). "Predicting online-purchasing behaviour". *European Journal of Operational Research*, 166 (2), 557–575.
- Van der Putten P., Ruiter M.d., Someren M.v. (2000), "CoIL Challenge 2000 Tasks and Results: Predicting and Explaining Caravan Policy Ownership", Διαθέσιμο στο: <<http://www.liacs.nl/~putten/library/cc2000/PUTTEN~1.pdf>>.
- Van der Putten P., Someren M.v., (2004). "A Bias-Variance Analysis of a Real World Learning Problem: The CoIL Challenge 2000", *Machine Learning*, 57, 177–195.
- Van der Sheer, H.R., (1998). "Quantitative approaches for profit maximization in direct marketing", PhD thesis, University of Groningen, Netherlands.
- Vapnik V.N. (1995). *The Nature of Statistical Learning Theory*, Springer, New York.
- Viaene, S., Baesens, B., Van den Poel, D., Dedene, G., Vanthiene, J., (2001a). "Wrapped input selection using multilayer perceptrons for repeat-purchase modeling in direct marketing". *International Journal of Intelligent Systems in Accounting, Finance & Management*, 10 (2), 115–126.
- Viaene, S., Baesens, B., Van Gestel, T., Suykens, J.A.K., Van den Poel, D., Vanthienen, J., De Moor, B., Dedene, G., (2001b). "Knowledge discovery in a direct marketing case using least squares support vector machines". *International Journal of Intelligent Systems*, 16 (9), 1023–1036.
- Wedel, M., Desarbo, W.S., Bult, J.R., Ramaswamy, V., (1993). "A latent class Poisson regression model for heterogeneous count data". *Journal of Applied Economics*, 8, 397–411.
- Weng, S.-S., Liu, M.-J., (2004). "Feature-based recommendation for one-to-one marketing". *Expert Systems with Applications*, 26 (4), 493–508.

- Wuensch, K. L. (2009). “Binary Logistic Regression with PASW/SPSS”, Διαθέσιμο στο: < core.ecu.edu/psyc/wuenschk/MV/MultReg/Logistic-SPSS.doc >.
- Yu, E., Cho, S. (2006). “Constructing response model using ensemble based on feature subset selection”, *Expert Systems with Applications*, 30, 352–360.
- Zahavi, J., Levin, N., (1997). “Applying neural computing to target marketing”. *Journal of Direct Marketing*, 11 (4), 76–93.

ΠΑΡΑΡΤΗΜΑ

Πίνακας Π1: Επεξήγηση Μεταβλητών

Αρ. Όνομα	Περιγραφή	Αρ. Όνομα	Περιγραφή
1 MOSTYPE	Υποκατηγορία Πελατών (L0)	44 PWAPART	Εισφορά ιδιωτικής ασφάλειας τρίτων (L4)
2 MAANTHUI	Αριθμός Σπιτιών (1-10)	45 PWABEDR	Εισφορά ασφάλειας τρίτων (εταιρίες)
3 MGEMOMV	Μ.Ο. Μεγέθους Οικογένειας (1-6)	46 PWAAND	Εισφορά ασφάλειας τρίτων (γεωργία)
4 MGEMLEEF	Μ.Ο Ηλικιών (L1)	47 PPERSAUT	Εισφορά ασφαλιστηρίου αυτοκινήτου
5 MOSHOOFD	Κατηγορία Πελατών (L2)	48 PBESAUT	Εισφορά ασφαλιστηρίου φορτηγού παράδοσης
6 MGODRK	Ρωμαιοκαθολικοί (L3)	49 PMOTSCO	Εισφορά ασφαλιστηρίου μοτοσυκλέτας
7 MGODPR	Προτεστάντες	50 PVRAAUT	Εισφορά ασφαλιστηρίου φορτηγού
8 MGODOV	Άλλη Θρησκεία	51 PAANHANG	Εισφορά ασφαλιστηρίου ρυμουλκού
9 MGODGE	Καμία Θρησκεία	52 PTRACTOR	Εισφορά ασφαλιστηρίου τρακτέρ
10 MRELGE	Παντρεμένοι	53 PWERKT	Εισφορά ασφαλιστηρίου γεωργικών μηχανημάτων
11 MRELSA	Συγκατοίκηση	54 PBROM	Εισφορά ασφαλιστηρίου μοτοποδηλάτου
12 MRELOV	Άλλου Είδους Σχέση	55 PLEVEN	Εισφορά ασφαλιστηρίου ζωής
13 MFALLEEN	Ανύπαντροι	56 PPERSONG	Εισφορά ιδιωτικού ασφαλιστηρίου ατυχήματος
14 MFGEKIND	Οικογένεια Χωρίς Παιδιά	57 PGEZONG	Εισφορά οικογενειακού ασφαλιστηρίου ατυχήματος
15 MFWEKIND	Οικογένεια Με Παιδιά	58 PWAOREG	Εισφορά ασφαλιστηρίου ανικανότητας
16 MOPLHOOG	Υψηλού επιπέδου εκπαίδευση	59 PBRAND	Εισφορά ασφαλιστηρίου φωτιάς
17 MOPLMIDD	Μεσαίου επιπέδου Εκπαίδευση	60 PZEILPL	Εισφορά ασφαλιστηρίου ιστιοσανίδας
18 MOPLLAAG	Χαμηλού επιπέδου εκπαίδευση	61 PPLEZIER	Εισφορά ασφαλιστηρίου βάρκας
19 MBERHOOG	Υψηλή θέση	62 PFIETS	Εισφορά ασφαλιστηρίου ποδηλάτου
20 MBERZELF	Επιχειρηματίας	63 PINBOED	Εισφορά ασφαλιστηρίου ιδιοκτησίας
21 MBERBOER	Αγρότης	64 PBYSTAND	Εισφορά ασφαλιστηρίου κοινωνικής ασφάλειας
22 MBERMIDD	Μεσαία στελέχη	65 AWAPART	Αριθμός ιδιωτικών ασφαλειών τρίτων (1-12)
23 MBERARBG	Ειδικευμένοι εργάτες	66 AWABEDR	Αριθμός ασφαλειών τρίτων (εταιρίες)
24 MBERARBO	Ανειδίκευτοι εργάτες	67 AWALAND	Αριθμός ασφαλειών τρίτων (γεωργία)
25 MSKA	Κοινωνική τάξη Α	68 APERSAUT	Αριθμός ασφαλιστηρίων αυτοκινήτου
26 MSKB1	Κοινωνική τάξη Β1	69 ABESAUT	Αριθμός ασφαλιστηρίων φορτηγού παράδοσης
27 MSKB2	Κοινωνική τάξη Β2	70 AMOTSCO	Αριθμός ασφαλιστηρίων μοτοσυκλέτας
28 MSKC	Κοινωνική τάξη C	71 AVRAAUT	Αριθμός ασφαλιστηρίων φορτηγού
29 MSKD	Κοινωνική τάξη D	72 AAANHANG	Αριθμός ασφαλιστηρίων ρυμουλκού
30 MHHUUR	Νοικιασμένο σπίτι	73 ATRACTOR	Αριθμός ασφαλιστηρίων τρακτέρ
31 MHKOOP	Ιδιοκτήτες σπιτιού	74 AWERKT	Αριθμός ασφαλιστηρίων γεωργικών μηχανημάτων
32 MAUT1	1 αυτοκίνητο	75 ABROM	Αριθμός ασφαλιστηρίων μοτοποδηλάτου
33 MAUT2	2 αυτοκίνητο	76 ALEVEN	Αριθμός ασφαλιστηρίων ζωής
34 MAUT0	Κανένα αυτοκίνητο	77 APERSONG	Αριθμός ιδιωτικών ασφαλιστηρίων ατυχήματος
35 MZFONDS	Εθνικό Σύστημα Υγείας	78 AGEZONG	Αριθμός οικογενειακών ασφαλιστηρίων ατυχήματος
36 MZPART	Ιδιωτική ασφάλεια υγείας	79 AWAOREG	Αριθμός ασφαλιστηρίων ανικανότητας
37 MINKM30	Εισόδημα < 30.000	80 ABRAND	Αριθμός ασφαλιστηρίων φωτιάς
38 MINK3045	Εισόδημα 30-45.000	81 AZEILPL	Αριθμός ασφαλιστηρίων ιστιοσανίδας
39 MINK4575	Εισόδημα 45-75.000	82 APLEZIER	Αριθμός ασφαλιστηρίων βάρκας
40 MINK7512	Εισόδημα 75-122.000	83 AFIETS	Αριθμός ασφαλιστηρίων ποδηλάτου
41 MINK123M	Εισόδημα >123.000	84 AINBOED	Αριθμός ασφαλιστηρίων ιδιοκτησίας
42 MINKGEM	Μέσο εισόδημα	85 ABYSTAND	Αριθμός ασφαλιστηρίων κοινωνικής ασφάλισης
43 MKOOPKLA	Κατηγορία αγοραστικής δύναμης	86 CARAVAN	Μεταβλητή Απόφασης-Ασφαλιστήριο τροχόσπιτου

L1:

- 1 20-30 χρονών
- 2 30-40 χρονών
- 3 40-50 χρονών
- 4 50-60 χρονών
- 5 60-70 χρονών
- 6 70-80 χρονών

L2:

- 1 Επιτυχημένοι επαγγελματίες
- 2 Καλλιεργητές
- 3 Μέση οικογένεια
- 4 Ανύπαντροι καριερίστες
- 5 Ζει καλά
- 6 Συνταξιούχοι ηλικιωμένοι
- 7 Συνταξιούχοι και θρησκόληπτοι
- 8 Οικογένεια με αυξημένες ανάγκες
- 9 Συντηρητικές οικογένειες
- 10 Αγρότες

L3:

- 0 0%
- 1 1 - 10%
- 2 11 - 23%
- 3 24 - 36%
- 4 37 - 49%
- 5 50 - 62%
- 6 63 - 75%
- 7 76 - 88%
- 8 89 - 99%
- 9 100%

L4:

- 0 $f^1 0$
- 1 $f 1 - 49$
- 2 $f 50 - 99$
- 3 $f 100 - 199$
- 4 $f 200 - 499$
- 5 $f 500 - 999$
- 6 $f 1000 - 4999$
- 7 $f 5000 - 9999$
- 8 $f 10.000 - 19.999$
- 9 $f 20.000 - ?$

L0:

- 1 Υψηλό εισόδημα, υψηλές δαπάνες παιδιών
- 2 Πολύ σημαντική θέση επαρχιωτών
- 3 Ανώτερη υψηλή θέση
- 4 Εύποροι ανώτεροι
- 5 Μικτοί ανώτεροι
- 6 Καριερίστας και φροντίδα παιδιών
- 7 Ευκατάστατο ζευγάρι νεαρών εργαζομένων (χωρίς παιδιά)
- 8 Οικογένειες μεσαίας τάξης
- 9 Σύγχρονες, πλήρεις οικογένειες
- 10 Σταθερές οικογένειες
- 11 Ξεκίνημα νέας οικογένειας
- 12 Εύπορες νέες οικογένειες
- 13 Νέα αμερικάνικη οικογένεια
- 14 Νεαροί κοσμοπολίτες
- 15 Γηραιότεροι κοσμοπολίτες
- 16 Φοιτητές με διαμέρισμα
- 17 Καινούργιοι στην πόλη
- 18 Ανύπαντρος νέος
- 19 Μικροαστός νέος
- 20 Διαφορετική εθνότητα
- 21 Νεαροί φτωχοί αστοί
- 22 Συγκατοίκηση
- 23 Νεαροί και ανερχόμενοι
- 24 Νεαροί, με χαμηλό επίπεδο εκπαίδευσης
- 25 Νεαροί ανώτερης τάξης
- 26 Ηλικιωμένοι με δικό τους σπίτι
- 27 Ηλικιωμένοι σε διαμερίσματα
- 28 Ηλικιωμένοι αριστοκράτες
- 29 Ηλικιωμένοι σε διαμέρισμα χωρίς μπαλκόνι/αυλή
- 30 Θρησκόληπτοι ανύπαντροι ηλικιωμένοι
- 31 Καθολικοί χαμηλού εισοδήματος
- 32 Μικτοί ανώτεροι ηλικιωμένοι
- 33 Μεγάλες οικογένειες εργατικής τάξης
- 34 Μεγάλη οικογένεια, εργαζόμενο παιδί
- 35 Οικογένειες σε χωριά
- 36 Παντρεμένοι με παιδιά
- 37 Μικτοί κάτοικοι χωριού
- 38 Παραδοσιακές οικογένειες
- 39 Θρησκόληπτες μεγάλες οικογένειες
- 40 Μεγάλες οικογένειες σε αγρόκτημα
- 41 Μικτοί αγρότες

f : Florin ή Florijn, νόμισμα της Ολλανδίας προ €

1 f = 0,45€

Πίνακας Π2: Ποσοστά βαθμίδων 60 μεταβλητών

Μεταβλητή	Βαθμίδα	Ποσοστό	Μεταβλητή	Βαθμίδα	Ποσοστό	Μεταβλητή	Βαθμίδα	Ποσοστό
1	1	89.52%	7	4	94.59%	16	0	95.76%
	2	92.68%		5	93.54%		1	94.48%
	3	89.96%		6	94.54%		2	93.88%
	4	96.15%		7	90.25%		3	92.87%
	5	95.56%		8	95.38%		4	88.65%
	6	89.92%		9	95.38%		5	88.77%
	7	93.18%	8	0	93.51%		6	88.06%
	8	84.96%		1	95.38%		7	88.24%
	9	95.68%		2	92.87%		8	90.91%
	10	94.55%		3	93.00%		9	88.89%
	11	94.12%		4	94.70%	17	0	96.69%
	12	85.59%		5	96.43%		1	94.26%
	13	92.74%	9	0	91.89%		2	94.88%
	15	100.00%		1	90.87%		3	93.98%
	16	100.00%		2	93.46%		4	93.97%
	17	100.00%		3	94.01%		5	92.68%
	18	100.00%		4	94.75%		6	94.54%
	19	100.00%		5	94.70%		7	88.54%
	20	92.00%		6	98.16%		8	91.89%
	21	100.00%		7	91.09%		9	90.70%
	22	95.92%	10	8	100.00%	18	0	90.30%
	23	98.41%		9	100.00%		1	88.89%
	24	97.22%		0	96.88%		2	90.25%
	25	97.56%		1	98.67%		3	93.09%
	26	97.92%		2	98.09%		4	94.01%
	27	98.00%		3	97.56%		5	95.14%
	28	100.00%		4	96.91%		6	96.26%
	29	97.67%		5	94.93%		7	95.78%
	30	96.61%		6	93.94%		8	94.88%
	31	97.07%		7	93.11%		9	97.52%
	32	94.33%	11	8	93.07%	19	0	95.21%
	33	94.32%		9	91.69%		1	95.10%
	34	95.05%		0	93.42%		2	94.28%
	35	96.26%		1	94.09%		3	92.99%
	36	92.89%		2	94.51%		4	91.69%
	37	92.42%		3	95.60%		5	94.78%
	38	93.22%		4	98.72%		6	86.96%
	39	94.21%		5	100.00%		7	84.78%
	40	100.00%		6	100.00%		8	88.46%
	41	97.56%		7	100.00%		9	93.55%
3	1	97.18%	12	0	92.07%	21	0	93.20%
	2	94.60%		1	93.88%		1	95.78%
	3	93.54%		2	93.74%		2	95.89%
	4	92.78%		3	93.58%		3	95.80%
	5	94.12%		4	95.99%		4	98.70%
5	1	91.30%	13	5	97.74%		5	98.31%
	2	86.85%		6	97.77%		6	100.00%
	3	93.34%		7	98.44%		7	100.00%
	4	100.00%		8	100.00%		8	100.00%
	5	97.36%		9	95.83%		9	100.00%
	6	98.05%	15	0	92.71%	22	0	94.75%
	7	96.36%		1	93.27%		1	95.29%
	8	94.31%		2	93.99%		2	94.30%
	9	93.70%		3	95.17%		3	95.12%
	10	98.19%		4	95.95%		4	93.60%
6	0	94.52%		5	94.98%		5	92.34%
	1	93.31%		6	96.85%		6	91.94%
	2	92.63%		7	98.51%		7	87.64%
	3	95.39%		8	95.83%		8	100.00%
	4	98.48%		9	100.00%		9	90.00%
	5	94.44%	23	0	96.73%		0	93.40%
	6	92.31%		1	96.58%		1	90.99%
	7	100.00%		2	93.86%		2	93.78%
	8	100.00%		3	93.94%		3	95.80%
	9	100.00%		4	94.20%		4	95.86%
7	0	98.72%		5	94.76%		5	95.81%
	1	96.27%		6	92.85%		6	96.45%
	2	93.94%		7	93.45%		7	92.65%
	3	95.08%		8	91.26%		8	87.50%
				9	92.47%		9	90.00%

Μεταβλητή	Βαθμίδα	Ποσοστό	Μεταβλητή	Βαθμίδα	Ποσοστό	Μεταβλητή	Βαθμίδα	Ποσοστό
24	0	91.94%	32	0	100.00%	42	0	100.00%
	1	92.24%		1	100.00%		1	97.96%
	2	94.30%		2	98.28%		2	96.93%
	3	95.49%		3	96.97%		3	96.43%
	4	95.08%		4	97.10%		4	92.50%
	5	95.47%		5	95.12%		5	90.45%
	6	95.08%		6	94.53%		6	93.24%
	7	96.97%		7	91.58%		7	87.02%
	8	88.89%		8	92.72%		8	88.57%
	9	100.00%		9	92.28%		9	100.00%
25	0	95.17%	34	0	91.66%	43	1	96.93%
	1	94.90%		1	93.81%		2	96.47%
	2	94.41%		2	93.35%		3	95.34%
	3	92.99%		3	95.40%		4	94.90%
	4	88.89%		4	97.79%		5	94.85%
	5	88.98%		5	97.70%		6	92.67%
	6	87.50%		6	94.38%		7	85.86%
	7	83.54%		7	100.00%		8	91.78%
	8	92.31%		8	100.00%	44	0	95.78%
	9	100.00%		9	100.00%		1	96.02%
26	0	94.68%	35	0	87.27%	46	2	91.02%
	1	94.53%		1	100.00%		3	81.82%
	2	94.34%		2	91.21%		0	93.95%
	3	91.61%		3	91.53%	47	2	100.00%
	4	93.62%		4	92.16%		3	96.49%
	5	94.87%		5	92.61%		4	98.33%
	6	80.00%		6	94.86%		0	97.47%
	7	100.00%		7	94.24%		4	100.00%
	8	87.50%		8	95.99%	49	5	97.72%
	9	100.00%		9	95.42%		6	88.70%
28	0	93.13%	36	0	95.42%		7	100.00%
	1	88.97%		1	95.99%		8	100.00%
	2	91.84%		2	94.24%	52	0	94.07%
	3	94.40%		3	94.70%		3	33.33%
	4	94.65%		4	92.74%		4	93.38%
	5	96.06%		5	92.31%		5	87.50%
	6	94.05%		6	91.57%		6	97.96%
	7	95.39%		7	91.21%		7	100.00%
	8	88.73%		8	100.00%	54	0	93.73%
	9	95.16%		9	87.27%		2	97.06%
29	0	92.79%	37	0	92.48%		3	97.87%
	1	93.73%		1	91.43%		4	100.00%
	2	95.31%		2	91.68%		5	92.86%
	3	96.83%		3	95.37%		6	88.89%
	4	97.76%		4	96.49%	55	0	93.73%
	5	99.00%		5	97.01%		2	97.06%
	6	95.45%		6	96.93%		3	97.87%
	7	92.31%		7	95.51%		4	100.00%
	9	100.00%		8	97.92%		5	93.75%
30	0	90.09%	39	9	100.00%		6	100.00%
	1	91.36%		0	95.74%	57	0	94.12%
	2	94.70%		1	96.35%		1	100.00%
	3	93.42%		2	95.88%		2	100.00%
	4	94.97%		3	93.09%		3	92.86%
	5	95.18%		4	91.49%		4	88.30%
	6	93.98%		5	91.77%		5	88.57%
	7	95.53%		6	91.20%		6	94.74%
	8	96.99%		7	92.47%		7	100.00%
	9	95.92%		8	94.34%		8	100.00%
31	0	95.92%	40	9	95.60%		9	100.00%
	1	96.98%		0	95.32%	58	0	94.09%
	2	95.54%		1	92.86%		2	92.00%
	3	93.98%		2	92.12%		3	69.23%
	4	94.99%		3	92.28%		0	94.07%
	5	95.00%		4	90.48%		4	100.00%
	6	93.54%		5	90.14%		5	100.00%
	7	94.75%		6	100.00%		6	78.95%
	8	91.36%		7	100.00%		7	100.00%
	9	90.09%		8	100.00%			
				9	75.00%			

Μεταβλητή	Βαθμίδα	Ποσοστό
59	0	95.91%
	1	98.14%
	2	98.88%
	3	92.61%
	4	87.68%
	5	94.63%
	6	98.06%
	7	100.00%
	8	100.00%
61	0	94.21%
	1	40.00%
	2	60.00%
	3	60.00%
	4	69.23%
	5	100.00%
	6	33.33%
62	0	94.13%
	1	89.80%
64	0	94.22%
	2	73.33%
	3	81.82%
	4	81.82%
	5	100.00%
65	0	95.78%
	1	91.39%
	2	100.00%
67	0	93.95%
	1	97.50%
68	0	97.47%
	1	91.26%
	2	84.55%
	3	91.67%
	4	100.00%
	6	100.00%
	7	100.00%
75	0	93.73%
	1	97.91%
	2	100.00%
76	0	94.12%
	1	95.38%
	2	90.00%
	3	81.82%
	4	62.50%
	8	100.00%
78	0	94.09%
	1	84.21%
79	0	94.07%
	1	78.95%
	2	100.00%
80	0	95.91%
	1	92.31%
	2	94.44%
	3	100.00%
	4	100.00%
	5	100.00%
	7	100.00%
82	0	94.21%
	1	61.29%
	2	50.00%
83	0	94.13%
	1	90.99%
	2	88.24%
	3	50.00%
84	0	94.06%
	1	88.64%
	2	100.00%
85	0	94.22%
	1	80.25%
	2	100.00%

Πίνακας Π3: Τελική Κωδικοποίηση

Μεταβλητή	Νέα Βαθμίδα	Μεταβλητή	Νέα Βαθμίδα	Μεταβλητή	Νέα Βαθμίδα
1	1	12	0	22	0
	2		1		1
	3		2		2
	4-30-31-35		3		3
	5-9-22-34		4		4
	6		5-6-7-8-9		5
	7-38	13	0		6
	8		1		7
	10		2		8-9
	11		3	23	0
	12		4		1
	13		5		2
	15-16-17-18-19- 21-23-24-25-26- 27-28-29-40-41		6-7-8-9		3
	20	15	0		4
	32		1		5
	33		2		6
	36		3		7
	37		4		8-9
	39		5	24	0
3	1		6		1
	2		7		2
	3		8		3
	4-5		9		4
5	1	16	0		5
	2		1		6-7-8-9
	3		2	25	0
	4-5-6-10		3		1
	7		4		2
	8		5		3
	9		6		4
6	0		7-8-9		5
	1	17	0		6
	2		1		7-8-9
	3-4-5-6-7-8-9		2	26	0
7	0-1		3		1
	2		4		2
	3		5		3
	4		6		4
	5		7		5-6-7-8-9
	6		8-9	28	0
	7	18	0		1
	8-9		1		2
8	0		2		3
	1		3		4
	2		4		5
	3		5		6
	4-5		6		7
9	0		7		8
	1		8		9
	2		9	29	0
	3	19	0		1
	4		1		2
	5		2		3
	6-7-8-9		3		4-5-6-7-8-9
10	0-1-2		4	30	0
	3		5		1
	4		6		2
	5		7		3
	6		8-9		4
	7	21	0		5
	8		1		6
	9		2		7
11	0		3-4-5-6-7-8-9		8
	1				9
	2				
	3-4-5-6-7				

Μεταβλητή	Νέα Βαθμίδα	Μεταβλητή	Νέα Βαθμίδα	Μεταβλητή	Νέα Βαθμίδα
31	0	39	0-1-2	82	0
	1		3		1-2
	2		4	83	0
	3		5		1
	4		6		2-3
	5		7	84	0
	6		1		1-2
	7	43	2	85	0
	8		3		1-2
	9		4		
32	0-1-2-3		5		
	4		6		
	5		7		
	6		8		
	7		0-1		
	8		2-3		
	9		0		
34	0		2-3-4		
	1	47	0-4		
	2		5		
	3	49	6-7-8		
	4		0-3		
35	5-6-7-8-9		4		
	0-1		5-6-7		
	2	52	0		
	3		3-4-5-6		
	4	54	0		
	5		2-3-4-5-6		
	6	55	0-1-2		
	7		3		
	8		4		
	9		5-6-7-8-9		
36	0	57	0		
	1		2-3		
	2	58	0		
	3		4-5-6-7		
	4	59	0		
	5		1-2		
	6		3		
	7		4		
	8-9		5-6-7-8		
37	0	61	0		
	1		1-2-3-4-5-6		
	2	62	0		
	3		1		
	4	64	0		
	5		2-3-4-5		
	6	65	0		
39	7-8-9		1-2		
	0	67	0		
	1		1		
	2	68	0		
	3		1		
	4		2-3-4-5-6-7		
	5	75	0		
	6		1-2		
	7	76	0-1		
	8-9		2		
40	0		3-4-8		
	1	78	0		
	2		1		
	3	79	0		
	4		1-2		
	5-6-7-8-9	80	0		
			1		
			2-3-4-5-7		

Πίνακας Π4: Μεταβλητές BINLOG27

Μεταβλητή	B	S.E.	Sig.	Δ.Ε. -	Δ.Ε. +	Μεταβλητή	B	S.E.	Sig.	Δ.Ε. -	Δ.Ε. +
newv1			,000			newv18			,000		
newv1(1)	,462	,374	,216	-0,501	1,426	newv18(1)	,145	,244	,552	-0,484	0,774
newv1(2)	1,021	,293	,000	0,266	1,776	newv18(2)	,425	,216	,048	-0,130	0,980
newv1(3)	-,272	,415	,512	-1,340	0,797	newv18(3)	,318	,225	,157	-0,261	0,898
newv1(4)	-,033	,415	,937	-1,103	1,037	newv18(4)	-,038	,230	,869	-0,630	0,554
newv1(5)	1,653	,521	,002	0,310	2,995	newv18(5)	,008	,239	,972	-0,607	0,624
newv1(6)	1,250	,303	,000	0,469	2,031	newv18(6)	,215	,257	,403	-0,446	0,876
newv1(7)	2,456	,508	,000	1,149	3,764	newv18(7)	,212	,272	,434	-0,488	0,913
newv1(8)	-,834	,485	,085	-2,082	0,414	newv18(8)	,716	,320	,025	-0,108	1,539
newv1(9)	-,251	,504	,618	-1,550	1,047	newv18(9)	-,969	,331	,003	-1,822	-0,115
newv1(10)	,732	,496	,140	-0,545	2,009	newv19			,000		
newv1(11)	-,371	,481	,441	-1,611	0,869	newv19(1)	-,201	,139	,148	-0,561	0,158
newv1(12)	-,650	,467	,164	-1,853	0,553	newv19(2)	-,005	,139	,970	-0,364	0,353
newv1(13)	,860	,483	,075	-0,383	2,103	newv19(3)	-,403	,168	,017	-0,836	0,031
newv1(14)	,654	,499	,190	-0,632	1,939	newv19(4)	,153	,195	,432	-0,349	0,656
newv1(15)	,610	,447	,172	-0,540	1,761	newv19(5)	-,711	,256	,005	-1,370	-0,052
newv1(16)	1,004	,467	,032	-0,199	2,207	newv19(6)	,270	,301	,369	-0,505	1,045
newv1(17)	,724	,305	,017	-0,060	1,509	newv19(7)	,492	,327	,132	-0,349	1,334
newv5			,001			newv19(8)	,076	,409	,854	-0,979	1,130
newv5(1)	-1,486	,420	,000	-2,567	-0,405	newv22			,000		
newv5(2)	,588	,371	,113	-0,367	1,543	newv22(1)	-,674	,226	,003	-1,257	-0,091
newv5(3)	-,045	,389	,908	-1,045	0,956	newv22(2)	-,396	,181	,028	-0,861	0,069
newv5(4)	,465	,375	,214	-0,500	1,430	newv22(3)	-,720	,196	,000	-1,224	-0,217
newv5(5)	,050	,351	,886	-0,853	0,953	newv22(4)	-,140	,201	,487	-0,659	0,379
newv6			,000			newv22(5)	-,560	,231	,015	-1,155	0,035
newv6(1)	,357	,111	,001	0,071	0,643	newv22(6)	-,321	,283	,257	-1,051	0,409
newv6(2)	,098	,126	,434	-0,225	0,422	newv22(7)	-,715	,320	,025	-1,538	0,109
newv6(3)	-,688	,225	,002	-1,269	-0,108	newv22(8)	-,130	,381	,732	-1,111	0,850
newv7			,000			newv23			,000		
newv7(1)	1,761	,313	,000	0,955	2,566	newv23(1)	,860	,142	,000	0,495	1,225
newv7(2)	1,794	,298	,000	1,026	2,561	newv23(2)	,356	,142	,012	-0,010	0,723
newv7(3)	1,963	,292	,000	1,212	2,714	newv23(3)	-,068	,157	,664	-0,474	0,337
newv7(4)	1,961	,294	,000	1,205	2,718	newv23(4)	-,063	,178	,725	-0,521	0,396
newv7(5)	1,651	,299	,000	0,880	2,422	newv23(5)	,331	,216	,125	-0,225	0,888
newv7(6)	2,683	,311	,000	1,881	3,484	newv23(6)	-,619	,322	,055	-1,449	0,211
newv7(7)	1,658	,357	,000	0,739	2,577	newv23(7)	2,059	,361	,000	1,129	2,988
newv8			,000			newv23(8)	1,546	,434	,000	0,428	2,664
newv8(1)	-,186	,109	,089	-0,468	0,096	newv26			,000		
newv8(2)	,256	,113	,023	-0,035	0,547	newv26(1)	,093	,144	,520	-0,278	0,463
newv8(3)	,835	,191	,000	0,341	1,328	newv26(2)	-,407	,140	,004	-0,768	-0,047
newv8(4)	,059	,288	,837	-0,682	0,801	newv26(3)	,002	,163	,990	-0,418	0,422
newv15			,000			newv26(4)	,086	,211	,684	-0,458	0,630
newv15(1)	-,159	,351	,650	-1,063	0,744	newv26(5)	,454	,272	,095	-0,246	1,154
newv15(2)	,526	,300	,080	-0,247	1,298	newv29			,000		
newv15(3)	,137	,305	,653	-0,647	0,921	newv29(1)	,049	,112	,664	-0,241	0,339
newv15(4)	,042	,310	,892	-0,756	0,839	newv29(2)	-,008	,135	,955	-0,355	0,340
newv15(5)	-,092	,311	,768	-0,894	0,710	newv29(3)	-,846	,191	,000	-1,338	-0,354
newv15(6)	,084	,315	,790	-0,727	0,894	newv29(4)	-,884	,249	,000	-1,526	-0,242
newv15(7)	-,197	,339	,561	-1,071	0,676	newv31			,001		
newv15(8)	,499	,357	,162	-0,420	1,417	newv31(1)	-,280	,198	,158	-0,792	0,231
newv15(9)	,903	,369	,014	-0,047	1,853	newv31(2)	-,449	,197	,023	-0,956	0,058
newv16			,000			newv31(3)	-,261	,200	,191	-0,776	0,253
newv16(1)	,632	,126	,000	0,306	0,957	newv31(4)	-,456	,192	,018	-0,951	0,039
newv16(2)	,149	,134	,264	-0,195	0,494	newv31(5)	-,520	,187	,005	-1,003	-0,038
newv16(3)	-,251	,175	,152	-0,703	0,201	newv31(6)	-,225	,189	,235	-0,713	0,263
newv16(4)	,370	,195	,058	-0,133	0,873	newv31(7)	-,767	,188	,000	-1,251	-0,282
newv16(5)	1,327	,250	,000	0,683	1,970	newv31(8)	-,167	,205	,413	-0,694	0,360
newv16(6)	,403	,382	,292	-0,582	1,387	newv31(9)	-,192	,187	,304	-0,674	0,290

newv32			,000		
newv32(1)	,247	,298	,406	-0,520	1,014
newv32(2)	1,225	,311	,000	0,423	2,027
newv32(3)	1,042	,318	,001	0,223	1,861
newv32(4)	1,402	,321	,000	0,574	2,231
newv32(5)	1,505	,366	,000	0,561	2,448
newv32(6)	,775	,345	,025	-0,115	1,664
newv34			,000		
newv34(1)	-,407	,160	,011	-0,818	0,005
newv34(2)	-,087	,133	,513	-0,430	0,256
newv34(3)	,342	,177	,054	-0,115	0,798
newv34(4)	-,791	,232	,001	-1,389	-0,194
newv34(5)	,773	,330	,019	-0,076	1,622
newv37			,000		
newv37(1)	,317	,146	,030	-0,060	0,693
newv37(2)	,660	,129	,000	0,329	0,992
newv37(3)	-,149	,156	,339	-0,550	0,252
newv37(4)	-,449	,190	,018	-0,939	0,040
newv37(5)	,048	,210	,820	-0,493	0,588
newv37(6)	,481	,280	,085	-0,239	1,202
newv37(7)	,206	,322	,523	-0,624	1,037
newv42			,000		
newv42(1)	1,180	,579	,041	-0,311	2,671
newv42(2)	1,070	,571	,061	-0,401	2,540
newv42(3)	1,502	,572	,009	0,028	2,976
newv42(4)	1,428	,574	,013	-0,051	2,907
newv42(5)	,764	,590	,195	-0,755	2,284
newv42(6)	2,129	,612	,000	0,554	3,704
newv42(7)	1,263	,635	,047	-0,374	2,900
newv44(1)	,469	,091	,000	0,235	0,702
newv47			,000		
newv47(1)	-,189	,151	,211	-0,578	0,200
newv47(2)	1,656	,082	,000	1,445	1,867
newv54(1)	-,861	,198	,000	-1,370	-0,352
newv58(1)	2,241	,462	,000	1,051	3,432
newv59			,000		
newv59(1)	-1,361	,186	,000	-1,840	-0,883
newv59(2)	,659	,107	,000	0,383	0,935
newv59(3)	,571	,108	,000	0,294	0,849
newv59(4)	-,455	,190	,017	-0,946	0,036
newv61(1)	3,559	,385	,000	2,567	4,551
newv62(1)	1,641	,610	,007	0,071	3,212
newv76	,164	,096	,090	-0,085	0,412
newv83	-,545	,440	,216	-1,678	0,589

Μεταβλητή	Ένωση	Μεταβλητή	Ένωση	Μεταβλητή	Ένωση
newv1	2-16	newv16	1 έως 4	newv29	1-2
	8-12		6-7		3-4
	3-4-9-11	newv18	1 έως 8	newv31	1 έως 5
	1-10-13-14-15-17	newv19	1 έως 4		6 έως 9
newv5	2-4		5 έως 8	newv32	2 έως 5
	3-5	newv22	1 έως 6	newv34	1-2
newv7	5-6-7	newv23	1 έως 6	newv37	1 έως 5
newv15	1 έως 8		7-8	newv42	1 έως 5
		newv26	1 έως 4	newv59	2-3

Πίνακας ενοποιημένων ψευδομεταβλητών.

(Αναφέρονται μόνο οι μεταβλητές στις οποίες έγινε επεξεργασία)