



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΕΞΕΛΙΚΤΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ ΓΙΑ ΤΗΝ ΑΝΑΠΤΥΞΗ ΜΟΝΤΕΛΩΝ  
ΠΡΟΒΛΕΨΗΣ ΤΗΣ ΑΠΩΛΕΙΑΣ ΠΕΛΑΤΩΝ

ΟΝΟΜΑ ΦΟΙΤΗΤΗ: ΚΛΑΔΑΣ ΙΑΣΩΝ

ΑΜ: 2013010102

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΜΙΧΑΗΛ ΔΟΥΜΠΟΣ

## Πίνακας περιεχομένων

Εισαγωγή .....	4
1. Διαχείριση πελατειακών σχέσεων .....	5
1.1 Ιστορική αναδρομή .....	6
1.2 Προσανατολισμός της διαχείρισης πελατειακών σχέσεων .....	7
1.3 Συστήματα CRM.....	8
1.3.1 Επιχειρησιακό CRM.....	9
1.3.2 Συνεργατικό CRM.....	12
1.3.3 Αναλυτικό CRM.....	12
1.4 Κύκλος ζωής του πελάτη .....	15
1.4.1 Απόκτηση πελατών.....	16
1.4.2 Ενεργοποίηση πελατών .....	17
1.4.3 Διαχείριση πελατειακών σχέσεων .....	18
1.4.4 Ανάκτηση πελατών.....	20
1.5 Απώλεια πελατών.....	21
1.5.1 Πρόβλεψη απώλειας πελατών .....	22
1.6 Παραδείγματα εφαρμογών πρόβλεψης απώλειας πελατών .....	24
2. Γενετικοί αλγόριθμοί .....	26
2.1 Δαρβινική θεωρία της εξέλιξης.....	26
2.2 Ιστορική αναδρομή .....	27
2.3 Αρχικοποίηση.....	28
2.3.1 Κωδικοποίηση .....	28
2.3.2 Αντικειμενική συνάρτηση .....	29
2.4 Λειτουργική δομή.....	30
2.4.1 Αρχικοποίηση πληθυσμού.....	30
2.4.2 Επιλογή.....	31
2.4.3 Διασταύρωση.....	31

2.4.4 Μετάλλαξη .....	32
2.4.5 Επανατοποθέτηση.....	32
2.5 Πλεονεκτήματα γενετικών αλγορίθμων.....	33
3. Υλοποίηση του γενετικού αλγορίθμου .....	35
3.1 Δεδομένα.....	35
3.2 Προσαρμογή των δεδομένων .....	38
3.3 Δημιουργία της αντικειμενικής συνάρτησης .....	40
3.4 Επιλογή διαθέσιμων παραμέτρων γενετικού αλγόριθμου .....	42
3.5 Εκτέλεση του γενετικού αλγορίθμου .....	44
3.6 Αποτελέσματα γενετικού αλγορίθμου .....	45
3.7 Λογιστική παλινδρόμηση.....	47
4. Συμπεράσματα .....	49
Βιβλιογραφία .....	<b>Error! Bookmark not defined.</b>

## Εισαγωγή

Η διαχείριση των πελατών είναι ο τρόπος με τον οποίο οι εταιρείες αντιμετωπίζουν τις σχέσεις που διαθέτουν με τους πελάτες. Με την πάροδο του χρόνου παρατηρείται πως όλο και περισσότεροι οργανισμοί επενδύουν στη διαχείριση των πελατειακών τους σχέσεων. Πέρα από τον προφανή στόχο που είναι η ικανοποίηση των πελατών μέσω της διαχείρισης αυτής, οι επιχειρήσεις δύνανται να επιτύχουν μεγαλύτερη κερδοφορία χρησιμοποιώντας τις διαθέσιμες τεχνικές. Μια τεχνική που οι εταιρείες χρησιμοποιούν είναι η πρόβλεψη της καταναλωτικής εμπιστοσύνης. Είναι ουσιαστικά η πρόβλεψη της απώλειας των πελατών με απώτερο σκοπό την αποφυγή τέτοιων καταστάσεων που θα ζημιώσουν τον οργανισμό.

Η εν λόγω διπλωματική εργασία αναφέρεται σε θέματα διαχείρισης πελατειακών σχέσεων, τα συστήματα της διαχείρισης αυτής καθώς και τις τεχνικές που δύνανται να επιφέρουν στις επιχειρήσεις το μεγαλύτερο κέρδος. Ωστόσο, η ανάλυση αυτή είναι το πρώτο μέρος της εργασίας.

Η δεύτερη και κυριότερη εννοιολογική ενότητα της παρούσας εργασίας είναι η εφαρμογή εξελικτικού αλγορίθμου για την πρόβλεψη της απώλειας των πελατών που αφορούν εταιρείες τηλεπικοινωνιών με έδρα τις Η.Π.Α. Πιο συγκεκριμένα, υλοποιείται ο απλός γενετικός αλγόριθμος στο περιβάλλον της Matlab με απώτερο σκοπό την πρόβλεψη των δέκα τοις εκατό των πιο πιθανών πελατών να εγκαταλείψουν την επιχείρηση. Στο δεύτερο κεφάλαιο αναλύεται το θεωρητικό υπόβαθρο του γενετικού αλγόριθμου.

Τέλος, το τρίτο και τελευταίο κεφάλαιο, που αφορά την υλοποίηση του αλγορίθμου, παρουσιάζονται αναλυτικότερα τα δεδομένα και οι μεταβλητές του προβλήματος καθώς και οι παράμετροι του αλγορίθμου. Στην τελευταία ενότητα του κεφαλαίου αναγράφονται τα συμπεράσματα που εξάχθηκαν από την εν λόγω εφαρμογή.

## 1. Διαχείριση πελατειακών σχέσεων

Είναι προφανές το επίκεντρο της προσοχής σε μια επιχείρηση είναι ο πελάτης. Είναι ο σημαντικότερος παράγοντας που χαρακτηρίζει αν η εταιρεία είναι επικερδής ή όχι. Κατά συνέπεια ένας από τους κύριους στόχους μιας επιχείρησης είναι να διαχειριστεί αυτή την πραγματικότητα με τον αποδοτικότερο τρόπο. Ο συνδυασμός της παραπάνω αλήθειας με το γεγονός πως ο παγκόσμιος ανταγωνισμός γίνεται ολοένα και μεγαλύτερος οδηγεί τις επιχειρήσεις να αναζητούν νέους τρόπους για να διαχειριστούν τις πελατειακές τους σχέσεις. Παύουν δηλαδή να εστιάζουν στο μαζικό μάρκετινγκ που αφορά το προϊόν και επικεντρώνονται στο μάρκετινγκ που είναι προσαρμοσμένο στον πελάτη (Lu, Lin, Lu, & Zhang, 2012).

Οι στρατηγικές αυτές και οι αναζητήσεις τους ανήκουν σε ένα ευρύτερο σύνολο που ονομάζεται διαχείριση πελατειακών σχέσεων (CRM, customer relationship management) ή πιο απλά διαχείριση πελατών. Πιο συγκεκριμένα, είναι μια στρατηγική προσέγγιση των επιχειρήσεων για την διαχείριση των σχέσεων με τους πελάτες τους χρησιμοποιώντας την γνώση που έχουν για τους ίδιους. Το CRM παρέχει την δυνατότητα κατανόησης του πελάτη χρησιμοποιώντας τις διαθέσιμες πληροφορίες. Με την πάροδο του χρόνου, όπως είναι φυσικό, τα τεχνολογικά συστήματα εξελίσσονται γεγονός που οδηγεί τις διαθέσιμες πληροφορίες να πληθαίνουν και ταυτόχρονα η μελέτη τους να γίνεται ευκολότερη. Η αξία της διαχείρισης των πελατειακών σχέσεων βασίζεται στην ικανότητα της να βοηθά τις επιχειρήσεις να κατανοήσουν καλύτερα τους καταναλωτές της. Οργανισμοί που διαχειρίζονται τις σχέσεις αυτές αποδοτικά είναι πιο ανταγωνιστικοί καθώς μπορούν να προσελκύσουν νέους πελάτες, να διατηρήσουν τους ήδη επικερδείς, να διαχειριστούν τους μη επικερδείς, να προσαρμόσουν τις προσφορές και τις προωθητικές τους ενέργειες αποδοτικότερα (Kourdi, 2011).

Η βιβλιογραφία διαθέτει αρκετούς ορισμούς για τον όρο διαχείριση πελατειακών σχέσεων. Παρακάτω παρατίθενται ενδεικτικά μερικοί από αυτούς για καλύτερη κατανόηση.

- Η διαχείριση πελατειακών σχέσεων αναφέρεται σε πρακτικές συλλογής, αποθήκευσης και ανάλυσης πληροφοριών του πελάτη και ενσωμάτωση των

αποτελεσμάτων σε διαδικασίες απόφασης της εταιρείας (Kumar & Petersen, 2012).

- Η διαχείριση πελατειακών σχέσεων είναι μια επιχειρηματική στρατηγική προσαρμοσμένη κάθε φορά σε συγκεκριμένο τμήμα της αγοράς με στόχο την αύξηση των κερδών της επιχείρησης μέσω της αναζήτησης, του εντοπισμού, της κατανόησης, της πρόβλεψης και της διαχείρισης των αναγκών και προτιμήσεων των πελατών της (σημερινών και μελλοντικών) (Αβραμούλη, Καραγεώργος, & Ράπτη, 2015).
- Η διαχείριση πελατειακών σχέσεων ορίζεται ως μια επιχειρηματική προσέγγιση για να κατανοηθεί και επηρεαστεί η καταναλωτική συμπεριφορά μέσω ουσιωδών επικοινωνιών με στόχο να βελτιωθεί η απόκτηση, διατήρηση, εμπιστοσύνη και η κερδοφορία του πελάτη (Swift, 2001).

## 1.1 Ιστορική αναδρομή

Η διαχείριση πελατειακών σχέσεων εμφανίστηκε την δεκαετία του 1980 ως μια στρατηγική αντιμετώπισης των ενδοεπιχειρησιακών σχέσεων και των σχέσεων των εταιρειών με τους πελάτες τους. Αρχικά αναπτύχθηκε για την αποτελεσματικότερη διαχείριση των υπάρχοντων πελατών και την προσέλκυση νέων. Την ίδια περίοδο, η ραγδαία εξέλιξη της τεχνολογίας επέτρεψε σταδιακά την αξιοποίηση μεγαλύτερου όγκου πληροφοριών. Η δημοτικότητα του όρου εκτοξεύθηκε την δεκαετία του 1990 καθώς οι επιχειρήσεις εστίασαν περισσότερο στον πελάτη. Τα αποτελέσματα της διατήρησης μακροχρόνιων και έμπιστων σχέσεων με τον καταναλωτή είναι η ρευστότητα και σταθερότητα για την εταιρεία (Kourdi, 2011).

Σύμφωνα με το βιβλίο Συστήματα Διαχείρισης Πελατειακών Σχέσεων (Αβραμούλη, Καραγεώργος, & Ράπτη, 2015) από την αρχή της δεκαετίας του 1990 η ιστορική εξέλιξη του CRM περιλαμβάνει τρεις γενιές.

- Η πρώτη γενιά, γνωστή ως «Λειτουργική Προσέγγιση» (1990 – 1996). Αφορούσε την αυτοματοποίηση των πωλήσεων και υποστήριξη της εξυπηρέτησης πελατών. Για παράδειγμα τήλε-μάρκετινγκ και τηλεφωνικά κέντρα.

- Η δεύτερη γενιά, γνωστή και ως «Front-end Λειτουργίες» (1996 – 2002), είχε ως στόχο να δημιουργήσει μια κοινή οπτική όλων των συναλλαγών των πελατών ανεξαρτήτως του λόγου και του μέσου της επικοινωνίας. Η εμφάνιση του διαδικτύου δημιούργησε υψηλές προσδοκίες αλλά είχε γίνει προφανές ότι οι προσδοκίες αυτές θα πραγματοποιούνταν μόνο εφόσον όλη η διαδικασία του CRM αποτελούσε τη βασική στρατηγική επιλογή.
- Κατά την τρίτη γενιά γνωστή και σαν «Στρατηγική Προσέγγιση» (2002 – σήμερα) οι εταιρείες συνειδητοποίησαν ότι το CRM πρέπει να αποτελεί κύρια στρατηγική επιλογή. Είχε γίνει αντιληπτή η διαφορά μεταξύ της προσδοκώμενης και της αντιλαμβανόμενης αξίας των πελατών. Οι εταιρείες εστίαζαν στη σύνδεση των συστημάτων εξυπηρέτησης πελατών της «πρώτης γραμμής» (front-end) και των «πίσω» τμημάτων (back-end). Η ανάπτυξη του διαδικτύου βοήθησε στην ενδυνάμωση του CRM. Οι εταιρείες αντιλήφθηκαν ότι σκοπός του CRM ήταν η δημιουργία εσόδων και όχι μόνο ο έλεγχος του κόστους.

## 1.2 Προσανατολισμός της διαχείρισης πελατειακών σχέσεων

Όπως αναφέρθηκε, η διαχείριση πελατειακών σχέσεων αποσκοπεί στην ανάπτυξη σχέσεων μεταξύ της επιχείρησης και του πελάτη διαμορφώνοντας μια επιχειρηματική προσέγγιση που είναι εστιασμένη στον πελάτη. Η «κουλτούρα» αυτή είναι αφοσιωμένη στο να κερδίζει νέους πελάτες αλλά και τους διατηρεί δημιουργώντας και προσφέροντας καλύτερα προϊόντα ή υπηρεσίες. Ο πελατειακός προσανατολισμός αυτός ανταγωνίζεται άλλες επιχειρησιακές λογικές. Αναγνωρίζονται τρεις κύριοι προσανατολισμοί: του προϊόντος, της παραγωγής και των πωλήσεων (Buttle, 2004).

Οι εταιρείες που υιοθετούν τον προσανατολισμό του προϊόντος θεωρούν πως οι πελάτες επιλέγουν τα προϊόντα που θα αγοράσουν σύμφωνα με κριτήρια καλύτερης ποιότητας, απόδοσης ή σχεδιασμού. Αυτές συνήθως είναι αρκετά καινοτόμοι οργανισμοί και το ποσοστό τους μικρό στην παγκόσμια αγορά.

Αντίθετα ο προσανατολισμός της παραγωγής πιστεύει πως οι καταναλωτές επιλέγουν με το κριτήριο του χαμηλού κόστους. Κατά συνέπεια οι ίδιες επιλέγουν να διατηρούν το κόστος χαμηλό. Ωστόσο σε αγορές που ο ανταγωνισμός είναι μεγάλος οδηγεί τους καταναλωτές να έχουν περισσότερες απαιτήσεις και το κριτήριο αυτό δεν αρκεί.

Τέλος ο προσανατολισμός των πωλήσεων που ενστερνίζονται ορισμένοι οργανισμοί τους οδηγεί στην παραδοχή πως αν επενδύσουν σε δημόσιες σχέσεις, διαφημίσεις προωθητικές ενέργειες και παρόμοιες τεχνικές θα προσελκύσουν πελάτες. Πολύ συχνά οι οργανισμοί αυτοί υιοθετούν και τον προσανατολισμό της παραγωγής, δηλαδή παράγουν τα προϊόντα τους σε χαμηλό κόστος και στην συνέχεια επενδύουν σε ενέργειες διαφήμισης.

Μια επιχείρηση που είναι εστιασμένη στον πελάτη μαζεύει και χρησιμοποιεί την πληροφορία των πελατών και των ανταγωνιστών με στόχο την ανάπτυξη καλύτερων και καταλληλότερων παροχών.

### 1.3 Συστήματα CRM

Τα τελευταία χρόνια, έχουν διαμορφωθεί διάφοροι τύποι διαχείρισης των πελατειακών σχέσεων με βάση τις διαφοροποιήσεις που εμφανίζουν επιχειρήσεις διαφόρων κλάδων στον τρόπο λειτουργίας τους. Οι εφαρμογές της διαχείρισης συνήθως αποτελούνται από διάφορα υποσυστήματα που περιλαμβάνουν όλα όσα χρειάζονται προκειμένου να εξυπηρετήσουν τις ανάγκες της επιχείρησης. Μια συνοπτική διάκριση των υποσυστημάτων αυτών συνίσταται σε τρεις κατηγορίες: Επιχειρησιακό CRM, αναλυτικό CRM και συνεργατικό CRM.

1. Το επιχειρησιακό σύστημα στοχεύει στην αυτοματοποίηση των διαδικασιών της διαχείρισης πελατών ώστε να βελτιώσει την αποτελεσματικότητα και την παραγωγικότητα. Η εξυπηρέτηση πελατών, η αυτοματοποίηση των πωλήσεων όπως τα POS (points of sale) και το αυτοματοποιημένο μάρκετινγκ όπως στοχευμένες διαφημίσεις στο διαδίκτυο, ανήκουν σε αυτή την κατηγορία (Khodakarami & Chan, 2014). Ουσιαστικά είναι το «τμήμα» το οποίο επικοινωνεί με τους πελάτες και γίνεται ανταλλαγή πληροφοριών από την εταιρεία προς τον πελάτη και αντίστροφα.



2. Το συνεργατικό σύστημα ασχολείται με την προσέλκυση πελατών εκμεταλλευόμενο διαύλους επικοινωνίας μεταξύ των εταιρειών και των πελατών, όπως διαδικτυακά καταστήματα και κέντρα τηλεφώνων (Mishra & Mishra, 2009).
3. Το αναλυτικό σύστημα παρέχει καλύτερη κατανόηση στις ανάγκες και συνήθειες των διαφορετικών πελατών. Δημιουργούνται μοντέλα πρόβλεψης προτιμήσεων και αναγνώρισης προτύπων αποσαφηνίζοντας τη καταναλωτική συμπεριφορά (Khodakarami & Chan, 2014). Πιο απλά χρησιμοποιούνται στατιστικά μοντέλα με σκοπό να δημιουργηθούν απτές πληροφορίες για τους καταναλωτές διευκολύνοντας το επιχειρησιακό σύστημα.

Τα συστήματα αυτά, δηλαδή, έχουν ως στόχο την απόκτηση και την παραγωγή γνώσης που αφορά τον πελάτη.

### 1.3.1 Επιχειρησιακό CRM

Το επιχειρησιακό σύστημα του CRM διαχειρίζεται το σχεδιασμό, προγραμματισμό και την εκτέλεση των επιχειρησιακών δραστηριοτήτων προσφέροντας εργαλεία για την υποστήριξη των εργασιών σε τομείς του μάρκετινγκ, πωλήσεων και υπηρεσιών. Τα παραπάνω επιτυγχάνονται αυτοματοποιώντας όλες τις διεργασίες που συνδέονται με τις σχέσεις της επιχείρησης και του πελάτη. Κύριος στόχος είναι η υποστήριξη των πελατών διαθέτοντας πληροφορίες από τη βάση δεδομένων του CRM ή των ενδοεπιχειρησιακών αποφάσεων. Στις περισσότερες περιπτώσεις το επιχειρησιακό και αναλυτικό σύστημα διαχείρισης πελατειακών σχέσεων μοιράζονται πληροφορίες των πελατών μεταξύ τους (Torggler, 2008). Παρακάτω αναπτύσσονται οι κύριες αρμοδιότητες επιχειρησιακού συστήματος: αυτοματοποίηση μάρκετινγκ, αυτοματοποίηση πωλήσεων και αυτοματοποίηση υπηρεσιών.

Η αυτοματοποίηση του μάρκετινγκ είναι υπεύθυνη για τη διαχείριση και υποστήριξη όλων των ενεργειών που αφορούν τον πελάτη. Η κυριότερη ενασχόληση του «τμήματος» αυτού είναι η αυτοματοποίηση του προγραμματισμού της κάθε καμπάνιας, όπου θα εστιάζει στην παροχή επαρκών πληροφοριών στον πελάτη στη σωστή στιγμή. Έτσι επιτυγχάνεται μια διαρκής αλληλεπίδραση με τον καταναλωτή.

Όσον αφορά την καμπάνια/εκστρατεία μιας επιχείρησης θα πρέπει το αντίστοιχο υποσύστημα CRM να είναι ικανό να την σχεδιάζει, εκτελεί και ελέγχει.

- Σχεδιασμός εκστρατείας: Ένα σύστημα διαχείρισης πελατειακών σχέσεων οφείλει να παρέχει υποστήριξη στο προγραμματισμό των ενεργειών του μάρκετινγκ (π.χ. στόχους, ημερομηνίες, κλπ.) και στην προετοιμασία της ίδιας της εκστρατείας (π.χ. οικονομικό πλάνο).
- Εκτέλεση εκστρατείας: Κατά τη διάρκεια της υλοποίησης μιας καμπάνιας τα συστήματα πληροφοριών προσφέρουν εφαρμογές που παρέχουν χρήσιμο υλικό για να χτιστούν διάλογοι επικοινωνίας με τον πελάτη. Απώτερος στόχος της επικοινωνίας αυτής είναι η απόκτηση γνώσης για τους πελάτες και των προτιμήσεων τους ούτως ώστε να αυτοματοποιηθεί η εκτέλεση και διαχείριση της εκστρατείας.
- Έλεγχος εκστρατείας: Συγκεκριμένα ο έλεγχος αντιμετωπίζει προβλήματα που αφορούν κυρίως την παρακολούθηση και ανάλυση των επικείμενων ή ολοκληρωμένων εκστρατειών. Τα αποτελέσματα και τα ευρήματα των αναλύσεων αυτών μπορούν να χρησιμοποιηθούν για τον σχεδιασμό των επόμενων.

Η αυτοματοποίηση των πωλήσεων περιλαμβάνει την υποστήριξη όλων των ενεργειών των πωλήσεων που αφορούν την επικοινωνία με τον κάθε πελάτη. Η υποστήριξη τόσο των εσωτερικών όσο και των εξωτερικών διαδικασιών των πωλήσεων πραγματοποιείται με την βοήθεια υπολογιστικών συστημάτων (π.χ. διαχείριση ευκαιριών και παραγγελιών) χρησιμοποιώντας πληροφοριακές και επικοινωνιακές τεχνολογίες. Παρακάτω παρουσιάζονται τα τρία κυριότερα ζητήματα της αυτοματοποίησης των πωλήσεων: διαχείριση παραγγελιών, υποστήριξη των πωλήσεων και διαμόρφωση προϊόντος.

- Διαχείριση παραγγελιών: Μια υπηρεσία τέτοιου συστήματος καλείται να διευθετήσει την τροφοδοσία και τις παραγγελίες σε έναν οργανισμό. Επιπλέον συμπεριλαμβάνεται σε αυτό η βοήθεια που παρέχει στην επεξεργασία των απαιτήσεων του πελάτη, της πορείας των πωλήσεων καθώς και την παρακολούθηση αυτής.
- Υποστήριξη των πωλήσεων: Η συγκεκριμένη αρμοδιότητα υφίσταται για να παρέχει ευκαιρίες στις εταιρείες μέσα από τα δεδομένα και τις

ενέργειες των πωλήσεων που είναι εύκολα διαθέσιμα από το τμήμα πωλήσεων. Επιπροσθέτως υποστηρίζεται από υπηρεσίες που διαχειρίζονται συναντήσεις και ενέργειες σε πραγματικό χρόνο, κατά συνέπεια ο χρόνος για επισκέψεις πελατών διαχειρίζεται βέλτιστα.

- Διαμόρφωση προϊόντος: Ο πρωταρχικός σκοπός της διαμόρφωσης είναι η προσαρμογή των διαφορετικών προϊόντων και υπηρεσιών, που παρέχει ο εκάστοτε οργανισμός, στις ανάγκες και επιθυμίες του καταναλωτή. Οι προσαρμογές αυτές μπορούν να χρησιμοποιηθούν τόσο για τον σχεδιασμό των προϊόντων όσο και για την παρουσίαση των δυνατοτήτων που έχουν οι συνδυασμοί των προϊόντων αλλά και των εναλλακτικών τους.

Η αυτοματοποίηση των υπηρεσιών υποβοηθά το συγκεκριμένο τμήμα, το οποίο είναι σημαντικός παράγοντας για την επιτυχία του εκάστοτε οργανισμού διατηρώντας μακροχρόνιες σχέσεις με τους πελάτες. Αναλόγως των διαδικασιών που ακολουθούνται τα συστήματα CRM προσφέρουν τις αντίστοιχες υπηρεσίες. Οι υπάλληλοι που διαχειρίζονται τέτοια θέματα χρειάζονται κατά μεγάλο βαθμό διοικητική υποστήριξη. Εργαζόμενοι σε εξυπηρέτηση πελατών χρησιμοποιούν ανεπτυγμένες τεχνολογίες επικοινωνίας απλουστεύοντας την επικοινωνία με τους πελάτες. Ενέργειες που συνεισφέρουν στη αυτοματοποίηση αυτή είναι: γραφεία υποστήριξης, διαχείριση παραπόνων και υπηρεσίες διαχείρισης αιτημάτων.

- Γραφεία υποστήριξης: Συνήθως, είναι το πρώτο σημείο επικοινωνίας του πελάτη με την εταιρεία για θέματα εξυπηρέτησης πελατών. Τα εισερχόμενα αιτήματα μπορούν αυτόματα να απονεμηθούν στο κατάλληλο τμήμα. Για να βρεθεί η κατάλληλη λύση στο πρόβλημα του εκάστοτε πελάτη πολλές υπηρεσίες είναι διαθέσιμες που πραγματοποιούν την αναγνώριση του προβλήματος ή επεξεργάζονται δεδομένα προηγούμενων παρόμοιων προβλημάτων και λύσεων τους.
- Διαχείριση παραπόνων: Η διαχείριση των παραπόνων είναι σημαντικός τομέας για έναν οργανισμό καθώς οι πληροφορίες που συλλέγονται και αποθηκεύονται ευνοούν την αύξηση της ικανοποίησης των πελατών.
- Υπηρεσίες διαχείρισης αιτημάτων: Με τον όρο αυτό, ο ρόλος του συστήματος CRM περιέχει όλες τις ενέργειες εξυπηρέτησης στο πλαίσιο

του κέντρου εξυπηρέτησης. Κύριος στόχος είναι η προσφορά γρήγορης και συγκεκριμένης λύσης στο πρόβλημα του καταναλωτή.

### 1.3.2 Συνεργατικό CRM

Το συνεργατικό CRM, όπως αναφέρθηκε, ελέγχει και ολοκληρώνει τις επικοινωνίες μεταξύ πελάτη και εταιρείας. Η επικοινωνία αυτή είναι η βάση για την αποκόμιση γνώσεων όσον αφορά τις σχέσεις μεταξύ τους. Εξαιτίας της χρήσης των διαφόρων δίαυλων επικοινωνίας βελτιστοποιείται η ανταλλαγή πληροφοριών (Torggler, 2008). Το συγκεκριμένο σύστημα εκμεταλλεύεται τα παρακάτω: διαχείριση επαφών, eCRM και internet και κέντρα αλληλεπίδρασης πελατών.

- Διαχείριση επαφών: Η διαχείριση επαφών περιλαμβάνει όλα τα εργαλεία για την διαχείριση και διατήρηση των στοιχείων που αφορούν την επικοινωνία με τους πελάτες (π.χ. διευθύνσεις, emails).
- eCRM και internet: Καθώς η τεχνολογία αναπτύσσεται και εισέρχεται στις διαδικασίες του CRM, τα δεδομένα των πελατών από διαδικτυακές εργασίες μπορούν να εισαχθούν απευθείας σε μια βάση δεδομένων. Κατά συνέπεια πληθώρα των διαδικασιών του CRM είναι υποστηριζόμενα. Επιπροσθέτως, ο σχεδιασμός και η εκτέλεση διαδικτυακών καταλόγων και καταστημάτων βασίζονται σε τέτοιες τεχνολογίες. Δηλαδή στηρίζονται σε πραγματικού χρόνου επικοινωνιακών δικτύων μέσω και έξω από τον οργανισμό.
- Κέντρα αλληλεπίδρασης πελατών: Ένα τέτοιο κέντρο είναι η βάση για την εξυπηρέτηση πελατών. Όλα οι δίαυλοι συγκεντρώνονται σε ένα κέντρο επικοινωνίας όπου χρησιμοποιεί ανεπτυγμένες τεχνολογίες όπως: αυτόματος διαχωριστής κλήσεων και τηλεφωνητές-οδηγούς.

### 1.3.3 Αναλυτικό CRM

Πιο συγκεκριμένα, το αναλυτικό σύστημα του CRM είναι ο συνδετικός κρίκος μεταξύ της εικόνας και συμπεριφοράς της εταιρείας με τις αποφάσεις που πραγματεύονται το εσωτερικό της. Εστιάζει δηλαδή στην προετοιμασία,

προσομοίωση, ανάλυση και βελτιστοποίηση των αποφάσεων που αφορούν τους πελάτες επεξεργάζοντας τα δεδομένα τους. Πληροφορίες που αφορούν πελάτες, προϊόντα και αγορές είναι συχνά αποθηκευμένα σε «αποθήκες» δεδομένων με απώτερο σκοπό να αναλυθούν από αντίστοιχα εργαλεία. Παρακάτω αναπτύσσονται τα συνηθέστερα πεδία που αναλύονται από το συγκεκριμένο σύστημα (Sean, 2000).

**Ανάλυση πωλήσεων:** Οι εφαρμογές της ανάλυσης αυτής παρέχουν την συνολική εικόνα των πωλήσεων συναρτήσει ποικίλων χαρακτηριστικών (διαστάσεων). Όπως για παράδειγμα πωλήσεις ανά προϊόν, ανά κατηγορία προϊόντος, ανά περιοχή, ανά εποχή και επίσης προσφέρει αναλύσεις χαρακτηριστικών των πελατών που αγοράζουν συγκεκριμένα προϊόντα ή αναλύσεις των σχέσεων μεταξύ ικανοποίησης πελατών και δυναμικότητα πωλήσεων. Παρέχεται έτσι στον οργανισμό μια ολοκληρωμένη οπτική των πωλήσεων αποσαφηνίζοντας τα μοτίβα των πολύπλοκων δεδομένων.

**Ανάλυση προφίλ του πελάτη:** Η εφαρμογή αυτή επιτρέπει στην επιχείρηση να ξεχωρίζει στην πληθώρα πελατών, τον πελάτη που την ενδιαφέρει καθώς και τα τμήματα που αποτελούν το σύνολο τους. Η κατάτμηση των πελατών μέσω της ανάλυσης αυτής γεννά ουσιώδη εργαλεία για το μάρκετινγκ καθώς οι αγορές διαχωρίζονται και τα τμήματα των καταναλωτών παρουσιάζουν διακριτά και ποικίλα χαρακτηριστικά. Ωφελεί όπως είναι προφανές την ανάπτυξη ουσιαστικών σχέσεων με τον πελάτη. Η συγκεκριμένη ανάλυση περιλαμβάνει τον προσδιορισμό ανάλογα με την χρήση του προϊόντος και της συχνότητας, τα δημογραφικά χαρακτηριστικά, ψυχογραφικά χαρακτηριστικά, γεωγραφικά χαρακτηριστικά κ.α.

**Ανάλυση καμπάνιας:** Η ικανότητα της επιχείρησης να αναγνωρίζει κατάλληλους πελάτες για τις καμπάνιες και προωθήσεις της καθώς και να αναλύει τις αποκρίσεις των στρατηγικών αυτών είναι το κλειδί για την μεταφορά από το μαζικό μάρκετινγκ στο προσαρμοσμένο μάρκετινγκ. Οι περισσότεροι οργανισμοί επιχειρούν πολλές διαφορετικές προωθητικές ενέργειες για διάφορα προϊόντα στα κατάλληλα μέσα. Η εφαρμογή αυτή ενισχύει την διαδικασία της μάθησης καθ' όλη την διάρκεια των ενεργειών, από την επιλογή των κατάλληλων πελατών έως και την ανάλυση της απόκρισης τους σε αυτές. Η ανάλυση αυτή περιλαμβάνει την επεξεργασία των δεδομένων των πελατών και σε ομάδες και σαν μεμονωμένα άτομα, καθώς και ανάλυση πελατών που έχουν θετική απόκριση στις καμπάνιες. Είναι ικανή να

επεξεργαστεί και να συγκρίνει χαρακτηριστικά καταναλωτών που δεν ανταποκρίνονται θετικά ακόμα και χαρακτηριστικά από καταναλωτές που έχουν αποκλειστεί από τις προωθητικές ενέργειες. Κατά συνέπεια παρέχει την ικανότητα να μετρά την αποδοτικότητα των μεμονωμένων προωθητικών κινήσεων, την αποδοτικότητα του κάθε μέσου στο οποίο είναι παρούσες και το κέρδος από αυτές.

**Ανάλυση καταναλωτικής εμπιστοσύνης:** Μια από τις σημαντικότερες παραμέτρους που απασχολεί έναν οργανισμό είναι η καταναλωτική εμπιστοσύνη. Η εμπιστοσύνη, δηλαδή, που έχει ο πελάτης προς την επιχείρηση και τον αποτρέπει από το να εγκαταλείψει τις υπηρεσίες ή τα προϊόντα της. Ουσιαστικά, η ανάλυση της καταναλωτικής εμπιστοσύνης μετρά, με δομημένο τρόπο την εμπιστοσύνη του κάθε πελάτη ή των συνόλων των πελατών και αποσκοπεί στην βαθύτερη κατανόηση των αιτιών που τον κάνουν να εγκαταλείπει τον οργανισμό. Πιο συγκεκριμένα, η εφαρμογή αυτή μετρά την εμπιστοσύνη συναρτήσει της διάρκειας ύπαρξης του καταναλωτή, του εύρους των υπηρεσιών ή προϊόντων που αυτός καταναλώνει και της συχνότητας όπου τα καταναλώνει. Επίσης επεξεργάζεται τα χαρακτηριστικά των πελατών που έχουν ή όχι εμπιστοσύνη, τον βαθμό φθοράς της εμπιστοσύνης των καταναλωτών, εποχιακές μεταβλητές που επηρεάζουν την φθορά αυτή καθώς και δημογραφικά, ψυχογραφικά και γεωγραφικά χαρακτηριστικά που την επηρεάζουν. Τα αποτελέσματα των αναλύσεων καθιστούν ικανή την ανάπτυξη προγραμμάτων επανένταξης των πελατών στην εταιρεία μετά από την εγκατάλειψη της.

**Ανάλυση της επικοινωνίας του πελάτη με την επιχείρηση:** Η αλληλεπίδραση των πελατών με τους οργανισμούς μπορεί να επιτευχθεί με ποικίλους τρόπους. Η αλληλεπίδραση αυτή μπορεί να περιέχει αναζήτηση πληροφοριών, κατάθεση παραπόνων ή φιλοφρονήσεων ακόμα και την αναφορά σφαλμάτων ή παροχή συμβουλών. Αυτές οι πληροφορίες αναβαθμίζουν τις γνώσεις και τις αντιλήψεις για την καταναλωτική συμπεριφορά. Η καταγραφή των επιπέδων ικανοποίησης ή δυσαρέστησης του πελάτη μπορεί να βρεθεί στο ιστορικό αλληλεπίδρασης του με την εταιρεία. Για το λόγο αυτό η παρούσα εφαρμογή επεξεργάζεται το ιστορικό επικοινωνίας του πελάτη, μελετά επιλυμένα ή μη παράπονα τους και αναλύει θετικά ή αρνητικά σχόλια. Επιπροσθέτως, η επεξεργασία αυτή λαμβάνει υπόψιν τον χρόνο επίλυσης των προβλημάτων και την ανταπόκριση σε αυτές της λύσεις. Απώτερος σκοπός της εφαρμογής αυτής είναι η διατήρηση των σχέσεων πελάτη-επιχείρησης και η αποτροπή της εγκατάλειψης της εταιρείας από τον καταναλωτή.

**Ανάλυση της κερδοφορίας:** Σε οποιαδήποτε επιχείρηση είναι σημαντική η κατανόηση και η ποσοτικοποίηση της κερδοφορίας ούτως ώστε να προσδιορίζονται οι τιμολογήσεις των προϊόντων ή υπηρεσιών, οι εκπτώσεις, οι προμηθευτές και οι στρατηγικές της εταιρείας με τον πιο αποτελεσματικό τρόπο. Ωστόσο η κερδοφορία είναι μια πολύπλευρη έννοια καθώς η ερμηνεία της αλλάζει ανάλογα με το πλαίσιο στο οποίο αξιολογείται. Για παράδειγμα, η κερδοφορία της εταιρείας είναι διαφορετική με την κερδοφορία ενός προϊόντος ή κατηγορίας προϊόντος. Συνεπώς οι οργανισμοί επιθυμούν να μετρούν το κέρδος σε όρους όπως τα ακαθάριστα κέρδη, καθαρά κέρδη. Η ανάλυση των διαφόρων διαστάσεων που διέπουν την κερδοφορία είναι αντικειμενικός στόχος της εν λόγω εφαρμογής.

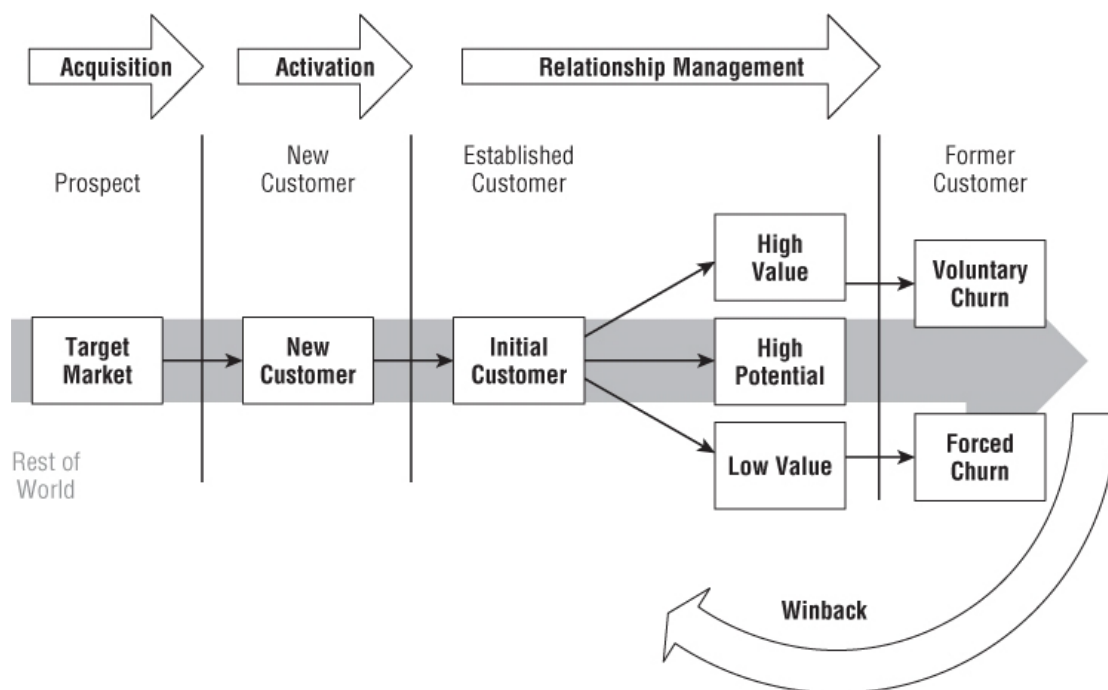
Εν κατακλείδι, η παραπάνω εφαρμογές του αναλυτικού CRM παρέχουν ουσιαστική γνώση για την αγορά στα πλαίσια του καταναλωτή μετατρέποντας τις επιχειρήσεις που είναι εστιασμένες στο προϊόν ή το μαζικό μάρκετινγκ σε επιχειρήσεις εστιασμένες στον πελάτη και το προσαρμοσμένο μάρκετινγκ.

#### 1.4 Κύκλος ζωής του πελάτη

Ο όρος κύκλος ζωής μπορεί να σημαίνει δύο διαφορετικά πράγματα όταν αναφερόμαστε σε επιχειρήσεις. Το πρώτο είναι ο κύκλος ζωής του πελάτη και το δεύτερο ο κύκλος ζωής του πελάτη συναρτήσει της σχέσης του με την εταιρεία.

Η πρώτη έννοια αναφέρεται στην πορεία του ατόμου στην προσωπική του ζωή και πως αυτή επηρεάζει την επιχείρηση. Με την έννοια πορεία στην προσωπική ζωή εννοούνται γεγονότα όπως η αποφοίτηση από δευτεροβάθμια ή τριτοβάθμια εκπαίδευση, απόκτηση παιδιών, γάμος, εύρεση εργασίας κτλ. Παραδείγματος χάριν, ένα σημαντικό γεγονός για τον κύκλο ζωής του πελάτη είναι η μετακόμιση. Όταν κάποιος μετακομίζει είναι πολύ πιθανό να χρειάζεται έπιπλα ή να ανοίξει κάποιον λογαριασμό σε τράπεζα. Συνεπώς είναι ιδιαίτερα χρήσιμη πληροφορία να ξέρει κανείς ποιος μετακομίζει και που, ειδικότερα μια εταιρεία επιπλών ή μια τράπεζα. Η ανάλυση με στόχο την κατανόηση, όπως αναφέρθηκε στο αναλυτικό CRM, των σταδίων αυτών είναι καθοριστικός παράγοντας για την κερδοφορία μια επιχείρησης.

Η σχέση μεταξύ των πελατών και της επιχείρησης εξελίσσεται με την πάροδο του χρόνου και ο δεύτερος όρος του κύκλου ζωής του πελάτη αναφέρεται στην εξέλιξη αυτή. Από εδώ και στο εξής με τον όρο κύκλος ζωής του πελάτη εννοείται ο κύκλος που αφορά την σχέση του πελάτη με την εταιρεία. Ο κύκλος αυτός παρουσιάζεται παρακάτω στο Σχήμα 1. Ο κύκλος δηλαδή, χωρίζεται σε 4 κύριες φάσεις: απόκτηση, ενεργοποίηση, διαχείριση πελατειακών σχέσεων και ανάκτηση (Linoff & Berry, 2011).



Σχήμα 1. Κύκλος ζωής του πελάτη (Linoff & Berry, 2011).

#### 1.4.1 Απόκτηση πελατών

Η φάση της απόκτησης πελατών είναι η διαδικασία όπου ο οργανισμός προσπαθεί να ελκύσει εν δυνάμει πελάτες και να τους μετατρέψει σε πραγματικούς. Η απόκτηση νέων πελατών δεν ακολουθεί, όπως αναφέρθηκε στη παραπάνω ενότητα, το μαζικό μάρκετινγκ αλλά το στοχευμένο. Ανάλογα με τις ανάγκες του εκάστοτε πελάτη το μάρκετινγκ των επιχειρήσεων οφείλει να προσφέρει τις κατάλληλες υπηρεσίες ή



προϊόντα. Η διαδικασία αυτή, όπως είναι προφανές, βελτιστοποιείται με αυτόν τον τρόπο παράγοντας καταλληλότερα αποτελέσματα. Επιπροσθέτως, λόγω της ραγδαίας ανάπτυξης της τεχνολογίας όπως η ανάλυση στα κοινωνικά δίκτυα, γεωγραφικές αναλύσεις και τεχνικές γρήγορης ανάλυσης το στοχευμένο μάρκετινγκ καθίσταται εφικτό (Krishna & Ravi, 2016).

Η αναγνώριση και κατανόηση των αναγκών των εν δυνάμει πελατών είναι αναπόσπαστο κομμάτι αυτού του βήματος. Η αναγνώριση των πελατών περιλαμβάνει την στόχευση ενός πληθυσμού που είναι πιο πιθανό να οδηγηθούν στην εταιρεία και ταυτόχρονα να προσφέρουν κέρδος. Στοιχεία του βήματος αυτού είναι η ομαδοποίηση των πελατών και η ανάλυση τους (Ngai, Xiu, & Chau, 2009). Η ομαδοποίηση των πελατών, πιο συγκεκριμένα, αποτελεί χρήσιμο εργαλείο για την επιχείρηση καθώς διαχωρίζει τους καταναλωτές σε τμήματα με κοινά χαρακτηριστικά. Δηλαδή δημιουργεί ομάδες που οι πελάτες μοιράζονται τις ίδιες ανάγκες και επιθυμίες. Το συγκεκριμένο εργαλείο είναι ικανό να αναγνωρίζει μη ικανοποιημένους πελάτες. Όπως είναι προφανές οι οργανισμοί το χρησιμοποιούν για να νικήσουν τον ανταγωνισμό προσφέροντας προϊόντα και υπηρεσίες στα μέτρα τους βελτιστοποιώντας όλη τη διαδικασία. Η κατάτμηση αυτή μπορεί να οριστεί επιτυχής όταν ο οργανισμός εξυπηρετεί τα σύνολα αυτά που είναι πιο επικερδή (Krishna & Ravi, 2016).

Ύστερα από την αναγνώριση των κατάλληλων πελατών και την κατάτμηση τους ακολουθεί η έλξη τους από την εταιρεία. Ένας τρόπος που χρησιμοποιείται για την προσέλκυση αυτή είναι το απευθείας ή στοχευμένο μάρκετινγκ. Είναι, ουσιαστικά, προωθητικές ενέργειες που ενθαρρύνουν των καταναλωτή να έρθει στην εταιρεία χρησιμοποιώντας διάφορα μέσα (Ngai, Xiu, & Chau, 2009). Η διαδικασία υλοποίησης των ενεργειών αυτών πραγματοποιείται με την απόκτηση βάσεων δεδομένων με τα προφίλ των πελατών όπου μετά την ανάλυση τους γίνεται προσφορά κατάλληλων προϊόντων στους κατάλληλους πελάτες (Krishna & Ravi, 2016).

Παραδείγματα των διάφορων αυτών μέσων που γίνονται οι προφορές είναι τα email, κουπόνια ακόμα και διαφημίσεις στα μέσα κοινωνικής δικτύωσης.

#### 1.4.2 Ενεργοποίηση πελατών

Όπως είναι λογικό, η ενεργοποίηση του πελάτη ακολουθεί την απόκτηση του. Είναι ουσιαστικά η αφετηρία της σχέσης του με την επιχείρηση. Η διαδικασία της ενεργοποίησης θα μπορούσε να είναι τόσο απλή όσο η συμπλήρωση από τον καταναλωτή μιας φόρμας σε ένα ιστότοπο ή η αγορά ενός προϊόντος ή υπηρεσίας (Linoff & Berry, 2011). Αξίζει να σημειωθεί πως η κατανόηση της διαδικασίας αυτής, δηλαδή η εύρεση απαντήσεων στο πώς, τότε και γιατί ο καταναλωτής επέλεξε την συγκεκριμένη επιχείρηση μπορεί να οδηγήσει στην απόκτηση πολύτιμων πληροφοριών που αφορούν την προηγούμενη και παρούσα φάση του κύκλου ζωής των πελατών. Στόχος, φυσικά, των πληροφοριών είναι να βελτιστοποιηθούν οι παραπάνω διαδικασίες.

#### 1.4.3 Διαχείριση πελατειακών σχέσεων

Μοναδικός στόχος της συγκεκριμένης φάσης είναι η διατήρηση και ανάπτυξη της κερδοφορίας από τους πελάτες. Η διαδικασία αυτή λοιπόν αναλύεται σε δύο κομμάτια: ανάπτυξη και διατήρηση πελατών.

**Ανάπτυξη πελατών:** Η διαδικασία αυτή περιέχει την διεύρυνση των συχνοτήτων και μεγεθών των συναλλαγών και κατά συνέπεια την κερδοφορία από τον εκάστοτε καταναλωτή. Στοιχεία που αποτελούν την πρόοδο αυτή είναι η ανάλυση της συνολικής αξίας στον κύκλο ζωής του πελάτη (Customer Lifetime Value, CLV), ανάλυση καλαθιού αγοράς του πελάτη (Market Basket Analysis, MBA), ομαδοποίηση πελατών με βάση τη δανειοληπτική τους ικανότητα (Customer profiling based on credit scoring) και ανίχνευση απάτης (Fraud detection) (Linoff & Berry, 2011) (Krishna & Ravi, 2016) (Ngai, Xiu, & Chau, 2009). Παρακάτω αναλύονται τα στοιχεία αυτά.

- Συνολική αξία κύκλου ζωής του πελάτη (CLV). Όπως είναι φυσικό, η ανάπτυξη της κερδοφορίας της επιχείρησης από κάθε πελάτη είναι ωφέλιμο αποτέλεσμα για την ίδια. Έτσι το συγκεκριμένο στοιχείο είναι ένα από τα σημαντικότερα της διαδικασίας αυτής και γενικά για την διαχείρισης πελατειακών σχέσεων. Ουσιαστικά είναι μια πρόβλεψη των καθαρών εσόδων

σε βάθος χρόνου από την εν δυνάμει συνεργασία της επιχείρησης με τον πελάτη. Η πρόβλεψη αυτή μπορεί να πραγματοποιηθεί λαμβάνοντας υπόψιν παράγοντες όπως η συχνότητα συναλλαγών, τελευταίων συναλλαγών και χρηματικών ποσών που αφορούν τον πελάτη.

- Ομαδοποίηση πελατών με βάση την δανειοληπτικής τους ικανότητα. Ο όρος οικονομική δυναμική αναφέρεται στην αριθμητική έκφραση των πληροφοριών του πελάτη σε σχέση με το οικονομικό του υπόβαθρο. Μια τέτοια ομαδοποίηση εκτιμά την αξία του πελάτη και κατά συνέπεια την αξία του στον κύκλο ζωής όπως αναφέρθηκε παραπάνω.
- Ανάλυση καλαθιού αγοράς του πελάτη (MBA). Η ανάλυση αυτή βασίζεται στην παραδοχή πως πολλοί πελάτες τείνουν να αγοράζουν ορισμένα προϊόντα ή υπηρεσίες σε ομάδες, ένα τέτοιο παράδειγμα είναι η αγορά κινητού μαζί με θήκη ή ακουστικά. Κατά συνέπεια είναι χρήσιμη η εφαρμογή αλγορίθμων σε δεδομένα συναλλαγών από πελάτες με στόχο την απόκτηση γνώσης για τις συνήθειες αυτές. Χρησιμοποιώντας την γνώση αυτή, μπορεί να επιτευχθεί και η πώληση περισσότερων προϊόντων αυξάνοντας το κέρδος (cross-sold). Δυο τεχνικές που χρησιμοποιούνται είναι: Cross-sell και Up-sell. Cross-sell είναι η πώληση των σχετικών προϊόντων που δεν έχουν ακόμα πιστωθεί από τον πελάτη και up-sell είναι η πώληση αναβαθμίσεων ενός προϊόντος ή υπηρεσίας. Και οι δύο περιπτώσεις χρησιμοποιούν αλγορίθμους για ανάλυση δεδομένων που αφορούν τους καταναλωτές και παράγουν γνώση για την εφαρμογή τους αυξάνοντας το κέρδος.
- Ανίχνευση απάτης. Ο όρος απάτη αναφέρεται στις ενέργειες των ατόμων που διεκδικούν στοιχεία (ταυτότητα κ.α.) που δεν τους ανήκουν προς όφελος τους. Ωστόσο τα τεχνολογικά άλματα ενώ παρέχουν την άνεση και την ευκολία στον πελάτη, δίνουν ταυτόχρονα την δυνατότητα να διαπράξουν απάτες μέσω αυτής. Η ανίχνευση αυτών γίνεται όλο και πιο απαραίτητη με το πέρασμα του χρόνου για την διασφάλιση της κερδοφορίας και της φήμης του οργανισμού.

**Διατήρηση πελατών.** Η διατήρηση πελατών είναι αρκετά πιο εύκολη και πιο φθηνή διαδικασία από την απόκτηση νέων πελατών και για το λόγο αυτό είναι από τους σημαντικότερους στόχους της διαχείρισης πελατειακών σχέσεων ενός οργανισμού. Η

παρούσα διαδικασία αποτελείται από δύο κύρια στοιχεία: την αναγνώριση απώλειας του πελάτη και την εξόρυξη γνώμης.

**Αναγνώριση απώλειας του πελάτη.** Μια επιχείρηση, ειδικά όταν παρέχει υπηρεσίες ζημιώνεται όταν κάποιος πελάτης σταματήσει να χρησιμοποιεί τις υπηρεσίες αυτές και όπως είναι φυσικό εταιρείες με πιστούς πελάτες είναι πολύ πιο ανταγωνιστικές. Κατά συνέπεια χρησιμοποιώντας τεχνικές πρόβλεψης μέσω αλγορίθμων οι οργανισμοί πρέπει να μπορούν να προβλέψουν τις ομάδες των πελατών που πρόκειται να τους εγκαταλείψουν και έτσι το αρμόδιο τμήμα του μάρκετινγκ να τους μεταπείσει. Κατά μια έννοια η αναγνώριση των πελατειακών εγκαταλείψεων και η διατήρηση πελατών είναι δύο πλευρές από το ίδιο νόμισμα.

**Εξόρυξη γνώμης.** Είναι μια πιο συγκεκριμένη διαδικασία που ουσιαστικά αποτελεί ένα στόχο της διαχείρισης των πελατειακών σχέσεων. Ο στόχος αυτός είναι η ανάλυση των προτιμήσεων των καταναλωτών που πραγματοποιείται σε δεδομένα που προέρχονται από ιστοσελίδες δικτύωσης, blogs, κα. Τέτοια δεδομένα που εκφράζονται από τους καταναλωτές δεν είναι δομημένα και πρέπει να αναλύονται μαζί με δημογραφικά δεδομένα και δεδομένα συναλλαγών με απώτερο σκοπό την δημιουργία ακριβέστερων προβλέψεων. Αξίζει να σημειωθεί πως η ανάλυση και εξόρυξη γνώμης μπορεί να χρησιμοποιηθεί ως εργαλείων των τεχνικών cross-selling και up-selling.

#### 1.4.4 Ανάκτηση πελατών

Η τελευταία διαδικασία που εμπεριέχεται στον κύκλο ζωής του πελάτη είναι η επανένταξη του στην περίπτωση που επέλεξε να εγκαταλείψει την επιχείρηση. Η διαδικασία επαναφοράς του πελάτη στην επιχείρηση υλοποιείται παρέχοντας σε επικερδής πελάτες κίνητρα και προϊόντα σε ειδικές τιμές. Σε γενικές γραμμές, δηλαδή, ακολουθείται η ίδια στρατηγική με την διατήρηση των πελατών (Linoff & Berry, 2011).

Η διαδικασία αυτή είναι γεγονός πως βασίζεται περισσότερο σε τεχνικές επιχειρησιακού CRM, ωστόσο τεχνικές του αναλυτικού CRM μπορούν να φανούν ιδιαίτερα χρήσιμες. Γνωρίζοντας τους παράγοντες που οδήγησαν τον πελάτη σε ανταγωνιστή του οργανισμού οι στρατηγικές που θα ακολουθηθούν στη συνέχεια δύναται να είναι προσαρμοσμένες κατάλληλα και ,όπως είναι φυσικό,

αποτελεσματικότερες. Η προσπάθεια ανάκτησης πελατών είναι αρκετά δύσκολη, έτσι η προσπάθεια διατήρησης τους είναι σαφώς προτιμητέα.

### 1.5 Απώλεια πελατών

Παραπάνω έγινε αναφορά σε όρους όπως η απώλεια πελατών και καταναλωτική εμπιστοσύνη, στην πραγματικότητα και οι δύο αυτοί όροι είναι αλληλένδετοι. Το φαινόμενο της απώλειας του πελάτη (customer churn) περιγράφει την αλλαγή της προτίμησης του καταναλωτή συναρτήσει της επιχείρησης που απευθύνεται για την πίστωση προϊόντων ή υπηρεσιών (Farquad, Ravi, & Raju, 2014). Στην πραγματικότητα, η απώλεια ενός πελάτη μπορεί να είναι εθελοντική ή μη. Με τον όρο της μη εθελοντικής απώλειας εννοείται η απώλεια με πρωτοβουλία της επιχείρησης για λόγους όπως η καθυστέρηση πληρωμής ή και η μεταφορά καταστήματος σε άλλη γεωγραφική τοποθεσία. Ωστόσο η έννοια αυτή δε θα αναλυθεί περισσότερο στην παρούσα διπλωματική εργασία και με τον όρο απώλεια πελατών περιγράφεται η εθελοντική απώλεια. Οι λόγοι που εγκαταλείπει ο πελάτης έναν οργανισμό θα μπορούσε να είναι η διαθεσιμότητα των τελευταίων τεχνολογιών, επίπεδο ευγένειας προσωπικού, γεωγραφική τοποθεσία, κα. Κατά συνέπεια οι οργανισμοί έχουν την ανάγκη να μπορούν ανά πάσα στιγμή να ξέρουν τα επίπεδα ικανοποίησης του εκάστοτε πελάτη ακόμα και να μπορούν να την προβλέψουν. Η ανάγκη αυτή είναι επιτακτική καθώς η απώλεια ενός πελάτη είναι μείζων πρόβλημα. Μερικοί από τους λόγους αναφέρονται παρακάτω (Saradhi & Palshikar, 2011).

1. Η απόκτηση νέου πελάτη είναι πολύ πιο δύσκολη και ακριβή από τη διατήρηση του.
2. Η διαδικασία τερματισμού της υπηρεσίας ενός πελάτη έχει κόστος.
3. Η απώλεια ενός πελάτη έχει αρνητικό αντίκτυπο στα έσοδα της επιχείρησης καθώς και στη φήμη της.
4. Η απώλεια του πελάτη επηρεάζει πολλές μεταβλητές που αφορούν την επιχείρηση. Ένα παράδειγμα είναι πως οι πελάτες που εγκατέλειψαν την επιχείρηση ζημιώνουν την αξία της επιχείρησης και μπορούν να επηρεάσουν και τους εν δυνάμει πελάτες.

Οι (Saradhi & Palshikar, 2011) αναφέρουν πως οι κυρίαρχοι λόγοι που ο καταναλωτής αποφασίζει να εγκαταλείψει μια επιχείρηση και να απευθυνθεί σε ανταγωνιστή της είναι: η ευελιξία/ευκολία των συναλλαγών, υπηρεσιών ή προϊόντων και ο ανταγωνισμός συναρτήσει των τιμών και της εξυπηρέτησης πελατών. Ταυτόχρονα, ο μεγαλύτερος ανταγωνισμός οδηγεί τον πελάτη σε εγκατάλειψη για πιο ασήμαντους λόγους. Οι οργανισμοί, συνεπώς, καλούνται να αντιμετωπίσουν δύο κύρια προβλήματα όσο αναφορά την απώλεια πελατών: την μείωση του φαινομένου αυτού και την πρόβλεψη του.

### 1.5.1 Πρόβλεψη απώλειας πελατών

Όπως έχει ήδη αναφερθεί στην προηγούμενη ενότητα η πρόβλεψη της απώλειας των πελατών αποτελεί μια από τις βασικότερες εφαρμογές του αναλυτικού CRM. Έχει μεγάλη εφαρμογή σε οργανισμούς που ασχολούνται με χρηματο-οικονομικές υπηρεσίες, εταιρείες τηλεπικοινωνιών, αεροπορικών εταιρειών, οργανισμούς κοινωνική δικτύωσης, κ.ά. Οι (Tsai & Lu, 2009) αναφέρουν δύο τύπους προσέγγισης για τη διαχείριση της απώλειας των πελατών: αντιδραστική προσέγγιση και εκ των προτέρων προσέγγιση. Όταν μια επιχείρηση επιλέγει τη αντιδραστική προσέγγιση ουσιαστικά περιμένει από τον πελάτη το αίτημα αποχώρησης και στη συνέχεια προσπαθεί να το δελεάσει με κάποια προσφορά προκειμένου να παραμείνει. Σε αντίθεση με αυτή τη μέθοδο, η επιχείρηση επιλέγει να προβλέψει πιθανούς πελάτες που πρόκειται να αποχωρήσουν και να τους προσφέρει κίνητρα παραμονής στον οργανισμό. Αν και οι προσεγγίσεις που χρησιμοποιούν τέτοιες προβλέψεις έχουν το πλεονέκτημα σε σχέση με τις αντιδραστικές, διατρέχουν μεγάλο κίνδυνο να είναι αδύνατες για χρήση αν δεν είναι ακριβείς. Ο κίνδυνος αυτός εκφράζεται σε κόστος της επιχείρησης να ξοδεύει χρήματα για πελάτες που δεν πρόκειται να εγκαταλείψουν την επιχείρηση (false positive prediction).

Για τη διαμόρφωση αποτελεσματικών και ακριβώς προβλέψεων για την απώλεια πελατών, έχουν χρησιμοποιηθεί διάφορες μεθοδολογίες. Κύριος ακόμα στόχος των μεθοδολογιών αυτών, πέρα από πρόβλεψη, είναι και η περιγραφή. Η περιγραφή επικεντρώνεται στην εύρεση προτύπων στα δεδομένα που αφορούν τον πελάτη. Πρότυπα δηλαδή που είναι ικανά να απαντήσουν στο γιατί ο καταναλωτής επέλεξε να

φύγει ή ποια είναι τα χαρακτηριστικά που επηρεάζουν τους πελάτες. Αντίθετα, η πρόβλεψη αναλύει τα χαρακτηριστικά αυτά και στη βάση δεδομένων για να εξάγει τα αποτελέσματα της (Tsai & Lu, 2009).

Μέθοδοι που χρησιμοποιούνται για την επίτευξη τέτοιων στόχων, μεταξύ άλλων, είναι η ομαδοποίηση, συσταδοποίηση, παλινδρόμηση, κ.ά. Συνήθεις αλγόριθμοι που αναφέρονται στη βιβλιογραφία είναι τα νευρωνικά δίκτυα, τα δέντρα αποφάσεων, η παλινδρόμηση, ο αλγόριθμος naïve Bayes, οι εξελικτικοί αλγόριθμοι και οι μηχανές διανυσμάτων υποστήριξης (support vector machines) (Saradhi & Palshikar, 2011) (Farquad, Ravi, & Raju, 2014) (Tsai & Lu, 2009).

Τρεις είναι οι βασικές ενέργειες που μπορούν να υλοποιηθούν για την μείωση της απώλειας πελατών σύμφωνα με τους (Saradhi & Palshikar, 2011):

1. Εξέλιξη των υπηρεσιών και προϊόντων που προσφέρονται είτε σε τεχνολογικό πλαίσιο είτε σε επικοινωνιακό. Γενικός κανόνας για την ενέργεια αυτή είναι η ενημέρωση αγαθών της εταιρείας ούτως ώστε να συμβαδίζει ή να καινοτομεί σε σχέση με τον ανταγωνισμό.
2. Δημιουργία προσφορών ή εκπτώτικών ενεργειών για τον καταναλωτή και παροχή πολλών διαφορετικών αλλά και εύκολων πληρωμών.
3. Βελτίωση της εξυπηρέτησης πελατών όπως ο χρόνος εξυπηρέτησης και αναμονής.

Για την καλύτερη εφαρμογή των ενεργειών αυτών το αντίστοιχο τμήμα του οργανισμού οφείλει να λάβει υπόψιν τα στοιχεία που παρέχει η πρόβλεψη.

Ένα ιδιαίτερα σημαντικό στοιχείο είναι η αναγνώριση των αιτιών που οδηγούν τον πελάτη σε ανταγωνιστή. Η άγνοια της επιχείρησης για τους παράγοντες αυτούς μπορεί να οδηγήσει σε λανθασμένους υπολογισμούς και συμπεράσματα. Για παράδειγμα αν η ποιότητα του προϊόντος δεν επηρεάζει την απώλεια των πελατών δεν υπάρχει λόγος να επενδύσει η επιχείρηση στην εξέλιξη του. Συχνά, οι αιτίες που οδηγούν τον πελάτη σε απώλεια δεν είναι άμεσα εμφανείς και να πρέπει να εκμαιευτούν από τα δεδομένα που σχετίζονται με τις παρελθοντικές συνεργασίες με πελάτες που έφυγαν.

Στην περίπτωση που έχει γίνει η πρόβλεψη και η πολιτική της επιχείρησης για τους πελάτες που σκοπεύουν να αποχωρήσουν είναι να προσφέρει ειδικά πακέτα ή

εκπτώσεις για την αποφυγή τέτοιων ενεργειών, πρέπει η πρόβλεψη αυτή να μπορεί να γίνει σε κάθε πελάτη ξεχωριστά. Ο τρόπος αυτός σε συνδυασμό με την πρόβλεψη σε πρώιμο στάδιο έχει τα καλύτερα δυνατά αποτελέσματα (Saradhi & Palshikar, 2011).

Επίσης θα πρέπει να συμπεριληφθεί και η αξία ζωής του πελάτη στην παραπάνω διαδικασία. Όπως αναφέρθηκε παραπάνω, δεν είναι όλοι οι πελάτες το ίδιο επικερδείς για την επιχείρηση. Για παράδειγμα, σε μια εταιρεία τηλεπικοινωνιών υπάρχουν πελάτες που πραγματοποιούν περισσότερες κλήσεις από άλλους ή πελάτες που κάνουν υπεραστικές κλήσεις. Δεδομένου ότι είναι αδύνατο να αποτρέψει κανείς την αποχώρηση κάθε πελάτη, θα ήταν συνετό μια επιχείρηση να εστιάσει σε πελάτες που είναι περισσότερο επικερδείς. Έτσι τα μοντέλα πρόβλεψης οφείλουν να συμπεριλαμβάνουν την αξία του κύκλου ζωής του πελάτη και να προσφέρουν όσο πιο ακριβείς αναλύσεις είναι δυνατό (Saradhi & Palshikar, 2011).

Τέλος, όπως έγινε κατανοητό η απώλεια των πελατών είναι κρίσιμος παράγοντας για τον οργανισμό και η ανάλυση του έχει εξελιχθεί με τον καιρό. Η πρόβλεψη της απώλειας αυτής είναι πολύ σημαντικός κλάδος της αναλυτικής διαχείρισης πελατών και κάνει τις επιχειρήσεις ανταγωνιστικές δίνοντας τη δυνατότητα να εκτελούν έγκαιρα ενέργειες που διατηρούν τους επικερδείς πελάτες στην επιχείρηση.

## 1.6 Παραδείγματα εφαρμογών πρόβλεψης απώλειας πελατών

Εφόσον, όπως αναφέρθηκε, η πρόβλεψη της καταναλωτικής εμπιστοσύνης είναι ιδιαίτερα σημαντική για τους οργανισμούς, υπάρχει πληθώρα εφαρμογών διαθέσιμη στη βιβλιογραφία. Ενδεικτικά, οι (Tsai & Lu, 2009) εφαρμόζουν τρία διαφορετικά μοντέλα για την πρόβλεψη της απώλειας πελατών, απλό νευρωνικό δίκτυο, τον συνδυασμό δυο νευρωνικών δικτύων (ND + ND) και τον συνδυασμό ενός μοντέλου αυτοοργανομένων χαρτών (self-organizing map) και ενός νευρωνικού δικτύου (AX + ND). Το πρώτο μέρος των συνδυασμών αποσκοπεί στο φιλτράρισμα μη αντιπροσωπευτικών δεδομένων. Στη συνέχεια τα δεδομένα που παράγονται χρησιμοποιούνται για τη δημιουργία του μοντέλου πρόβλεψης. Τα δεδομένα που χρησιμοποιήσαν είναι ένα αρχείο δεδομένων από αμερικάνικες εταιρείες τηλεπικοινωνιών. Το αρχείο αυτό περιέχει 51,306 πελάτες από τους οποίους οι 34,761 αποχώρησαν από την εταιρεία και οι 16,545 παρέμειναν στο διάστημα



Ιουλίου 2001-Ιανουαρίου 2002. Σημαντικές πληροφορίες είναι πως για να θεωρηθεί κάποιος πελάτης πρέπει να βρίσκεται τουλάχιστον 6 μήνες στην εταιρεία και θεωρείται πως εγκατέλειψε τον οργανισμό αν δεν έχει ανανεώσει την συνδρομή για 31-60 μέρες. Τα αποτελέσματα που παρουσίασε το συγκεκριμένο άρθρο δείχνουν πως ο συνδυασμός των νευρωνικών δικτύων ξεπερνούν σε απόδοση και το συνδυασμό των αυτοοργανομένων χαρτών με το νευρωνικό δίκτυο και το απλό μοντέλο. Είναι επίσης πιο σταθερό μοντέλο από τα υπόλοιπα και παρουσιάζει τα μικρότερα σφάλματα.

Ένα ακόμα παράδειγμα είναι το άρθρο των (Sharma & Panigrahi, 2013) όπου εφαρμόζεται νευρωνικό δίκτυο για την πρόβλεψη της απώλειας των πελατών. Η εφαρμογή αυτή γίνεται σε αρχείο δεδομένων από το πανεπιστήμιο της Καλιφόρνια και αφορά εταιρείες τηλεπικοινωνιών. Είναι επίσης αρχείο πελατών και δείχνει τα χαρακτηριστικά που αφορούν τις παροχές καθώς και το ιστορικό των πελατών που εγκατέλειψαν την εταιρεία. Η ανάλυση με νευρωνικά δίκτυα κατάφερε να προβλέψει με ποσοστό ακρίβειας 92.35%.

## 2. Γενετικοί αλγόριθμοι

Στην παρούσα διπλωματική εργασία, η οποία πραγματεύεται το πρόβλημα της πρόβλεψης απώλειας πελατών, επιλέχθηκε ο γενετικός αλγόριθμος. Οι γενετικοί αλγόριθμοι αποτελούν μια μέθοδο αναζήτησης βέλτιστων λύσεων σε πολύπλοκα προβλήματα βελτιστοποίησης. Οι εν λόγω αλγόριθμοι αποτελούν βασικά στοιχεία της τεχνητής νοημοσύνης και είναι ευρέως γνωστοί. Πιο συγκεκριμένα, μαζί με τον εξελικτικό προγραμματισμό, στρατηγικές εξέλιξης, συστήματα ταξινόμησης και το γενετικό προγραμματισμό αποτελούν μια κατηγορία συστημάτων επίλυσης προβλημάτων που είναι γνωστοί με τον όρο εξελικτικοί αλγόριθμοι (Γεωργόπουλος & Λυκοθανάσης, 1999), οι οποίοι βασίζονται εξελικτική θεωρία (Whitley, 1994). Δηλαδή η βασική ιδέα για τον τρόπο λειτουργίας τους είναι η μίμηση της βιολογικής θεωρίας που συναντάται στην φύση.

### 2.1 Δαρβινική θεωρία της εξέλιξης

Η Δαρβινική θεωρία ή αλλιώς η θεωρία της εξέλιξης των ειδών αναπτύχθηκε από τον Δαρβίνο. Η εν λόγω θεωρία, που με το πέρασ του χρόνου είναι πλέον ευρέως αποδεκτή, έχει ως στόχο την επεξήγηση του φαινομένου της ζωής και την προέλευση της. Πιο συγκεκριμένα, παρακάτω αναλύονται τα κυριότερα στοιχεία της θεωρίας αυτής που σχετίζονται με τους γενετικούς αλγορίθμους σύμφωνα με (Λυκοθανάσης, 2001).

Πρώτο κύριο στοιχείο είναι πως δεν υπάρχει αντικειμενική βάση διαχωρισμού των ζωντανών οργανισμών σε ανώτερους και κατώτερους. Σε κάθε είδος μερικοί οργανισμοί του πληθυσμού αφήνουν περισσότερους απογόνους σε σύγκριση με τους υπόλοιπους και έτσι τα κληροδοτούμενα χαρακτηριστικά των οργανισμών αυτών πληθαίνουν στην επόμενη γενιά. Το περιβάλλον με τις αντιξοότητες του, καθορίζει ποιοι οργανισμοί θα καταφέρουν να επιζήσουν και να αναπαραχθούν ανάλογα με τα χαρακτηριστικά που διαθέτουν. Κατά συνέπεια μια αλλαγή στο περιβάλλον δοκιμάζει τους οργανισμούς στην ικανότητα τους να προσαρμοστούν και να επιβιώσουν.

Η αλλαγή αυτή, που αποτελεί το δεύτερο κύριο στοιχείο, επέρχεται και στα χρωμοσώματα τα οποία αποτελούνται από γονίδια. Το σύνολο της γενετικής πληροφορίας είναι κωδικοποιημένο στα γονίδια και ονομάζεται γονότυπος.

Κυρίαρχες επίσης λειτουργίες του φαινομένου της εξέλιξης είναι η αναπαραγωγή και η μετάλλαξη. Η μετάλλαξη στα χρωμοσώματα γίνεται με τυχαίο τρόπο και μπορεί να επιφέρει βελτιώσεις και λάθη στη γενετική πληροφορία.

Το τελευταίο κύριο στοιχείο είναι ο τρόπος που κληροδοτείται η πληροφορία. Το προϊόν της αναπαραγωγής είναι ένας νέος οργανισμός ο οποίος έχει παραλάβει τα μισά του γονίδια από τον πατέρα και τα άλλα μισά από τη μητέρα. Οι περιπτώσεις, κατά συνέπεια, είναι δύο. Τα γονίδια για μια «τιμή» θα συμφωνούν ή όχι. Στην περίπτωση που δεν συμφωνούν, ένα από τα δύο θα επικρατήσει και θα καθορίσει το χαρακτηριστικό. Το άλλο αν και δύναται να κληρονομηθεί, θα αγνοηθεί. Το γονίδιο που επικρατεί ονομάζεται κυρίαρχο και αυτό που δεν επικρατεί υπολειπόμενο.

Οι γενετικοί αλγόριθμοι χρησιμοποιούν ορολογίες δανεισμένες από την παραπάνω θεωρία. Αναφέρονται δηλαδή σε άτομα ή γονότυπους, χρωμοσώματα που αποτελούν το γονίδιο και αφομοιώνουν τους τρόπους λειτουργίας της.

## 2.2 Ιστορική αναδρομή

Η ιδέα της εφαρμογής της Δαρβινικής θεωρίας στην επίλυση προβλημάτων χρονολογείται κατά τη δεκαετία του 1940. Στις αρχές του 1948 ο Allan Turing πρότεινε τη μηχανική μάθηση, η οποία θα βασιζόταν στην θεωρία του Δαρβίνου. Η υπολογιστική προσομοίωση της εξέλιξης που ξεκίνησε το 1954 με το έργο του Nils Aall Barricelli, ο οποίος χρησιμοποιούσε τον υπολογιστή στο πανεπιστήμιο Princeton. Το 1962 ο Bremermann πραγματοποιούσε πειράματα στη βελτιστοποίηση μέσω της εξέλιξης. Ακολούθως, στη δεκαετία του 1960 αναπτύχθηκαν τρεις διαφορετικές εφαρμογές της βασικής ιδέας. Στις ΗΠΑ, οι Fogel, Walsh και Owens ανέπτυξαν τον εξελικτικό προγραμματισμό και ο Holland ονόμασε τη μέθοδό του γενετικό αλγόριθμο. Ταυτόχρονα, στη Γερμανία, οι Rechenberg και Schwefel ανέπτυξαν τις εξελικτικές στρατηγικές (Eiben & Smith, Introduction to Evolutionary Computing, 2013).

Ο εξελικτικός προγραμματισμός χρησιμοποιεί μηχανές πεπερασμένων καταστάσεων για την πρόβλεψη του περιβάλλοντος. Ξεκινώντας από το 1980 και μέχρι τον 21<sup>ο</sup> αιώνα ο εξελικτικός προγραμματισμός έγινε αντικείμενο μελέτης και πολλές εκδοχές των προγραμμάτων του συναγωνίστηκαν για την επίτευξη γρήγορων και καλύτερων λύσεων σε διάφορα προβλήματα (Linoff & Berry, 2011).

Οι γενετικοί αλγόριθμοι αποτελούν τα πιο γνωστά στοιχεία των εξελικτικών αλγορίθμων. Αρχικά ο (Holland, 1992) τους παρουσίασε ως μέσο μελέτης προσαρμοστικής λειτουργίας όπως περιγράφει ο τίτλος του βιβλίου του: «*Adaptation in natural and artificial systems*». Ωστόσο οι εν λόγω αλγόριθμοι έχουν σε μεγάλο βαθμό καθιερωθεί ως αλγόριθμοι βελτιστοποίησης. Αυτό θα μπορούσε να οφείλεται στις πολύ πρώτες επιτυχίες στην επίλυση προβλημάτων βελτιστοποίησης (Eiben & Smith, *Introduction to Evolutionary Computing*, 2013). Τα παραπάνω σε συνδυασμό με τη διδακτορική διατριβή του (De Jong, 1975) οδήγησαν στην επίτευξη του ορισμού του κλασικού γενετικού αλγόριθμου.

## 2.3 Αρχικοποίηση

Όπως έχει ήδη αναφερθεί η γενική ιδέα πίσω από τον τρόπο λειτουργίας όλων των εξελικτικών αλγορίθμων είναι η ίδια. Δηλαδή, δίνεται ο πληθυσμός και στη συνέχεια οι συνθήκες θα οδηγήσουν σε φυσική επιλογή των χαρακτηριστικών που θα επιβιώσουν από τον πληθυσμό. Έτσι μπορεί κανείς να δει τη λειτουργία αυτή σαν διαδικασία βελτιστοποίησης. Πιο συγκεκριμένα, δίνεται μια αντικειμενική συνάρτηση για μεγιστοποίηση ή ελαχιστοποίηση. Τυχαία θα επιλεγεί ο πληθυσμός των λύσεων και το κριτήριο της προσαρμογής θα είναι η μέτρηση της συνάρτησης με τις λύσεις αυτές. Στοιχεία που χρησιμοποιεί ο απλός γενετικός αλγόριθμος είναι η κωδικοποίηση και η αντικειμενική συνάρτηση.

### 2.3.1 Κωδικοποίηση

Η κωδικοποίηση είναι το πρώτο βήμα για την έναρξη του αλγορίθμου. Αποτελεί τη γέφυρα επικοινωνίας του «πραγματικού κόσμου» και του «υπολογιστικού κόσμου»

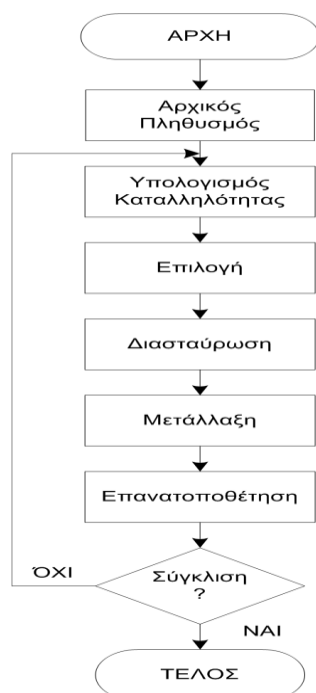
του αλγορίθμου. Το πρόβλημα που χριζει βελτιστοποίηση μπορεί να είναι πολύπλοκο. Μπορεί επίσης οι μεταβλητές του προβλήματος να μην είναι ποσοτικές αλλά ποιοτικές, δηλαδή να μην αποτελούνται από μετρήσιμους αριθμούς αλλά από χαρακτηριστικά όπως «άνδρας» ή «έγγαμος». Κύριος, λοιπόν, στόχος είναι η αναπαράσταση με ικανοποιητικό τρόπο των επιμέρους χαρακτηριστικών των λύσεων, ώστε να διευκολύνει τις επόμενες λειτουργίες του αλγορίθμου. Αποτελέσματα της κωδικοποίησης οφείλει να είναι η ύπαρξη ομοιοτήτων ανάμεσα στα άτομα με σκοπό την κατάλληλη εκμετάλλευσή τους, διότι οι ομοιότητες βοηθούν στην κατεύθυνση της εύρεσης των λύσεων. Η κωδικοποίηση του προβλήματος εξαρτάται από τη φύση του εν λόγω προβλήματος και των μεταβλητών του. Το πιο απλό παράδειγμα κωδικοποίησης είναι η κωδικοποίηση με δυαδικά ψηφία. Η παρούσα διαδικασία είναι κρίσιμο βήμα στην εφαρμογή του αλγορίθμου καθώς επηρεάζει την αποτελεσματικότητά του. (Λυκοθανάσης, 2001), (Eiben & Smith, Introduction to Evolutionary Computing, 2013), (Eiben & Schoenauer, Evolutionary computing, 2002).

### 2.3.2 Αντικειμενική συνάρτηση

Δεύτερο βασικό στοιχείο για την υλοποίηση ενός γενετικού αλγορίθμου είναι η αντικειμενική συνάρτηση. Η συνάρτηση αυτή λαμβάνει ως είσοδο μια κωδικοποιημένη σειρά αριθμών και επιστρέφει μια τιμή. Η εν λόγω τιμή αποτελεί το βασικότερο παράγοντα που καθορίζει την καταλληλότητα των λύσεων. Με βάση την τιμή αυτή κρίνεται αν οι λύσεις που έχουν βρεθεί είναι ικανοποιητικές. Στόχος της είναι να παρουσιάσει τις συνθήκες στις οποίες ο πληθυσμός πρέπει να προσαρμοστεί για να επιβιώσει. Εφόσον ο πληθυσμός είναι οι λύσεις, ουσιαστικά δημιουργεί τις προδιαγραφές για την βελτιστοποίηση των λύσεων. Είναι πιο απλά οι «συνθήκες» που καθορίζουν ποια άτομα θα περάσουν στην επόμενη γενεά όπως αναφέρθηκε παραπάνω. Είναι η μόνη πληροφορία που ο αλγόριθμος είναι ικανός να επεξεργαστεί (Eiben & Smith, Introduction to Evolutionary Computing, 2013) (Λυκοθανάσης, 2001).

## 2.4 Λειτουργική δομή

Είναι σαφές πως οι παραπάνω λειτουργίες αποτελούν βασικά στοιχεία σύνδεσης του γενετικού αλγορίθμου με το πρόβλημα. Παρόλο που είναι απαραίτητα βήματα για την εφαρμογή του αλγορίθμου δεν ανήκουν στο βασικό κομμάτι της διαδικασίας. Είναι



δηλαδή η προεργασία για την υλοποίηση της βελτιστοποίησης. Τα βήματα της υλοποίησης του γενετικού αλγορίθμου είναι: αρχικοποίηση πληθυσμού, επιλογή, διασταύρωση, μετάλλαξη και επανατοποθέτηση.

Σχήμα 2. Η βασική λειτουργική δομή ενός τυπικού γενετικού αλγορίθμου (Καμπουρλάζος & Παπακώστας, 2015).

### 2.4.1 Αρχικοποίηση πληθυσμού

Όταν η διαδικασία της κωδικοποίησης τερματιστεί, πραγματοποιείται η αρχικοποίηση πληθυσμού που είναι και το πρώτο βήμα της διαδικασίας του γενετικού αλγορίθμου. Ο ρόλος του πληθυσμού είναι να περιέχει τις πιθανές λύσεις του προβλήματος. Ο πληθυσμός αποτελεί, δηλαδή, ένα σύνολο γονότυπων. Συνήθως στην αρχή ορίζεται με τυχαίο τρόπο με τη χρήση διάφορων γεννητριών τυχαίων αριθμών και το μέγεθός του παραμένει σταθερό καθ' όλη τη διάρκεια λειτουργίας του

αλγορίθμοι. Βασικός στόχος της αρχικοποίησης είναι να δημιουργήσει όσο γίνεται πιο ποικιλόμορφο πληθυσμό. Με τον όρο ποικιλόμορφο πληθυσμό εννοείται πως οι λύσεις, δηλαδή τα χρωμοσώματα να είναι κατανεμημένα ομοιόμορφα στον χώρο της αναζήτησης. Ο στόχος αυτός καθορίζεται για να μην εγκλωβίζεται η αναζήτηση σε μια μόνο περιοχή (Καμπουρλάζος & Παπακώστας, 2015).

#### 2.4.2 Επιλογή

Όπως φαίνεται και στο σχήμα 2, η διαδικασία της επιλογής απαιτεί ένα επιπλέον βήμα για την έναρξη της διαδικασίας. Το βήμα αυτό είναι ο υπολογισμός καταλληλότητας. Ο υπολογισμός αυτός επιτυγχάνεται με τη χρήση της αντικειμενικής συνάρτησης που όπως αναφέρθηκε παραπάνω ποσοτικοποιεί την υποψήφια λύση του κάθε χρωμοσώματος. Με την ποσοτικοποίηση αυτή δύναται να περάσουν τα χρωμοσώματα από τη διαδικασία της επιλογής και να επιλεγθούν αυτά που θα περάσουν στην επόμενη γενιά. Πιο συγκεκριμένα, η επιλογή αυτή ονομάζεται επιλογή γονέα και διαχωρίζει ποιοι γονότυποι θα γίνουν γονείς (Eiben & Smith, *Introduction to Evolutionary Computing*, 2013). Ωστόσο, σύμφωνα με τους (Καμπουρλάζος & Παπακώστας, 2015) η επιλογή των γονέων δεν πρέπει να βασίζεται αποκλειστικά στα καλύτερα χρωμοσώματα, διότι μπορεί να αποκλείσει τη βέλτιστη λύση. Οι μέθοδοι που έχουν προταθεί ποικίλουν. Ενδεικτικά μερικοί από αυτούς είναι: επιλογή ρουλέτας, η επιλογή καλύτερου συζύγου, η επιλογή με διαγωνισμό.

#### 2.4.3 Διασταύρωση

Η διαδικασία της διασταύρωσης ακολουθεί τη διαδικασία της επιλογής. Η εν λόγω διαδικασία αναλαμβάνει την αναπαραγωγή, όπως και στη φύση, δύο γονέων για να προκύψει η νέα γενεά του πληθυσμού. Ο τρόπος επιλογής των γονότυπων επηρεάζει την ικανότητα και την ταχύτητα του αλγορίθμου. Είναι απαραίτητη λειτουργία για την καλή επίδοση του αλγορίθμου. Εξαιτίας της σπουδαιότητας της είναι αντικείμενο μελέτης και έχουν επινοηθεί πολλοί μέθοδοι για την υλοποίηση της (Λυκοθανάσης,

2001). Σημαντικό πλεονέκτημα της λειτουργίας αυτής είναι η ικανότητα της να ανακατευθύνει τις λύσεις του αλγορίθμου σε περιοχές που δεν έχουν διερευνηθεί. Διευρύνεται με αυτόν τον τρόπο η ποικιλία των πιθανών λύσεων. Ενδεικτικά, κάποιιοι από τους τρόπους διασταύρωσης που είναι διαθέσιμοι είναι η αριθμητική διασταύρωση και η ομοιόμορφη διασταύρωση.

#### 2.4.4 Μετάλλαξη

Τέλος, λαμβάνει χώρα η διαδικασία της μετάλλαξης. Στη βιολογία με τον όρο μετάλλαξη ορίζεται η μεταβολή που μπορεί να συμβεί στο γενετικό υλικό ενός οργανισμού και συμβαίνει με τυχαίο τρόπο. Η διαδικασία που χρησιμοποιείται στον αλγόριθμο αφορά την αλλαγή των χρωμοσωμάτων σε κάποιο γονότυπο του πληθυσμού με τυχαίο τρόπο. Η επιλογή του γονιδίου στο οποίο θα γίνει η μετάλλαξη γίνεται με βάση μιας πιθανότητας μετάλλαξης. Το πλεονέκτημα της ανακατεύθυνσης συναντάται και στη συγκεκριμένη περίπτωση. Ωστόσο η πιθανότητα μετάλλαξης, σε περίπτωση που είναι μεγαλύτερη από όσο πρέπει, εκφυλίζει τον αλγόριθμο. Για τον λόγο αυτό η εφαρμογή της μετάλλαξης πρέπει να γίνει με προσοχή. Ενδεικτικά, ορισμένοι μέθοδοι μετάλλαξης είναι η ομοιόμορφη μετάλλαξη και η Γκαουσιανή μετάλλαξη.

#### 2.4.5 Επανατοποθέτηση

Με το πέρας των παραπάνω διαδικασιών ο τελικός πληθυσμός που περιέχει τις πιθανές λύσεις περνά πάλι από τον υπολογισμό καταλληλότητας και αν τα αποτελέσματα δεν είναι ικανοποιητικά η διαδικασία επαναλαμβάνεται. Έτσι ολοκληρώνεται μια επανάληψη του αλγορίθμου. Τα κριτήρια τερματισμού του αλγορίθμου είναι είτε ο αριθμός των επιτρεπτών γενεών είτε η προεπιλεγμένη επιθυμητή ακρίβεια του αλγορίθμου.



## 2.5 Πλεονεκτήματα γενετικών αλγορίθμων

Γενικά, η χρήση των γενετικών αλγορίθμων είναι πολύ συχνή. Ενδεικτικά, μερικοί λόγοι που καθιστούν τους εν λόγω αλγορίθμους τόσο ελκυστικούς αναλύονται παρακάτω σύμφωνα με τους (Γεωργόπουλος & Λυκοθανάσης, 1999) και (Λυκοθανάσης, 2001).

- Σημαντικό προσόν αυτών των αλγορίθμων είναι η ικανότητα που διαθέτουν να επιλύουν δύσκολα προβλήματα γρήγορα και αξιόπιστα. Η αποδοτικότητα μιας στρατηγικής αποτελεί βασικότατο κριτήριο της αξιολόγησης της. Οι πολυάριθμες εφαρμογές των γενετικών αλγορίθμων έχουν δείξει πως προβλήματα που περιέχουν πολλές και δυσεύρετες λύσεις αντιμετωπίζονται καλύτερα από τέτοιους αλγορίθμους σε σχέση με άλλες στρατηγικές βελτιστοποίησης.
- Στην αποδοτικότητα του αλγορίθμου σημαντικό στοιχείο αποτελεί η ικανότητά του να υλοποιεί ταυτόχρονα τη διερεύνηση του χώρου των λύσεων και την εκμετάλλευση της ήδη υπάρχουσας πληροφορίας. Ο συνδυασμός αυτός είναι δύσκολος καθώς συνήθως η τυχαία αναζήτηση έχει σαν αποτέλεσμα την επαρκή αναζήτηση αλλά δεν εκμεταλλεύεται την υπάρχουσα πληροφορία και το αντίστροφο. Ο γενετικός αλγόριθμος, ωστόσο, πετυχαίνει αυτό το συνδυασμό.
- Επίσης, οι γενετικοί αλγόριθμοι μπορούν να συνεργαστούν με τα υπάρχοντα μοντέλα και συστήματα. Τα μοντέλα των γενετικών αλγορίθμων προσφέρουν το πλεονέκτημα της χρήσης τους με προσθετικό τρόπο στα μοντέλα που χρησιμοποιούνται σήμερα, μη απαιτώντας την επανασχεδίαση τους.
- Οι γενετικοί αλγόριθμοι είναι επεκτάσιμοι και εξελίξιμοι. Δηλαδή, δεν αντιστέκονται σε αλλαγές και επεκτάσεις στον τρόπο λειτουργίας τους που καθορίζονται από τον προγραμματιστή. Οι διαφοροποιήσεις στη διάταξη των αλγορίθμων δεν είναι μόνο υλοποιήσιμες, αλλά και πολλές φορές αναγκαίες.
- Επιπροσθέτως, δεν θέτουν περιορισμούς στις συναρτήσεις που επεξεργάζονται. Άλλες μέθοδοι θέτουν στις αντικειμενικές συναρτήσεις περιορισμούς, όπως οι ύπαρξη συνέχειας ή απαιτούν μη θορυβώδεις

συναρτήσεις. Έτσι κρίνονται ακατάλληλες για ορισμένα προβλήματα σε αντίθεση με τους γενετικούς αλγορίθμους.

- Το παραπάνω πλεονέκτημα δίνει τη δυνατότητα στα μοντέλα των γενετικών αλγορίθμων να μπορούν να συμμετέχουν σε υβριδικές μορφές με άλλα μοντέλα. Σε ορισμένες περιπτώσεις υπάρχει η ανάγκη του συνδυασμού τεχνικών. Ο γενετικός αλγόριθμος είναι ικανός να υπηρετήσει την ανάγκη αυτή.
- Τέλος, οι γενετικοί αλγόριθμοι μπορούν να αξιοποιήσουν τα πλεονεκτήματα των παράλληλων μηχανών καθώς η φύση τους επιτρέπει την παράλληλη υλοποίηση. Το γεγονός αυτό αυξάνει την ταχύτητα και την απόδοση τους.

### 3. Υλοποίηση του γενετικού αλγορίθμου

Στο πρώτο κεφάλαιο αναφέρθηκε η απώλεια πελατών καθώς και η υψηλή αξία που έχει για έναν οργανισμό η πρόβλεψη της. Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η υλοποίηση του γενετικού αλγορίθμου στο περιβάλλον MATLAB με στόχο την πρόβλεψη της απώλειας των πελατών. Πιο συγκεκριμένα, ο στόχος του αλγορίθμου είναι να προβλέπει το 10% των πιο πιθανών πελατών που πρόκειται να εγκαταλείψουν την επιχείρηση.

#### 3.1 Δεδομένα

Για να επιτευχθεί ο παραπάνω στόχος, χρησιμοποιήθηκε το αρχείο δεδομένων από την ιστοσελίδα (kaggle, 2018), το οποίο διατίθεται δωρεάν στο διαδίκτυο. Το εν λόγω αρχείο παρέχεται για την πρόβλεψη της απώλειας των πελατών σε μια αμερικάνικη εταιρεία τηλεπικοινωνιών. Πιο συγκεκριμένα, διαθέτει 7043 πελάτες, τα 21 χαρακτηριστικά που τους διέπουν και την κατάληξη των πελατών όσον αφορά την επιλογή τους να φύγουν από την εταιρεία ή όχι. Τα χαρακτηριστικά είναι κυρίως κατηγορικά καθώς το αρχείο διαθέτει και μερικά ποσοτικά. Τα γνωρίσματα αυτά όπως αναγράφονται στο αρχείο είναι: customerID, gender, Senior Citizen, Partner, Dependents, tenure, Phone Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device Protection, Tech Support, Streaming TV, Streaming Movies, Contract, Paperless Billing, Payment Method, Monthly Charges, Total Charges, Churn.

Το πρώτο χαρακτηριστικό είναι ο κωδικός του πελάτη και καθώς δεν κατέχει κάποια χρησιμότητα στη διαδικασία της εκτέλεσης του αλγορίθμου διαγράφεται από το αρχείο και δεν θα γίνει καμία άλλη αναφορά σε αυτό. Τα επόμενα τέσσερα χαρακτηριστικά είναι δημογραφικά.

**Gender:** Το φύλο του πελάτη.

**Senior Citizen:** Η ηλικία του πελάτη (αν είναι ηλικιωμένος ή όχι).

**Partner:** Αν ο πελάτης έχει σύντροφο (ναι ή όχι).

**Dependents:** Αν ο πελάτης έχει παιδιά (ναι ή όχι).

Τα υπόλοιπα χαρακτηριστικά συνδέονται αυστηρά με την επιχείρηση. Παρακάτω αναγράφονται οι μεταβλητές αυτές με την ερμηνεία τους.

**Tenure:** Ο χρόνος (σε μήνες) που έμεινε ή είναι ο πελάτης στην επιχείρηση.

**Phone Service:** Εάν διαθέτει ο πελάτης υπηρεσίες τηλεφώνου (ναι ή όχι).

**Multiple Line:** Εάν διαθέτει ο πελάτης πολλαπλές γραμμές (ναι, όχι ή δεν έχει υπηρεσίες τηλεφώνου).

**Internet Service:** Αν παρέχονται στον πελάτη υπηρεσίες διαδικτύου (DSL, οπτικές ίνες ή δεν του παρέχονται τέτοιες υπηρεσίες).

**Online Security:** Αν διαθέτει ο πελάτης υπηρεσίες προστασίας διαδικτύου (ναι, όχι ή μη διαθέσιμες υπηρεσίες διαδικτύου).

**Online Backup:** Αν ο ίδιος διαθέτει υπηρεσίες αντιγράφων ασφαλείας που αφορούν το διαδίκτυο (ναι, όχι ή μη διαθέσιμες υπηρεσίες διαδικτύου).

**Device Protection:** Εάν ο πελάτης διαθέτει υπηρεσίες προστασίας συσκευής (ναι, όχι ή μη διαθέσιμες υπηρεσίες διαδικτύου).

**Tech Support:** Παροχές τεχνικής υποστήριξης (ναι, όχι ή μη διαθέσιμες υπηρεσίες διαδικτύου).

**Streaming TV:** Παροχές διαδικτυακής τηλεόρασης (ναι, όχι ή μη διαθέσιμες υπηρεσίες διαδικτύου).

**Streaming Movies:** Παροχές διαδικτυακών ταινιών (ναι, όχι ή μη διαθέσιμες υπηρεσίες διαδικτύου).

**Contract:** Διάστημα συμβολαίου (μηνιαίο, ετήσιο ή δύο χρόνων).

**Paperless Billing:** Εάν η πίστωση πραγματοποιείται με έγγραφη απόδειξη (ναι ή όχι).

**Payment Method:** Μέθοδος πίστωσης (ηλεκτρονική επιταγή, ταχυδρομική επιταγή, αυτόματη τραπεζική μεταφορά ή αυτόματη πληρωμή μέσω πιστωτικής κάρτας).

**Monthly Charges:** Το ποσό της μηνιαίας χρέωσης.

**Total Charges:** Το συνολικό ποσό χρέωσης.

**Churn:** Αν ο πελάτης τελικά εγκατέλειψε την επιχείρηση (ναι ή όχι).

<b>Gender</b>	<b>% churned</b>	<b>Senior Citizen</b>	<b>% churned</b>	<b>PaymentMethod</b>	<b>% churned</b>
Male	26,16%	No	23,61%	Bank transfer (automatic)	16,71%
Female	26,92%	Yes	41,68%	Credit card (automatic)	15,24%
<b>PhoneService</b>	<b>% churned</b>	<b>Partner</b>	<b>% churned</b>	Electronic check	45,29%
No	24,93%	No	32,96%	Mailed check	19,11%
Yes	26,71%	Yes	19,66%	<b>OnlineSecurity</b>	<b>% churned</b>
<b>Dependents</b>	<b>% churned</b>	<b>PaperlessBilling</b>	<b>% churned</b>	No	41,77%
No	31,28%	No	16,33%	No internet service	7,40%
Yes	15,45%	Yes	33,57%	Yes	14,61%
<b>MultipleLines</b>	<b>% churned</b>	<b>InternetService</b>	<b>% churned</b>	<b>TechSupport</b>	<b>% churned</b>
No	25,04%	DSL	18,96%	No	41,64%
No phone service	24,93%	Fiber optic	41,89%	No internet service	7,40%
Yes	28,61%	No	7,40%	Yes	15,17%
<b>OnlineBackup</b>	<b>% churned</b>	<b>DeviceProtection</b>	<b>% churned</b>	<b>Contract</b>	<b>% churned</b>
No	39,93%	No	39,13%	Month-to-month	42,71%
No internet service	7,40%	No internet service	7,40%	One year	11,27%
Yes	21,53%	Yes	22,50%	Two year	2,83%
<b>StreamingTV</b>	<b>% churned</b>	<b>StreamingMovies</b>	<b>% churned</b>		
No	33,52%	No	33,68%		
No internet service	7,40%	No internet service	7,40%		
Yes	30,07%	Yes	29,94%		

Πίνακας 1. Ποσοστό των πελατών που εγκατέλειψαν την εταιρεία για κάθε μεταβλητή.

	<b>Churners</b>	<b>Non churners</b>
<b>Average of tenure</b>	17,979	37,570
<b>Count of MonthlyCharges</b>	76,281	60,806
<b>Average of TotalCharges</b>	1556,603	2567,244

Πίνακας 2. Μέσοι όροι των ποσοτικών μεταβλητών για του πελάτες που εγκατέλειψαν ή όχι την εταιρεία.

Σύμφωνα με τους παραπάνω πίνακες εξάγονται οι παρακάτω πληροφορίες. Όσον αφορά τα δημογραφικά χαρακτηριστικά, φαίνεται το ποσοστό των πελατών που φεύγουν από την εταιρεία είναι υψηλότερο για άγαμους καταναλωτές και πελάτες χωρίς παιδιά. Η ηλικία των πελατών που εγκαταλείπουν την επιχείρηση είναι κατά μέσο όρο μεγαλύτερη. Εν συνεχεία, το ποσοστό αποχώρησης των πελατών είναι μεγαλύτερο σε άτομα που δεν είχαν το πακέτο online security σε σχέση με αυτούς που είχαν. Το ίδιο φαινόμενο συναντάται και στα πρόσθετα πακέτα online backup, device protection και tech support. Επίσης το 41,89% των πελατών που διέθεταν το πακέτο των οπτικών ινών (fiber optic) είναι πελάτες που εγκατέλειψαν την εταιρεία. Οι καταναλωτές που προτιμούν να πληρώνουν μήνα με τον μήνα και να πιστώνουν

τις υπηρεσίες ηλεκτρονικά έχουν το υψηλότερο ποσοστό αποχώρησης σε σχέση με τις άλλες διαθέσιμες επιλογές. Θα μπορούσε, ενδεχομένως, να σχηματιστεί ένα προφίλ καταναλωτή που εγκαταλείπει την επιχείρηση. Το προφίλ αυτό παρουσιάζει άτομα με φαινομενικά λιγότερες οικογενειακές υποχρεώσεις, χωρίς να διαθέτουν τις πρόσθετες υπηρεσίες. Το παραπάνω προφίλ συμπληρώνεται με τον δεύτερο πίνακα, ο οποίος δείχνει ότι η μέση μηνιαία πίστωση είναι μεγαλύτερη σε άτομα που έφυγαν από την εταιρεία. Η μέση συνολική πίστωση δικαιολογείται από τον χρόνο παραμονής της κάθε κατηγορίας (πελάτες που παρέμειναν και που δεν παρέμειναν) στον οργανισμό.

### 3.2 Προσαρμογή των δεδομένων

Όπως έχει ήδη αναφερθεί στην προηγούμενη ενότητα, για την εφαρμογή του γενετικού αλγορίθμου είναι πολλές φορές απαραίτητη η προσαρμογή (ή κωδικοποίηση) των δεδομένων. Στην εν λόγω διπλωματική εργασία η προσαρμογή που πραγματοποιήθηκε έγινε για τρεις βασικούς λόγους.

Ο πρώτος και βασικότερος λόγος είναι για να γίνει η μετατροπή των δεδομένων σε τέτοια μορφή ώστε να δύναται να αναλυθούν από τον αλγόριθμο. Έτσι, όλα τα χαρακτηριστικά που διαθέτουν κατηγορικές μεταβλητές μετατράπηκαν σε αριθμούς από το ένα μέχρι το τέσσερα ανάλογα με τον αριθμό των επιλογών τους. Για παράδειγμα το χαρακτηριστικό που αφορά το φύλο του πελάτη μετατράπηκε από άντρας/γυναίκα σε ένα/δύο. Οι μεταβλητές *tenure*, *Monthly Charges* και *Total Changer* είναι αριθμητικές μεταβλητές και κατά συνέπεια δεν χρήζουν επεξεργασία στο συγκεκριμένο βήμα.

Ο δεύτερος βασικός λόγος που τα δεδομένα επιδέχθηκαν μετατροπή είναι για να γίνει η επεξεργασία τους ευκολότερη. Ανάλογα με την σημαντικότητα των χαρακτηριστικών στο αν ο πελάτης εν τέλει έφυγε ή όχι, έγινε μείωση των τιμών. Για καλύτερη κατανόηση θα χρησιμοποιηθεί το παρακάτω παράδειγμα. Έστω πως στα δεδομένα υπάρχει το χαρακτηριστικό *A* και περιέχει τις τιμές 1,2 και 3. Εάν το ποσοστό της σημαντικότητας του 1 είναι πολύ χαμηλό, δηλαδή το ποσοστό που έχουν οι πελάτες οι οποίοι εγκατέλειψαν την επιχείρηση, τότε υπάρχει πιο απλή σύνδεση μεταξύ των μεταβλητών του χαρακτηριστικού *A* και της εγκατάλειψης των πελατών.

Έτσι, ενώνεται η μεταβλητή 2 και 3 απλοποιώντας το χαρακτηριστικό και στη πρόβλεψη του αλγορίθμου εξετάζεται αν το χαρακτηριστικό A είναι 1 ή όχι. Στη συγκεκριμένη περίπτωση τα δεδομένα που επιδέχθηκαν μια τέτοια μετατροπή είναι: Multiple Lines, Streaming TV, Streaming Movies, Contract και Payment Method.

Ο τρίτος και τελευταίος λόγος μετατροπής των δεδομένων αφορά την μείωση του θορύβου που εν δυνάμει μπορούν να προκαλέσουν τα δεδομένα στην έκβαση του αλγορίθμου. Για να αποφευχθεί ένα τέτοιο γεγονός έγινε κατάλληλη μετατροπή των τιμών όλων των μεταβλητών. Πιο συγκεκριμένα, οι κατηγορικές μεταβλητές κωδικοποιήθηκαν σε δυαδικές (0 ή 1). Ένα χαρακτηριστικό που έχει τις τιμές 1,2 και 3 μετατράπηκε σε τρία χαρακτηριστικά άρα σε τρεις στήλες που η κάθε μια αντιπροσωπεύει την μια τιμή. Έτσι, η κάθε στήλη έχει 1 για κάθε πελάτη που διαθέτει την αντίστοιχη τιμή. Φυσικά, η τελευταία στήλη μπορεί να αποκοπεί καθώς η τιμή της εξαρτάται από τα προηγούμενα. Για παράδειγμα, αν οι δύο πρώτες στήλες έχουν το 0 και οι τιμές (στήλες) είναι τρεις είναι προφανές πως η τρίτη στήλη θα έχει την τιμή 1. Η διαγραφή αυτή αποσκοπεί στην ταχύτερη και αποτελεσματικότερη λειτουργία του αλγορίθμου. Τέλος, οι τιμές που έχουν αριθμητικές μεταβλητές ισοσταθμίσθηκαν με τις υπόλοιπες μετατρέποντας αυτές σε ανάλογες τιμές στο διάστημα μεταξύ 0 και 1.

Με το πέρας των διαδικασιών αυτών το αρχείο δεδομένων έχει 7043 γραμμές και 25 στήλες (οι 24 αφορούν τα χαρακτηριστικά και η 25<sup>η</sup> την έκβαση των πελατών όσον αφορά την επιλογή τους να εγκαταλείψουν την εταιρεία). Όλες οι τιμές του πίνακα αποτελούνται από αριθμούς, είτε δυαδικών μεταβλητών είτε δεκαδικά ψηφία.

<b>Gender</b>	<b>Value</b>	<b>InternetService_DSL</b>	<b>Value</b>	<b>DeviceProtection_No</b>	<b>Value</b>	<b>Contract</b>	<b>Value</b>
Female	1	TRUE	1	TRUE	1	Month-to-month	1
Male	0	FALSE	0	FALSE	0	One year -Two year	0
<b>SeniorCitizen</b>	<b>Value</b>	<b>InternetService_No</b>	<b>Value</b>	<b>DeviceProtection_Yes</b>	<b>Value</b>	<b>PaperlessBilling</b>	<b>Value</b>
Yes	0	TRUE	1	TRUE	1	Yes	0
No	1	FALSE	0	FALSE	0	No	1
<b>Partner</b>	<b>Value</b>	<b>OnlineSecurity_No</b>	<b>Value</b>	<b>TechSupport_No</b>	<b>Value</b>	<b>PaymentMethod</b>	<b>Value</b>
Yes	0	TRUE	1	TRUE	1	Electronic check	1
No	1	FALSE	0	FALSE	0	Mailed check	
<b>Dependents</b>	<b>Value</b>	<b>OnlineSecurity_Yes</b>	<b>Value</b>	<b>TechSupport_Yes</b>	<b>Value</b>	Bank transfer (automatic)	0
Yes	0	TRUE	1	TRUE	1	Credit card (automatic)	
No	1	FALSE	0	FALSE	0	<b>Churn</b>	<b>Value</b>
<b>PhoneService</b>	<b>Value</b>	<b>OnlineBackup_No</b>	<b>Value</b>	<b>StreamingTV</b>	<b>Value</b>	Yes	1
Yes	0	TRUE	1	No - Yes	1	No	0
No	1	FALSE	0	No internet service	0		
<b>MultipleLines</b>	<b>Value</b>	<b>OnlineBackup_Yes</b>	<b>Value</b>	<b>StreamingMovies</b>	<b>Value</b>		
No - Yes	1	TRUE	1	No - Yes	1		
No phone service	0	FALSE	0	No internet service	0		

Πίνακας 3. Οι ποιοτικές μεταβλητές του αρχείου ύστερα από τις προσαρμογές.

	<b>Value</b>
<b>tenure</b>	0-1
<b>MonthlyCharges</b>	0-1
<b>TotalCharges</b>	0-1

Πίνακας 4. Οι κανονικοποιημένες ποσοτικές μεταβλητές του αρχείου ύστερα από τις προσαρμογές.

Πρέπει να σημειωθεί πως η έλλειψη ορισμένων επιλογών των μεταβλητών στον πίνακα οφείλεται στο γεγονός ότι παραλήφθηκαν από την διαδικασία καθώς εννοούνται από τις τιμές των άλλων επιλογών. Η παράλειψη αυτή πραγματοποιήθηκε για την γρηγορότερη και αποτελεσματικότερη εκτέλεση του αλγορίθμου.

### 3.3 Δημιουργία της αντικειμενικής συνάρτησης

Όπως εξηγήθηκε παραπάνω, η πρόβλεψη που η διπλωματική εργασία πραγματοποιεί αφορά το 10% των πιο πιθανών πελάτων. Πιο συγκεκριμένα, χωρίζεται το δείγμα σε 10 τμήματα (δέκατα) που το καθένα περιέχει τον ίδιο αριθμό πελατών. Στόχος του αλγορίθμου είναι να εντάξει όσο το δυνατόν αποτελεσματικότερα τους πελάτες που εγκατέλειψαν την εταιρεία στα πρώτα τμήματα. Το ποσοστό των πελατών αυτών σε σχέση με αυτούς που παρέμειναν θα φθίνει από τα πρώτα τμήματα στα επόμενα (π.χ. ο αριθμός των πελατών που εγκατέλειψαν την εταιρεία στο πρώτο τμήμα θα είναι μεγαλύτερος από το δεύτερο τμήμα και κ.ο.κ.). Επομένως, το αντικείμενο μελέτης του αλγορίθμου είναι το πρώτο τμήμα όπου μετά το πέρας της διαδικασίας θα



περιέχει τους περισσότερους πελάτες που προτίμησαν να εγκαταλείψουν την εταιρεία σε σχέση με αυτούς που παρέμειναν.

Η διαδικασία που ακολουθείται στο αλγόριθμο όσον αφορά την αντικειμενική συνάρτηση περιγράφεται παρακάτω. Αρχικά, η αντικειμενική συνάρτηση ως συνάρτηση προγραμματισμού δέχεται τον πληθυσμό του γενετικού αλγορίθμου ο οποίος αναλύεται σε επόμενη ενότητα. Η συνάρτηση υπολογίζει με τη χρήση του πληθυσμού, δηλαδή των τυχαίων υποψήφιων λύσεων, ένα διάνυσμα βαθμολογιών που περιέχει τις βαθμολογίες του εκάστοτε πελάτη και αποτελείται από 7043 γραμμές. Το διάνυσμα βαθμολογιών είναι μια γραμμική συνάρτηση που υπολογίζεται ως εξής:

$$f = a_0 + \sum_{i=1}^n a_i x_i$$

Όπου  $x_1, \dots, x_n$  είναι οι ανεξάρτητες μεταβλητές (τα χαρακτηριστικά των πελατών που χρησιμοποιούνται για την πρόβλεψη),  $a_0$  είναι ο σταθερός όρος του μοντέλου, και  $a_0, \dots, a_n$  οι συντελεστές των μεταβλητών στο μοντέλο. Το μοντέλο αυτό δίνει μια αριθμητική βαθμολογία για κάθε πελάτη, τέτοια ώστε όσο υψηλότερη είναι η βαθμολογία, τόσο μεγαλύτερη είναι η πιθανότητα ο πελάτης να εγκαταλείψει την εταιρεία.

Όπως θα αναφερθεί και παρακάτω, στην ανάλυση χρησιμοποιούνται 25 μεταβλητές, οπότε στον αλγόριθμο ο πληθυσμός περιλαμβάνει 25 στοιχεία για τους συντελεστές  $a_1, \dots, a_{24}$  και τον σταθερό όρο  $a_0$ .

Ανάλογα με τις βαθμολογίες των πελατών που προκύπτουν από την παραπάνω συνάρτηση, οι πελάτες ταξινομούνται σε φθίνουσα σειρά και αποθηκεύεται μόνο το πρώτο 10% (704 γραμμές).

Η αντικειμενική συνάρτηση ως μαθηματική έκφραση η οποία αποτελεί το κριτήριο της βελτιστοποίησης του αλγορίθμου είναι ουσιαστικά το ποσοστό των χαμένων πελατών μεταξύ εκείνων που βρίσκονται στο άνω 10% της παραπάνω ταξινομημένης λίστας. Το μέγεθος αυτό είναι γνωστό ως top decile lift και χρησιμοποιείται εκτενώς για την αξιολόγηση της προβλεπτικής ικανότητας ανάλογων μοντέλων (Tsiptsis & Chorianoopoulos, 2009), (Linoff & Berry, 2011). Ωστόσο για να αποφευχθεί η εύρεση των λύσεων αυστηρά και μόνο στα δεδομένα που εκπαιδεύεται ο αλγόριθμος καθιστώντας τον ανίκανο να προβλέπει με έναν γενικό κανόνα προστέθηκε και μια

συνάρτηση ποινής στην αντικειμενική συνάρτηση, η οποία πριμοδοτεί λύσεις στις οποίες οι συντελεστές των μεταβλητών  $\alpha_1, \dots, \alpha_{24}$  είναι μικροί (κατά απόλυτη τιμή).

Συνεπώς, η αντικειμενική συνάρτηση  $F$  (σε μορφή μεγιστοποίησης) διαμορφώνεται τελικά ως εξής:

$$F = \pi - \frac{1}{24} \sum_{i=1}^n a_i^2$$

όπου  $\pi$  είναι το ποσοστό των πελατών που εγκαταλείπουν τη εταιρεία στο άνω 10% των βαθμολογιών. Η προσθήκη του επιπλέον όρου ποινής βασίζεται στην αρχή του Tikhonov regularization (Tikhonov, Goncharsky, & Stepanov, 1995), η οποία αποσκοπεί στην ανάπτυξη ευσταθών μοντέλων πρόβλεψης. Ανάλογη προσέγγιση χρησιμοποιείται σε διάφορες δημοφιλείς τεχνικές μηχανικής μάθησης, όπως οι μηχανές διανυσμάτων υποστήριξης (support vector machines, (Schölkopf & Smola, 2002) ) και το μοντέλο LASSO (least absolute shrinkage and selection operator, (Tibshirani, 1996) ). Πρέπει να σημειωθεί ότι για πριν την εφαρμογή του αλγορίθμου όλες οι μεταβλητές κανονικοποιούνται μεταξύ 0 και 1, ώστε όλες οι μεταβλητές να είναι στην ίδια κλίμακα και να έχει νόημα ο υπολογισμός της παραπάνω συνάρτησης ποινής.

### 3.4 Επιλογή διαθέσιμων παραμέτρων γενετικού αλγόριθμου

Όπως αναφέρθηκε και στο δεύτερο κεφάλαιο της διπλωματικής εργασίας, η δομή του αλγορίθμου περιέχει τρία βασικά στάδια την επιλογή, τη διασταύρωση και τη μετάλλαξη. Επίσης απαραίτητο στοιχείο του αλγορίθμου είναι η αρχικοποίηση του πληθυσμού.

Αν και η αρχικοποίηση του πληθυσμού γίνεται αυτόματα στο περιβάλλον MATLAB με το κάλεσμα της συνάρτησης που υλοποιεί τον γενετικό αλγόριθμο πρέπει να καθοριστούν από το χρήστη δυο μεταβλητές. Και τα δύο στοιχεία που ορίζονται αφορούν το μέγεθος του πληθυσμού. Ο πρώτος παράγοντας είναι οι στήλες -δηλαδή το μέγεθος των λύσεων- που θα έχει και ο δεύτερος το μέγεθος του, δηλαδή το αριθμό των χρωμοσωμάτων. Ο τελικός πίνακας των δεδομένων περιέχει 24 χαρακτηριστικά (στήλες) συν μια στήλη που αφορά την κατάσταση του εκάστοτε

πελάτη, αν δηλαδή έφυγε από την εταιρεία ή όχι. Στην παρούσα εργασία ο πληθυσμός ορίστηκε 250 και το πλήθος των χαρακτηριστικών 25. Η λύση που θα προκύψει και θα αφορούν τα χαρακτηριστικά θα αποτελείται από 25 στοιχεία.

Για την επιλογή, διασταύρωση και μετάλλαξη του αλγορίθμου δοκιμάστηκαν διάφοροι συνδυασμοί των συναρτήσεων που διαθέτει η υλοποίηση του MATLAB για την τις εν λόγω διαδικασίες. Για την επιλογή του αλγορίθμου ελέγχθηκαν οι συναρτήσεις: uniform, roulette και remainder. Για τη μετάλλαξη ελέγχθηκαν οι συναρτήσεις: gaussian και uniform. Και τέλος για την διασταύρωσης δοκιμάστηκαν οι συναρτήσεις: scattered, heuristic, singlepoint, twopoint και arithmetic.

Παρακάτω αναλύονται οι συναρτήσεις επιλογής:

Uniform: Αυτή η συνάρτηση δημιουργεί μια γραμμή στην οποία τοποθετούνται κατά μήκος της όλες οι υποψήφιες λύσεις (χρωμοσώματα). Το τμήμα της γραμμής που καταλαμβάνει το κάθε χρωμόσωμα είναι ανάλογο των τιμών του. Ο αλγόριθμος κινείται κατά μήκος της γραμμής με σταθερό βήμα και επιλέγει για γονείς τα χρωμοσώματα που αντιστοιχούν στα τμήματα που συναντά. Το διάστημα του πρώτου βήματος που πραγματοποιεί ο αλγόριθμος είναι ένας τυχαίος αριθμός και η τιμή του είναι μικρότερη από αυτή των επόμενων βημάτων.

Remainder: Η συνάρτηση επιλέγει τους γονείς από το ακέραιο τμήμα της προσαρμοσμένης τιμής (scaled value) της αντικειμενικής που φέρει το χρωμόσωμα.<sup>1</sup> Για παράδειγμα αν έχει την τιμή 2.3 θα επιλεγθεί δύο φορές σαν γονέας επειδή το ακέραιο κομμάτι του είναι δύο. Αφού καθοριστούν οι γονείς με αυτόν τον τρόπο οι υπόλοιποι επιλέγονται στοχαστικά. Η πιθανότητα επιλογής ενός γονέα σε αυτό το βήμα είναι ανάλογη του κλασματικού τμήματος της προσαρμοσμένης τιμής της αντικειμενικής συνάρτησης.

Roulette: Η συνάρτηση της ρουλέτας επιλέγει τους γονείς με την προσομοίωση του τυχερού παιγνίου. Για κάθε χρωμόσωμα υπολογίζεται η τιμή του μέσω της αντικειμενικής συνάρτησης. Η τιμή αυτή κανονικοποιείται με τις τιμές των υπόλοιπων χρωμοσωμάτων (αθροίζουν στο 1). Η περιοχή του τμήματος του τροχού που καταλαμβάνει κάθε χρωμόσωμα είναι ανάλογη της τιμής αυτής.

---

<sup>1</sup> Η προσαρμογή των τιμών της αντικειμενικής συνάρτησης μίας λύσης (χρωμόσωμα) γίνεται ανάλογα με την κατάταξη της λύσης σε σχέση με τις υπόλοιπες ως προς την τιμή της αντικειμενικής συνάρτησης.

Παρακάτω αναλύονται οι συναρτήσεις μετάλλαξης:

**Gaussian:** Η συνάρτηση αυτή προσθέτει έναν τυχαίο αριθμό σύμφωνα με την κανονική κατανομή με μέσο όρο μηδέν στον γονέα που επιλέχθηκε για μετάλλαξη.

**Uniform:** Η συνάρτηση, αυτή τη φορά, χωρίζεται σε δύο βήματα. Αρχικά ο αλγόριθμος επιλέγει ένα μέρος του χρωμοσώματος που θα μεταλλαχτεί με μια πιθανότητα μετάλλαξης (η προεπιλεγμένη πιθανότητα είναι 1%). Στο δεύτερο βήμα απλά αντικαθιστά τις αντίστοιχες τιμές με έναν ομοιόμορφα κατανομημένο τυχαίο αριθμό επιλεγμένο από το εύρος των τιμών που διαθέτει το χρωμόσωμα.

Παρακάτω αναλύονται οι συναρτήσεις διασταύρωσης:

**Scattered:** Η συνάρτηση αυτή δημιουργεί ένα τυχαίο διάνυσμα που έχει το μέγεθος του χρωμοσώματος με δυαδικές τιμές. Οι θέσεις των τιμών που έχουν τον αριθμό ένα είναι οι αντίστοιχες που θα κληρονομηθούν από τον πρώτο γονέα και οι θέσεις των τιμών που έχουν μηδέν είναι οι αντίστοιχες που θα κληρονομηθούν από τον δεύτερο.

**Heuristic:** Η συνάρτηση επιστρέφει το χρωμόσωμα με βάση την τιμή της αντικειμενικής συνάρτησης που οι γονείς δημιουργούν. Προσθέτει στον γονέα με την χειρότερη τιμή της αντικειμενικής συνάρτησης τη διαφορά των τιμών των δύο γονέων πολλαπλασιασμένη με ένα ποσοστό (η προεπιλογή αυτού του ποσοστού είναι 1.2).

**Singlepoint:** Η εν λόγω συνάρτηση επιλέγει έναν αριθμό  $k$  από το ένα μέχρι και το μέγεθος του χρωμοσώματος. Το νέο χρωμόσωμα θα περιέχει τις τιμές από τον πρώτο γονέα για τις πρώτες  $k$  θέσεις και τις τιμές από τον δεύτερο για τις υπόλοιπες.

**Twopoint:** Η συνάρτηση αυτή είναι παρόμοια της προηγούμενης μόνο που αυτή τη φορά ο αλγόριθμος επιλέγει δύο αριθμούς. Οι θέσεις του νέου χρωμοσώματος που είναι ανάμεσα σε αυτούς τους δύο αριθμούς επιλέγονται από τον δεύτερο γονέα και οι υπόλοιπες από τον πρώτο.

**Arithmetic:** Είναι η πιο απλή συνάρτηση που χρησιμοποιεί ο αλγόριθμος. Ουσιαστικά το νέο χρωμόσωμα είναι ο σταθμισμένος μέσος όρος των δύο γονέων.

### 3.5 Εκτέλεση του γενετικού αλγορίθμου

Στη εκτέλεση του γενετικού αλγορίθμου περιλαμβάνονται όλες οι μεταβλητές που αναφέρονται στην παραπάνω ενότητα δοκιμάζοντας όλους τους πιθανούς συνδυασμούς των μεταβλητών αυτών. Το δείγμα χωρίστηκε σε δύο τμήματα, το τμήμα εκπαίδευσης και το τμήμα ελέγχου. Δηλαδή, ο αλγόριθμος εκπαιδεύεται και παράγει τις βέλτιστες λύσεις αξιοποιώντας το τμήμα εκπαίδευσης και τα αποτελέσματα του αξιολογούνται από το τμήμα ελέγχου.

Με κύριο στόχο την εύρεση αξιόπιστης λύσης, ο διαχωρισμός αυτός έγινε δέκα φορές και ήταν τυχαίος χρησιμοποιώντας τη διαδικασία 10-fold cross validation. Επομένως ο αλγόριθμος εκτελέστηκε δέκα φορές, μια εκτέλεση για κάθε διαχωρισμό. Κατά συνέπεια το κριτήριο αξιολόγησης για ένα συνδυασμό των παραμέτρων του αλγορίθμου είναι ο μέσος όρος του ποσοστού των πελατών που εγκατέλειψαν την εταιρεία για τις δέκα αυτές φορές.

Με το πέρας της ανωτέρω διαδικασίας τα στοιχεία που αποθηκεύονται ανά εκτέλεση είναι τα ποσοστά των χαμένων πελατών για τα τμήματα εκπαίδευσης και ελέγχου, ο μέσος όρος των ποσοστών αυτών για τις δέκα εκτελέσεις, και τα διανύσματα των λύσεων για κάθε εκτέλεση.

### 3.6 Αποτελέσματα γενετικού αλγορίθμου

Το σημαντικότερο κριτήριο αξιολόγησης του γενετικού αλγορίθμου είναι το ποσοστό των χαμένων πελατών στο ανώτερο 10% του δείγματος. Παρακάτω παρατίθενται ο πίνακας αποτελεσμάτων για κάθε συνδυασμό παραμέτρων του αλγορίθμου.

Selection	Mutation	Crossover	Mean_train	Mean_test
uniform	uniform	scattered	0,987	0,9786
uniform	uniform	heuristic	0,9972	0,9757
uniform	uniform	singlepoint	0,9842	0,9843
uniform	uniform	twopoint	0,97	0,9814
uniform	uniform	arithmetic	0,9954	0,9814
uniform	gaussian	scattered	0,77	0,8
uniform	gaussian	heuristic	0,9618	0,97
uniform	gaussian	singlepoint	0,8258	0,8271
uniform	gaussian	twopoint	0,8671	0,8558
uniform	gaussian	arithmetic	0,9796	0,9757
roulette	uniform	scattered	0,9843	0,987
roulette	uniform	heuristic	0,987	0,9843
roulette	uniform	singlepoint	0,9875	0,9771
roulette	uniform	twopoint	0,9888	0,98
roulette	uniform	arithmetic	0,9956	0,9729
roulette	gaussian	scattered	0,982	0,9729
roulette	gaussian	heuristic	0,9962	0,9857
roulette	gaussian	singlepoint	0,9429	0,9474
roulette	gaussian	twopoint	0,9788	0,9786
roulette	gaussian	arithmetic	0,9957	0,9886
remainder	uniform	scattered	0,9891	0,9829
remainder	uniform	heuristic	0,9948	0,9757
remainder	uniform	singlepoint	0,9875	0,9843
remainder	uniform	twopoint	0,9861	0,9771
remainder	uniform	arithmetic	0,9926	0,9743

Πίνακας 5. Αποτελέσματα του γενετικού αλγορίθμου ανά συνδυασμό για πληθυσμό ίσο με 250.

Παρατηρείται πως ο καλύτερος συνδυασμός που επέφερε στο τμήμα τη εκπαίδευσης είναι ο 2<sup>ος</sup> και ο καλύτερος συνδυασμός όσον αφορά το αποτέλεσμα στο τμήμα ελέγχου είναι ο 20<sup>ος</sup>. Τα ποσοστά αυτά είναι αρκετά ικανοποιητικά και έχουν σχετικά μικρές διαφορές πλην ορισμένων εξαιρέσεων. Ο 6<sup>ος</sup> συνδυασμός για παράδειγμα, διαμορφώνει το πρώτο δέκατο με 80% πελάτες που έφυγαν από την εταιρεία και με 20% που δεν έφυγαν. Για να γίνουν τα μεγέθη πιο κατανοητά, η διαφορά με τον καλύτερο συνδυασμό (στο τμήμα των ελέγχου) είναι 0,1886 (18,86%). Στους 704 πελάτες που διαθέτει κάθε δέκατο το ποσοστό αυτό αντιστοιχεί σε 132 πελάτες. Δηλαδή, ο χειρότερος συνδυασμός κατέταξε λανθασμένα στο δέκατο που μελετά 132 παραπάνω πελάτες σε σχέση με τον καλύτερο συνδυασμό. Είναι προφανές πως το ποσοστό επιτυχίας στο τμήμα των ελέγχων είναι η πιθανότητα με την οποία ο αλγόριθμος προβλέπει τους 10% πιο πιθανούς πελάτες. Συνεπώς, ο καλύτερος

συνδυασμός προβλέπει τους 10% πιο πιθανούς πελάτες με πιθανότητα επιτυχίας 98,86%.

Οι λύσεις που παρέχει ο καλύτερος συνδυασμός για κάθε μια από τις δέκα εκτελέσεις του φαίνονται στους παρακάτω πίνακες.

Συντελεστές	# Εκτέλεσης									
	1	2	3	4	5	6	7	8	9	10
α0	1,091	0,124	1,717	0,120	-1,471	1,154	0,386	1,299	0,087	0,787
α1	0,001	0,001	0,002	0,005	0,003	0,001	-0,001	0,012	-0,011	-0,001
α2	0,010	-0,012	0,018	0,006	-0,013	0,000	0,007	-0,016	0,002	-0,003
α3	-0,006	-0,021	-0,019	-0,023	-0,016	-0,018	-0,004	-0,009	0,006	0,012
α4	-0,002	0,014	0,011	0,004	0,018	0,013	-0,005	0,001	-0,007	-0,010
α5	0,054	0,026	0,032	0,028	0,014	0,028	0,023	0,037	0,016	0,019
α6	-0,003	0,007	-0,016	-0,003	-0,018	0,011	0,011	0,044	-0,002	0,011
α7	0,018	0,009	0,001	-0,009	0,021	0,045	-0,010	-0,007	-0,022	0,001
α8	0,036	-0,019	0,053	0,022	0,001	-0,010	0,017	0,003	0,005	0,043
α9	0,049	-0,001	0,014	0,030	-0,025	0,008	0,016	0,013	-0,022	0,008
α10	0,021	-0,006	-0,001	-0,013	0,022	0,002	0,003	-0,044	0,001	-0,053
α11	0,008	0,043	-0,023	0,021	-0,013	0,026	0,012	-0,014	-0,019	0,013
α12	-0,010	0,001	-0,034	-0,029	-0,009	-0,036	-0,008	0,008	0,004	-0,016
α13	0,020	-0,016	-0,029	-0,018	0,001	0,017	-0,001	-0,001	-0,006	-0,028
α14	-0,001	0,020	-0,008	0,014	-0,016	-0,031	0,000	0,000	-0,005	0,004
α15	0,004	-0,033	0,021	-0,006	0,013	-0,015	0,006	0,006	-0,004	0,018
α16	-0,037	-0,047	-0,014	-0,014	-0,005	-0,037	-0,012	-0,005	-0,004	-0,020
α17	0,027	-0,015	-0,008	0,016	-0,009	-0,009	0,000	0,014	-0,020	0,022
α18	-0,072	-0,059	-0,009	-0,027	-0,023	-0,009	-0,015	-0,035	-0,023	-0,025
α19	-0,003	-0,040	0,016	-0,040	-0,033	-0,049	-0,012	-0,008	-0,005	0,018
α20	-0,031	-0,051	-0,062	-0,050	-0,028	-0,044	-0,009	-0,066	-0,021	-0,050
α21	0,028	0,011	0,002	0,001	0,001	0,007	0,010	0,008	-0,007	-0,001
α22	-0,013	-0,014	-0,020	-0,008	-0,016	-0,019	-0,026	-0,007	-0,019	-0,016
α23	-0,060	-0,021	0,002	0,013	-0,034	-0,004	-0,008	0,009	-0,016	-0,015
α24	-0,003	0,048	0,028	0,021	-0,016	-0,005	0,006	-0,038	0,018	0,009

Πίνακας 6. Διανύσματα λύσεων του αλγορίθμου με τον βέλτιστο συνδυασμό.

Άξιο αναφοράς είναι το γεγονός πως ο σταθερός όρος των λύσεων είναι αισθητά μεγαλύτερος κατά απόλυτη τιμή από τις υπόλοιπες τιμές σε κάθε εκτέλεση.

### 3.7 Λογιστική παλινδρόμηση

Με στόχο την καλύτερη αξιολόγηση των αποτελεσμάτων του γενετικού αλγορίθμου εφαρμόστηκε και το μοντέλο της λογιστικής παλινδρόμησης. Η λογιστική παλινδρόμηση είναι μια διαδοσμένη στατιστική μέθοδος που εκτιμά τις πιθανότητες των κατηγορικών στόχων, όπου στην παρούσα εργασία ο κατηγορικός στόχος είναι η επιλογή του πελάτη να εγκαταλείψει την εταιρεία ή όχι. Η εν λόγω μέθοδος είναι

ανάλογη με τη γραμμική παλινδρόμηση αλλά για τα κατηγορικά αποτελέσματα. Χρησιμοποιεί το γενικευμένο γραμμικό μοντέλο και υπολογίζει τους συντελεστές παλινδρόμησης που αντιπροσωπεύουν την επίδραση των μεταβλητών στις πιθανότητες εμφάνισης των κατηγορικών στόχων (Hosmer Jr, Lemeshow, & Sturdivant, 2013).

Η λογιστική παλινδρόμηση για να είναι συγκρίσιμη με τον γενετικό αλγόριθμο δέχτηκε σαν είσοδο τα ίδια δέκα χωρίσματα των δεδομένων που είχε και ο γενετικός αλγόριθμος. Τα τυχαία αυτά χωρίσματα που αφορούν το δείγμα εκπαίδευσης και το τμήμα ελέγχου χρησιμοποιήθηκαν αντίστοιχα με την ίδια σειρά δέκα φορές.

Ο μέσος όρος των ποσοστών των χαμένων πελατών για το πρώτο 10% του δείγματος για τις δέκα εκτελέσεις είναι ίδιος με το αποτέλεσμα που είχε ο γενετικός αλγόριθμος, δηλαδή 98,86%. Κατά συνέπεια η λογιστική παλινδρόμηση προβλέπει με την ίδια πιθανότητα επιτυχώς τους 10% πιο πιθανούς πελάτες με τον γενετικό αλγόριθμο.



## 4. Συμπεράσματα

Το φαινόμενο της απώλειας των πελατών αναλύεται εκτενώς από ερευνητικές εργασίες τα τελευταία χρόνια. Οι λόγοι που οι οργανισμοί απαιτούν την διάθεση τέτοιων προβλέψεων είναι αναγκαίοι για τη βιωσιμότητα και εξέλιξη τους. Η ανάγκη αυτή εκφράζεται σε οικονομικούς παράγοντες. Η απόκτηση νέου πελάτη είναι ακριβότερη από τη διατήρηση του και ο τερματισμός των υπηρεσιών επιφέρει ζημία στην επιχείρηση. Οι ανωτέρω λόγοι σε συνδυασμό με τον μεγάλο ανταγωνισμό που οδηγούν τους πελάτες σε ευκολότερη εγκατάλειψη μετατρέπουν την εν λόγω ανάγκη σε επιτακτική.

Η συγκεκριμένη πρόβλεψη της απώλειας πελατών αποτελεί ένα από τα πολλά εργαλεία που διαθέτει η διαχείριση των πελατών σαν σύστημα. Αυτά τα εργαλεία αφορούν κατά κύριο λόγο τον κύκλο ζωής του πελάτη. Η απόκτηση νέων πελατών και η κερδοφορία από τους ήδη υπάρχοντες πελάτες αποτελούν λειτουργίες του συστήματος της διαχείρισης πελατών και διαθέτουν εργαλεία παρόμοιων τεχνικών.

Η διπλωματική εργασία αναλαμβάνει την πρόβλεψη των χαμένων πελατών για μια εταιρία τηλεπικοινωνιών. Για την εφαρμογή αυτή υλοποιήθηκε ο απλός γενετικός αλγόριθμος. Οι γενετικοί αλγόριθμοι ανήκουν στους εξελικτικούς αλγορίθμους, όπως ο εξελικτικός προγραμματισμός, στρατηγικές εξέλιξης, συστήματα επίλυσης προβλημάτων κ.α. που μιμούνται στον τρόπο λειτουργίας τους τη βιολογική θεωρία που συναντάται στη φύση.

Τα αποτελέσματα της εργασίας αποτελούν πολύτιμη πληροφορία για την εταιρεία καθώς δύναται να χρησιμοποιηθούν και σε μελλοντικά δεδομένα. Η εφαρμογή των αποτελεσμάτων καθιστά τον οργανισμό ικανό να αξιολογήσει τους πελάτες του με στόχο την εύρεση των πιθανότερων προς αποχώρηση.

Οι εικοσιπέντε διαθέσιμοι συνδυασμοί της επιλογής, διασταύρωσης και μετάλλαξης που δοκιμάστηκαν επέφεραν ως επί των πλείστων παρόμοια αποτελέσματα στη διαδικασία ελέγχου. Ωστόσο, ο συνδυασμός με το καλύτερο αποτέλεσμα καθιστά τον αλγόριθμο ικανό να προβλέπει με πιθανότητα επιτυχίας 98,86% τους 10% πιο πιθανούς πελάτες που θα εγκαταλείψουν την εταιρεία. Η πρόβλεψη αυτή επαληθεύεται από τα αποτελέσματα της εφαρμογής της λογιστικής παλινδρόμησης

στα δεδομένα. Η προβλεπτική ικανότητα της παλινδρόμησης είναι ίδια με αυτή του γενετικού αλγορίθμου.

Φυσικά, η πρόβλεψη που κάνει ο γενετικός αλγόριθμος είναι ένα από τα πολυάριθμα προβλήματα που καλείται να επιλύσει. Ο εν λόγω αλγόριθμος αποτελεί βασικό εργαλείο της μηχανικής μάθησης. Ο ίδιος όπως και τα υπόλοιπα εργαλεία υπηρετούν στόχους που θέτει η διαχείριση πελατών σε μια επιχείρηση. Τα εργαλεία αυτά εξελίσσονται και θα συνεχίσουν να εξελίσσονται απασχολώντας του ερευνητές ακόμα και σε τομείς που δεν ανήκουν στον χώρο των επιχειρήσεων.

Εν κατακλείδι, η παρούσα διπλωματική εργασία είχε ως απώτερο σκοπό την πρόβλεψη των πιθανότερων χαμένων πελατών. Ο στόχος αυτός αποτελεί χρήσιμο εργαλείο για έναν οργανισμό καθώς η πρόβλεψη αυτή θα μπορούσε ενδεικτικά να επιφέρει την δυνατότητα στη εταιρεία να διατηρήσει τους υποψήφιους προς εγκατάλειψη πελάτες χρησιμοποιώντας διάφορα μέσα όπως καμπάνιες προσφορών, εκπτωτικά πακέτα κ.α.

## References

- Buttle, F. (2004). *Customer relationship management*. Oxford: Routledge.
- De Jong, A. K. (1975). *Analysis of the behavior of a class of genetic adaptive systems*. Michigan: Department of Computer and Communication Sciences, University of Michigan.
- Eiben, A. E., & Schoenauer, M. (2002). Evolutionary computing. *Information Processing Letters*, 82(1), 1-6.
- Eiben, A. E., & Smith, J. E. (2013). *Introduction to Evolutionary Computing*. Berlin: Springer.
- Farquad, M. H., Ravi, V., & Raju, S. B. (2014). Churn prediction using comprehensible support vector machine: And analytical CRM application. *Applied Soft Computing*, 19, 31-40.
- Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Hoboken, NJ: John Wiley & Sons, Inc.
- kaggle. (2018, 2 23). Ανάκτηση από kaggle: <https://www.kaggle.com/blastchar/telco-customer-churn>
- Khodakarami, F., & Chan, Y. E. (2014). Exploring the role of customer relationship management (CRM) systems in customer knowledge creation. *Information & Management*, 51(1), 27-42.
- Kourdi, J. (2011). *The marketing century: How marketing drives business and shapes society*. John Wiley & Sons.
- Krishna, G. J., & Ravi, V. (2016). Evolutionary computing applied to customer relationship management: A survey. *Engineering Applications of Artificial Intelligence*, 56, 30-59.
- Kumar, V., & Petersen, J. A. (2012). *Statistical methods in customer relationship management*. New York: 2 John Wiley & Sons.

- Linoff & Berry. (2011). *Data Mining Techniques: For Marketing, Sales and Customer Relationship Management*. Indianapolis, Indiana: WILEY.
- Linoff, G. S., & Berry, A. M. (2011). *For Marketing, Sales and Customer Relationship Management*. Indianapolis: Wiley Publishing, Inc.
- Lu, N., Lin, H., Lu, J., & Zhang, G. (2012). A customer churn prediction model in telecom industry using boosting. *IEEE Transactions on Industrial Informatics*, 10(2), 1659-1665.
- Mishra, A., & Mishra, D. (2009). Customer Relationship Management: Implementation Process Perspective. *Acta Polytechnica Hungarica*, 6(4), 83-99.
- Ngai, E. W., Xiu, L., & Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592-2602.
- Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. *Expert Systems with Applications*, 38(3), 1999-2006.
- Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge: Massachusetts: MIT Press.
- Sean, K. (2000). Analytical CRM: The fusion of data and intelligence. *Interactive Marketing*, 1(3), 262-267.
- Sharma, A., & Panigrahi, P. K. (2013). A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. *International Journal of Computer Applications*, 26-31.
- Swift, R. S. (2001). *Accelerating customer relationships: Using CRM and relationship technologies*. USA: Prentice Hall Professional.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the lasso. *Journal of the Royal Statistical Society*, 58(1), 88-267.
- Tikhonov, A., Goncharsky, A., & Stepanov, V. A. (1995). *Numerical Methods for the Solutions of ill-Posed Problems*. Dordrecht: Springer.

- Torggler, M. (2008). The Functionality and Usage of CRM Systems. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, 2(5), 771-779.
- Tsai, C.-F., & Lu, Y.-H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36(10), 12547-12553.
- Tsiptsis, K. K., & Chorianopoulos, A. (2009). *Data Mining Techniques in CRM: Inside Customer Segmentation*. John Wiley & Sons.
- Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and computing*, 4, 65-85.
- Αβραμούλη, Καραγεώργος, & Ράπτη. (2015). Στο *Συστήματα Διαχείρισης Πελατειακών Σχέσεων (Customer Relationship Management–CRM)*.
- Γεωργόπουλος, Φ. Ε., & Λυκοθανάσης, Σ. Δ. (1999). *Εισαγωγή στους Γενετικούς Αλγόριθμους*. Πάτρα: Πανεπιστήμιο Πατρών Πολυτεχνική Σχολή.
- Καμπουρλάζος, Β., & Παπακώστας, Γ. (2015). *Εξελικτικός Υπολογισμός*. Κάλλιπος.
- Λυκοθανάσης, Σ. (2001). *Γενετικοί αλγόριθμοι και εφαρμογές*. Πάτρα: ΕΣΠΠ ΕΚΔΟΤΙΚΗ Ε.Π.Ε.