2025-01

þÿ — Ç Á ® Ã · ‘ ™ Ã µ ¼ µ ³ ¬ » ± ´ µ ´ ¿ ¼

þÿ Ä · ½ À Á Ì ² » µ È · À µ » ± Ä µ ¹ ± º ® Â

þÿ Ã Å ¼ À µ Á ¹ Æ ¿ Á ¬ Â Ã Ä ¿ ½ Ä ¿ ¼ - ±

þÿ œ µ Ã Ã ± Á ¹ Ä ¬ º · Â , • ¼ ¼ ± ½ ¿ Å ® »

þÿ œ µ Ä ± À Ä Å Ç ¹ ± º Ì Ã Ä ± » · Á ¿ Æ ¿ Á ¹ ± º ¬ £ Å Ã Ä ® ¼ ± Ä ± º ± ¹ ¨ · Æ ¹ ± º ® š ± ¹ ½ ¿ Ä ¿ ¼ ¯ ± ,
þÿ º ± ¹ • À ¹ Ã Ä ® ¼ · Â ¥ À ¿ » ¿ ³ ¹ Ã Ä Î ½ , ± ½ µ À ¹ Ã Ä ® ¼ ¹ ¿ • µ ¬ À ¿ » ¹ Â ¬ Æ ¿ Å

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΨΗΦΙΑΚΗ ΚΑΙΝΟΤΟΜΙΑ

Η ΧΡΗΣΗ ΑΙ ΣΕ ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΠΕΛΑΤΕΙΑΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΣΤΟΝ ΤΟΜΕΑ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΜΕΣΣΑΡΙΤΑΚΗΣ ΕΜΜΑΝΟΥΗΛ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ ΛΑΒΔΑΣ ΣΠΥΡΟΣ

ΠΑΦΟΣ ΙΑΝΟΥΑΡΙΟΣ 2025

Πανεπιστήμιο Νεάπολις Πάφος

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΨΗΦΙΑΚΗ ΚΑΙΝΟΤΟΜΙΑ

Η ΧΡΗΣΗ ΑΙ ΣΕ ΜΕΓΑΛΑ ΔΕΔΟΜΕΝΑ ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΠΕΛΑΤΕΙΑΚΗΣ ΣΥΜΠΕΡΙΦΟΡΑΣ ΣΤΟΝ ΤΟΜΕΑ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Διπλωματική Εργασία η οποία υποβλήθηκε προς απόκτηση μεταπτυχιακού τίτλου σπουδών στα πληροφοριακά συστήματα και στην ψηφιακή καινοτομία στο Πανεπιστήμιο Νεάπολις Πάφος.

ΜΕΣΣΑΡΙΤΑΚΗΣ ΕΜΜΑΝΟΥΗΛ

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

ΛΑΒΑΔΑ ΣΠΥΡΟΣ

ΣΑΒΒΑΣ ΧΑΤΖΗΧΡΙΣΤΟΦΗΣ

ΕΛΕΝΑ ΚΑΚΟΥΛΛΗ

ΠΑΦΟΣ ΙΑΝΟΥΑΡΙΟΣ 2025

Πνευματικά δικαιώματα

Σελίδα Εγκυρότητας

Ονοματεπώνυμο Φοιτητή: Μεσσαριτάκης Εμμανουήλ

Τίτλος Διπλωματικής Εργασίας: Η χρήση AI σε μεγάλα δεδομένα για την πρόβλεψη πελατειακής συμπεριφοράς στον τομέα τηλεπικοινωνιών.

Η παρούσα Διπλωματική Εργασία εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση εξ αποστάσεως μεταπτυχιακού τίτλου στο Πανεπιστήμιο Νεάπολις και εγκρίθηκε στις 10 Φεβρουαρίου 2025 από τα μέλη της Εξεταστικής Επιτροπής.

Εξεταστική Επιτροπή:

Πρώτος επιβλέπων (Πανεπιστήμιο Νεάπολις Πάφος)……[Λάβδας Σπύρος, βαθμίδα]

Μέλος Εξεταστικής Επιτροπής: ……[Χατζηχριστοφής Σάββας, βαθμίδα]

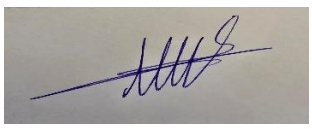
Μέλος Εξεταστικής Επιτροπής: ……[Κακουλλή Ελενα, βαθμίδα]

Ή ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

Ο Μεσσαριτάκης Εμμανουήλ, γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα ότι η παρούσα εργασία με τίτλο «Η χρήση AI σε μεγάλα δεδομένα για την πρόβλεψη πελατειακής συμπεριφοράς στον τομέα τηλεπικοινωνιών», αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές που έχω χρησιμοποιήσει, έχουν δηλωθεί κατάλληλα στις βιβλιογραφικές παραπομπές και αναφορές. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή.

Ο Δηλών

Μεσσαριτάκης Εμμανουήλ



Πρόλογος και ευχαριστίες

Θεωρώ πως αυτή η Διπλωματική θα με βοηθήσει να κατανοήσω και να εμβαθύνω περισσότερο και στον κλάδο που εργάζομαι εδώ και 15 χρόνια αλλά και τον τρόπο χρήση της AI στα μεγάλα Δεδομένα.

Θα ήθελα να εκφράσω τις ευχαριστίες μου σε όλους όσους με υποστήριξαν καθ' όλη τη διάρκεια αυτής της προσπάθειας.

Πρώτα απ' όλα, όμως ευχαριστώ τον **επιβλέποντα καθηγητή μου, κο Λάβδα Σπύρο**, για την πολύτιμη καθοδήγησή του, τις εποικοδομητικές παρατηρήσεις και την αδιάκοπη υποστήριξη του καθ' όλη τη διάρκεια της εργασίας. Οι γνώσεις και οι συμβουλές του αποτέλεσαν καθοριστικό παράγοντα για την ολοκλήρωση αυτής της μελέτης.

Δευτερευόντως θα ήθελα να ευχαριστήσω τα άλλα 2 μέλη της τριμελής μου επιτροπής, την

κα Κακουλλή Έλενα και κο Χατζηχριστοφή Σάββα.

Ιδιαίτερες ευχαριστίες απευθύνω στους **γονείς μου**, οι οποίοι με στήριξαν αδιάκοπα με την αγάπη, την υπομονή και την κατανόησή τους.

Επίσης, εκφράζω την ευγνωμοσύνη μου στη **σύζυγό μου, Μαρία**, για τη συνεχή παρουσία της, τη στήριξη και την αφοσίωσή της. Η κατανόηση, η υπομονή και η ενθάρρυνσή της μου έδωσαν τη δύναμη να συνεχίσω και να ολοκληρώσω αυτό το έργο, παρά τις δυσκολίες.

Περίληψη

Η διατήρηση πελατών αποτελεί μία από τις σημαντικότερες στρατηγικές προκλήσεις για την τηλεπικοινωνιακή βιομηχανία, έναν από τους πιο ανταγωνιστικούς κλάδους της σύγχρονης εποχής. Η πρόβλεψη της πελατειακής συμπεριφοράς, με ιδιαίτερη έμφαση στο φαινόμενο της αποχώρησης πελατών (churn), αναγνωρίζεται ως κρίσιμος παράγοντας για τη βιωσιμότητα και την ανάπτυξη των εταιρειών. Το churn δεν αντικατοπτρίζει μόνο απώλειες εσόδων αλλά και την ποιότητα των παρεχόμενων υπηρεσιών, την εμπιστοσύνη και την ικανοποίηση των πελατών. Οι τηλεπικοινωνιακές εταιρείες, αντιμέτωπες με το υψηλό κόστος απόκτησης νέων πελατών, επιδιώκουν να μειώσουν το churn μέσω της ανάπτυξης προγνωστικών συστημάτων που αξιοποιούν σύγχρονες τεχνολογίες και δεδομένα.

Στην παρούσα εργασία, διερευνώνται οι δυνατότητες πρόβλεψης του churn μέσω της χρήσης αλγορίθμων μηχανικής μάθησης, όπως το XGBoost, το Random Forest και το Decision Tree. Τα εργαλεία αυτά εφαρμόστηκαν σε μεγάλα δεδομένα από πηγές όπως τα Call Detail Records (CDRs), τα κοινωνικά δίκτυα και έρευνες ικανοποίησης, επιτρέποντας την εξαγωγή κρίσιμων μοτίβων συμπεριφοράς πελατών. Η σύνδεση αυτών των τεχνολογιών με τις στρατηγικές μάρκετινγκ είναι ιδιαίτερα σημαντική, καθώς επιτρέπει τη δημιουργία προσωποποιημένων προσφορών και την ανάπτυξη στοχευμένων καμπανιών που ενισχύουν τη δέσμευση των πελατών και μειώνουν την πιθανότητα αποχώρησης.

Επιπλέον, η εργασία αναδεικνύει τις πρακτικές εφαρμογές των προγνωστικών μοντέλων για τη βελτίωση της εμπειρίας πελάτη, μέσα από την έγκαιρη ανίχνευση δυσαρέσκειας και την προσαρμογή των παρεχόμενων υπηρεσιών. Παρουσιάζονται προτάσεις για στρατηγικές διατήρησης, όπως η ανάπτυξη loyalty προγραμμάτων, η μείωση των χρεώσεων για πελάτες με υψηλή χρήση και η βελτίωση της εξυπηρέτησης μέσω της ανάλυσης των αιτίων δυσαρέσκειας. Η ανάλυση περιπτώσεων δείχνει πώς οι εταιρείες τηλεπικοινωνιών μπορούν να χρησιμοποιήσουν αυτές τις προσεγγίσεις για τη μεγιστοποίηση της ικανοποίησης και της εμπιστοσύνης.

Τέλος, επισημαίνονται οι προκλήσεις της εφαρμογής, όπως η ανάγκη για ποιοτικά και ισορροπημένα δεδομένα, τα ηθικά ζητήματα γύρω από τη διαχείριση ιδιωτικότητας και η προσαρμογή των συστημάτων σε ένα διαρκώς μεταβαλλόμενο περιβάλλον. Η εργασία καταλήγει με προτάσεις για μελλοντική έρευνα, με στόχο την ενσωμάτωση real-time δεδομένων, τη χρήση πιο καινοτόμων μοντέλων AI και την περαιτέρω ανάπτυξη της διασύνδεσης CRM συστημάτων με τις τεχνολογίες πρόβλεψης.

Περιεχόμενα

1. Εισαγωγή

Στη σύγχρονη εποχή, η τηλεπικοινωνιακή βιομηχανία αναδεικνύεται ως ένας από τους πιο δυναμικούς και ανταγωνιστικούς κλάδους, με την αποχώρηση πελατών (customer churn) να αποτελεί μία από τις σημαντικότερες προκλήσεις για τη βιωσιμότητα και την ανάπτυξη των εταιρειών. Το πρόβλημα της αποχώρησης πελατών επηρεάζει όχι μόνο την οικονομική απόδοση, αλλά και τη φήμη και την εμπιστοσύνη των πελατών προς μια επιχείρηση. Το υπό μελέτη ερώτημα της παρούσας εργασίας είναι η κατανόηση και η πρόβλεψη της πελατειακής συμπεριφοράς, εστιάζοντας στη μείωση της αποχώρησης μέσω της αξιοποίησης τεχνολογιών τεχνητής νοημοσύνης (AI) και ανάλυσης δεδομένων.

Υπό μελέτη πρόβλημα και σημασία του θέματος

Το πρόβλημα του churn, δηλαδή η αποχώρηση πελατών, είναι μια πολυδιάστατη πρόκληση που σχετίζεται με διάφορους παράγοντες, όπως η ποιότητα υπηρεσιών, η τιμολογιακή πολιτική και η ανταγωνιστική πίεση. Στον κλάδο των τηλεπικοινωνιών, όπου η απόκτηση νέων πελατών είναι έως και επτά φορές πιο δαπανηρή από τη διατήρηση των υπαρχόντων, η ικανότητα πρόβλεψης του churn μπορεί να αποτελέσει καθοριστικό παράγοντα ανταγωνιστικού πλεονεκτήματος.

Η επιλογή του θέματος βασίζεται στη διαρκώς αυξανόμενη ανάγκη των εταιρειών να μειώσουν τις οικονομικές και λειτουργικές απώλειες που προκαλεί το churn. Επιπλέον, οι πρόσφατες εξελίξεις στον τομέα της τεχνητής νοημοσύνης και της ανάλυσης μεγάλων δεδομένων (Big Data) προσφέρουν νέες δυνατότητες πρόβλεψης και διαχείρισης της πελατειακής συμπεριφοράς, συμβάλλοντας τόσο στην επιστήμη όσο και στην επιχειρηματική πρακτική.

Όρια του θέματος και στόχος της έρευνας

Η παρούσα εργασία εστιάζει αποκλειστικά στη χρήση τεχνολογιών AI και συγκεκριμένων αλγορίθμων, όπως το XGBoost, το Random Forest και το Decision Tree, για την πρόβλεψη του churn στην τηλεπικοινωνιακή βιομηχανία. Παράλληλα, περιλαμβάνει την ανάλυση δεδομένων από πηγές όπως τα Call Detail Records (CDRs), τα κοινωνικά δίκτυα και οι έρευνες ικανοποίησης πελατών. Σε αντίθεση με άλλες έρευνες που επικεντρώνονται στην απλή καταγραφή μοτίβων, η παρούσα εργασία επιχειρεί να προσφέρει πρακτικές λύσεις, όπως η προσωποποίηση υπηρεσιών και η ανάπτυξη στρατηγικών διατήρησης πελατών.

Τρέχουσα κατάσταση της έρευνας

Μέχρι στιγμής, η έρευνα στον τομέα της πρόβλεψης του churn έχει επιτύχει σημαντικά αποτελέσματα, κυρίως μέσω της εφαρμογής τεχνικών μηχανικής μάθησης και της ανάλυσης δεδομένων. Παρ' όλα αυτά, οι περισσότερες μελέτες περιορίζονται στην περιγραφική ανάλυση των δεδομένων ή στη χρήση παραδοσιακών στατιστικών μοντέλων. Η παρούσα εργασία επιχειρεί να

προχωρήσει ένα βήμα παραπέρα, εστιάζοντας στη χρήση προγνωστικών μοντέλων και στην πρακτική εφαρμογή αυτών των τεχνικών στη στρατηγική μάρκετινγκ.

Δομή της μελέτης

Η εργασία οργανώνεται σε επτά κύρια κεφάλαια. Στην Εισαγωγή, παρουσιάζεται το υπό μελέτη πρόβλημα, η σημασία του θέματος και η μεθοδολογία που ακολουθείται. Στο Θεωρητικό Πλαίσιο, αναλύεται η τρέχουσα βιβλιογραφία σχετικά με την πρόβλεψη πελατειακής συμπεριφοράς και το μάρκετινγκ σχέσεων. Το κεφάλαιο Μεθοδολογία και Εργαλεία περιγράφει τα χρησιμοποιούμενα μοντέλα και δεδομένα, καθώς και τις τεχνικές επεξεργασίας. Στα επόμενα κεφάλαια, Αποτελέσματα και Επιπτώσεις Μάρκετινγκ, παρουσιάζονται τα ευρήματα της ανάλυσης και οι προτάσεις στρατηγικής. Τέλος, στο κεφάλαιο Συμπεράσματα και Μελλοντικές Τάσεις, συνοψίζονται τα ευρήματα και προτείνονται κατευθύνσεις για περαιτέρω έρευνα.

Με αυτό τον τρόπο, η εργασία επιδιώκει να συμβάλει στην κατανόηση και επίλυση του προβλήματος του churn, παρέχοντας στις επιχειρήσεις τηλεπικοινωνιών τα εργαλεία που χρειάζονται για να διατηρήσουν και να ενισχύσουν τη βάση των πελατών τους.

2. Θεωρητικό Πλαίσιο

2.1 Τι είναι το Μάρκετινγκ

Έννοιες, Εξέλιξη και Σύγχρονες Προκλήσεις

Το μάρκετινγκ (marketing) αποτελεί ένα από τα θεμελιώδη πεδία των επιχειρηματικών σπουδών και πρακτικών. Αναγνωρίζεται ως η διαδικασία με την οποία οι επιχειρήσεις κατανοούν τις ανάγκες των καταναλωτών, προσαρμόζουν τα προϊόντα ή τις υπηρεσίες τους σε αυτές τις ανάγκες, και τελικά προσελκύουν και διατηρούν τους πελάτες τους μέσω της αξίας που προσφέρουν. Στον ακαδημαϊκό και επιχειρηματικό κόσμο, το μάρκετινγκ ορίζεται συχνά ως η «δημιουργία αξίας» και η «ικανοποίηση των αναγκών του καταναλωτή» με κεντρικό σκοπό την επίτευξη ανταγωνιστικού πλεονεκτήματος και τη διατήρηση βιώσιμων επιχειρηματικών σχέσεων.

Ιστορική Εξέλιξη του Μάρκετινγκ

Η μελέτη και η πρακτική του μάρκετινγκ, όπως την αντιλαμβανόμαστε σήμερα, έχει διαμορφωθεί μέσα από αρκετές δεκαετίες ακαδημαϊκής έρευνας και επιχειρηματικής εμπειρίας. Το μάρκετινγκ εξελίχθηκε από την απλή διαδικασία της εμπορικής συναλλαγής στην οργανωμένη και σύνθετη επιστήμη που αποτελεί την σημερινή παρουσία του Μαρκετινγκ. Οι πρώτες μορφές μάρκετινγκ επικεντρώνονταν στο **προσανατολισμό παραγωγής** (production orientation), όπου οι εταιρείες έδιναν έμφαση στη μεγιστοποίηση της παραγωγής και την αποτελεσματικότητα των διαδικασιών, εκτιμώντας ότι η υψηλή προσφορά θα οδηγούσε στη ζήτηση.

Με τη μετάβαση στο **προσανατολισμένο στην πώληση μάρκετινγκ** (sales orientation), οι επιχειρήσεις άρχισαν να συνειδητοποιούν την ανάγκη για ενεργή προώθηση και πειθώ, αναπτύσσοντας στρατηγικές πωλήσεων που περιλάμβαναν διαφήμιση και προσωπική πώληση. Το κίνημα αυτό εξελίχθηκε περαιτέρω στη δεκαετία του 1950 με την άνοδο του **προσανατολισμού στον καταναλωτή** (customer orientation), όπου το επίκεντρο του μάρκετινγκ μετατοπίστηκε στην κατανόηση και ικανοποίηση των αναγκών του καταναλωτή, με σκοπό την οικοδόμηση μακροχρόνιων σχέσεων.(Alshamari, M. A. 2023)

Η εισαγωγή του **στρατηγικού μάρκετινγκ** στις δεκαετίες του 1970 και 1980 υποστήριξε ότι οι επιχειρήσεις πρέπει να διαμορφώνουν τις στρατηγικές τους όχι μόνο με βάση τις ανάγκες του καταναλωτή, αλλά και σύμφωνα με τις ευρύτερες συνθήκες της αγοράς, τον ανταγωνισμό και τις εσωτερικές τους δυνατότητες. Σήμερα, το μάρκετινγκ έχει ενσωματώσει καινοτόμες μεθόδους όπως το **ψηφιακό μάρκετινγκ**, η **ανάλυση δεδομένων** και οι **προσωποποιημένες εμπειρίες πελατών**, γεγονός που το καθιστά πιο περίπλοκο και αποτελεσματικό από ποτέ.

2.2 Θεμελιώδεις Έννοιες και Αρχές του Μάρκετινγκ

Οι βασικές αρχές του μάρκετινγκ συνδέονται άρρηκτα με τη θεωρία των καταναλωτικών αναγκών και την προσπάθεια να δημιουργηθεί αξία για τον πελάτη. Μερικές από τις κεντρικές έννοιες που διέπουν το μάρκετινγκ είναι οι εξής:

1. **Ανάγκες, Επιθυμίες και Απαιτήσεις**: Οι ανάγκες είναι οι θεμελιώδεις απαιτήσεις των ανθρώπων, οι οποίες, σύμφωνα με τον **Maslow (1943)**, περιλαμβάνουν βασικές ανάγκες όπως τροφή, ασφάλεια και κοινωνικές σχέσεις. Οι επιθυμίες αποτελούν πιο σύνθετες και συνειδητές εκφράσεις των αναγκών, ενώ οι απαιτήσεις προκύπτουν όταν οι επιθυμίες συνοδεύονται από την αγοραστική δυνατότητα του ατόμου.

2. **Ανταλλαγή και Αξία**: Το μάρκετινγκ επικεντρώνεται στην **αξία** που παρέχει ένα προϊόν ή μια υπηρεσία στον καταναλωτή. Η αξία αυτή δεν είναι απλώς η χρηστική ιδιότητα του προϊόντος, αλλά και η ψυχολογική και κοινωνική σημασία του για τον καταναλωτή. Η **ανταλλαγή** είναι η θεμελιώδης αρχή μέσω της οποίας πραγματοποιείται αυτή η διαδικασία, δημιουργώντας μια σχέση μεταξύ του καταναλωτή και του παραγωγού που βασίζεται στην αμοιβαία ωφέλεια.

3. **Μίγμα Μάρκετινγκ (4Ps)**: Το μάρκετινγκ συνήθως υλοποιείται μέσω του **μίγματος μάρκετινγκ**, γνωστού και ως 4Ps: Προϊόν (Product), Τιμή (Price), Προώθηση (Promotion), και Διανομή (Place). Ο συνδυασμός των 4Ps διαμορφώνει τη συνολική στρατηγική που ακολουθεί μια επιχείρηση για την ανάπτυξη και προώθηση των προϊόντων ή των υπηρεσιών της στην αγορά.

    - **Προϊόν**: Αναφέρεται στα χαρακτηριστικά, την ποιότητα και τη μοναδικότητα του προϊόντος ή της υπηρεσίας.
    - **Τιμή**: Η τιμολόγηση καθορίζει την οικονομική ανταλλαγή που απαιτείται για την απόκτηση του προϊόντος και επηρεάζει την αντίληψη των καταναλωτών για την αξία του.
    - **Προώθηση**: Αφορά την προβολή του προϊόντος και περιλαμβάνει τη διαφήμιση, τις δημόσιες σχέσεις και την άμεση επικοινωνία με τον καταναλωτή.
    - **Διανομή**: Περιγράφει τη διαδικασία διάθεσης και την επιλογή των καναλιών μέσω των οποίων φτάνει το προϊόν στους καταναλωτές.

Σύγχρονες Προσεγγίσεις και Εργαλεία του Μάρκετινγκ

Η σύγχρονη πρακτική του μάρκετινγκ περιλαμβάνει τεχνικές και στρατηγικές που βασίζονται σε αναλυτικά εργαλεία και τεχνολογικές καινοτομίες:

1. **Έρευνα Αγοράς**: Η έρευνα αγοράς είναι ο ακρογωνιαίος λίθος για την κατανόηση των αναγκών του καταναλωτή, καθώς και του ανταγωνιστικού περιβάλλοντος. Χωρίζεται σε **ποσοτική** και **ποιοτική** έρευνα, και περιλαμβάνει μεθόδους όπως οι δημοσκοπήσεις, οι ομάδες εστίασης και η ανάλυση δεδομένων.

2. **Στρατηγικές Branding**: Το branding αναφέρεται στη δημιουργία και διαχείριση της εικόνας και της ταυτότητας μιας μάρκας στην αγορά. Ο **Aaker (1991)** και άλλοι ερευνητές έχουν τονίσει τη σημασία της διαφοροποίησης και της δημιουργίας συναισθηματικής σύνδεσης με τους καταναλωτές, καθώς οι ισχυρές μάρκες αυξάνουν την αναγνωρισιμότητα και την πίστη του κοινού.

3. **Ψηφιακό Μάρκετινγκ και Χρήση Κοινωνικών Δικτύων**: Το ψηφιακό μάρκετινγκ περιλαμβάνει τη χρήση ηλεκτρονικών μέσων, όπως το διαδίκτυο, τα κοινωνικά δίκτυα, το email και το μάρκετινγκ περιεχομένου, για την προσέγγιση και αλληλεπίδραση με το κοινό. Οι σύγχρονες πλατφόρμες κοινωνικών δικτύων, όπως το Facebook, το Instagram και το Twitter, προσφέρουν νέες δυνατότητες άμεσης επικοινωνίας και προσωποποιημένης προβολής.

4. **Ανάλυση Δεδομένων και Προσωποποιημένο Μάρκετινγκ**: Η ανάλυση μεγάλων δεδομένων (big data) και η χρήση εργαλείων μηχανικής μάθησης επιτρέπει στις επιχειρήσεις να αναλύουν σε βάθος τις προτιμήσεις και τις συνήθειες των καταναλωτών, με στόχο την εξατομίκευση των προσφορών τους. Σύμφωνα με τους **Davenport και Harris (2007)**, οι τεχνικές αυτές παρέχουν τη δυνατότητα βελτιστοποίησης της εμπειρίας του καταναλωτή και της αλληλεπίδρασης με την επιχείρηση.

2.3 Προκλήσεις και Περιορισμοί του Σύγχρονου Μάρκετινγκ

Το σύγχρονο μάρκετινγκ δεν είναι χωρίς προκλήσεις, ειδικά στο συνεχώς μεταβαλλόμενο ψηφιακό και παγκοσμιοποιημένο περιβάλλον:

1. **Αυξημένος Ανταγωνισμός και Εξατομίκευση**: Οι καταναλωτές έχουν πλέον πρόσβαση σε πληθώρα επιλογών, γεγονός που καθιστά τον ανταγωνισμό πιο έντονο από ποτέ. Οι επιχειρήσεις πρέπει να προσαρμόζουν συνεχώς τις στρατηγικές τους για να παραμένουν ανταγωνιστικές και να καλύπτουν τις συνεχώς μεταβαλλόμενες προτιμήσεις των καταναλωτών.

2. **Προστασία Δεδομένων και Ασφάλεια**: Οι ανησυχίες για την ιδιωτικότητα και την ασφάλεια των δεδομένων έχουν αυξηθεί, καθώς οι επιχειρήσεις συλλέγουν και αναλύουν δεδομένα για να βελτιώσουν τις στρατηγικές μάρκετινγκ τους. Η ανάγκη για συμμόρφωση με κανονισμούς, όπως ο **Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR)**, αποτελεί πρόκληση για τις επιχειρήσεις.

3. **Διαχείριση Φήμης σε Πραγματικό Χρόνο**: Στην εποχή των κοινωνικών δικτύων, η διαχείριση της φήμης γίνεται άμεσα και απαιτεί προσοχή στις λεπτομέρειες, καθώς οι καταναλωτές μπορούν να εκφράσουν τις απόψεις τους δημόσια και σε πραγματικό χρόνο. Αυτό αυξάνει την πίεση για άμεση και αποτελεσματική διαχείριση της επικοινωνίας.

4. **Τεχνολογική Καινοτομία και Προσαρμοστικότητα**: Ο ρυθμός της τεχνολογικής εξέλιξης είναι ταχύς και οι επιχειρήσεις πρέπει να επενδύουν συνεχώς σε νέες τεχνολογίες και να προσαρμόζουν τις στρατηγικές τους, για να εκμεταλλεύονται τις νέες δυνατότητες και να διατηρούν την ανταγωνιστικότητά τους.

Το μάρκετινγκ συνεχίζει να εξελίσσεται και να διαμορφώνει τις επιχειρηματικές πρακτικές, παρέχοντας στις επιχειρήσεις ένα πλαίσιο για την επίτευξη των στόχων τους μέσω της κατανόησης και της ικανοποίησης των αναγκών των καταναλωτών. Σε έναν κόσμο που αλλάζει διαρκώς, η καινοτομία, η προσαρμοστικότητα και η αξία για τον πελάτη αποτελούν βασικούς πυλώνες για την επιτυχία. Ο ακαδημαϊκός και επιχειρηματικός τομέας, εστιάζοντας στις ανάγκες και τις επιθυμίες των καταναλωτών, ενισχύει τη σημασία της στρατηγικής του μάρκετινγκ, καθιστώντας το έναν από τους κεντρικούς μοχλούς οικονομικής ανάπτυξης και κοινωνικής προόδου. Etim, G. S., Etuk, I. U., James, E. E., & Ekpe, S. 2021)

2.4 Μάρκετινγκ στις Τηλεπικοινωνίες

Προκλήσεις Ανταγωνισμού και Αλλαγών στην Καταναλωτική Συμπεριφορά

Ο τομέας των τηλεπικοινωνιών βρίσκεται αντιμέτωπος με μια νέα εποχή προκλήσεων και ανακατατάξεων, που οδηγούνται από την ταχύτατη τεχνολογική εξέλιξη και τη μεταβολή των καταναλωτικών απαιτήσεων. Οι εταιρείες τηλεπικοινωνιών αναζητούν τρόπους να διατηρήσουν την ανταγωνιστικότητά τους, ενώ ταυτόχρονα αντιμετωπίζουν σοβαρές πιέσεις που σχετίζονται με τη συμπίεση των τιμών και την αύξηση του ανταγωνισμού, ειδικά λόγω της ανάπτυξης των Over-The-Top (OTT) εφαρμογών όπως τα WhatsApp, Skype, και Viber. Αυτές οι προκλήσεις ενισχύονται από την ανάγκη των εταιρειών να ανταποκριθούν στις νέες καταναλωτικές απαιτήσεις για υπηρεσίες υψηλής ποιότητας με χαμηλότερο κόστος, καθιστώντας επιτακτική τη διαρκή προσαρμογή των επιχειρηματικών τους στρατηγικών.

Αύξηση Ανταγωνισμού και Συμπίεση Τιμών στις Τηλεπικοινωνίες

Η αγορά των τηλεπικοινωνιών, ιστορικά μονοπωλιακή ή ολιγοπωλιακή σε πολλές περιοχές, έχει εξελιχθεί σε έναν από τους πιο ανταγωνιστικούς τομείς παγκοσμίως. Ο ανταγωνισμός αυτός προέρχεται τόσο από νέες επιχειρήσεις που εισέρχονται στην αγορά, όσο και από τεχνολογικές εξελίξεις που προσφέρουν εναλλακτικές λύσεις στις παραδοσιακές υπηρεσίες τηλεπικοινωνιών.

Ο Ρόλος των OTT Εφαρμογών

Οι OTT εφαρμογές, όπως το WhatsApp, το Skype, και το Viber, έχουν επιφέρει τεράστιες αλλαγές στην αγορά, καθώς προσφέρουν υπηρεσίες επικοινωνίας που παρακάμπτουν τις παραδοσιακές υποδομές των τηλεπικοινωνιακών παρόχων. Αυτές οι εφαρμογές επιτρέπουν στους χρήστες να επικοινωνούν δωρεάν ή με πολύ χαμηλό κόστος, αρκεί να έχουν πρόσβαση σε Wi-Fi ή δεδομένα κινητής τηλεφωνίας. Η **ανάπτυξη των OTT εφαρμογών** έχει επηρεάσει την κερδοφορία των τηλεπικοινωνιακών εταιρειών, καθώς περιορίζει την ανάγκη για παραδοσιακές υπηρεσίες, όπως τα SMS και οι διεθνείς κλήσεις, οι οποίες αποτελούσαν στο παρελθόν σημαντικές πηγές εσόδων.

Η παρουσία των OTT εφαρμογών επιβάλλει πιέσεις στις τηλεπικοινωνιακές εταιρείες να μειώσουν τις τιμές τους για να παραμείνουν ανταγωνιστικές. Σύμφωνα με την **Θεωρία του Δικτύου** (Network Theory), η αξία των υπηρεσιών αυτών αυξάνεται με τον αριθμό των χρηστών τους, και αυτό έχει οδηγήσει σε ταχεία υιοθέτησή τους σε παγκόσμιο επίπεδο. Οι τηλεπικοινωνιακές εταιρείες πρέπει πλέον να ανταγωνίζονται με υπηρεσίες που δεν χρειάζονται παραδοσιακή υποδομή, μειώνοντας έτσι την ικανότητά τους να προσφέρουν διαφοροποιημένες υπηρεσίες στις ίδιες τιμές με το παρελθόν.

Συμπίεση Τιμών και Διατήρηση Κερδοφορίας

Η συνεχής συμπίεση των τιμών αποτελεί μία από τις μεγαλύτερες προκλήσεις στον τομέα των τηλεπικοινωνιών. Καθώς οι καταναλωτές αποκτούν περισσότερες επιλογές για επικοινωνία μέσω των OTT εφαρμογών, οι τηλεπικοινωνιακές εταιρείες αναγκάζονται να μειώσουν τις τιμές των παραδοσιακών υπηρεσιών τους, όπως η φωνητική κλήση και τα δεδομένα κινητής τηλεφωνίας. Ωστόσο, η μείωση των τιμών οδηγεί σε συμπίεση των περιθωρίων κέρδους, γεγονός που καθιστά δύσκολη τη διατήρηση της κερδοφορίας.

Οι εταιρείες τηλεπικοινωνιών αντιμετωπίζουν το δίλημμα της μείωσης των τιμών για την ανταγωνιστικότητα, ενώ πρέπει ταυτόχρονα να συνεχίσουν να επενδύουν σε νέες τεχνολογίες και υποδομές, όπως το **5G**, προκειμένου να παρέχουν βελτιωμένες υπηρεσίες στους καταναλωτές. Αυτή η ανάγκη για συνεχή επένδυση σε τεχνολογική καινοτομία με παράλληλη πίεση για μείωση των τιμών θέτει τις εταιρείες σε μια δύσκολη οικονομική θέση, καθώς τα κόστη αυξάνονται, ενώ τα έσοδα παραμένουν σταθερά ή μειώνονται. (Etim, G. S., Etuk, I. U., James, E. E., & Ekpe, S. 2021)

Στρατηγικές Αντιμετώπισης του Ανταγωνισμού

Η πίεση για συμπίεση των τιμών λόγω του ανταγωνισμού και της παρουσίας των OTT εφαρμογών αναγκάζει τις τηλεπικοινωνιακές εταιρείες να στραφούν σε νέες στρατηγικές. Μία από τις πιο διαδεδομένες στρατηγικές είναι η **διαφοροποίηση των υπηρεσιών**, όπου οι εταιρείες παρέχουν πακέτα που περιλαμβάνουν ειδικές προσφορές δεδομένων, σύνθετα πακέτα με προγράμματα ψυχαγωγίας ή πρόσβαση σε OTT υπηρεσίες. Αυτή η στρατηγική έχει ως στόχο να διατηρήσει τους πελάτες στη βάση της εταιρείας, προσφέροντας προστιθέμενη αξία.

Επιπλέον, πολλές εταιρείες τηλεπικοινωνιών προσπαθούν να προσελκύσουν τους καταναλωτές μέσω της παροχής **υπηρεσιών κινητής τραπεζικής** και **συνεργασιών με εταιρείες τεχνολογίας**, επεκτείνοντας έτσι τις προσφερόμενες υπηρεσίες πέρα από την παραδοσιακή τηλεφωνία και τα δεδομένα. Η στρατηγική αυτή, ωστόσο, απαιτεί σημαντικές επενδύσεις σε νέα προϊόντα και αναβάθμιση των δικτύων, πράγμα που δεν είναι εύκολο όταν τα έσοδα συμπιέζονται από την έντονη πίεση για μείωση των τιμών.

Αλλαγή Συμπεριφοράς Καταναλωτών και Απαιτήσεις για Εξατομικευμένες Υπηρεσίες

Παράλληλα με τον αυξανόμενο ανταγωνισμό και τη συμπίεση τιμών, η συμπεριφορά των καταναλωτών έχει αλλάξει δραστικά. Οι πελάτες των τηλεπικοινωνιών αναζητούν πλέον υψηλή ποιότητα υπηρεσιών, άμεση εξυπηρέτηση, και εξατομικευμένες λύσεις που ανταποκρίνονται στις προσωπικές τους ανάγκες και προτιμήσεις. Η τάση αυτή είναι αποτέλεσμα της αυξημένης τεχνολογικής ευαισθητοποίησης των καταναλωτών και της ευρείας διάδοσης των κινητών συσκευών, που επιτρέπει στους χρήστες να διαχειρίζονται τις επικοινωνίες τους με βάση τις προσωπικές τους προτιμήσεις. (Ameen, N., Tarhini, A., Reppel, A., & Anand, A. 2021)

Η Δύναμη της Προσωποποίησης

Σύμφωνα με τη **Θεωρία της Προσωποποίησης των Υπηρεσιών** (Service Personalization Theory), οι καταναλωτές είναι πιθανότερο να δείξουν αφοσίωση σε εταιρείες που προσαρμόζουν τις υπηρεσίες τους στις ατομικές ανάγκες και προτιμήσεις τους. Στον τομέα των τηλεπικοινωνιών, αυτό σημαίνει ότι οι πάροχοι πρέπει να προσφέρουν πακέτα που επιτρέπουν στους χρήστες να διαμορφώνουν τον τρόπο χρήσης των δεδομένων τους, να επιλέγουν πρόσθετες υπηρεσίες, και να ελέγχουν τα κόστη τους μέσω εξατομικευμένων προγραμμάτων.

Οι τηλεπικοινωνιακές εταιρείες αναπτύσσουν πλέον εξελιγμένα **συστήματα διαχείρισης πελατειακών σχέσεων (CRM)**, τα οποία βασίζονται σε αναλύσεις μεγάλων δεδομένων (big data) και τεχνικές μηχανικής μάθησης. Αυτά τα συστήματα τους επιτρέπουν να αναγνωρίζουν πρότυπα συμπεριφοράς, να προβλέπουν τις ανάγκες των πελατών και να προσφέρουν στοχευμένες προσφορές και πακέτα. Παρά τα οφέλη, αυτή η στρατηγική προσωποποίησης έχει κόστος, καθώς απαιτεί επενδύσεις σε υποδομές δεδομένων και τεχνολογίες ανάλυσης. (Ameen, N., Tarhini, A., Reppel, A., & Anand, A. 2021)

Οι Περιορισμοί στην ανοχή των Καταναλωτών στις Τιμές

Παράλληλα με την απαίτηση για υψηλής ποιότητας υπηρεσίες, οι καταναλωτές σήμερα είναι λιγότερο διατεθειμένοι να πληρώσουν υψηλές τιμές. Η εύκολη πρόσβαση στις τιμές και τις προσφορές άλλων παρόχων μέσω διαδικτύου έχει καταστήσει τους καταναλωτές πιο συνειδητοποιημένους και ευαίσθητους στην τιμή. Οι τηλεπικοινωνιακές εταιρείες πρέπει να εξισορροπήσουν την ανάγκη για προστιθέμενη αξία και υψηλή ποιότητα υπηρεσιών με την επιθυμία των καταναλωτών για χαμηλές τιμές, κάτι που αποτελεί σημαντική πρόκληση.

Η πίεση για χαμηλές τιμές σε συνδυασμό με την υψηλή προσδοκία για ποιοτικές υπηρεσίες αποτελεί μια διπλή πρόκληση που απαιτεί στρατηγικές **αναδιάρθρωσης κόστους**, αλλά και **καινοτομίας**. Οι εταιρείες τηλεπικοινωνιών πρέπει να προσαρμόσουν τη στρατηγική τους για να προσφέρουν ελκυστικές τιμές, ενώ παράλληλα βελτιώνουν την ποιότητα των υπηρεσιών τους, γεγονός που μπορεί να επιτευχθεί μέσω τεχνολογικής αναβάθμισης και βελτίωσης των εσωτερικών διαδικασιών.

Το μάρκετινγκ στις τηλεπικοινωνίες σήμερα πρέπει να ισορροπήσει ανάμεσα στις απαιτήσεις για μειωμένες τιμές και υψηλή ποιότητα υπηρεσιών, αντιμετωπίζοντας τις προκλήσεις που προκύπτουν από την αύξηση του ανταγωνισμού και τις τεχνολογικές αλλαγές. Οι τηλεπικοινωνιακές εταιρείες βρίσκονται σε ένα μεταβατικό στάδιο, όπου η ανάγκη για διαφοροποίηση και προσαρμογή στις νέες καταναλωτικές απαιτήσεις είναι επιτακτική. Η επιτυχία στον τομέα αυτό θα εξαρτηθεί από την ικανότητα των εταιρειών να καινοτομούν, να προσωποποιούν τις υπηρεσίες τους, και να προσφέρουν προστιθέμενη αξία στους πελάτες τους, διατηρώντας ταυτόχρονα την κερδοφορία τους. (Ahmad, A. K., Jafar, A., & Aljoumaa, K. 2019)

Ένα από τα πιο κρίσιμα προβλήματα που αναδεικνύουν τα άρθρα είναι το **customer churn** (αποχώρηση πελατών), το οποίο επηρεάζει σοβαρά τον κλάδο των τηλεπικοινωνιών. Το **churn** αναφέρεται στους πελάτες που διακόπτουν τη συνδρομή τους ή αλλάζουν πάροχο, και αποτελεί μια σημαντική απειλή για την κερδοφορία των εταιρειών, ιδιαίτερα σε έναν κορεσμένο και άκρως ανταγωνιστικό τομέα όπως αυτός των τηλεπικοινωνιών.

Το πρόβλημα του churn οφείλεται σε διάφορους παράγοντες:

1. **Αυξημένος Ανταγωνισμός και Εναλλακτικές Υπηρεσίες**: Με την είσοδο νέων παρόχων και τη διάδοση των OTT εφαρμογών, οι καταναλωτές έχουν περισσότερες επιλογές για φθηνότερες ή ακόμη και δωρεάν επικοινωνίες, καθιστώντας εύκολο το να μετακινηθούν από έναν πάροχο σε άλλον.

2. **Αλλαγή στις Προσδοκίες των Πελατών**: Οι πελάτες επιθυμούν εξατομικευμένες υπηρεσίες υψηλής ποιότητας με χαμηλό κόστος. Εάν οι προσδοκίες αυτές δεν ικανοποιούνται, είναι πιθανότερο να αλλάξουν πάροχο.

3. **Αδυναμία Διατήρησης Πελατών λόγω Έλλειψης Εμπιστοσύνης και Αξιοπιστίας**: Τα προβλήματα ποιότητας, η ανεπαρκής εξυπηρέτηση και η απουσία προστιθέμενης αξίας στους πελάτες οδηγούν στην απογοήτευση και στην επιθυμία για εναλλακτικές λύσεις.

Το **customer churn** αποτελεί κρίσιμη πρόκληση, καθώς η απόκτηση νέων πελατών είναι συνήθως πιο δαπανηρή από τη διατήρηση των υφιστάμενων. Για να αντιμετωπίσουν το churn, οι τηλεπικοινωνιακές εταιρείες εφαρμόζουν **στρατηγικές διαχείρισης πελατειακών σχέσεων (CRM)**, χρησιμοποιούν **ανάλυση δεδομένων** και αναπτύσσουν προγράμματα **εξατομίκευσης** και **βελτίωσης της εμπειρίας πελάτη** με στόχο την αύξηση της αφοσίωσης των πελατών και τη μείωση της αποχώρησής τους.

2.5 Μάρκετινγκ και Customer Churn στις Τηλεπικοινωνίες

Αιτίες και Αντιμετώπιση

Το φαινόμενο του **customer churn** ή αποχώρησης πελατών αποτελεί μια από τις πιο επίμονες προκλήσεις για τις εταιρείες τηλεπικοινωνιών παγκοσμίως. Ο όρος **churn** αναφέρεται στους πελάτες που διακόπτουν τη συνδρομή τους ή μετακινούνται σε ανταγωνιστικές εταιρείες. Ο ρυθμός αποχώρησης των πελατών, ειδικά σε έναν τομέα όπως οι τηλεπικοινωνίες, όπου οι υπηρεσίες είναι κορεσμένες και οι πελάτες έχουν πολλές εναλλακτικές, αποτελεί έναν κρίσιμο δείκτη που αντικατοπτρίζει την ανταγωνιστικότητα και την ποιότητα των υπηρεσιών μιας εταιρείας. Το μάρκετινγκ διαδραματίζει καίριο ρόλο στην κατανόηση, ανάλυση και τελικά, μείωση του churn, με στρατηγικές που εστιάζουν στη διατήρηση των πελατών και στην ενίσχυση της αφοσίωσης. (Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system.)

Ο Ρόλος του Μάρκετινγκ στη Διαχείριση του Churn

Το μάρκετινγκ, μέσω της εστίασης στις ανάγκες και προτιμήσεις των καταναλωτών, έχει τον κεντρικό ρόλο στη στρατηγική διαχείρισης του churn. Σύμφωνα με τη θεωρία του **Customer Relationship Management (CRM)**, η διατήρηση των πελατών είναι πιο αποδοτική από την προσέλκυση νέων. Οι στρατηγικές μάρκετινγκ που επικεντρώνονται στην πρόληψη του churn περιλαμβάνουν:

Αναγνώριση Προτύπων Συμπεριφοράς και Πρόβλεψη του Churn: Μέσω της ανάλυσης των δεδομένων πελατών, οι εταιρείες μπορούν να εντοπίσουν μοτίβα που υποδηλώνουν τάση αποχώρησης. Τα εργαλεία του μάρκετινγκ που αξιοποιούν δεδομένα (data-driven marketing) επιτρέπουν την αναγνώριση αυτών των προτύπων και τη λήψη προληπτικών μέτρων. (Etim, G. S., Etuk, I. U., James, E. E., & Ekpe, S. 2021)

Εξατομίκευση Προσφορών και Εμπειριών: Οι τηλεπικοινωνιακές εταιρείες προσαρμόζουν τις υπηρεσίες και τις προσφορές τους, βάσει των αναγκών του πελάτη, με στόχο την ενίσχυση της αφοσίωσης. Αυτή η προσέγγιση επιτρέπει την καλύτερη ανταπόκριση στις εξατομικευμένες ανάγκες του πελάτη και την ενίσχυση της σχέσης με τον πάροχο. (Etim, G. S., Etuk, I. U., James, E. E., & Ekpe, S. 2021)

Ενίσχυση της Εμπειρίας Πελάτη (Customer Experience - CX): Το μάρκετινγκ συμβάλλει στη διασφάλιση θετικής εμπειρίας για τον πελάτη, η οποία είναι βασικός παράγοντας για τη μείωση του churn. Μια καλή εμπειρία, από τη στιγμή της ένταξης έως και την εξυπηρέτηση μετά την πώληση, αυξάνει την αφοσίωση και μειώνει την πιθανότητα αποχώρησης. (Etim, G. S., Etuk, I. U., James, E. E., & Ekpe, S. 2021)

2.6 Παράγοντες που Συντελούν στην Αύξηση του Churn

Η αποχώρηση των πελατών στις τηλεπικοινωνίες μπορεί να αποδοθεί σε μια σειρά παραγόντων, οι οποίοι σχετίζονται τόσο με τις προσδοκίες και την εμπειρία των πελατών, όσο και με τις εσωτερικές στρατηγικές και πρακτικές των εταιρειών. Οι κύριοι παράγοντες περιλαμβάνουν:

Αύξηση του Ανταγωνισμού και Παρουσία OTT Εφαρμογών

Η εντατικοποίηση του ανταγωνισμού στον κλάδο των τηλεπικοινωνιών, σε συνδυασμό με την άνοδο των εφαρμογών **Over-The-Top (OTT)**, όπως το WhatsApp, το Viber και το Skype, έχουν μεταβάλει το τοπίο της αγοράς. Οι εφαρμογές OTT παρέχουν υπηρεσίες επικοινωνίας χωρίς να απαιτείται πρόσθετη χρέωση από τον πάροχο, γεγονός που οδηγεί σε μείωση της χρήσης των παραδοσιακών υπηρεσιών τηλεφωνίας και SMS. Η παρουσία αυτών των εφαρμογών ωθεί τους πελάτες να αναζητούν πιο οικονομικές λύσεις, ειδικά εάν ο τρέχων πάροχος δεν προσφέρει πακέτα που καλύπτουν αυτή την ανάγκη.

Ελλιπής Εξυπηρέτηση Πελατών και Υποστήριξη

Η εμπειρία του πελάτη όσον αφορά την εξυπηρέτηση παίζει κρίσιμο ρόλο στη διατήρησή του. Εάν η υποστήριξη δεν ανταποκρίνεται επαρκώς στις ανάγκες του πελάτη, αυτό μπορεί να οδηγήσει σε απογοήτευση και αποχώρηση. Σύμφωνα με τη θεωρία της **Εμπειρίας Πελάτη (Customer Experience Theory)**, η αρνητική εμπειρία εξυπηρέτησης οδηγεί σε απώλεια εμπιστοσύνης και αυξάνει τις πιθανότητες αλλαγής παρόχου. Η έλλειψη διαθεσιμότητας, η αργή εξυπηρέτηση ή η ανικανότητα επίλυσης προβλημάτων είναι βασικές αιτίες αποχώρησης. (Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system.)

Συμπίεση Τιμών και Μείωση Περιθωρίων Κέρδους

Οι τηλεπικοινωνιακές εταιρείες βρίσκονται υπό συνεχή πίεση για τη μείωση των τιμών, εξαιτίας τόσο του ανταγωνισμού όσο και της ύπαρξης των OTT υπηρεσιών. Οι πελάτες είναι πλέον εξοικειωμένοι με την ευρεία διάδοση των φθηνών ή δωρεάν εφαρμογών και, συνεπώς, είναι λιγότερο διατεθειμένοι να πληρώσουν για παραδοσιακές υπηρεσίες. Ωστόσο, η μείωση των τιμών δημιουργεί περιορισμούς στη δυνατότητα των εταιρειών να επενδύουν σε νέες υποδομές και στην παροχή ποιοτικής εξυπηρέτησης, κάτι που οδηγεί σε αυξημένο churn.

Ανεπαρκής Προσαρμογή στις Προσδοκίες και τις Αλλαγές στις Προτιμήσεις των Πελατών.Οι καταναλωτές τηλεπικοινωνιών σήμερα επιθυμούν εξατομικευμένες λύσεις και πακέτα που να ανταποκρίνονται στις προσωπικές τους ανάγκες, καθώς και διαφάνεια στη χρέωση των υπηρεσιών. Οι εταιρείες που δεν προσαρμόζουν τα προϊόντα τους και δεν προσφέρουν επαρκή ευελιξία, συχνά χάνουν πελάτες. Σύμφωνα με έρευνες, οι πελάτες είναι πιο πιθανό να εγκαταλείψουν έναν πάροχο εάν δεν βρίσκουν τις υπηρεσίες που θέλουν ή αισθάνονται ότι οι ανάγκες τους δεν λαμβάνονται υπόψη.

Έλλειψη Επένδυσης σε Υποδομές και Καινοτομία

Η τεχνολογική καινοτομία, όπως η διάδοση του δικτύου 5G, δημιουργεί νέες προσδοκίες για υψηλή ταχύτητα, αξιοπιστία και ποιότητα στις υπηρεσίες. Οι πάροχοι που δεν επενδύουν στις νέες τεχνολογίες ενδέχεται να μην μπορούν να καλύψουν τις απαιτήσεις των καταναλωτών για βελτιωμένες δυνατότητες, με αποτέλεσμα την απογοήτευση και την αναζήτηση εναλλακτικών παρόχων.

2.7 Στρατηγικές Μάρκετινγκ για τη Μείωση του Churn

Ο στόχος του μάρκετινγκ στις τηλεπικοινωνίες είναι να δημιουργήσει σχέσεις εμπιστοσύνης και αφοσίωσης με τους πελάτες, αποτρέποντας την αποχώρησή τους. Υπάρχουν αρκετές στρατηγικές που μπορούν να εφαρμοστούν για να μειωθεί το churn και να βελτιωθεί η εμπειρία πελάτη:

a. **Προσωποποίηση και Εξατομικευμένες Προσφορές**: Η εξατομίκευση των υπηρεσιών είναι κεντρική στη στρατηγική μάρκετινγκ για τη μείωση του churn. Οι τηλεπικοινωνιακές εταιρείες μπορούν να χρησιμοποιήσουν δεδομένα για την αναγνώριση των προτιμήσεων των πελατών και την προσαρμογή των πακέτων ανάλογα. Προσφορές και εκπτώσεις σε πελάτες που παρουσιάζουν κίνδυνο αποχώρησης, καθώς και προγράμματα επιβράβευσης της πιστότητας, μπορούν να βοηθήσουν στη διατήρηση των πελατών. (Zulaikha, S., Mohamed, H., Kurniawati, M., Rusgianto, S., & Rusmita, S. A. 2021)

b. **Ενίσχυση της Εμπειρίας Πελάτη**: Η επένδυση στην εμπειρία πελάτη μέσω βελτίωσης της υποστήριξης και της εξυπηρέτησης αποτελεί θεμελιώδες στοιχείο για τη μείωση του churn. Η παροχή άμεσης και αποτελεσματικής εξυπηρέτησης, η απλοποίηση της διαδικασίας επίλυσης προβλημάτων και η τακτική επικοινωνία με τους πελάτες βοηθούν στην αύξηση της αφοσίωσης.

c. **Προγνωστική Ανάλυση για Έγκαιρη Παρέμβαση**: Η χρήση ανάλυσης δεδομένων επιτρέπει στις εταιρείες να εντοπίζουν πελάτες που είναι πιθανό να αποχωρήσουν. Αυτή η προγνωστική ανάλυση επιτρέπει την έγκαιρη παρέμβαση μέσω ειδικών προσφορών, επικοινωνίας ή επιπλέον παροχών, προτού οι πελάτες φτάσουν σε σημείο απόφασης για αποχώρηση.

d. **Δημιουργία Συνδυαστικών Υπηρεσιών και Πακέτων**: Η ανάπτυξη πακέτων που ενσωματώνουν υπηρεσίες πέραν της παραδοσιακής τηλεφωνίας, όπως πρόσβαση σε περιεχόμενο ψυχαγωγίας, streaming, και δεδομένα για OTT εφαρμογές, μπορεί να αποτελέσει κίνητρο για τη διατήρηση των πελατών. Τέτοιες υπηρεσίες αυξάνουν την αξία της συνδρομής και δυσχεραίνουν τη μετακίνηση σε ανταγωνιστικές εταιρείες.

e. **Αναβάθμιση της Υποδομής και Τεχνολογικής Υπεροχής**: Οι τηλεπικοινωνιακές εταιρείες πρέπει να επενδύουν συνεχώς σε νέες τεχνολογίες, όπως το 5G, για να ανταποκρίνονται στις αυξημένες απαιτήσεις των πελατών και να προσφέρουν προηγμένες υπηρεσίες. Η βελτίωση της ποιότητας του δικτύου και η παροχή καινοτόμων υπηρεσιών μειώνουν την ανάγκη των πελατών να στραφούν σε άλλους παρόχους.

Η αποχώρηση πελατών είναι ένας από τους σημαντικότερους κινδύνους για τη βιωσιμότητα των τηλεπικοινωνιακών εταιρειών. Οι στρατηγικές μάρκετινγκ μπορούν να βοηθήσουν αποτελεσματικά στη μείωση του churn, εστιάζοντας στη βελτίωση της εμπειρίας πελάτη, στην εξατομίκευση των υπηρεσιών και στην αξιοποίηση των δεδομένων για προγνωστική ανάλυση. Σε έναν κλάδο με έντονο ανταγωνισμό και συνεχώς αυξανόμενες προσδοκίες, οι εταιρείες τηλεπικοινωνιών πρέπει να διατηρούν αδιάλειπτη επαφή με τις ανάγκες των πελατών τους και να προσαρμόζονται διαρκώς για να τους διατηρήσουν. Η επιτυχία στη διαχείριση του churn είναι καθοριστική για τη μακροπρόθεσμη κερδοφορία και ανταγωνιστικότητα των επιχειρήσεων στον κλάδο των τηλεπικοινωνιών.

Ο ρόλος του μάρκετινγκ σχέσεων

Το μάρκετινγκ σχέσεων αποτελεί θεμελιώδη στρατηγική για τη διατήρηση των πελατών και τη μακροχρόνια κερδοφορία. Η δέσμευση των πελατών, η οποία περιλαμβάνει την ικανοποίηση, την εμπιστοσύνη και την αφοσίωση, είναι ένας από τους πιο σημαντικούς παράγοντες που επηρεάζουν την επιτυχία μιας επιχείρησης. Στη βιομηχανία τηλεπικοινωνιών, όπου ο ανταγωνισμός είναι έντονος και οι πελάτες έχουν εύκολη πρόσβαση σε εναλλακτικούς παρόχους, η ενίσχυση της δέσμευσης πελατών είναι κρίσιμη.

Η εμπιστοσύνη αποτελεί βασικό συστατικό του μάρκετινγκ σχέσεων. Οι πελάτες που εμπιστεύονται την εταιρεία είναι πιο πιθανό να συνεχίσουν να χρησιμοποιούν τις υπηρεσίες της, ακόμα και όταν εμφανίζονται προβλήματα. Η επικοινωνία παίζει επίσης κεντρικό ρόλο, καθώς επιτρέπει στις εταιρείες να κατανοούν τις ανάγκες των πελατών τους, να προσφέρουν εξατομικευμένες λύσεις και να επιλύουν ζητήματα γρήγορα και αποτελεσματικά. Η προσωποποίηση, από την άλλη, δίνει στις επιχειρήσεις τη δυνατότητα να δημιουργήσουν πιο ισχυρούς δεσμούς με τους πελάτες τους, προσφέροντας μοναδικές εμπειρίες που ικανοποιούν τις συγκεκριμένες ανάγκες και προσδοκίες τους.

Διαχείριση πελατειακής εμπειρίας (Customer Experience)

Η εμπειρία των πελατών (Customer Experience - CX) είναι ένας από τους σημαντικότερους παράγοντες που επηρεάζουν την αφοσίωση και τη διατήρηση των πελατών. Η διαχείριση της εμπειρίας των πελατών δεν περιλαμβάνει μόνο την παροχή ποιοτικών υπηρεσιών αλλά και τη συνεχή βελτίωση των διαδικασιών, την παροχή εξατομικευμένων λύσεων και την ενίσχυση της επικοινωνίας με τους πελάτες.

Ο αντίκτυπος της CX στις στρατηγικές μάρκετινγκ είναι τεράστιος. Οι επιχειρήσεις που δίνουν προτεραιότητα στην εμπειρία των πελατών είναι πιο πιθανό να αυξήσουν την ικανοποίηση και την εμπιστοσύνη, οδηγώντας σε υψηλότερα επίπεδα δέσμευσης και, συνεπώς, σε μεγαλύτερη κερδοφορία. Για παράδειγμα, η ανάλυση των δεδομένων πελατών μπορεί να αποκαλύψει ποιες πτυχές της εμπειρίας τους χρειάζονται βελτίωση, επιτρέποντας στις εταιρείες να σχεδιάσουν στοχευμένες παρεμβάσεις.

Η ολοκλήρωση των δεδομένων πελατών με στρατηγικές μάρκετινγκ είναι κρίσιμη για τη βελτίωση της εμπειρίας τους. Η συλλογή και ανάλυση δεδομένων από διάφορες πηγές, όπως τα Call Detail Records (CDRs), τα κοινωνικά δίκτυα και οι έρευνες ικανοποίησης, επιτρέπει στις επιχειρήσεις να δημιουργήσουν ένα πιο ολοκληρωμένο προφίλ των πελατών τους. Αυτά τα δεδομένα μπορούν να χρησιμοποιηθούν για την ανάπτυξη πιο αποδοτικών στρατηγικών μάρκετινγκ, όπως η προσωποποίηση προσφορών, η ανάπτυξη loyalty προγραμμάτων και η στόχευση συγκεκριμένων ομάδων πελατών.

Θεωρίες πρόβλεψης συμπεριφοράς πελατών

Η πρόβλεψη της συμπεριφοράς πελατών είναι ένας από τους βασικούς στόχους του σύγχρονου μάρκετινγκ. Στη βιομηχανία τηλεπικοινωνιών, οι εταιρείες βασίζονται σε προηγμένα αναλυτικά εργαλεία και τεχνικές μηχανικής μάθησης για την κατανόηση και πρόβλεψη της συμπεριφοράς των πελατών τους.

Μία από τις πιο σημαντικές θεωρίες είναι το **loyalty marketing**, το οποίο επικεντρώνεται στη δημιουργία και διατήρηση της αφοσίωσης των πελατών. Οι στρατηγικές loyalty marketing περιλαμβάνουν την παροχή ανταμοιβών, τη δημιουργία συναισθηματικών δεσμών και την παροχή υψηλής ποιότητας υπηρεσιών. Η κατανόηση της συμπεριφοράς των πιστών πελατών μπορεί να βοηθήσει τις εταιρείες να αναγνωρίσουν μοτίβα που σχετίζονται με τη διατήρηση και την απώλεια πελατών.

Η θεωρία του churn είναι επίσης κρίσιμη. Το churn αναφέρεται στην αποχώρηση ενός πελάτη από μια υπηρεσία και συνήθως προκαλείται από παράγοντες όπως η δυσαρέσκεια με τις υπηρεσίες, η εμφάνιση καλύτερων προσφορών από ανταγωνιστές ή η απουσία συναισθηματικής σύνδεσης με την εταιρεία. Οι τεχνικές πρόβλεψης churn χρησιμοποιούν δεδομένα όπως το ιστορικό χρήσης, τα σχόλια πελατών και τα κοινωνικά δεδομένα για να εντοπίσουν πελάτες που είναι πιθανό να αποχωρήσουν. Εργαλεία όπως το **XGBoost** και το **Random Forest** έχουν αποδειχθεί εξαιρετικά αποδοτικά στη βελτίωση της ακρίβειας των προβλέψεων. (Etim, G. S., Etuk, I. U., James, E. E., & Ekpe, S. 2021)

Η ανάλυση συναισθημάτων (sentiment analysis) αποτελεί επίσης βασικό συστατικό της πρόβλεψης συμπεριφοράς πελατών. Μέσω της ανάλυσης δεδομένων από κοινωνικά δίκτυα και άλλες πηγές, οι επιχειρήσεις μπορούν να εντοπίσουν σημάδια δυσαρέσκειας ή ικανοποίησης, τα οποία μπορεί να επηρεάσουν την πιθανότητα εγκατάλειψης. Για παράδειγμα, οι αρνητικές αναφορές σε κοινωνικά μέσα μπορούν να είναι πρώιμα προειδοποιητικά σημάδια για πιθανούς πελάτες που σκοπεύουν να αποχωρήσουν.(Kumar, A., & Singh, R. 2022)

Το μάρκετινγκ σχέσεων, η διαχείριση της εμπειρίας των πελατών και οι θεωρίες πρόβλεψης συμπεριφοράς είναι άρρηκτα συνδεδεμένα. Οι επιχειρήσεις που επενδύουν σε στρατηγικές που ενσωματώνουν εμπιστοσύνη, επικοινωνία και προσωποποίηση είναι πιο πιθανό να διατηρήσουν τους πελάτες τους και να αυξήσουν την κερδοφορία τους. Ταυτόχρονα, η χρήση τεχνικών πρόβλεψης, όπως το churn και το sentiment analysis, ενισχύει την ικανότητα των επιχειρήσεων να ανταποκρίνονται στις ανάγκες των πελατών τους, καθιστώντας το μάρκετινγκ πιο αποτελεσματικό και αποδοτικό.

3. Τι είναι η Ανάλυση Συναισθήματος (Sentiment Analysis)

Η ανάλυση συναισθήματος, γνωστή και ως **Sentiment Analysis**, είναι ένα πεδίο της **Επεξεργασίας Φυσικής Γλώσσας** (Natural Language Processing - NLP) και της **Μηχανικής Μάθησης** (Machine Learning) που επιδιώκει να κατανοήσει και να κατηγοριοποιήσει τα συναισθήματα ή τις απόψεις που εκφράζονται σε κείμενα. Η ανάλυση συναισθήματος εφαρμόζεται κυρίως σε δεδομένα από ψηφιακές πηγές, όπως κοινωνικά δίκτυα, κριτικές προϊόντων, φόρουμ και σχόλια, και έχει στόχο να ανιχνεύσει τον συναισθηματικό τόνο των κειμένων. Ο τόνος αυτός μπορεί να είναι **θετικός**, **αρνητικός** ή **ουδέτερος**, ενώ συχνά αναγνωρίζονται και πιο λεπτομερείς συναισθηματικές καταστάσεις, όπως η έκπληξη, η χαρά, ο θυμός ή η λύπη.

Ιστορικό Υπόβαθρο και Εξέλιξη της Ανάλυσης Συναισθήματος

Αρχικά, η ανάλυση συναισθήματος βασιζόταν σε απλές μεθόδους λεξικογραφικής κατηγοριοποίησης, όπου οι λέξεις επισημαίνονταν ως θετικές, αρνητικές ή ουδέτερες, και στη συνέχεια αναλύονταν βάσει της παρουσίας αυτών των χαρακτηριστικών σε ένα κείμενο. Ωστόσο, καθώς τα δεδομένα έγιναν πιο πολύπλοκα και οι απαιτήσεις ανάλυσης αυξήθηκαν, οι ερευνητές ανέπτυξαν πιο σύνθετες τεχνικές, οι οποίες στηρίζονται στη **Μηχανική Μάθηση (Machine Learning)** και τη **Βαθιά Μάθηση (Deep Learning)**. Οι σύγχρονες τεχνικές μπορούν να εντοπίζουν και να αναλύουν το συναισθηματικό περιεχόμενο με μεγαλύτερη ακρίβεια, λαμβάνοντας υπόψη το συμφραζόμενο και την πολυπλοκότητα της γλώσσας.

Οι πρόοδοι στην βαθιά μάθηση, ιδιαίτερα μέσω των νευρωνικών δικτύων, όπως τα **Συγκλινοντικά Νευρωνικά Δίκτυα (CNN)** και τα **Αναδρομικά Νευρωνικά Δίκτυα (RNN)**, καθώς και τα πιο πρόσφατα μοντέλα όπως το **BERT** (Bidirectional Encoder Representations from Transformers), έχουν αλλάξει ριζικά το πεδίο της ανάλυσης συναισθήματος, επιτρέποντας σε υπολογιστικά συστήματα να διακρίνουν με ακρίβεια ακόμη και τις πιο λεπτές αποχρώσεις του συναισθήματος. (Σταλίδης, Γ., & Καρδαράς, Δ. 2015).

Βασικές Προσεγγίσεις στην Ανάλυση Συναισθήματος

Η ανάλυση συναισθήματος περιλαμβάνει διάφορες μεθοδολογικές προσεγγίσεις, οι οποίες μπορούν να κατηγοριοποιηθούν σε τρεις κύριες κατηγορίες: **μεθόδους βασισμένες σε λεξικά** (lexicon-based), **μεθόδους μηχανικής μάθησης** και **μεθόδους βαθιάς μάθησης**.

- **Μεθόδοι Βασισμένες σε Λεξικά**: Αυτές οι μέθοδοι περιλαμβάνουν τη χρήση λεξικών συναισθήματος, όπως το **SentiWordNet**, το οποίο περιέχει λέξεις με προκαθορισμένα συναισθηματικά χαρακτηριστικά (π.χ. θετικά, αρνητικά, ουδέτερα). Κάθε λέξη συνοδεύεται από μια συναισθηματική βαθμολογία, η οποία χρησιμοποιείται για την κατηγοριοποίηση

του συνολικού συναισθήματος ενός κειμένου. Παρόλο που οι λεξικογραφικές μέθοδοι είναι απλές στην υλοποίηση και ερμηνεία, παρουσιάζουν περιορισμούς ως προς την αδυναμία τους να αντιληφθούν το συμφραζόμενο. Για παράδειγμα, η λέξη "καταπληκτικός" μπορεί να είναι θετική σε ένα περιβάλλον, αλλά αρνητική σε ένα άλλο, αν συνδυαστεί με σαρκασμό.

- **Μέθοδοι Μηχανικής Μάθησης**: Οι μέθοδοι αυτές χρησιμοποιούν αλγορίθμους όπως ο **Naïve Bayes**, ο **Support Vector Machine (SVM)** και το **Random Forest**, οι οποίοι εκπαιδεύονται σε δεδομένα που έχουν κατηγοριοποιηθεί από ανθρώπους. Οι μέθοδοι μηχανικής μάθησης απαιτούν την εξαγωγή χαρακτηριστικών (feature extraction) από τα δεδομένα, όπως η συχνότητα εμφάνισης λέξεων ή η πολικότητα των εκφράσεων, και στη συνέχεια δημιουργούν ένα μοντέλο για την πρόβλεψη του συναισθήματος.
- **Μέθοδοι Βαθιάς Μάθησης (Deep Learning)**: Η βαθιά μάθηση χρησιμοποιεί τεχνικές νευρωνικών δικτύων, όπως τα **Συγκλινοντικά Νευρωνικά Δίκτυα (CNN)**, που είναι αποτελεσματικά στην αναγνώριση προτύπων σε δεδομένα κειμένου και τα **Αναδρομικά Νευρωνικά Δίκτυα (RNN)**, που μπορούν να κατανοούν τις αλληλουχίες λέξεων σε ένα κείμενο. Επιπλέον, τα πιο εξελιγμένα μοντέλα, όπως το **BERT** και το **GPT**, προσφέρουν τη δυνατότητα διπλής κατεύθυνσης επεξεργασίας κειμένου (bidirectional processing), γεγονός που επιτρέπει την κατανόηση του συμφραζόμενου με μεγαλύτερη ακρίβεια. Αυτά τα μοντέλα μπορούν να μάθουν από τεράστιους όγκους δεδομένων, επιτυγχάνοντας ακρίβεια που προσεγγίζει αυτήν των ανθρώπων. (Kumar, A., & Singh, R. 2022).

3.1 Εφαρμογές της Ανάλυσης Συναισθήματος

Η ανάλυση συναισθήματος βρίσκει εφαρμογή σε πολλούς τομείς, με κύριους κλάδους τη **Διαχείριση Σχέσεων με Πελάτες (CRM)**, το **Μάρκετινγκ**, την **Πολιτική Ανάλυση** και την **Υποστήριξη Πελατών**.

- **Διαχείριση Σχέσεων με Πελάτες (CRM)**: Οι επιχειρήσεις χρησιμοποιούν την ανάλυση συναισθήματος για να αξιολογήσουν τα σχόλια των πελατών στα μέσα κοινωνικής δικτύωσης, εντοπίζοντας θετικές και αρνητικές απόψεις. Αυτό τους επιτρέπει να ανταποκρίνονται γρήγορα σε παράπονα, ενισχύοντας την ικανοποίηση και αφοσίωση των πελατών. Για παράδειγμα, μια τηλεπικοινωνιακή εταιρεία μπορεί να χρησιμοποιήσει την ανάλυση συναισθήματος για να αναλύσει παράπονα πελατών σχετικά με τη ποιότητα σήματος, λαμβάνοντας άμεσα μέτρα για τη βελτίωση των υπηρεσιών.
- **Μάρκετινγκ και Ανάλυση Ανταγωνισμού**: Οι εταιρείες χρησιμοποιούν την ανάλυση συναισθήματος για να κατανοήσουν τη φήμη του προϊόντος ή της υπηρεσίας τους και να συγκρίνουν τις απόψεις των πελατών για τους ανταγωνιστές τους. Η ανάλυση των σχολίων των πελατών στο Twitter ή το Facebook μπορεί να δώσει πληροφορίες σχετικά με την

απόδοση των διαφημιστικών καμπανιών και να βοηθήσει στην ανάπτυξη στρατηγικών μάρκετινγκ.

- **Πολιτική Ανάλυση και Κοινωνική Έρευνα**: Στην πολιτική ανάλυση, η ανάλυση συναισθήματος μπορεί να χρησιμοποιηθεί για την κατανόηση της κοινής γνώμης σχετικά με ένα συγκεκριμένο ζήτημα. Οι πολιτικοί οργανισμοί και οι κυβερνήσεις μπορούν να αναλύσουν τα σχόλια στα κοινωνικά μέσα για να κατανοήσουν τη στάση των πολιτών και να προσαρμόσουν τις πολιτικές τους ανάλογα.
- **Υποστήριξη Πελατών και Chatbots**: Η ανάλυση συναισθήματος ενσωματώνεται σε chatbots και συστήματα υποστήριξης πελατών, τα οποία προσαρμόζουν τις απαντήσεις τους ανάλογα με τον τόνο των χρηστών. Εάν το σύστημα αναγνωρίσει ότι ένας πελάτης είναι απογοητευμένος, μπορεί να τον παραπέμψει σε έναν εκπρόσωπο για πιο άμεση και εξειδικευμένη υποστήριξη. (Πιζάνιας, Δ. Ε. 2018)

3.2 Προκλήσεις και Περιορισμοί

Η ανάλυση συναισθήματος παρουσιάζει αρκετές προκλήσεις, οι οποίες επηρεάζουν την ακρίβεια και τη χρησιμότητά της. Αυτές περιλαμβάνουν τη **διαχείριση του σαρκασμού και της ειρωνείας**, την **πολυπλοκότητα των συναισθημάτων**, και τη **διαθεσιμότητα δεδομένων υψηλής ποιότητας**.

- **Σαρκασμός και Ειρωνεία**: Ένα από τα μεγαλύτερα προβλήματα της ανάλυσης συναισθήματος είναι η αδυναμία των συστημάτων να κατανοήσουν σαρκαστικά ή ειρωνικά σχόλια. Για παράδειγμα, η φράση «Η εξυπηρέτηση ήταν εξαιρετική, όπως πάντα!» μπορεί να έχει αρνητικό τόνο όταν χρησιμοποιείται σαρκαστικά. Αυτό αποτελεί σημαντικό περιορισμό, ειδικά στα μέσα κοινωνικής δικτύωσης, όπου ο σαρκασμός είναι συχνός.

- **Πολυπλοκότητα και Ποικιλία Συναισθημάτων**: Τα συναισθήματα είναι σύνθετα και συχνά δεν περιορίζονται απλά σε θετικά, αρνητικά ή ουδέτερα. Συναισθήματα όπως η «έκπληξη», ο «θυμός» ή η «απογοήτευση» μπορεί να εμφανίζονται σε συνδυασμό, καθιστώντας δύσκολη την κατηγοριοποίηση τους σε μια ενιαία κατηγορία. Για την αντιμετώπιση αυτού του ζητήματος, έχουν αναπτυχθεί πιο εξελιγμένα μοντέλα που προσπαθούν να κατανοήσουν την πολυπλοκότητα των συναισθημάτων.

- **Γλωσσικές και Πολιτισμικές Ιδιαιτερότητες**: Οι πολιτισμικές διαφορές επηρεάζουν τον τρόπο έκφρασης των συναισθημάτων και η ίδια λέξη μπορεί να έχει διαφορετικό συναίσθημα σε διαφορετικά συμφραζόμενα. Αυτό καθιστά δύσκολη τη χρήση μοντέλων σε διαφορετικές γλώσσες χωρίς προσαρμογή. Για παράδειγμα, η λέξη "πολύ" μπορεί να έχει διαφορετική ένταση και συναισθηματική αξία σε διαφορετικές γλώσσες και κουλτούρες.

- **Διαθεσιμότητα Δεδομένων και Χρήση Πολύ-γλωσσικών Μοντέλων**: Η ποιότητα των αποτελεσμάτων της ανάλυσης συναισθήματος εξαρτάται σε μεγάλο βαθμό από την ποιότητα και το μέγεθος των δεδομένων. Τα πολύ-γλωσσικά μοντέλα, όπως το **M-BERT**, έχουν αναπτυχθεί για να ανταποκριθούν στην ανάγκη για ανάλυση δεδομένων σε διαφορετικές γλώσσες, αλλά εξακολουθούν να παρουσιάζουν περιορισμούς στη διατήρηση της ακρίβειας.

Η ανάλυση συναισθήματος αποτελεί ένα ισχυρό εργαλείο που μπορεί να ενισχύσει την κατανόηση των απόψεων και της εμπειρίας των πελατών. Παρόλο που έχει καθιερωθεί ως σημαντική τεχνολογία στον επιχειρηματικό και κοινωνικό τομέα, εξακολουθεί να αντιμετωπίζει προκλήσεις, ειδικά όσον αφορά την πολυπλοκότητα της γλώσσας και την ακρίβεια της συναισθηματικής κατηγοριοποίησης. Με τη συνεχιζόμενη εξέλιξη των τεχνολογιών βαθιάς μάθησης και την αύξηση της διαθεσιμότητας δεδομένων, αναμένεται ότι η ανάλυση συναισθήματος θα συνεχίσει να βελτιώνεται, επιτρέποντας την ολοένα και πιο ακριβή και αξιόπιστη κατανόηση των συναισθημάτων και της στάσης των ανθρώπων. (Kumar, A., & Singh, R. 2022)

3.3 Ανάλυση Συναισθημάτων (Sentiment Analysis)

Η ανάλυση συναισθημάτων παρέχει τη δυνατότητα κατανόησης των συναισθημάτων πελατών μέσω δεδομένων από κοινωνικά δίκτυα. Περιλαμβάνει:

- **Κατηγοριοποίηση συναισθημάτων:** Θετικά, αρνητικά και ουδέτερα σχόλια.
- **Εφαρμογές:** Προσαρμογή μηνυμάτων μάρκετινγκ βάσει των συναισθημάτων που εκφράζουν οι πελάτες.

Αξιολόγηση και Βελτιστοποίηση Στρατηγικών Μάρκετινγκ

Μετρικές Αξιολόγησης

- **Lifetime Value (CLV):**
- Υπολογίζει τη συνολική αξία ενός πελάτη κατά τη διάρκεια της σχέσης του με την εταιρεία.

Return on Investment (ROI):

  - Αξιολογεί την απόδοση των καμπανιών μάρκετινγκ.

Βελτιστοποίηση Στρατηγικών

✓ **Ανάλυση Αποτελεσμάτων:**

Χρήση εργαλείων όπως Google Analytics για την παρακολούθηση καμπανιών.

✓ **Δυναμική Προσαρμογή:**

Προσαρμογή καμπανιών σε πραγματικό χρόνο βάσει δεδομένων χρήσης.

4. Μεθοδολογία και Εργαλεία

Η ανάλυση της πελατειακής συμπεριφοράς στις τηλεπικοινωνίες βασίζεται σε μεγάλο βαθμό στη διαχείριση μεγάλων δεδομένων (Big Data), την τεχνητή νοημοσύνη (AI), και τη χρήση προχωρημένων τεχνικών μηχανικής μάθησης, όπως το **XGBoost, Decision Tree** και το **Random Forest**. Επιπλέον, η ανάλυση συναισθημάτων (sentiment analysis) δίνει τη δυνατότητα στις επιχειρήσεις να κατανοήσουν καλύτερα τις ανάγκες και τις αντιλήψεις των πελατών, οδηγώντας σε βελτιωμένες στρατηγικές μάρκετινγκ και διατήρησης πελατών.

Για να μπορέσουμε να εξάγουμε ασφαλή συμπεράσματα θα κάνουμε χρήση των Τεχνικών Μηχανικής Μάθησης σε πραγματικά σύνολα μεγάλων δεδομένων (**Big Data**).

4.1 Τα Μεγάλα Δεδομένα (Big Data)

Τι είναι τα μεγάλα δεδομένα όμως. Τα **Big Data** όπως ονομάζονται δεν είναι τίποτε άλλο από ένα σύνολο πληροφοριών, συγκεκριμένα είναι ένας τεράστιος όγκος δεδομένων που συλλέγονται, αποθηκεύονται, και επεξεργάζονται από ποικίλες πηγές. Αυτά τα δεδομένα διακρίνονται από κάποια χαρακτηριστικά. Αυτά τα χαρακτηριστικά είναι τα λεγόμενα 5 Vs.

i. **Όγκος (Volume):** Ο τεράστιος όγκος δεδομένων περιλαμβάνει εκατομμύρια εγγραφές πελατών, πληροφορίες από Call Detail Records (CDRs) και κοινωνικά μέσα. Για παράδειγμα, οι τηλεπικοινωνιακές εταιρείες επεξεργάζονται petabytes δεδομένων καθημερινά.

ii. **Ποικιλία (Variety):** Τα δεδομένα είναι δομημένα (π.χ., πίνακες συναλλαγών), ημιδομημένα (π.χ., δεδομένα JSON), και μη δομημένα (π.χ., σχόλια πελατών, εικόνες, βίντεο).

iii. **Ταχύτητα (Velocity):** Η συνεχής παραγωγή δεδομένων σε πραγματικό χρόνο, όπως τα δεδομένα χρήσης δικτύου, απαιτεί προηγμένα συστήματα για άμεση επεξεργασία.

iv. **Ακρίβεια (Veracity):** Η αξιοπιστία των δεδομένων είναι κρίσιμη για τη λήψη επιχειρηματικών αποφάσεων.

v. **Αξία (Value):** Η εξαγωγή αξίας από τα Big Data επιτρέπει τη δημιουργία εξατομικευμένων εμπειριών και την ενίσχυση της δέσμευσης πελατών. (Βελισσαρίου, Α., & Πράπας, Δ. 2010).

4.2 Πηγές Big Data στις Τηλεπικοινωνίες

Τα Big Data στις τηλεπικοινωνίες συλλέγονται από πολλαπλές πηγές:

- **Call Detail Records (CDRs):** Περιλαμβάνουν δεδομένα όπως διάρκεια κλήσεων, συχνότητα χρήσης SMS, και κατανάλωση Internet.
- **Social Media:** Τα σχόλια και οι δημοσιεύσεις στα κοινωνικά δίκτυα παρέχουν ανεκτίμητη πληροφορία για τις αντιλήψεις των πελατών.
- **Υπηρεσίες Εξυπηρέτησης Πελατών:** Παράπονα, ερωτήματα, και αιτήματα προσφέρουν δεδομένα για τη βελτίωση των υπηρεσιών.

4.3 Εφαρμογές Big Data

Πρόβλεψη Συμπεριφοράς Πελατών:

Αναγνώριση μοτίβων που σχετίζονται με την εγκατάλειψη (churn), όπως μειωμένη χρήση υπηρεσιών.

- ✓ **Ανάλυση Εμπειρίας Πελατών:**

Τα δεδομένα από κοινωνικά δίκτυα και έρευνες ικανοποίησης βοηθούν στον εντοπισμό προβλημάτων.

- ✓ **Εξατομικευμένες Προσφορές:**

Εξαγωγή προτάσεων βασισμένων στις ατομικές προτιμήσεις και ανάγκες πελατών.

4.4 Διαχείριση Δεδομένων και Χαρακτηριστικών

Συλλογή Δεδομένων

Η συλλογή δεδομένων είναι το πρώτο βήμα για την ανάλυση της πελατειακής συμπεριφοράς. Οι κυριότερες πηγές δεδομένων περιλαμβάνουν:

Δεδομένα μάρκετινγκ

- ✓ Ανάλυση καμπανιών όπως click-through rates (CTR) και conversion rates.
- ✓ Email και SMS καμπάνιες για την αξιολόγηση της αποτελεσματικότητας.

Συμπεριφορικά δεδομένα

- ✓ Ιστορικό χρήσης υπηρεσιών (π.χ., διάρκεια κλήσεων, κατανάλωση δεδομένων).
- ✓ Συνήθειες πληρωμών και επιλογές προγραμμάτων.

Κοινωνικά Μέσα Δικτύωσης (Social Media)

- ✓ Αναφορές πελατών από πλατφόρμες όπως το Twitter και το Facebook.
- ✓ Ανάλυση hashtags και σχολίων για ανίχνευση τάσεων.

4.5 Προεπεξεργασία Δεδομένων

Η προεπεξεργασία δεδομένων είναι ζωτικής σημασίας για τη δημιουργία ενός ποιοτικού συνόλου δεδομένων. Περιλαμβάνει:

- Καθαρισμός δεδομένων, γίνεται επεξεργασίας των δεδομένων με σκοπό την αφαίρεση θορύβου και μη έγκυρων δεδομένων που δεν θα είναι κατανοητά στον αλγόριθμο.
- Συμπλήρωση ελλείψεων μέσω μεθόδων όπως η μέση τιμή ή η προγνωστική ανάλυση ανάλογα την περίπτωση .
- Κανονικοποίηση μετατρέπουμε τα δεδομένα σε μια ενιαία κλίμακα για καλύτερη ανάλυση.
- Ομαδοποίηση χαρακτηριστικών, κατηγοριοποιούμε τα δεδομένων σε δημογραφικά, συμπεριφορικά, και αγοραστικά χαρακτηριστικά.
- Εξαγωγή νέων χαρακτηριστικών, δημιουργία δεικτών όπως "πιθανότητα αποχώρησης" ή "συχνότητα αλληλεπίδρασης."

Χρήση AI στο Μάρκετινγκ

Η τεχνητή νοημοσύνη προσφέρει τις απαραίτητες τεχνικές για την κατανόηση και αξιοποίηση των Big Data στο Μάρκετινγκ, οι αλγόριθμοι που θα δούμε στις μελέτες Περίπτωσης είναι η XGBoost, Random Forest και Decision Tree.

Αλγόριθμοι για Πρόβλεψη Συμπεριφοράς

1. **XGBoost (Extreme Gradient Boosting):**

   Τι είναι το XGBoost (Extreme Gradient Boosting); Το XGBoost είναι ένας αλγόριθμος μηχανικής μάθησης που βασίζεται στην τεχνική του Gradient Boosting. Χρησιμοποιείται για προβλήματα ταξινόμησης και παλινδρόμησης, και είναι γνωστός για την υψηλή ακρίβεια και την απόδοσή του, ιδιαίτερα σε μεγάλα και ανισόρροπα δεδομένα. Το XGBoost χρησιμοποιείται ευρέως στην πρόβλεψη churn για τον εντοπισμό πελατών που είναι πιθανό να αποχωρήσουν, με βάση χαρακτηριστικά όπως το ιστορικό χρήσης και οι δημογραφικές πληροφορίες.

   Χαρακτηριστικά XGBoost:

a. **Boosting:** Συνδυάζει πολλά "ασθενή" μοντέλα για τη δημιουργία ενός "ισχυρού" μοντέλου.

b. **Παράλληλη Επεξεργασί**α: Υποστηρίζει ταχύτερη εκπαίδευση σε μεγάλα δεδομένα.

c. **Αντιμετώπιση Σπανιότητας:** Μπορεί να διαχειριστεί δεδομένα με πολλές ελλείψεις.

d. **Υποστήριξη Τακτικοποίησης:** Περιλαμβάνει ενσωματωμένες μεθόδους για την αποφυγή υπερ-προσαρμογής (regularization).

Ο XGBoost είναι ιδιαίτερα αποτελεσματικός για προβλήματα ταξινόμησης, όπως η πρόβλεψη churn , εφαρμόζεται επίσης για την κατηγοριοποίηση πελατών βάσει κινδύνου (υψηλού, μεσαίου, χαμηλού). (Chen, T., & Guestrin, C. 2016)

Random Forest:

Τι είναι το Random Forest; Το Random Forest είναι ένας αλγόριθμος μηχανικής μάθησης που χρησιμοποιεί πολλαπλά Decision Trees για να παράγει πιο ακριβείς και σταθερές προβλέψεις. Χρησιμοποιείται τόσο για ταξινόμηση όσο και για παλινδρόμηση. Το Random Forest μπορεί να αναλύσει μεγάλα και πολύπλοκα δεδομένα, όπως τα δημογραφικά και τα οικονομικά δεδομένα πελατών, για να προβλέψει την πιθανότητα churn. (Σταλίδης, Γ., & Καρδαράς, Δ. 2015)

Χαρακτηριστικά Random Forest:

a. **Απόφαση μέσω Ψηφοφορίας:** Συνδυάζει τις προβλέψεις πολλαπλών Decision Trees για να παράγει το τελικό αποτέλεσμα.

b. **Τυχαιότητα:** Επιλέγει τυχαία χαρακτηριστικά και δείγματα δεδομένων για κάθε δέντρο, μειώνοντας την πιθανότητα υπερ-προσαρμογής.

c. **Ανθεκτικότητα:** Είναι λιγότερο ευαίσθητο σε θόρυβο ή ελλείψεις στα δεδομένα.

Decision Tree:

Τι είναι το Decision Tree; Το Decision Tree είναι ένας βασικός αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται τόσο για προβλήματα ταξινόμησης όσο και παλινδρόμησης. Είναι δομημένο σαν δέντρο, με κόμβους που αντιπροσωπεύουν αποφάσεις. (Σταλίδης, Γ., & Καρδαράς, Δ. 2015) Το Decision Tree χρησιμοποιείται για την κατηγοριοποίηση πελατών σε ομάδες, όπως "υψηλού κινδύνου" και "χαμηλού κινδύνου", με βάση χαρακτηριστικά όπως η συχνότητα χρήσης και οι καθυστερήσεις στις πληρωμές. Το Decision Tree είναι ένα από τα πιο κατανοητά και απλά εργαλεία ανάλυσης δεδομένων.

a. **Δομή Δέντρου:** Κάθε κόμβος του δέντρου αντιπροσωπεύει μια ερώτηση, και κάθε κλάδος μια πιθανή απάντηση.

b. **Διαισθητική Κατανόηση:** Είναι εύκολο να εξηγηθεί και να οπτικοποιηθεί.

c. **Ευαισθησία στον Θόρυβο:** Μπορεί να οδηγήσει σε υπερ-προσαρμογή αν δεν εφαρμοστεί τακτικοποίηση.

Με αυτούς τους αλγορίθμους, η Τεχνητή Νοημοσύνη μπορεί να βοηθήσει τις επιχειρήσεις να προβλέψουν τη συμπεριφορά των πελατών, να βελτιώσουν τις στρατηγικές μάρκετινγκ και να αυξήσουν την αποδοτικότητα των καμπανιών τους. Θεωρούμε ότι είναι οι πλέον κατάλληλοι για να μπορέσουμε να προβλέψουμε την πιθανότητα Churn.

Η ολοκληρωμένη χρήση των Big Data, της AI και των προηγμένων αλγορίθμων προσφέρει ισχυρά εργαλεία για την πρόβλεψη και ανάλυση πελατειακής συμπεριφοράς. Μέσω των τεχνικών αυτών, οι τηλεπικοινωνιακές εταιρείες μπορούν να μειώσουν την αποχώρηση πελατών (churn), να εξατομικεύσουν τις υπηρεσίες τους, και να ενισχύσουν τη συνολική εμπειρία πελατών.(Ahmad, A. K., Jafar, A., & Aljoumaa, K. 2019)

5. Μελέτες Περίπτωσης

5.1 Πρώτη Μελέτη Περίπτωσης

Δεδομένα προς ανάλυση

Για να μπορέσουμε να έχουμε όσο των δυνατών μια πιο ολοκληρωμένη εικόνα και ανάλυση για το Churn θα δούμε παρακάτω 2 δείγματα για ανάλυση, με δεδομένα που θα χρησιμοποιήσουμε μετά από αναζήτηση στο διαδίκτυο τα βρήκαμε από την δημοφιλή πλατφόρμα Kaggle.com που επικεντρώνεται στην ανάλυση δεδομένων και τη μηχανική μάθηση .

Το πρώτο μας παράδειγμα θα περιλαμβάνει δεδομένα με στοιχεία για περισσότερους από 3000 πελάτες. Τα στοιχεία αυτά είναι πραγματικά από την Orange Telecom και περιλαμβάνουν πολλά στοιχεία Μεταβλητών (Variables). https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets

Από τα δεδομένα που βρήκαμε κατά την αναζήτηση μας έγινε καθαρισμός και ομογενοποίηση των δεδομένων πριν ξεκινήσουμε να χρησιμοποιήσουμε τους αλγόριθμους μηχανικής μάθησης.

Ακόμη διαχωρίστηκαν τα δεδομένα αυτά με τυχαία σειρά ώστε να δημιουργήσουμε δεδομένα για εκπαίδευση (Data Train) των αλγορίθμων και δεδομένα επαλήθευση των τιμών (Data Test) επιλέξαμε τις παρακάτω μεταβλητές ως τις κυριότερες για να μπορέσουμε να εξάγουμε τα αποτελέσματα μας.

Variables - Μεταβλητές

1. Account_length

- **Περιγραφή:** Ο αριθμός των μηνών που ο πελάτης έχει λογαριασμό στην εταιρεία τηλεπικοινωνιών.
- **Σημασία:** Μεγάλη διάρκεια μπορεί να υποδηλώνει πιστότητα, ενώ μικρή διάρκεια ενδέχεται να σχετίζεται με μεγαλύτερο κίνδυνο αποχώρησης.

2. Voice_mail_plan

- **Περιγραφή:** Αν ο πελάτης έχει εγγραφεί σε κάποιο πλάνο voicemail (ναι ή όχι).
- **Σημασία:** Η ύπαρξη ή μη voicemail plan μπορεί να επηρεάσει την ικανοποίηση του πελάτη. Οι πελάτες που δεν χρησιμοποιούν voicemail ίσως είναι λιγότερο δεσμευμένοι.

3. Number_vmail_messages

- **Περιγραφή:** Ο αριθμός των μηνυμάτων που αποθηκεύονται στο voicemail.
- **Σημασία:** Πελάτες που χρησιμοποιούν συχνά το voicemail μπορεί να δείχνουν αυξημένη δέσμευση με την υπηρεσία.

4. Total_day_minutes

- **Περιγραφή:** Ο συνολικός αριθμός λεπτών κλήσεων κατά τη διάρκεια της ημέρας.
- **Σημασία:** Υψηλή χρήση κατά τη διάρκεια της ημέρας δείχνει πελάτες που βασίζονται πολύ στις υπηρεσίες τηλεφωνίας, αλλά και μεγαλύτερο κόστος για αυτούς.

5. Total_day_calls

- **Περιγραφή:** Ο συνολικός αριθμός κλήσεων κατά τη διάρκεια της ημέρας.
- **Σημασία:** Η ποσότητα κλήσεων, σε συνδυασμό με τη διάρκεια, δίνει μια εικόνα για τη συχνότητα και την ένταση χρήσης.

6. Total_day_charge

- **Περιγραφή:** Το συνολικό κόστος για κλήσεις κατά τη διάρκεια της ημέρας.
- **Σημασία:** Υψηλές χρεώσεις μπορεί να σχετίζονται με δυσαρέσκεια, ειδικά αν συνδυάζονται με χαμηλή αντιληπτή αξία.

7. Total_eve_minutes

- **Περιγραφή:** Ο συνολικός αριθμός λεπτών κλήσεων κατά τη διάρκεια της βραδιάς.
- **Σημασία:** Δείχνει τη χρήση των υπηρεσιών σε ώρες εκτός αιχμής, οι οποίες συνήθως έχουν χαμηλότερες χρεώσεις.

8. Total_eve_calls

- **Περιγραφή:** Ο συνολικός αριθμός κλήσεων κατά τη διάρκεια της βραδιάς.
- **Σημασία:** Μπορεί να υποδηλώνει προτιμήσεις πελατών για επικοινωνία σε χαλαρότερους χρόνους.

9. Total_eve_charge

- **Περιγραφή:** Το συνολικό κόστος για βραδινές κλήσεις.
- **Σημασία:** Χαμηλές χρεώσεις μπορεί να ενισχύσουν την ικανοποίηση του πελάτη.

10. Total_night_minutes

- **Περιγραφή:** Ο συνολικός αριθμός λεπτών κλήσεων κατά τη διάρκεια της νύχτας.
- **Σημασία:** Μπορεί να υποδεικνύει χρήστες που αξιοποιούν τις οικονομικές χρεώσεις της νύχτας.

11. Total_night_calls

- **Περιγραφή:** Ο συνολικός αριθμός κλήσεων κατά τη διάρκεια της νύχτας.

- **Σημασία:** Παρόμοιο με το Total_night_minutes, δείχνει προτιμήσεις επικοινωνίας.

12. Total_night_charge

- **Περιγραφή:** Το συνολικό κόστος για νυχτερινές κλήσεις.
- **Σημασία:** Οι νυχτερινές χρεώσεις είναι συνήθως χαμηλότερες και σχετίζονται με πελάτες που είναι πιο προσαρμοστικοί στις τιμολογιακές πολιτικές.

13. Total_intl_minutes

- **Περιγραφή:** Ο συνολικός αριθμός λεπτών διεθνών κλήσεων.
- **Σημασία:** Πελάτες με υψηλή διεθνή χρήση μπορεί να είναι πιο σημαντικοί για την εταιρεία λόγω υψηλών χρεώσεων.

14. Total_intl_calls

- **Περιγραφή:** Ο συνολικός αριθμός διεθνών κλήσεων.
- **Σημασία:** Υποδεικνύει πελάτες που έχουν ανάγκη διεθνούς επικοινωνίας.

15. Total_intl_charge

- **Περιγραφή:** Το συνολικό κόστος για διεθνείς κλήσεις.
- **Σημασία:** Οι υψηλές διεθνείς χρεώσεις μπορεί να είναι πηγή δυσαρέσκειας ή ευκαιρία για εξατομικευμένες προσφορές.

16. Churn

- **Περιγραφή:** Δυαδική μεταβλητή που δείχνει αν ένας πελάτης αποχώρησε ή παρέμεινε στην εταιρεία (0 = παραμονή, 1 = αποχώρηση).
- **Σημασία:** Είναι η βασική εξαρτημένη μεταβλητή (target variable) για την πρόβλεψη churn. Όλες οι υπόλοιπες μεταβλητές χρησιμοποιούνται για να προβλεφθεί αυτή η έκβαση.

Έπειτα από τα παραπάνω βήματα τα στοιχεία αναλύθηκαν με του αλγόριθμους μηχανικής μάθησης όπως αναφέραμε παραπάνω **XGBoost, Decision Tree & Random Forest**.

5.1.1 Αποτελέσματα Ανάλυσης

Πριν «τρέξουμε» το αρχείο με τα δεδομένα έγινε μία στατιστική ανάλυση και παρακάτω παραθέτουμε του πίνακες με τα αποτελέσματα της ανάλυσης

Train Data								
A/A	Variables	mean	median	std	min	max	skewness	kurtosis
0	Account_length	100,7723862	100	39,56739197	1	243	0,092263832	-0,18160884
1	Voice_mail_plan	0,275637819	0	0,446835106	0	1	1,004228084	-0,991525955
2	Number_vmail_messages	7,992996498	0	13,56694446	0	50	1,278780551	-0,007592806
3	Total_day_minutes	179,8141571	180,5	53,9357486	0	346,8	-0,094418656	0,021371024
4	Total_day_calls	100,4907454	101	19,92825524	0	160	-0,079445365	0,210493033
5	Total_day_charge	30,56897949	30,69	9,169105922	0	58,96	-0,094393329	0,021468798
6	Total_eve_minutes	199,7637819	200,2	50,74511752	0	363,7	-0,035256804	-0,037592631
7	Total_eve_calls	99,85092546	100	20,1584466	0	170	-0,038800426	0,131501107
8	Total_eve_charge	16,9802001	17,02	4,313272032	0	30,91	-0,035189942	-0,037632381
9	Total_night_minutes	201,103902	202,1	50,80517009	43,7	395	-0,030328463	0,035417025
10	Total_night_calls	100,3376688	101	20,01604001	33	166	0,009955659	-0,068797764
11	Total_night_charge	9,049794897	9,09	2,286262825	1,97	17,77	-0,03036637	0,034920829
12	Total_intl_minutes	10,22156078	10,2	2,801914654	0	20	-0,193675674	0,581400342
13	Total_intl_calls	4,488244122	4	2,462675223	0	19	1,267131178	2,580630918
14	Total_intl_charge	2,760275138	2,75	0,756489003	0	5,4	-0,193860731	0,582229342
15	Churn	0,142571286	0	0,349635116	0	1	2,044581326	2,180312801

Πίνακας 1: Στατιστικά Δεδομένα εκπαίδευσης των αλγόριθμων

Επεξήγηση και ορισμός των τιμών

V**ariables:** Τα ονόματα των μεταβλητών του dataset.

- Κάθε γραμμή αντιπροσωπεύει μια διαφορετική μεταβλητή.(Εχουν αναλυθεί οι μεταβλητές παραπάνω

mean (Μέσος Όρος):

- Ο μέσος όρος των τιμών για κάθε μεταβλητή.
- Για παράδειγμα, ο μέσος χρόνος ομιλίας στη διάρκεια της ημέρας (Total_day_min) είναι **178,39 λεπτά**.

median (Διάμεσος):

- Η μέση τιμή, δηλαδή η τιμή που χωρίζει το δείγμα σε δύο ίσα μέρη.
- Για τη μεταβλητή "Total_day_min", η διάμεσος είναι **177**, υποδεικνύοντας ότι οι μισοί χρήστες έχουν λιγότερο από 177 λεπτά ομιλίας την ημέρα.

std (Τυπική Απόκλιση):

- Δείχνει πόσο διασκορπισμένες είναι οι τιμές γύρω από το μέσο όρο.
- Η υψηλή τυπική απόκλιση, όπως για τη "Total_day_charge" (**30,32**), δείχνει μεγάλη ποικιλία στη χρήση.

min (Ελάχιστο):

- Η μικρότερη τιμή της μεταβλητής.
- Για τη "Total_day_min", το ελάχιστο είναι **0**, που σημαίνει ότι υπάρχουν χρήστες χωρίς ομιλία στη διάρκεια της ημέρας.

max (Μέγιστο):

- Η μεγαλύτερη τιμή της μεταβλητής.
- Για παράδειγμα, η "Total_day_min" φτάνει μέχρι **350 λεπτά**.

skewness (Σκέψη):

- Μέτρο της ασυμμετρίας της κατανομής.
- Αν το skewness είναι:
    - **> 0**, η κατανομή είναι ασύμμετρη προς τα δεξιά (tail προς υψηλές τιμές).
    - **< 0**, η κατανομή είναι ασύμμετρη προς τα αριστερά (tail προς χαμηλές τιμές).
    - Για παράδειγμα, το skewness για τη "Total_intl_min" (**0,715**) δείχνει ότι η κατανομή είναι ελαφρώς ασύμμετρη προς τα δεξιά.

kurtosis (Κύρτωση):

- Μετρά την "αιχμηρότητα" της κατανομής.
- Αν το kurtosis είναι:
    - **> 0**, η κατανομή έχει περισσότερες τιμές στα άκρα (heavy tails).
    - **< 0**, η κατανομή έχει λιγότερες τιμές στα άκρα (light tails).
    - Για παράδειγμα, το kurtosis για τη "Total_intl_charge" (**5,41**) δείχνει έντονη αιχμηρότητα.

Παρακάτω είναι οι πίνακας με τα αποτελέσματα εξαγωγής των μεταβλητών

Test Data								
A/A	Variables	mean	median	std	min	max	skewness	kurtosis
0	Account_length	100,0930931	101	39,5066845	1	224	0,041850937	-0,012727491
1	Voice_mail_plan	0,271771772	0	0,444872876	0	1	1,026037956	-0,947246113
2	Number_vmail_messages	8,070570571	0	13,71315337	0	45	1,255721607	-0,121749025
3	Total_day_minutes	178,3900901	177	54,95964059	0	350,8	0,069015527	0,018508006
4	Total_day_calls	99,79279279	101	20,1479291	0	151	-0,271115436	0,486496069
5	Total_day_charge	30,32671171	30,09	9,342945139	0	59,64	0,069022692	0,018771752
6	Total_eve_minutes	202,3061562	202,5	51,50158347	49,2	354,2	0,046401923	-0,014304729
7	Total_eve_calls	100,5660661	101	20,1513639	12	154	-0,147338705	0,365863603
8	Total_eve_charge	17,19618619	17,21	4,37776396	4,18	30,11	0,04633246	-0,014468546
9	Total_night_minutes	201,4250751	198,25	50,68052607	54	381,9	0,182893274	0,082734406
10	Total_night_calls	99,44144144	100	17,46812697	49	153	-0,027975274	-0,170528102
11	Total_night_charge	9,064114114	8,92	2,280556912	2,43	17,19	0,18285375	0,083054574
12	Total_intl_minutes	10,28183183	10,3	2,74642278	0	17,9	-0,318135786	0,715109309
13	Total_intl_calls	4,406906907	4	2,434958185	0	20	1,637634398	5,414179572
14	Total_intl_charge	2,776711712	2,78	0,741447026	0	4,83	-0,31811117	0,716629444
15	Churn	0,154654655	0	0,361575155	0	1	1,910227187	1,648967908

Πίνακας 2 : Δεδομένα επαλήθευσης των αλγόριθμων

Percentage Differences								
A/A	Variables	mean	median	std	min	max	skewness	kurtosis
0	Account_length	0,676366195	-0,995024876	0,153545806	0	8,137044968	75,17873819	173,8031673
1	Voice_mail_plan	1,412487908	0	0,440105888	0	0	-2,14847428	4,567823407
2	Number_vmail_messages	-0,965838683	0	-1,07190899	0	10,52631579	1,819603399	-176,518638
3	Total_day_minutes	0,795114519	1,958041958	-1,880505682	0	-1,146788991	1286,724825	14,35851077
4	Total_day_calls	0,6969645	0	-1,09628131	0	5,78778135	-109,3505439	-79,19866636
5	Total_day_charge	0,795681183	1,974333662	-1,878119463	0	-1,146711636	1288,229547	13,40461515
6	Total_eve_minutes	-1,264642802	-1,142289546	-1,479687742	-200	2,646608163	-1465,371949	89,74599734
7	Total_eve_calls	-0,713652668	-0,995024876	0,035141318	-200	9,87654321	-116,620593	-94,2417068
8	Total_eve_charge	-1,263949233	-1,110137306	-1,4841022	-200	2,622091118	-1463,267161	88,91909
9	Total_night_minutes	-0,159577642	1,923317098	0,245638591	-21,08495394	3,372377397	-279,5162739	-80,09616192
10	Total_night_calls	0,897218325	0,995024876	13,59460939	-39,02439024	8,150470219	-420,9960618	-85,01407715
11	Total_night_charge	-0,158101895	1,887840089	0,249885623	-20,90909091	3,318077803	-279,6560877	-81,59962787
12	Total_intl_minutes	-0,587912962	-0,975609756	2,000306389	0	11,0817942	-48,63514087	-20,62598877
13	Total_intl_calls	1,828798969	0	1,131854318	0	-5,128205128	-25,51002557	-70,88469845
14	Total_intl_charge	-0,593701036	-1,084990958	2,008360504	0	11,14369501	-48,53799138	-20,69510625
15	Churn	-8,130763418	0	-3,357667739	0	0	6,794470001	27,7516815

Πίνακας 3 : % διαφοράς Train & Test Data

Από την πρώτη ανάγνωση των στατιστικών στοιχείων που μας έδωσαν τα δεδομένα της εκπαίδευση και της επαλήθευσης αντιλαμβανόμαστε ότι:

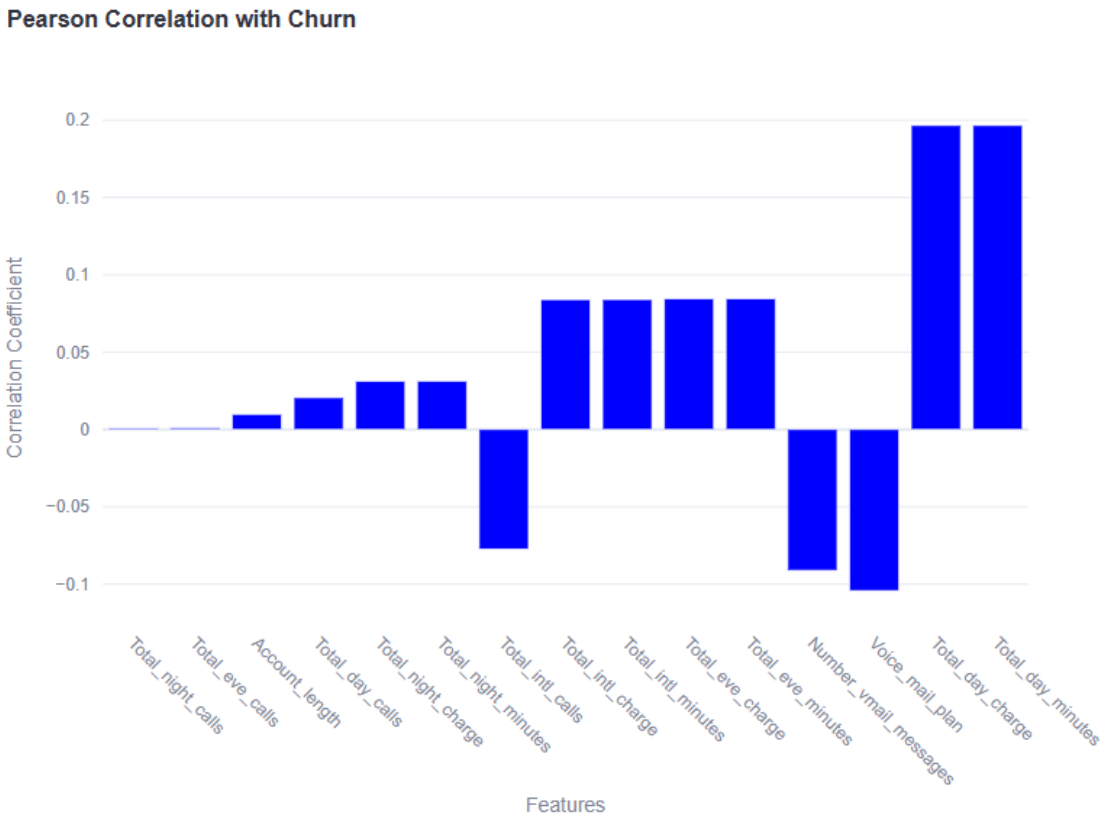
1. Το **Churn** που είναι η βασική εξαρτημένη μεταβλητή και ουσιαστικά αναφέρεται στο ποσοστό αποχώρησης που είναι και η μεταβλητή που αναζητούμε. Εμφανίζεται ελαφρώς υψηλότερη στα δεδομένα των Test με μια απόκλιση μόλις 8% συγκριτικά με το Set δεδομένων εκπαίδευσης .

   Το ποσοστό 15% Churn είναι ένα ποσοστό που υποδεικνύει ότι εταιρεία δραστηριοποιείται σε μία χώρα που έχει ώριμη αγορά και είναι ένα φυσιολογικό ποσοστό μεταβολής.

2. Τα δεδομένα train και test είναι **παρόμοια**, γεγονός που δείχνει ότι τα train data αντιπροσωπεύουν σωστά το συνολικό πληθυσμό.
3. Τα στατιστικά για τις χρεώσεις όπως το **Total_day_charge** ή το **Total_night_charge** είναι σταθερά, γεγονός που δείχνει καλή ποιότητα δεδομένων.

Pearson Correlation

Στη στατιστική, ο συντελεστής συσχέτισης Pearson είναι ένας συντελεστής συσχέτισης που μετρά τη γραμμική συσχέτιση μεταξύ δύο συνόλων δεδομένων. Είναι ο λόγος μεταξύ της συνδιακύμανσης δύο μεταβλητών και του γινόμενου των τυπικών αποκλίσεων τους.



Εικόνα 1: Ραβδόγραμμα Συσχέτισης Χαρακτηριστικών με την μεταβλητή Churn

Τι δείχνει το Ραβδόγραμμα:

Το ραβδόγραμμα παρουσιάζει τον συντελεστή συσχέτισης Pearson για κάθε χαρακτηριστικό (feature) σε σχέση με τη μεταβλητή στόχο **Churn**. Ο συντελεστής Pearson κυμαίνεται από -1 έως 1:

- **Θετική συσχέτιση (0 έως 1):** Το χαρακτηριστικό αυξάνεται όταν αυξάνεται το churn.
- **Αρνητική συσχέτιση (-1 έως 0):** Το χαρακτηριστικό μειώνεται όταν αυξάνεται το churn.
- **Μηδενική συσχέτιση (0):** Καμία γραμμική συσχέτιση.

Ανάλυση Σημαντικών Χαρακτηριστικών

1. **Total_day_minutes και Total_day_charge**
    - **Συσχέτιση:** ~0.2 (ισχυρή θετική συσχέτιση).
    - **Ερμηνεία:** Πελάτες που καταναλώνουν – χρησιμοποιούν τον χρόνο ομιλίας τους και καθημερινά μιλάνε πολλά περισσότερα λεπτά από τον χρόνο που τους προσφέρει το πρόγραμμα τους κατά συνέπεια να αυξάνουν την καθημερινή τους χρέωση κατά την διάρκεια της ημέρας είναι πιο πιθανό να αποχωρήσουν. Αυτό μπορεί να υποδηλώνει δυσαρέσκεια λόγω υψηλών χρεώσεων.
2. **Voice_mail_plan**
    - **Συσχέτιση:** Αρνητική (~-0.1).
    - **Ερμηνεία:** Οι πελάτες με voicemail plan είναι λιγότερο πιθανό να αποχωρήσουν. Αυτό υποδηλώνει ότι τέτοια πρόσθετα πλάνα αυξάνουν τη δέσμευση.
3. **Total_intl_minutes, Total_intl_calls, και Total_intl_charge**
    - **Συσχέτιση:** Θετική (περίπου 0.1).
    - **Ερμηνεία:** Πελάτες με υψηλή χρήση διεθνών κλήσεων τείνουν να αποχωρούν, πιθανώς λόγω κόστους ή έλλειψης ανταγωνιστικών προσφορών.
4. **Number_vmail_messages**
    - **Συσχέτιση:** Αρνητική (~-0.1).
    - **Ερμηνεία:** Πελάτες που χρησιμοποιούν περισσότερο το voicemail για την αποστολή μηνυμάτων είναι λιγότερο πιθανό να αποχωρήσουν, υποδεικνύοντας μεγαλύτερη ικανοποίηση ή δέσμευση.

Χαρακτηριστικά με Αδύναμη Συσχέτιση

1. **Total_night_minutes και Total_night_charge**
    - **Συσχέτιση:** Ελαφρώς αρνητική (~-0.05).
    - **Ερμηνεία:** Η χρήση τη νύχτα έχει μικρή επίδραση στο churn, πιθανώς λόγω χαμηλών χρεώσεων που ικανοποιούν τους πελάτες.
2. **Total_eve_calls και Total_eve_minutes**
    - **Συσχέτιση:** Ελαφρώς θετική (~0.05).
    - **Ερμηνεία:** Η χρήση των υπηρεσιών το βράδυ έχει μικρή επίδραση στο churn.

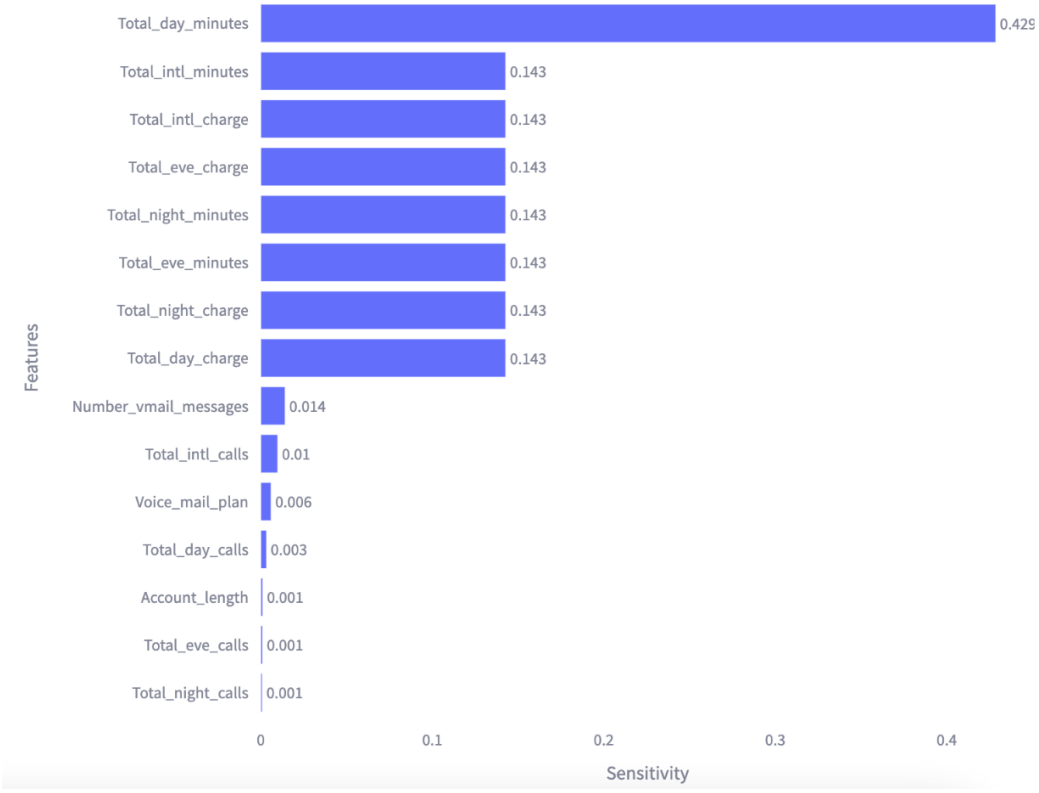
3. **Account_length**

    - **Συσχέτιση:** Πολύ χαμηλή (~0.02).
    - **Ερμηνεία:** Η διάρκεια του λογαριασμού φαίνεται να μην επηρεάζει σημαντικά την πιθανότητα αποχώρησης.

Συμπεράσματα

- **Σημαντικά χαρακτηριστικά:** Τα χαρακτηριστικά **Total_day_minutes**, **Total_day_charge**, **Voice_mail_plan**, και *Total_intl_ (minutes, calls, charge)** είναι τα πιο σημαντικά για την πρόβλεψη του churn.
- **Μη σημαντικά χαρακτηριστικά:** Χαρακτηριστικά όπως **Total_night_minutes** και **Account_length** έχουν πολύ μικρή επίδραση στη μεταβλητή στόχο.

Αναλύοντας βαθύτερα και αναλυτικότερα τα χαρακτηριστικά του Set των δεδομένων μας ο παρακάτω πίνακας μας δείχνει την συσχέτιση του churn με τα χαρακτηριστικά των δεδομένων που χρησιμοποιήσαμε.



Εικόνα 2 : Ραβδόγραμμα ευαισθησίας χαρακτηριστικών (Feature Sensitivity)

Τα χαρακτηριστικά με τη μεγαλύτερη ευαισθησία (sensitivity) υποδεικνύουν τη μεγαλύτερη επίδραση στην πρόβλεψη της εξαρτημένης μεταβλητής (**Churn**).

- **Χαρακτηριστικό Total_day_minutes** (0.429)

Είναι το χαρακτηριστικό με τη μεγαλύτερη ευαισθησία- σημαντικότητα. Οι πελάτες που χρησιμοποιούν περισσότερο το κινητό τους, μιλάνε περισσότερο κατά τη διάρκεια της ημέρας είναι σημαντικά πιο πιθανό να αποχωρήσουν. Αυτό συνδέεται με υψηλά κόστη που δεν έχουν ϋπολογίσει οι πελάτες.

- **Ομάδα Χαρακτηριστικών με Ευαισθησία 0.143**

Total_intl_minutes, Total_intl_charge, Total_eve_charge, Total_night_minutes, Total_eve_minutes, Total_night_charge, Total_day_charge.

Όλα αυτά τα χαρακτηριστικά σχετίζονται με τη χρήση και τις χρεώσεις υπηρεσιών σε διαφορετικές χρονικές περιόδους ή τύπους επικοινωνίας. Η χρήση υπηρεσιών και οι αντίστοιχες χρεώσεις αποτελούν βασικούς παράγοντες που επηρεάζουν το churn.

- **Λιγότερο Σημαντικά Χαρακτηριστικά**

Τα χαρακτηριστικά με χαμηλή ευαισθησία επηρεάζουν λιγότερο την πρόβλεψη του churn.

Number_vmail_messages (0.014)

Ο αριθμός των μηνυμάτων στο voicemail δεν έχει σημαντική επίδραση, πιθανώς λόγω της μικρής χρήσης αυτής της υπηρεσίας.

1. **Total_intl_calls (0.01)**

Ο αριθμός των διεθνών κλήσεων επηρεάζει λιγότερο την πιθανότητα churn σε σύγκριση με τη διάρκεια ή το κόστος των διεθνών κλήσεων.

2. **Voice_mail_plan (0.006)**

Η ύπαρξη voicemail plan έχει μικρή επίδραση στη δέσμευση των πελατών.

3. **Account_length, Total_day_calls, Total_eve_calls, Total_night_calls (0.001)**

Αυτά τα χαρακτηριστικά έχουν πολύ χαμηλή ευαισθησία, δείχνοντας ότι η συχνότητα των κλήσεων ή η διάρκεια του λογαριασμού δεν είναι κρίσιμα για την πρόβλεψη του churn.

Συμπεράσματα

- **Ισχυρή Συσχέτιση:**

Η διάρκεια και το κόστος των υπηρεσιών, ειδικά κατά τη διάρκεια της ημέρας (**Total_day_minutes, Total_day_charge**), επηρεάζουν σημαντικά το churn.

Οι διεθνείς κλήσεις (**Total_intl_minutes, Total_intl_charge**) έχουν επίσης αξιοσημείωτη επίδραση.

- **Αδύναμη Συσχέτιση:**

Τα χαρακτηριστικά όπως το **Account_length** και ο αριθμός κλήσεων (day, eve, night) έχουν ελάχιστη επιρροή. Αυτή η συσχέτιση με το churn μας δίνει μια σημαντική πληροφορία ότι ανεξάρτητα το χρονικό διάστημα που ο πελάτης βρίσκεται στο δίκτυο τηλεπικοινωνιών συγκεκριμένα της Orange στο παράδειγμα μας δεν διστάζει να αποχωρήσει και να αλλάξει πάροχο από την στιγμή που θα έχει καλύτερη πρόταση ή οικονομικότερη.

Επειδή το πρόβλημα του churn είναι ένα πρόβλημα Classification από τα αποτελέσματα του αλγορίθμου θα χρησιμοποιήσουμε τις παρακάτω τιμές- μετρήσεις.

1. Acc (Accuracy - Ακρίβεια)

Ο ορισμός του FP είναι το ποσοστό των σωστών προβλέψεων (τόσο θετικών όσο και αρνητικών) σε σχέση με το σύνολο των προβλέψεων.

Αν το μοντέλο έχει υψηλή ακρίβεια, σημαίνει ότι προβλέπει σωστά τις περισσότερες περιπτώσεις.

2. FP (False Positives - Ψευδώς Θετικά)

Ο ορισμός του FP είναι Ο αριθμός των περιπτώσεων όπου το μοντέλο προέβλεψε **"churn"** ενώ στην πραγματικότητα δεν υπήρχε churn.

Ένα υψηλό FP σημαίνει ότι το μοντέλο προκαλεί περιττές παρεμβάσεις, π.χ., δίνοντας προσφορές σε πελάτες που δεν πρόκειται να αποχωρήσουν.

Σε ένα πρόβλημα όπως το churn, στόχος είναι να διατηρηθεί το FP χαμηλό.

3. TP (True Positives - Αληθώς Θετικά)

Ο ορισμός του TP είναι Ο αριθμός των περιπτώσεων όπου το μοντέλο προέβλεψε σωστά **"churn"** όταν υπήρχε πραγματικά churn.

Ένα υψηλό TP δείχνει ότι το μοντέλο εντοπίζει σωστά τους πελάτες που είναι πιθανό να αποχωρήσουν. Είναι μια κρίσιμη μετρική για προβλήματα churn, καθώς αυτοί οι πελάτες είναι που πρέπει να εντοπιστούν για να εφαρμοστούν στρατηγικές διατήρησης.

4. FN (False Negatives - Ψευδώς Αρνητικά)

Ο ορισμός του FN είναι Ο αριθμός των περιπτώσεων όπου το μοντέλο προέβλεψε **"όχι churn"**, αλλά στην πραγματικότητα υπήρχε churn.

Ένα υψηλό FN σημαίνει ότι το μοντέλο χάνει σημαντικούς πελάτες που αποχωρούν, γεγονός που μπορεί να είναι πολύ κοστοβόρο. Ο στόχος είναι να διατηρηθεί το FN όσο το δυνατόν χαμηλότερο.

5. TN (True Negatives - Αληθώς Αρνητικά)

Ο ορισμός του TN είναι ο αριθμός των περιπτώσεων όπου το μοντέλο προέβλεψε σωστά **"όχι churn"**, όταν στην πραγματικότητα δεν υπήρχε churn. Ένα υψηλό TN δείχνει ότι το μοντέλο είναι αποτελεσματικό στον εντοπισμό πελατών που δεν θα αποχωρήσουν, είναι σημαντικό για να αποφεύγονται περιττές παρεμβάσεις.

6. RunTimeSec (Χρόνος Εκτέλεσης)

Ο ορισμός του RunTimeSec είναι Ο χρόνος (σε δευτερόλεπτα) που χρειάστηκε το μοντέλο για να ολοκληρώσει τις προβλέψεις του ή την εκπαίδευσή του.

Αντιπροσωπεύει την ταχύτητα του μοντέλου, είναι σημαντικό για συστήματα που λειτουργούν σε πραγματικό χρόνο, όπου απαιτούνται γρήγορες προβλέψεις.

5.1.2 Αλγόριθμος XGBOOST

Παρακάτω τα αποτελέσματα του αλγορίθμου XGBoost που θα χρειαστεί να αναλύσουμε για να κατανοήσουμε περισσότερο πως λειτουργεί ο αλγόριθμος πάνω στα δεδομένα που του δώσαμε.

XGBOOST							
A/A	Data	Acc	FP	TP	FN	TN	RunTimeSec
0	Train	0,93996998	0	0,08254127	0,06003002	0,85742871	26,2186649
1	Test	0,89189189	0,0015015	0,04804805	0,10660661	0,84384384	0,00205741

Πίνακας 4: Αποτελέσματα από αλγορίθμου XGBoost

Αποτελέσματα Train Data

1. **Acc (Accuracy): 93,99%**
    - Υψηλή ακρίβεια στην εκπαίδευση, γεγονός που δείχνει ότι το μοντέλο αποδίδει καλά στα δεδομένα εκπαίδευσης.
2. **FP (False Positives): 0**
    - Το μοντέλο δεν έκανε ψευδώς θετικές προβλέψεις (δηλαδή δεν προέβλεψε churn όταν δεν υπήρχε).

3. **TP (True Positives): 0,0825 (~8,25%)**
    - Το ποσοστό των πελατών που το μοντέλο εντόπισε σωστά ότι θα αποχωρήσουν.
4. **FN (False Negatives): 0,0600 (~6%)**
    - Το ποσοστό των πελατών που αποχώρησαν αλλά το μοντέλο δεν τους εντόπισε.
5. **TN (True Negatives): 0,8574 (~85,74%)**
    - Το ποσοστό των πελατών που σωστά προβλέφθηκαν ότι δεν θα αποχωρήσουν.
6. **RunTimeSec: 26,22 δευτερόλεπτα**
    - Ο χρόνος που χρειάστηκε το μοντέλο για την εκπαίδευση ή τις προβλέψεις.

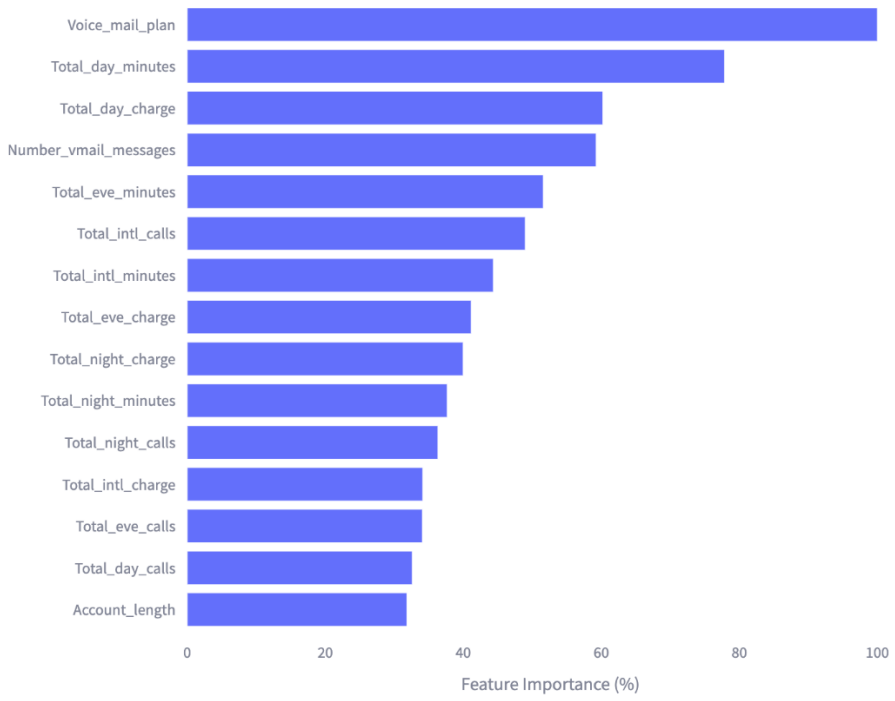
Αποτελέσματα Test Data

1. **Acc (Accuracy): 89,19%**
    - Σημαντικά χαμηλότερη ακρίβεια σε σχέση με το training, υποδεικνύοντας πιθανή υπερπροσαρμογή (overfitting) στο train set.
2. **FP (False Positives): 0,0015 (~0,15%)**
    - Πολύ χαμηλό ποσοστό ψευδώς θετικών προβλέψεων, γεγονός που δείχνει καλό φιλτράρισμα των πελατών που δεν αποχωρούν.
3. **TP (True Positives): 0,0480 (~4,8%)**
    - Σημαντική μείωση των σωστών προβλέψεων churn σε σχέση με το train set.
4. **FN (False Negatives): 0,1066 (~10,66%)**
    - Υψηλό ποσοστό πελατών που αποχωρούν αλλά δεν εντοπίστηκαν από το μοντέλο, το οποίο αποτελεί πρόβλημα για το επιχειρηματικό σενάριο.
5. **TN (True Negatives): 0,8438 (~84,38%)**
    - Πολύ καλό ποσοστό σωστών προβλέψεων ότι οι πελάτες δεν θα αποχωρήσουν.
6. **RunTimeSec: 0,0020 δευτερόλεπτα**
    - Πολύ γρήγορη απόδοση στις προβλέψεις για το test set.

Συγκρίσεις και Συμπεράσματα

1. **Ακρίβεια (Accuracy):**
    - Η ακρίβεια μειώθηκε από 93,99% στο train set σε 89,19% στο test set. Αυτό μπορεί να υποδεικνύει **υπερπροσαρμογή**.
2. **True Positives (TP):**
    - Η απόδοση μειώθηκε σημαντικά στο test set (8,25% → 4,8%), γεγονός που σημαίνει ότι το μοντέλο δυσκολεύεται να εντοπίσει πελάτες που αποχωρούν σε μη εκπαιδευμένα δεδομένα.
    -
3. **False Negatives (FN):**
    - Τα FN αυξήθηκαν στο test set (6% → 10,66%), που αποτελεί κρίσιμο πρόβλημα, καθώς οι πελάτες που αποχωρούν είναι αυτοί που θέλουμε να ανιχνεύσουμε.
4. **False Positives (FP):**
    - Παραμένουν χαμηλά και στα δύο sets, γεγονός που δείχνει ότι το μοντέλο δεν "σημαίνει συναγερμό" για πελάτες που δεν πρόκειται να αποχωρήσουν.

Το παρακάτω διάγραμμα παρουσιάζει τη σημασία (feature importance) των χαρακτηριστικών όπως προέκυψε από την εκπαίδευση του μοντέλου **XGBoost**.



Εικόνα 3: Ραβδόγραμμα σημαντικότητας χαρακτηριστικών (Feature Importance) XGBoost

Υψηλής Σημασίας Χαρακτηριστικά

1. **Voice_mail_plan**
    - Το πιο σημαντικό χαρακτηριστικό στο μοντέλο η ύπαρξη ή μη voicemail plan φαίνεται να παίζει κρίσιμο ρόλο στην πρόβλεψη του churn, πιθανώς επειδή επηρεάζει την ικανοποίηση ή τη δέσμευση του πελάτη με την υπηρεσία.
2. **Total_day_minutes**
    - Το δεύτερο πιο σημαντικό χαρακτηριστικό, η υψηλή χρήση κατά τη διάρκεια της ημέρας μπορεί να σχετίζεται με δυσαρέσκεια λόγω κόστους ή υπερφόρτωσης χρήσης.
3. **Total_day_charge**
    - Το συνολικό κόστος για ημερήσιες κλήσεις είναι επίσης κρίσιμο, πελάτες που επιβαρύνονται με υψηλά κόστη είναι πιθανότερο να αποχωρήσουν.
4. **Number_vmail_messages**
    - Η συχνότητα χρήσης voicemail φαίνεται να επηρεάζει σημαντικά την πιθανότητα churn , οι πελάτες που χρησιμοποιούν περισσότερο voicemail μπορεί να είναι πιο δεσμευμένοι.

Μεσαίας Σημασίας Χαρακτηριστικά

5. **Total_eve_minutes**
    - Η χρήση το βράδυ παίζει μέτριο ρόλο στην πρόβλεψη, πιθανώς λόγω χαμηλότερων χρεώσεων που ικανοποιούν τους πελάτες.
6. **Total_intl_calls και Total_intl_minutes**
    - Η χρήση διεθνών κλήσεων έχει επίδραση, πιθανώς επειδή οι διεθνείς χρήστες είναι πιο απαιτητικοί ή επηρεάζονται από το κόστος.
7. **Total_eve_charge**
    - Το κόστος των βραδινών κλήσεων είναι σημαντικό, αλλά λιγότερο σε σχέση με άλλες χρονικές ζώνες.

Χαμηλής Σημασίας Χαρακτηριστικά

8. **Total_night_minutes και Total_night_charge**
    - Η χρήση και το κόστος κατά τη διάρκεια της νύχτας έχουν μικρότερη επίδραση, πιθανώς λόγω χαμηλότερων χρεώσεων που δεν οδηγούν εύκολα σε δυσαρέσκεια.
9. **Account_length**
    - Η διάρκεια του λογαριασμού δεν φαίνεται να παίζει σημαντικό ρόλο, γεγονός που υποδεικνύει ότι η πιθανότητα churn εξαρτάται περισσότερο από την τρέχουσα χρήση παρά από τη διάρκεια σχέσης με την εταιρεία.

Συμπεράσματα

- **Υψηλή Σημασία:** Τα χαρακτηριστικά **Voice_mail_plan**, **Total_day_minutes**, και **Total_day_charge** είναι τα πιο κρίσιμα για την πρόβλεψη του churn. Αυτά πρέπει να ληφθούν υπόψη για τη βελτίωση της απόδοσης του μοντέλου και των στρατηγικών διατήρησης.
- **Χαμηλή Σημασία:** Χαρακτηριστικά όπως το **Account_length** και οι κλήσεις κατά τη νύχτα έχουν χαμηλή επίδραση και μπορούν να απορριφθούν σε περιπτώσεις όπου απαιτείται απλοποίηση του μοντέλου.

Συμπεράσματα και Προτάσεις

Υψηλή Σημασία:

- Θα πρέπει να εστιάσει στην εξατομίκευση των υπηρεσιών **voicemail** και τη βελτίωση της εμπειρίας για χρήστες με **υψηλή ημερήσια χρήση** (Total_day_charge, Total_day_minutes).
- Να δημιουργήσει και να προσφέρει ειδικά πακέτα με περισσότερα λεπτά και χαμηλότερες χρεώσεις για να μειώσεις τη δυσαρέσκεια.

Διατήρηση Πελατών:

- Να εντοπίσει πελάτες με υψηλό κόστος ή έντονη χρήση και προτείνει προσωποποιημένες προσφορές για μείωση του churn.
- Να δημιουργήσει προγράμματα επιβράβευσης (loyalty programs) για πελάτες με συχνή χρήση voicemail ή διεθνείς κλήσεις, ώστε να ενισχύσεις τη δέσμευσή τους.

5.1.3 Αλγόριθμος Random Forest

Παρακάτω τα αποτελέσματα του αλγορίθμου **Random Forest** που θα χρειαστεί να αναλύσουμε για να κατανοήσουμε περισσότερο πως λειτουργεί ο αλγόριθμος πάνω στα δεδομένα που του δώσαμε.

RANDOM FOREST							
	Data	Acc	FP	TP	FN	TN	RunTimeSec
0	Train	0,90245123	0,002001	0,04702351	0,09554777	0,85542771	61,4003823
1	Test	0,88738739	0	0,04204204	0,11261261	0,84534535	0,00461769

Πίνακας 5: Αποτελέσματα από αλγορίθμου Random Forest

Αποτελέσματα Train Data

1. **Acc (Accuracy): 90,25%**
    - Υψηλή ακρίβεια στα δεδομένα εκπαίδευσης, γεγονός που δείχνει ότι το μοντέλο εκπαιδεύτηκε καλά.
2. **FP (False Positives): 0,002 (~0,2%)**
    - Πολύ χαμηλό ποσοστό ψευδώς θετικών προβλέψεων. Το μοντέλο εντόπισε σωστά σχεδόν όλους τους πελάτες που δεν αποχωρούν.
3. **TP (True Positives): 0,047 (~4,7%)**
    - Το ποσοστό των πελατών που το μοντέλο εντόπισε σωστά ως "churn" είναι χαμηλό, κάτι που μπορεί να υποδεικνύει πρόβλημα στην ανάκληση (recall).
4. **FN (False Negatives): 0,095 (~9,5%)**
    - Το ποσοστό των πελατών που αποχωρούν αλλά το μοντέλο δεν τους εντόπισε είναι σχετικά υψηλό, κάτι που δείχνει ότι το μοντέλο χάνει πελάτες που αποχωρούν.
5. **TN (True Negatives): 0,855 (~85,5%)**
    - Πολύ καλό ποσοστό σωστών προβλέψεων για πελάτες που δεν αποχωρούν.
6. **RunTimeSec: 61,40 δευτερόλεπτα**
    - Ο χρόνος εκπαίδευσης του μοντέλου, που είναι σχετικά υψηλός λόγω της φύσης του Random Forest.
    -

Αποτελέσματα Test Data

1. **Acc (Accuracy): 88,74%**
    - Η ακρίβεια μειώθηκε ελαφρώς σε σχέση με το Train set, αλλά παραμένει υψηλή.

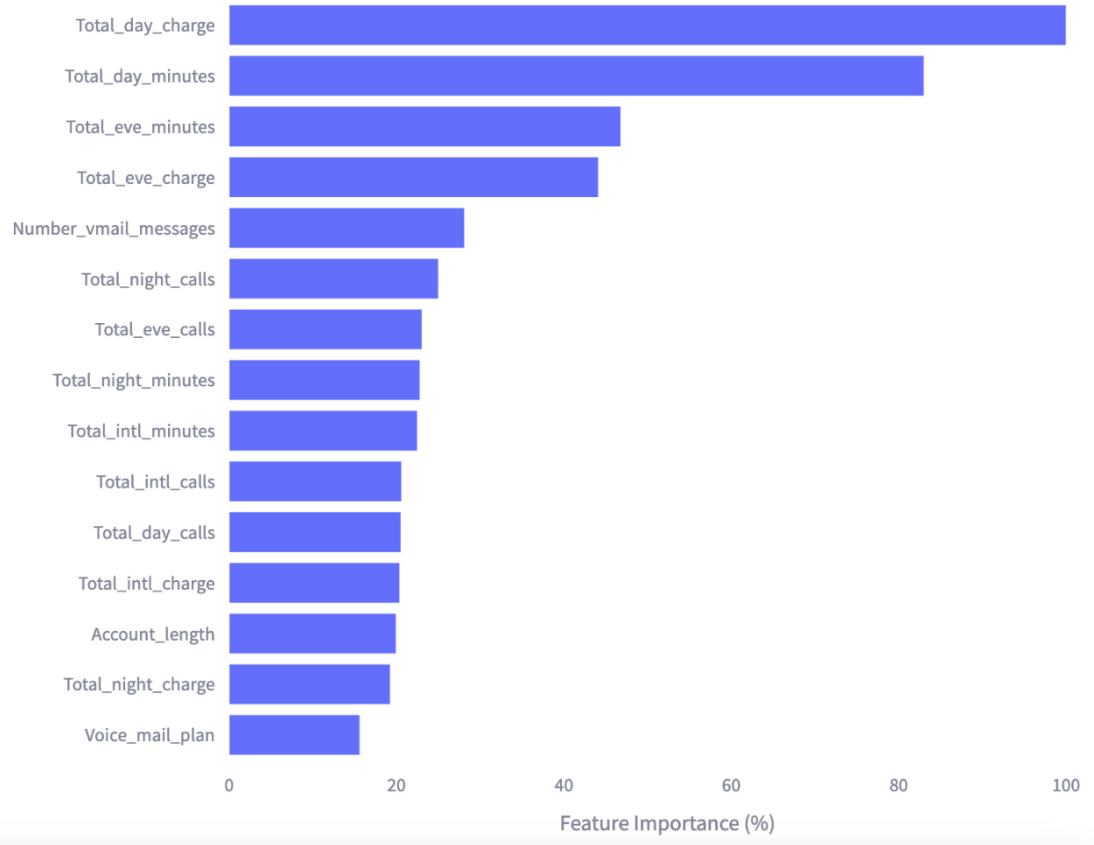
2. **FP (False Positives): 0**
    - Δεν υπήρχαν ψευδώς θετικές προβλέψεις στο test set, γεγονός που δείχνει εξαιρετική απόδοση στην αποφυγή ακατάλληλων προβλέψεων.
3. **TP (True Positives): 0,042 (~4,2%)**
    - Μικρό ποσοστό εντοπισμένων πελατών που αποχωρούν, κάτι που υποδεικνύει ότι το μοντέλο εξακολουθεί να δυσκολεύεται να ανιχνεύσει πελάτες που θα αποχωρήσουν.
4. **FN (False Negatives): 0,112 (~11,2%)**
    - Υψηλό ποσοστό πελατών που αποχωρούν αλλά δεν εντοπίστηκαν από το μοντέλο.
5. **TN (True Negatives): 0,845 (~84,5%)**
    - Το μοντέλο έχει καλή απόδοση στην αναγνώριση πελατών που δεν αποχωρούν.
6. **RunTimeSec: 0,0046 δευτερόλεπτα**
    - Πολύ γρήγορος χρόνος πρόβλεψης στο test set, γεγονός που δείχνει την αποδοτικότητα του Random Forest για προβλέψεις.

Συγκρίσεις και Συμπεράσματα

1. **Ακρίβεια (Accuracy):**
    - Η ακρίβεια είναι υψηλή και στα δύο σύνολα δεδομένων (90,25% για train και 88,74% για test), υποδεικνύοντας ότι το μοντέλο έχει καλή γενίκευση.
2. **False Negatives (FN):**
    - Το ποσοστό FN είναι αρκετά υψηλό και στα δύο σύνολα δεδομένων (9,5% στο train και 11,2% στο test). Αυτό σημαίνει ότι το μοντέλο αποτυγχάνει να ανιχνεύσει πολλούς πελάτες που αποχωρούν, κάτι που αποτελεί κρίσιμο ζήτημα για την πρόβλεψη churn.
3. **True Positives (TP):**
    - Τα TP παραμένουν χαμηλά (~4,7% στο train και ~4,2% στο test), υποδεικνύοντας ότι το μοντέλο χρειάζεται βελτίωση στην ανίχνευση πελατών που αποχωρούν.
4. **False Positives (FP):**

- Τα FP είναι σχεδόν μηδενικά και στα δύο datasets, κάτι που είναι καλό γιατί το μοντέλο αποφεύγει περιττές ενέργειες για πελάτες που δεν πρόκειται να αποχωρήσουν.

Το διάγραμμα παρουσιάζει τη σημασία των χαρακτηριστικών όπως προέκυψε από την εκπαίδευση του μοντέλου **Random Forest**.



Εικόνα 4 : Ραβδόγραμμα σημαντικότητας χαρακτηριστικών (Feature Importance) Random Forest

Υψηλής Σημασίας Χαρακτηριστικά

1. **Total_day_charge**
    - Το πιο σημαντικό χαρακτηριστικό, οι ημερήσιες χρεώσεις φαίνεται να έχουν τη μεγαλύτερη επίδραση στο churn. Οι πελάτες που πληρώνουν υψηλά ποσά είναι πιθανότερο να αποχωρήσουν.
2. **Total_day_minutes**
    - Το δεύτερο πιο σημαντικό χαρακτηριστικό, η διάρκεια κλήσεων κατά τη διάρκεια της ημέρας παίζει σημαντικό ρόλο.

Μεσαίας Σημασίας Χαρακτηριστικά

3. **Total_eve_minutes και Total_eve_charge**
    - Οι βραδινές χρεώσεις και διάρκεια είναι σημαντικοί παράγοντες, αλλά λιγότερο από τις ημερήσιες.
4. **Number_vmail_messages**
    - Οι Πελάτες που χρησιμοποιούν περισσότερο voicemail ενδέχεται να είναι πιο δεσμευμένοι.
5. **Total_night_calls και Total_night_minutes**
    - Η χρήση τη νύχτα φαίνεται να έχει μικρότερη επίδραση, αλλά παραμένει σημαντική.

Χαμηλής Σημασίας Χαρακτηριστικά

6. **Account_length**

    Η διάρκεια του λογαριασμού δεν επηρεάζει πολύ την πρόβλεψη churn.
7. **Voice_mail_plan**
    - Η ύπαρξη voicemail plan έχει χαμηλότερη σημασία από άλλα χαρακτηριστικά.

Συμπεράσματα και Προτάσεις

- **Υψηλή Σημασία:**
    - Θα χρειαστεί η Orange και κατ' επέκταση το τμήμα marketing να εστιάσει στη μείωση του κόστους (Total_day_charge, Total_eve_charge) και στην αύξηση της ικανοποίησης για πελάτες με υψηλή ημερήσια και βραδινή χρήση.
    - Να δημιουργήσει δυναμικά προγράμματα που ανταποκρίνονται στις ανάγκες χρήσης.
- **Διατήρηση Πελατών:**
    - Το τμήμα μάρκετινγκ θα χρειαστεί να χρησιμοποιήσει τα insights για να εντοπίσει πελάτες που πληρώνουν υψηλά ποσά ή έχουν έντονη χρήση και να προτείνει ειδικές προσφορές.
    - Να αναπτύξει loyalty προγράμματα για πελάτες που χρησιμοποιούν voicemail και άλλες υπηρεσίες.

5.1.4 Αλγόριθμος Decision Tree

Decision Tree							
	Data	Acc	FP	TP	FN	TN	RunTimeSec
0	Train	0,867433717	0,025012506	0,035017509	0,107553777	0,832416208	0,953969955
1	Test	0,86036036	0,021021021	0,036036036	0,118618619	0,824324324	0,000284457

Πίνακας 6: Αποτελέσματα αλγορίθμου Decision Tree

Αποτελέσματα Train Data

1. Accuracy (Acc): 86,74%

- Το μοντέλο προβλέπει σωστά το **86,74%** των περιπτώσεων στο σύνολο εκπαίδευσης. Αυτό δείχνει ικανοποιητική ακρίβεια, αν και υπάρχει περιθώριο βελτίωσης.

2. False Positives (FP): 2,5%

- Το μοντέλο αναγνωρίζει λανθασμένα το **2,5%** των πελατών που δεν αποχωρούν ως πελάτες που θα αποχωρήσουν.
- Αυτό σημαίνει ότι υπάρχει κάποιο κόστος από περιττές παρεμβάσεις προς αυτούς τους πελάτες.

3. True Positives (TP): 3,5%

- Το μοντέλο εντοπίζει σωστά το **3,5%** των πελατών που αποχωρούν.
- Αν και προβλέπει σωστά αρκετές περιπτώσεις, το ποσοστό TP είναι χαμηλό.

4. False Negatives (FN): 10,75%

- Το **10,75%** των πελατών που αποχωρούν δεν αναγνωρίζονται από το μοντέλο.
- Αυτό αποτελεί σημαντικό πρόβλημα, καθώς οι πελάτες αυτοί μπορεί να αποχωρήσουν χωρίς να αντιμετωπιστεί το ζήτημά τους.

5. True Negatives (TN): 83,24%

- Το μοντέλο αναγνωρίζει σωστά το **83,24%** των πελατών που δεν αποχωρούν.

6. RunTimeSec: 0,953 δευτερόλεπτα

- Ο χρόνος εκπαίδευσης του μοντέλου είναι κάτω από ένα δευτερόλεπτο, γεγονός που δείχνει ότι το Decision Tree είναι αποδοτικό κατά την εκπαίδευση.

Αποτελέσμτα Test Data

1. Accuracy (Acc): 86,03%

- Στο σύνολο ελέγχου, το μοντέλο προβλέπει σωστά το **86,03%** των περιπτώσεων. Η ακρίβεια είναι παρόμοια με το Train, γεγονός που υποδεικνύει ότι το μοντέλο έχει καλή γενίκευση.

2. False Positives (FP): 2,1%

- Μόλις το **2,1%** των πελατών που δεν αποχωρούν αναγνωρίζονται λανθασμένα ως πελάτες που θα αποχωρήσουν.
- Αυτό δείχνει ότι το μοντέλο διατηρεί χαμηλά τα FP και στο test set.

3. True Positives (TP): 3,6%

- Το μοντέλο εντοπίζει σωστά το **3,6%** των πελατών που αποχωρούν στο test set.
- Το TP παραμένει χαμηλό, κάτι που δείχνει περιορισμένη ικανότητα ανίχνευσης churn.

4. False Negatives (FN): 11,86%

- Το **11,86%** των πελατών που αποχωρούν δεν αναγνωρίζονται από το μοντέλο.
- Το FN είναι ελαφρώς αυξημένο σε σχέση με το Train, κάτι που μπορεί να επηρεάσει την απόδοση του μοντέλου στις προβλέψεις churn.

5. True Negatives (TN): 82,43%

- Το μοντέλο αναγνωρίζει σωστά το **82,43%** των πελατών που δεν αποχωρούν.

6. RunTimeSec: 0,00028 δευτερόλεπτα

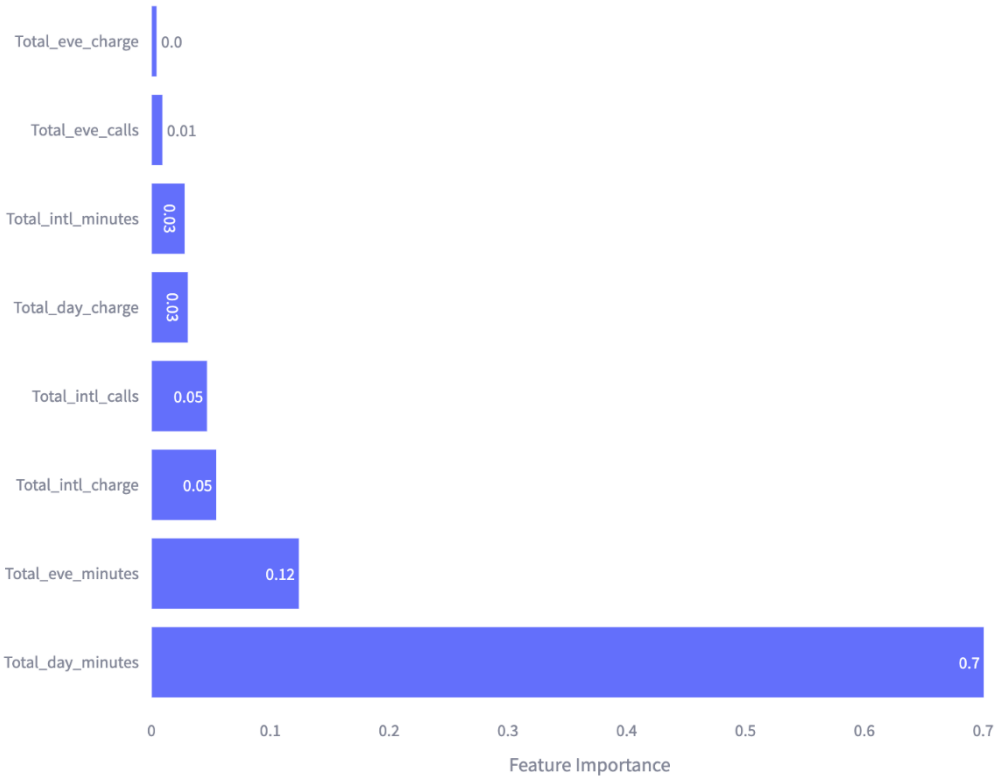
- Ο χρόνος εκτέλεσης για το test set είναι εξαιρετικά χαμηλός, καθιστώντας το Decision Tree γρήγορο για προβλέψεις.

Συγκριτικά Συμπεράσματα (Train vs Test)

1. **Σταθερότητα:**
    - Η απόδοση του μοντέλου είναι παρόμοια και στα δύο σύνολα δεδομένων, γεγονός που υποδεικνύει καλή γενίκευση χωρίς σημαντικό overfitting.
2. **False Negatives:**
    - Το υψηλό FN (~10-12%) παραμένει το κύριο πρόβλημα και στα δύο datasets, καθώς το μοντέλο χάνει αρκετούς πελάτες που αποχωρούν.

3. **True Positives:**
    - Το TP είναι χαμηλό (~3,5-3,6%), κάτι που υποδεικνύει την ανάγκη βελτίωσης στην ικανότητα του μοντέλου να προβλέπει το churn.
4. **False Positives:**
    - Τα FP είναι χαμηλά (~2-2,5%), κάτι που είναι θετικό, καθώς μειώνονται οι περιττές παρεμβάσεις.

Το παρακάτω διάγραμμα παρουσιάζει τη σημασία των χαρακτηριστικών όπως προέκυψε από την εκπαίδευση του **Decision Tree** μοντέλου



Εικόνα 5 : Ραβδόγραμμα σημαντικότητας χαρακτηριστικών (Feature Importance) Decision Tree

Υψηλής Σημασίας Χαρακτηριστικά

1. **Total_day_minutes (0.7):**
    - Το πιο σημαντικό χαρακτηριστικό , η διάρκεια ημερήσιων κλήσεων φαίνεται να έχει τη μεγαλύτερη επίδραση στο churn. Πελάτες που χρησιμοποιούν πολλές ημερήσιες κλήσεις ενδέχεται να είναι πιο δυσαρεστημένοι λόγω κόστους ή περιορισμών.

2. **Total_eve_minutes (0.12):**
    - Το δεύτερο πιο σημαντικό χαρακτηριστικό, οι βραδινές συνομιλίες επηρεάζουν μέτρια την πιθανότητα churn, πιθανότατα επειδή οι χρεώσεις είναι χαμηλότερες από τις ημερήσιες.

Μεσαίας Σημασίας Χαρακτηριστικά

3. **Total_intl_calls και Total_intl_charge (0.05):**
    - Τα χαρακτηριστικά που σχετίζονται με τις διεθνείς κλήσεις και τις χρεώσεις τους έχουν μέτρια σημασία, οι πελάτες που πραγματοποιούν διεθνείς κλήσεις μπορεί να είναι πιο ευαίσθητοι στις χρεώσεις και να επηρεάζονται από αυτές.
4. **Total_day_charge (0.03):**
    - Η ημερήσια χρέωση, αν και έχει μικρότερη επίδραση από τα **Total_day_minutes**, παραμένει σημαντική, οι πελάτες μπορεί να επηρεάζονται λιγότερο από τη συνολική χρέωση και περισσότερο από την εμπειρία χρήσης.

Χαμηλής Σημασίας Χαρακτηριστικά

5. **Total_eve_calls (0.01):**
    - Πολύ χαμηλή σημασία, που δείχνει ότι ο αριθμός των βραδινών κλήσεων δεν επηρεάζει σημαντικά το churn.
6. **Total_eve_charge (0.00):**
    - Καμία ουσιαστική επίδραση στο churn, οι βραδινές χρεώσεις είναι πιθανώς αμελητέες λόγω των χαμηλότερων τιμολογίων.

Συμπεράσματα και Προτάσεις

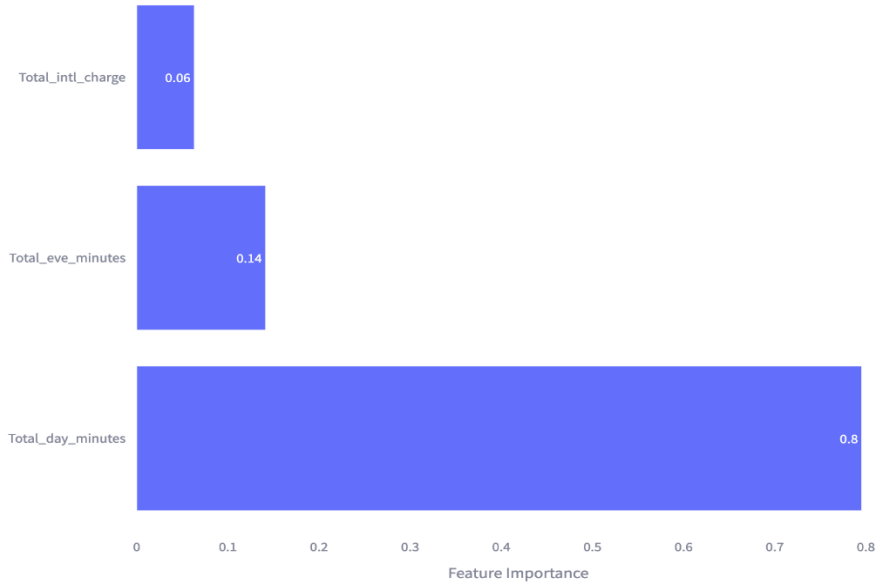
1. **Total_day_minutes:**
    - Το μάρκετινγκ πρέπει να εστιάσει στη μείωση του κόστους για πελάτες με υψηλή ημερήσια χρήση.
    - Να προωθήσει δυναμικά πακέτα που προσφέρουν περισσότερα λεπτά ή χαμηλότερες ημερήσιες χρεώσεις.
    -

2. **Total_eve_minutes:**
    - Να αναπτύξει ειδικές προσφορές για βραδινές κλήσεις, εστιάζοντας σε πελάτες με έντονη χρήση.

Διατήρηση Πελατών:

- Να χρησιμοποιήσει insights από τα χαρακτηριστικά για να εντοπίσεις πελάτες υψηλής χρήσης και να προτείνει προσωποποιημένες προσφορές.
- Να ενισχύσει loyalty προγράμματα για πελάτες που χρησιμοποιούν συχνά διεθνείς κλήσεις ή άλλες ειδικές υπηρεσίες.

Ραβδόγραμμα: Σημασίας Χαρακτηριστικών (Feature Importance)



Εικόνα 6 : Ραβδόγραμμα σημαντικότητας τριών χαρακτηριστικών (Feature Importance) Decision Tree

- **Σημαντικά χαρακτηριστικά για το Decision Tree:**

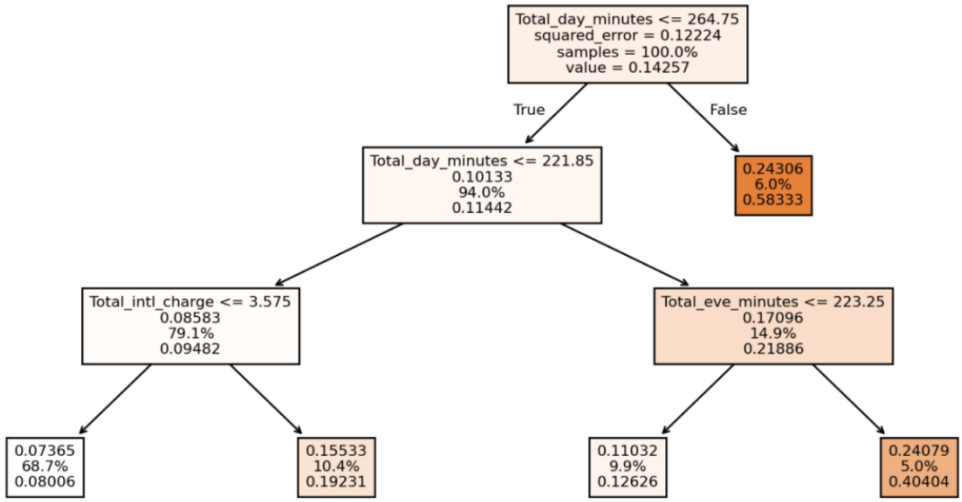
Το διάγραμμα δείχνει πόσο συνεισφέρει κάθε χαρακτηριστικό στη λήψη αποφάσεων του δέντρου. Το **διάγραμμα σημασίας χαρακτηριστικών** παρέχει στρατηγική καθοδήγηση για το ποιες μεταβλητές να στοχεύσει το μάρκετινγκ για να μειώσει το churn.

- **Total_day_minutes:** Κατέχει τη μεγαλύτερη σημασία (0.8), ακολουθούμενο από το **Total_eve_minutes (0.14)** και το **Total_intl_charge (0.06)**.

Τα χαρακτηριστικά κατατάσσονται με φθίνουσα σειρά σημαντικότητας.

Αντικατοπτρίζεται η συνολική συνεισφορά κάθε χαρακτηριστικού σε όλες τις διακλαδώσεις του δέντρου.

Παρέχει συνοπτική εικόνα των χαρακτηριστικών στα οποία πρέπει να εστιάσει το μάρκετινγκ και επιτρέπει την απλοποίηση του μοντέλου, εστιάζοντας στα σημαντικότερα χαρακτηριστικά.



Εικόνα 7 : Δέντρο Απόφασης (decision tree) που απεικονίζει τη διαδικασία πρόβλεψης

Διάγραμμα Δέντρου Αποφάσεων (Decision Tree Depth 3)

Το διάγραμμα δείχνει την ιεραρχία των διακλαδώσεων στο Decision Tree, το οποίο επιλέγει τα χαρακτηριστικά που διαχωρίζουν καλύτερα τους πελάτες με βάση την πιθανότητα churn. Το **Decision Tree** εξηγεί πώς το μοντέλο καταλήγει στις προβλέψεις, υποδεικνύοντας συγκεκριμένα μονοπάτια και υποομάδες πελατών.

Σημαντική διακλάδωση: Το χαρακτηριστικό **Total_day_minutes** είναι η ρίζα του δέντρου, που σημαίνει ότι έχει τη μεγαλύτερη σημασία για τον διαχωρισμό των πελατών.

Στις επόμενες διακλαδώσεις, χαρακτηριστικά όπως το **Total_eve_minutes** και το **Total_intl_charge** συμβάλλουν περαιτέρω στην πρόβλεψη.

Κάθε κόμβος δείχνει:

- **Συνολικό δείγμα πελατών** που ανήκει σε αυτόν τον κόμβο.
- **Squared error:** Το σφάλμα στην πρόβλεψη για αυτόν τον κόμβο.
- **Value:** Η μέση τιμή churn (πιθανότητα) για τους πελάτες σε αυτόν τον κόμβο.

- **Χρωματισμός:** Οι κόμβοι με πιο σκούρο πορτοκαλί χρώμα δείχνουν υψηλότερη πιθανότητα churn. Αναδεικνύει ποιες υποκατηγορίες πελατών έχουν τη μεγαλύτερη πιθανότητα να αποχωρήσουν, π.χ., πελάτες με υψηλή ημερήσια χρήση (>264.75 λεπτά) έχουν πιθανότητα churn 58.33%.

Συνολική Ανάλυση του Data Set για το Churn

Το churn (αποχώρηση πελατών) είναι ένα κρίσιμο ζήτημα στις τηλεπικοινωνίες και επηρεάζεται από διάφορους παράγοντες που αφορούν τη χρήση υπηρεσιών, το κόστος και τη συμπεριφορά των πελατών. Από την ανάλυση των δεδομένων με τα μοντέλα **XGBoost**, **Random Forest** και **Decision Tree**, προκύπτει η παρακάτω ανάλυση:

1. **Κορυφαίοι Παράγοντες που Επηρεάζουν το Churn:**
    - **Total_day_minutes:** Η διάρκεια ημερήσιων κλήσεων είναι ο πιο σημαντικός παράγοντας. Πελάτες με υψηλή ημερήσια χρήση έχουν αυξημένη πιθανότητα αποχώρησης.
    - **Total_day_charge:** Η συσχέτιση με τις ημερήσιες χρεώσεις ενισχύει τη σημασία της ημερήσιας χρήσης ως προγνωστικού παράγοντα.
    - **Total_eve_minutes και Total_eve_charge:** Οι βραδινές κλήσεις και χρεώσεις έχουν επίσης σημαντική επιρροή, αλλά σε μικρότερο βαθμό.
    - **Total_intl_charge και Total_intl_calls:** Οι διεθνείς κλήσεις και οι σχετικές χρεώσεις επηρεάζουν εξειδικευμένες ομάδες πελατών.
2. **Μεσαίας Σημασίας Παράγοντες:**
    - Χαρακτηριστικά όπως το **Number_vmail_messages** και το **Voice_mail_plan** έχουν μικρότερη αλλά σημαντική επιρροή για την κατανόηση της συνολικής εμπειρίας του πελάτη.
    - Ο αριθμός κλήσεων (π.χ., **Total_day_calls, Total_eve_calls**) επηρεάζει λιγότερο το churn αλλά είναι χρήσιμος για τη βελτίωση των υπηρεσιών.
3. **Χαρακτηριστικά Χαμηλής Σημασίας:**
    - Χαρακτηριστικά όπως το **Account_length** (διάρκεια λογαριασμού) δεν έχουν σημαντική συμβολή στην πρόβλεψη του churn, κάτι που υποδεικνύει ότι οι πελάτες μπορεί να αποχωρήσουν ανεξάρτητα από το πόσο καιρό είναι στην εταιρεία.

Απόδοση και Σύγκριση Μοντέλων

XGBoost

- **Ακρίβεια (Test):** ~89%.
- **Σημασία Χαρακτηριστικών:**
    - Δίνει τη μεγαλύτερη σημασία στο **Total_day_minutes**, υποδεικνύοντας ότι η ημερήσια χρήση είναι ο κυρίαρχος παράγοντας.
- **Χρόνος Εκτέλεσης:**
    - Εξαιρετικά γρήγορο και αποδοτικό σε μεγάλα δεδομένα.
- **Πλεονεκτήματα:**
    - Καλύτερη διαχείριση μη ισορροπημένων δεδομένων.
    - Υψηλή ακρίβεια και δυνατότητα fine-tuning.
- **Μειονεκτήματα:**
    - Πολύπλοκο στην ερμηνεία και δυσκολότερο στην παρουσίαση προς μη τεχνικό κοινό.

Random Forest

- **Ακρίβεια (Test):** ~88%.
- **Σημασία Χαρακτηριστικών:**
    - Αναδεικνύει περισσότερα χαρακτηριστικά μεσαίας σημασίας (π.χ., **Number_vmail_messages**) σε σχέση με το XGBoost και το Decision Tree.
- **Χρόνος Εκτέλεσης:**
    - Σχετικά αργό σε σύγκριση με το XGBoost, αλλά παραμένει πρακτικό.
- **Πλεονεκτήματα:**
    - Ισορροπημένη προσέγγιση στη σημασία χαρακτηριστικών.
    - Ανθεκτικότητα σε μη γραμμικές σχέσεις.
- **Μειονεκτήματα:**
    - Υψηλότερος χρόνος επεξεργασίας και μεγαλύτερη πολυπλοκότητα από το Decision Tree.

Decision Tree

- **Ακρίβεια (Test):** ~86%.
- **Σημασία Χαρακτηριστικών:**
    - Εστιάζει κυρίως στα κορυφαία χαρακτηριστικά, όπως το **Total_day_minutes** και το **Total_eve_minutes**, αλλά παραλείπει λιγότερο σημαντικά χαρακτηριστικά.
- **Χρόνος Εκτέλεσης:**
    - Εξαιρετικά γρήγορο και εύκολο στην εφαρμογή.
- **Πλεονεκτήματα:**
    - Εύκολο στην κατανόηση και κατάλληλο για παρουσιάσεις.
- **Μειονεκτήματα:**
    - Ευάλωτο στο overfitting χωρίς περιορισμούς βάθους.
    - Χάνει τη συνολική εικόνα των δεδομένων.

5.1.5 Σύγκριση Αποτελεσμάτων

Κριτήριο	**XGBoost**	**Random Forest**	**Decision Tree**
Ακρίβεια (Test)	~89%	~88%	~86%
Χρόνος Εκτέλεσης (Test)	0.002 sec	0.0046 sec	0.0003 sec
Σημασία Χαρακτηριστικών	Υψηλή ακρίβεια σημασίας	Ισορροπημένη σημασία	Εστιάζει στα κορυφαία
Χρήση σε Μάρκετινγκ	Προσαρμοσμένες καμπάνιες	Ανάλυση μεσαίας σημασίας παραγόντων	Εύκολο στην παρουσίαση
Πολυπλοκότητα	Υψηλή	Μέτρια	Χαμηλή

Πίνακας 7 : Σύγκριση Αποτελεσμάτων Αλγορίθμων

Συμπεράσματα για το Μάρκετινγκ της Orange

Βασικά Σημεία Εστίασης:

1. **Μείωση Κόστους και Εξατομικευμένα Πακέτα:**
    - Πρέπει να εστιάσει στους πελάτες με υψηλή ημερήσια χρήση, προσφέροντας χαμηλότερες χρεώσεις για να μειώθεί το churn.
2. **Διατήρηση Πελατών:**
    - Θα χρειαστεί να δημιουργήσει loyalty προγράμματα για πελάτες που κάνουν συχνή χρήση διεθνών κλήσεων ή voicemail.
3. **Προσαρμοσμένες Καμπάνιες:**
    - Να χρησιμοποιήσει το **XGBoost** για ακριβείς προβλέψεις και στοχευμένες καμπάνιες.
    - Το **Random Forest** μπορέι να το χρησιμοποιήσει για τον εντοπισμό μεσαίας σημασίας παραγόντων και δημιουργία νέων ευκαιριών.

Τελική Ανάλυση

- **XGBoost:** Ιδανικό για ακριβή πρόβλεψη churn και βέλτιστη χρήση πόρων.
- **Random Forest:** Κατάλληλο για στρατηγικές με έμφαση στη γενική βελτίωση.
- **Decision Tree:** Εξαιρετικό για παρουσίαση ευρημάτων και αρχική ανάλυση.

5.2 Δεύτερη Μελέτη Περίπτωσης

Το δεύτερο δείγμα προς ανάλυση πρόκειται για μια εταιρεία τηλεπικοινωνιών και περιλαμβάνει πάνω από 3300 πελάτες επίσης και αυτό το δείγμα το αναζητήσαμε και το βρήκαμε στο Kaggle.com. https://www.kaggle.com/datasets/mehmetsabrikunt/internet-service-churn

Από τα δεδομένα που βρήκαμε κατά την αναζήτηση μας έγινε καθαρισμός και ομογενοποίηση των δεδομένων πριν ξεκινήσουμε να χρησιμοποιήσουμε τους αλγόριθμους μηχανικής μάθησης.

Ακόμη διαχωρίστηκαν τα δεδομένα αυτά με τυχαία σειρά ώστε να δημιουργήσουμε δεδομένα για εκπαίδευση (Data Train) των αλγορίθμων και δεδομένα επαλήθευση των τιμών (Data Test) επιλέξαμε τις παρακάτω μεταβλητές ως τις κυριότερες για να μπορέσουμε να εξάγουμε τα αποτελέσματα μας.

Variables - Μεταβλητές

Οι μεταβλητές που θα χρησιμοποιήσουμε είναι οι παρακάτω.

1. AccountWeeks

Αντιπροσωπεύει τον αριθμό εβδομάδων που ένας πελάτης παραμένει ενεργός στην εταιρεία.

2. ContractRenewal

Αντιπροσωπεύει αν ένας πελάτης έχει ανανεώσει το συμβόλαιο του (1) ή όχι (0).

3. DataPlan

Υποδεικνύει αν ένας πελάτης έχει ενεργό πακέτο δεδομένων (1) ή όχι (0).

4. DataUsage

Ποσότητα δεδομένων που χρησιμοποιεί ένας πελάτης.

5. CustServCalls

Αριθμός κλήσεων που έγιναν στο τμήμα εξυπηρέτησης πελατών.

6. DayMins

Συνολικός αριθμός λεπτών που χρησιμοποιούνται κατά τη διάρκεια της ημέρας.

7. DayCalls

Αριθμός κλήσεων κατά τη διάρκεια της ημέρας.

8. MonthlyCharge

Μηνιαίο ποσό που χρεώνεται ο πελάτης.

9. OverageFee

Επιπλέον χρεώσεις λόγω υπέρβασης χρήσης.

10. RoamMins

Λεπτά περιαγωγής που χρησιμοποιούνται από τον πελάτη.

11. Churn

Εξαρτημένη μεταβλητή που δείχνει αν ένας πελάτης έχει αποχωρήσει (1) ή παραμένει (0).

5.2.1 Αποτελέσματα Ανάλυσης δεύτερου δείγματος

Πριν «τρέξουμε» το αρχείο με τα δεδομένα έγινε μία στατιστική ανάλυση και παρακάτω παραθέτουμε του πίνακες με τα αποτελέσματα της ανάλυσης όπως δόθηκαν και στο προηγούμενο δείγμα.

Train Data								
A/A	Variables	mean	median	std	min	max	skewness	kurtosis
1	AccountWeeks	100.5552	100	39.7522899083814	1	243	0.1171530986210808	-0.08391151802608121
2	ContractRenewal	0.9052	1	0.2929384918374439	0	1	-2.766451055704114	5.653251443605196
3	DataPlan	0.2724	0	0.4451946091317764	0	1	1.022474195920127	-0.9545465186773945
4	DataUsage	72.4908	0	121.6169014379164	0	475	1.467052299458082	0.596117967308238
5	CustServCalls	1.5736	1	1.331684286908885	0	9	1.103511781241021	1.776391952535231
6	DayMins	1642.1044	1730	703.5144745494866	0	3508	-0.5245312281030573	-0.08936015879162262
7	DayCalls	100.7288	101	20.24320257666754	0	165	-0.1354155181956976	0.371595440376796
8	MonthlyCharge	297.2408	61	317.5244261712158	14	1113	0.7133006062640528	-1.100875876379207
9	OverageFee	917.3376	979	368.1647148033607	0	1809	-0.835274501095425	0.4122166169184602
10	RoamMins	92.6652	98	38.10514281511101	0	184	-0.7557172650667205	0.1631138169065292
11	Churn	0.1396	0	0.3465715510540355	0	1	2.079801408418697	2.325573898459768

Πίνακας 8: Στατιστικά Δεδομένα εκπαίδευσης των αλγόριθμων

Test Data								
A/A	Variables	mean	median	std	min	max	skewness	kurtosis
1	AccountWeeks	102.594237695078	102	39.96812734280206	1	224	0.03475511743276825	-0.1729627943323284
2	ContractRenewal	0.8967587034813925	1	0.304273776740232	0	1	-2.607905996579631	4.801173686996145
3	DataPlan	0.2893157262905162	0	0.4534447450478479	0	1	0.9292610665815647	-1.136473870135679
4	DataUsage	76.73709483793517	0	125.0686122559619	0	464	1.364941517496604	0.2656245391862564
5	CustServCalls	1.530612244897959	1	1.264279489019993	0	7	1.036073238420928	1.477449099174233
6	DayMins	1646.524609843937	1722	719.1158150929548	27	3453	-0.4659207510399576	-0.1207301344500289
7	DayCalls	99.55582232893157	99	19.49838095478671	35	150	-0.04295082289092141	-0.2073902175582356
8	MonthlyCharge	303.0708283313325	63	321.0595768484545	16	1087	0.6848100567067729	-1.163666165188908
9	OverageFee	896.3277310924369	964	376.3884604295833	4	1819	-0.7785949922752619	0.1685511644233966
10	RoamMins	92.01800720288115	101	40.1042867708342	0	189	-0.6979547971566751	0.01509491085158832
11	Churn	0.1608643457382953	0	0.3674057811310067	0	1	1.84610951530334	1.408120342493548

Πίνακας 9 : Δεδομένα επαλήθευσης των αλγόριθμων

Percentage Differences								
A/A	Variables	mean	median	std	min	max	skewness	kurtosis
1	AccountWeeks	-2.00742637362214	-1.98019801980198	-0.5414859627254455	0	8.137044967880085	108.4839034106013	-69.33451265613233
2	ContractRenewal	0.9369023277058296	0	-3.796065653434832	0	0	5.900056791243956	16.3008055625157
3	DataPlan	-6.022877942985533	0	-1.836139465225871	0	0	9.551821000465386	-17.40082042543376
4	DataUsage	-5.691020224532086	0	-2.798470624690898	0	2.342917997870074	7.211229159481741	76.70352237036668
5	CustServCalls	2.769640199228973	0	5.19304610595144	0	25	6.30388997869762	18.37476684387783
6	DayMins	-0.2688177858132614	0.4634994206257242	-2.193309204373729	-200	1.580232725183163	11.83509716721698	-29.86332702417972
7	DayCalls	1.171310765079141	2	3.748323824546788	-200	9.523809523809524	103.67953364012	705.1976155165119
8	MonthlyCharge	-1.942333966622661	-3.225806451612903	-1.107184226514294	-13.33333333333333	2.363636363636364	4.075578609312764	-5.545517606396119
9	OverageFee	2.316840769615199	1.544004117344313	-2.209041852155094	-200	-0.5512679162072768	7.024051083806507	83.91149107206934
10	RoamMins	0.7008680506700139	-3.015075376884422	-5.112283688314873	0	-2.680965147453083	7.94711123796348	166.1185823130355
11	Churn	-14.15432216161619	0	-5.836104071598601	0	0	11.90510419904287	49.144546754957

Πίνακας 10 : % διαφοράς Train & Test Data

ContractRenewal

- **Train Data**: Το μέσο ποσοστό ανανέωσης συμβολαίων είναι **90.52%**, με τυπική απόκλιση 29.29%, που υποδηλώνει σχετικά υψηλή συγκέντρωση πελατών που ανανεώνουν

συμβόλαια. Η ασυμμετρία (**skewness**) είναι **-2.77**, υποδεικνύοντας ότι η πλειοψηφία των τιμών είναι συγκεντρωμένες στο ανώτερο άκρο (πελάτες που ανανεώνουν συμβόλαια).

- **Test Data**: Το μέσο ποσοστό είναι **89.68%**, πολύ κοντά στο train data, με παρόμοια κατανομή (skewness **-2.61**) και τυπική απόκλιση.
- **Percentage Differences**: Η μέση διαφορά είναι μόλις **0.94%**, κάτι που δείχνει σταθερότητα μεταξύ των συνόλων δεδομένων.
- **Συμπέρασμα**: Το **ContractRenewal** είναι ένας ισχυρός δείκτης δέσμευσης πελατών. Μια χαμηλή πιθανότητα ανανέωσης μπορεί να αποτελεί ένδειξη επικείμενου churn.

2. DataUsage

- **Train Data**: Η μέση χρήση δεδομένων είναι **72.49 MB**, με σημαντική τυπική απόκλιση (**121.62 MB**), γεγονός που υποδηλώνει μεγάλη διακύμανση μεταξύ των πελατών. Η skewness είναι **1.47**, κάτι που δείχνει ότι λίγοι πελάτες έχουν πολύ υψηλή χρήση δεδομένων.
- **Test Data**: Η μέση χρήση είναι ελαφρώς υψηλότερη (**76.73 MB**), με μικρότερη τυπική απόκλιση (**125.07 MB**) και παρόμοια skewness (**1.36**).
- **Percentage Differences**: Η μέση διαφορά είναι **-5.69%**, που υποδεικνύει ελαφρώς υψηλότερη χρήση στο test set.
- Συμπερασματικά η χρήση δεδομένων μπορεί να είναι ένδειξη της ικανοποίησης ή του ενδιαφέροντος για τις υπηρεσίες.

3. DayMins

- **Train Data**: Η μέση διάρκεια ημερήσιων κλήσεων είναι **1642.1 λεπτά**, με τυπική απόκλιση **703.51 λεπτά**. Η skewness είναι **-0.52**, δείχνοντας μια σχετικά ομαλή κατανομή γύρω από τη μέση τιμή.
- **Test Data**: Το μέσο όρο είναι πολύ κοντά (**1646.52 λεπτά**) με παρόμοια τυπική απόκλιση (**719.12 λεπτά**).
- **Percentage Differences**: Οι μέσες διαφορές είναι σχεδόν μηδενικές (**-0.27%**), δείχνοντας συνεπή κατανομή μεταξύ των συνόλων δεδομένων.
- Κατά συνέπεια **το DayMins** είναι κρίσιμο χαρακτηριστικό για την πρόβλεψη churn, καθώς οι πελάτες με υψηλή χρήση μπορεί να είναι πιο δεσμευμένοι.

4. MonthlyCharge

- **Train Data**: Η μέση μηνιαία χρέωση είναι **297.24 ευρώ**, με σημαντική τυπική απόκλιση (**317.52 ευρώ**). Η skewness είναι **0.71**, δείχνοντας ότι οι περισσότερες χρεώσεις συγκεντρώνονται σε χαμηλότερα επίπεδα, με λίγους πελάτες να έχουν υψηλές χρεώσεις.
- **Test Data**: Η μέση τιμή είναι ελαφρώς υψηλότερη (**303.07 ευρώ**), με παρόμοια κατανομή (skewness **0.68**).
- **Percentage Differences**: Η μέση διαφορά είναι **-1.94%**, που είναι σχετικά μικρή.
- Οι υψηλές **μηνιαίες χρεώσεις** ενδέχεται να σχετίζονται με churn, καθώς υποδηλώνουν δυσαρέσκεια ή αδυναμία πληρωμής.

5. Roam Mins

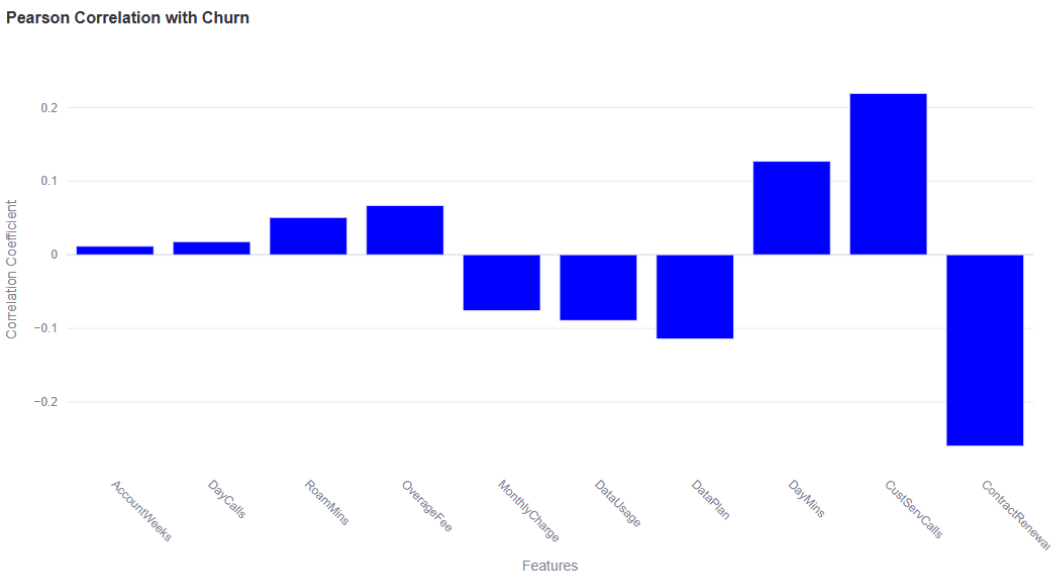
Train Data: Ο μέσος χρόνος περιαγωγής είναι **92.67 λεπτά**, με τυπική απόκλιση **38.10 λεπτά**. Η skewness είναι **-0.76**, που δείχνει ομοιόμορφη κατανομή.

Test Data: Το μέσο όρο είναι ελαφρώς χαμηλότερο (**92.02 λεπτά**), με παρόμοια τυπική απόκλιση (**40.10 λεπτά**).

- **Percentage Differences**: Η μέση διαφορά είναι **0.70%**, κάτι που υποδηλώνει μικρή μεταβλητότητα μεταξύ των δύο συνόλων δεδομένων.
- Το **RoamMins** μπορεί να είναι δευτερεύουσα ένδειξη churn, ιδιαίτερα για πελάτες που ταξιδεύουν συχνά.

Γενικά Συμπεράσματα

- Τα **ContractRenewal** και **DayMins** φαίνεται να είναι τα πιο κρίσιμα χαρακτηριστικά για την πρόβλεψη του churn, δεδομένης της υψηλής σταθερότητας μεταξύ των train και test sets και της συσχέτισής τους με τη συμπεριφορά πελατών.
- Οι **μηνιαίες χρεώσεις** και η **χρήση δεδομένων** παρέχουν πρόσθετες πληροφορίες για την κατανόηση της δέσμευσης ή δυσαρέσκειας των πελατών.
- Τα train και test datasets έχουν συνεπή κατανομή, όπως φαίνεται από τις μικρές ποσοστιαίες διαφορές, καθιστώντας το μοντέλο κατάλληλο για γενίκευση.



Εικόνα 8: Ραβδόγραμμα Συσχέτισης Χαρακτηριστικών με την μεταβλητή Churn

Το διάγραμμα Pearson μας δείχνει τη συσχέτιση των χαρακτηριστικών με το churn (αποχώρηση πελατών). Η ανάλυση του έχει ως εξής:

Ανάλυση Συσχετίσεων:

1. **ContractRenewal (-0.2):**
    - Έχει την πιο ισχυρή αρνητική συσχέτιση με το churn. Προσπαθώντας να ερμηνεύσουμε το αποτέλεσμα αυτής της συσχέτισης. Οι πελάτες που ανανεώνουν τα συμβόλαιά τους έχουν χαμηλότερες πιθανότητες αποχώρησης. Η μη ανανέωση αποτελεί βασικό δείκτη επικείμενου churn.
2. **CustServCalls (+0.15):**
    - Έχει την υψηλότερη θετική συσχέτιση με το churn. Αυτό συνεπάγεται ότι οι πελάτες που καλούν συχνά την εξυπηρέτηση πελατών είναι πιο πιθανό να αποχωρήσουν, πιθανόν λόγω δυσαρέσκειας ή μη επίλυσης προβλημάτων.
3. **DayMins (+0.1):**
    - Η θετική συσχέτιση αυτού του χαρακτηριστικού υποδηλώνει ότι οι πελάτες με υψηλή χρήση ημερήσιων λεπτών είναι πιθανότερο να αποχωρήσουν. Πιθανών να νιώθουν ότι πληρώνουν υπερβολικά ή δεν εξυπηρετούνται επαρκώς.

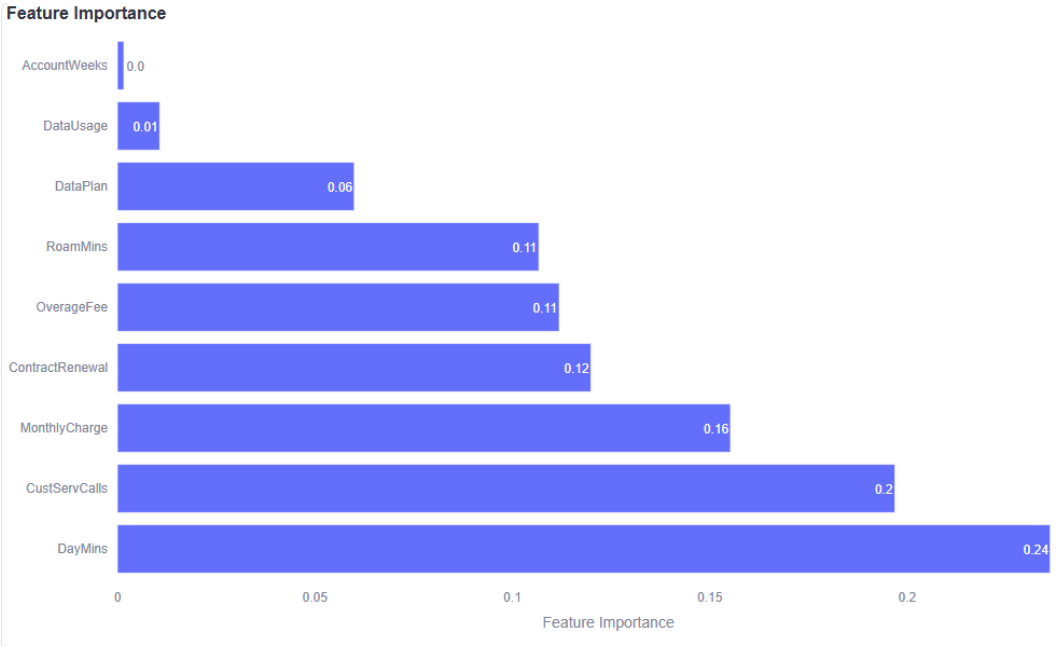
4. **DataPlan (-0.05):**
    - Εμφανίζει αρνητική συσχέτιση , δηλώνει ότι οι πελάτες που διαθέτουν data plan έχουν χαμηλότερες πιθανότητες να αποχωρήσουν, πιθανόν λόγω μεγαλύτερης δέσμευσης στις υπηρεσίες ή λόγω μεγαλύτερης ικανοποίησης.
5. **MonthlyCharge (-0.05):**
    - Παρόμοια αρνητική συσχέτιση, η μηνιαία χρέωση δεν φαίνεται να επηρεάζει σε μεγάλο βαθμό την απόφαση αποχώρησης.
6. **OverageFee (+0.05):**
    - Εμφανίζει μικρή θετική συσχέτιση. Οι πελάτες με υψηλά τέλη υπερχρήσης είναι ελαφρώς πιθανότερο να αποχωρήσουν.
7. **RoamMins (+0.05):**
    - Εμφανίζει μικρή θετική συσχέτιση, οι πελάτες που χρησιμοποιούν περιαγωγή μπορεί να είναι πιο δυσαρεστημένοι λόγω πρόσθετων χρεώσεων.
8. **DayCalls και AccountWeeks (κοντά στο 0):**
    - Πολύ χαμηλή συσχέτιση, δεν φαίνεται να επηρεάζουν σημαντικά την πιθανότητα churn.

Συμπεράσματα:

- Το **ContractRenewal** είναι ο ισχυρότερος παράγοντας με αρνητική συσχέτιση, υποδεικνύοντας ότι οι πελάτες που ανανεώνουν συμβόλαια είναι πιο σταθεροί.
- Το **CustServCalls** είναι ο πιο σημαντικός θετικός παράγοντας, πιθανώς δείκτης δυσαρέσκειας.
- Το **DayMins** και το **DataUsage** παρέχουν σημαντικές πληροφορίες, αλλά με μέτριες συσχετίσεις.

Προτάσεις για το Μάρκετινγκ:

- Θα χρειαστεί να εστίασει σε **προσφορές ανανέωσης συμβολαίων** για να ενθαρρυνθούν οι πελάτες να ανανεώσουν.
- Να βελτιώσει της **εξυπηρέτησης πελατών** για να μειωθούν οι δυσαρεστημένοι πελάτες.
- Να αναλύσει περεταίρω τις ανάγκες των πελατών με υψηλά **DayMins** και **OverageFee** για να προταθούν πιο οικονομικά πακέτα.



Εικόνα 9: Ραβδόγραμμα ευαισθησίας χαρακτηριστικών (Feature Sensitivity)

Αναλύοντας σε βάθος το παραπάνω διάγραμμα αναπαριστά τη σημασία των χαρακτηριστικών για το churn, όπως προσδιορίζεται από την στατιστική ανάλυση.

Υψηλής Σημασίας Σημαντικότητας:

1. **DayMins (0.24):**
    - Το πιο σημαντικό χαρακτηριστικό είναι οι ημερήσιες χρεώσεις/λεπτά επηρεάζουν περισσότερο την πιθανότητα churn. Πελάτες με υψηλή χρήση ημερήσιων λεπτών είναι πιθανότερο να αποχωρήσουν.
2. **CustServCalls (0.2):**
    - Το δεύτερο σημαντικότερο χαρακτηριστικό είναι οι τηλεφωνικές κλήσεις που κάνουν οι πελάτες στην εξυπηρέτηση πελατών που ενδέχεται να είναι δυσαρεστημένοι και να αποχωρήσουν.

Μεσαίας Σημαντικότητας:

3. **MonthlyCharge (0.16):**
    - Οι μηνιαίες χρεώσεις παίζουν σημαντικό ρόλο. Καθώς οι Πελάτες με υψηλές μηνιαίες χρεώσεις μπορεί να νιώθουν ότι πληρώνουν υπερβολικά, αυξάνοντας την πιθανότητα αποχώρησης.
4. **ContractRenewal (0.12):**
    - Σημαντική παραμένει η ανανέωση συμβολαίου, οι πελάτες που δεν ανανεώνουν το συμβόλαιό τους είναι πιθανότερο να αποχωρήσουν.

5. **OverageFee και RoamMins (0.11):**
    - Παρόμοια σημαντικότητα έχουν και οι πελάτες με υπερβολικές χρεώσεις ή αυξημένα λεπτά περιαγωγής ίσως νιώθουν δυσαρέσκεια λόγω πρόσθετων εξόδων.

Χαμηλής Σημαντικότητας:

6. **DataPlan (0.06):**
    - Οι πελάτες με πρόγραμμα δεδομένων έχουν ελαφρώς μειωμένη πιθανότητα churn, πιθανώς λόγω μεγαλύτερης δέσμευσης στις υπηρεσίες.
7. **DataUsage (0.01):**
    - Η χρήση δεδομένων έχει πολύ μικρή επίδραση, πιθανόν επειδή οι πελάτες έχουν σταθερή συμπεριφορά ως προς αυτήν.
8. **AccountWeeks (0.0):**
    - Η διάρκεια του λογαριασμού δεν φαίνεται να επηρεάζει την απόφαση αποχώρησης.

Συμπεράσματα και Προτάσεις:

1. **Υψηλή Σημασία:**
    - Τα χαρακτηριστικά με την μέγιστη σημασίας που θα χρειαστεί η εταιρεία τηλεπικοινωνιών να επικεντρωθεί είναι σε πελάτες με υψηλές ημερήσιες χρεώσεις και συχνή επικοινωνία με την εξυπηρέτηση πελατών, προσφέροντας ειδικά πακέτα ή προνόμια.
    - Καθώς επίσης θα πρέπει να μειωθούν τα προβλήματα που οδηγούν του πελάτες σε κλήση στην εξυπηρέτηση πελατών.
2. **Μεσαίας Σημαντικότητας:**
    - Να παρέχει κίνητρα για την ανανέωση συμβολαίων, όπως εκπτώσεις ή επιπλέον παροχές και
    - Να ενθαρρύνει τη χρήση πακέτων που μειώνουν τις υπερβολικές χρεώσεις και την περιαγωγή.
3. **Χαμηλής Σημαντικότητας:**
    - Να βελτιστοποιήσει τα προγράμματα δεδομένων ώστε να ενισχύσει τη δέσμευση πελατών, παρόλο που η επίδραση τους στο churn είναι λιγότερο σημαντική.

5.2.2 Αλγόριθμος XGBOOST

XGBOOST							
A/A	Data	Acc	FP	TP	FN	TN	RunTimeSec
0	Train	0.9472	0.0104	0.0972	0.0424	0.85	0.896453642845153
1	Test	0.9207683073229291	0.01080432172869148	0.09243697478991597	0.06842737094837935	0.8283313325330132	0.0004549980163574219

Πίνακας 11: Αποτελέσματα από αλγορίθμου XGBoost

Ο πίνακας περιγράφει την απόδοση του XGBoost μοντέλου για τα δεδομένα εκπαίδευσης (**Train**) και δοκιμών (**Test**). Ακολουθεί ανάλυση των αποτελεσμάτων:

1. **Accuracy (Acc):**
    - **Train:** 94.72% - Το μοντέλο αποδίδει εξαιρετικά καλά στα δεδομένα εκπαίδευσης.
    - **Test:** 92.08% - Παρόμοια καλή απόδοση και στα δεδομένα δοκιμών, υποδεικνύοντας ότι το μοντέλο δεν υπερπροσαρμόστηκε.
2. **False Positives (FP):**
    - **Train:** 0.0104 - Το ποσοστό των λανθασμένων θετικών στα δεδομένα εκπαίδευσης είναι πολύ χαμηλό.
    - **Test:** 0.0108 - Στα δεδομένα δοκιμών, η διαφορά είναι ελάχιστη, που δείχνει καλή γενίκευση.
3. **True Positives (TP):**
    - **Train:** 0.0972 - Το μοντέλο εντοπίζει σωστά το 9.72% των θετικών περιπτώσεων.
    - **Test:** 0.0924 - Ελαφρώς μειωμένο ποσοστό, αλλά εντός αποδεκτών ορίων.
4. **False Negatives (FN):**
    - **Train:** 0.0424 - Το 4.24% των πραγματικών θετικών περιπτώσεων δεν αναγνωρίστηκε.
    - **Test:** 0.0684 - Αυξημένο στα δεδομένα δοκιμών, πιθανώς λόγω μικρών αποκλίσεων στη γενίκευση.
5. **True Negatives (TN):**
    - **Train:** 0.85 - Το 85% των πραγματικών αρνητικών εντοπίστηκε σωστά.
    - **Test:** 0.828 - Λίγο χαμηλότερο, αλλά εξακολουθεί να είναι υψηλό.

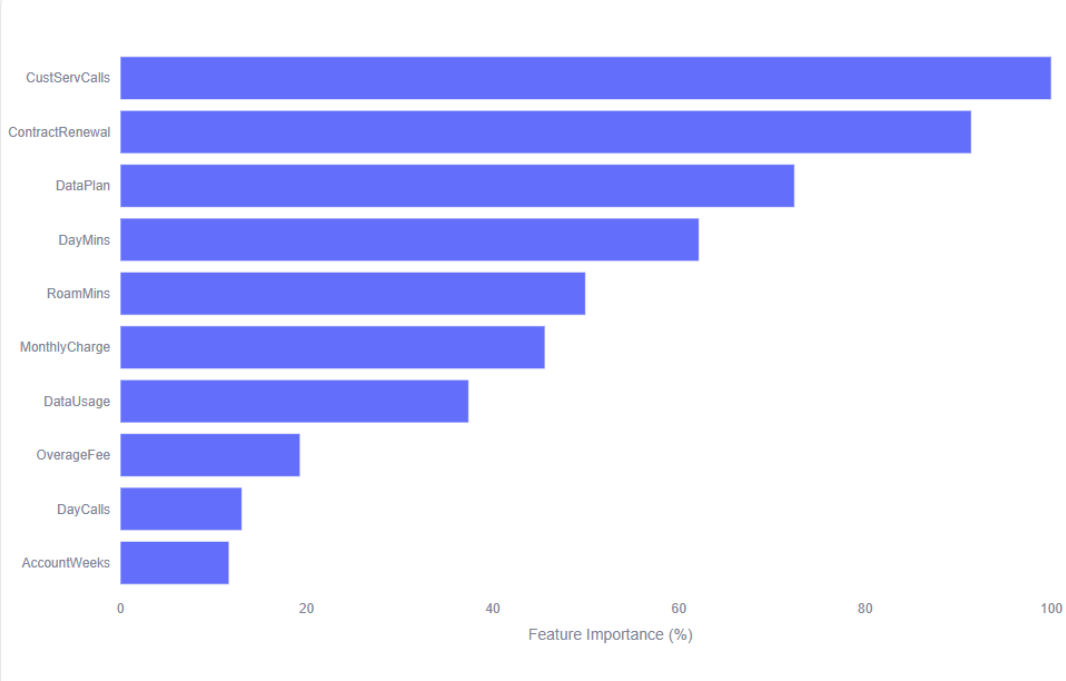
6. **RunTimeSec:**
    - **Train:** 0.896 δευτερόλεπτα - Χρόνος εκπαίδευσης αρκετά γρήγορος.
    - **Test:** 0.00045 δευτερόλεπτα - Εξαιρετικά ταχύς χρόνος δοκιμών.

Συμπεράσματα

1. **Σταθερή Απόδοση:**
    - Το XGBoost μοντέλο επιτυγχάνει υψηλή ακρίβεια και στις δύο φάσεις (εκπαίδευση και δοκιμές), δείχνοντας ότι είναι αξιόπιστο και δεν υπερπροσαρμόζεται.
2. **Αναγνώριση Churn:**
    - Τα ποσοστά TP και TN δείχνουν ότι το μοντέλο μπορεί να αναγνωρίσει με ακρίβεια τόσο τους πελάτες που αποχωρούν όσο και αυτούς που παραμένουν.
3. **Γρήγορη Εκτέλεση:**
    - Η ταχύτητα εκτέλεσης του XGBoost είναι εντυπωσιακή, κάνοντάς το κατάλληλο για πραγματικού χρόνου προβλέψεις.

Ανάλυση Χαρακτηριστικών XGBoost

Το παρακάτω διάγραμμα παρουσιάζει τη σημασία των χαρακτηριστικών που χρησιμοποιήθηκαν στο μοντέλο XGBoost για την πρόβλεψη του **Churn**. Ακολουθεί ανάλυση βασισμένη στη σειρά σημασίας των χαρακτηριστικών:



Εικόνα 10: Ραβδόγραμμα σημαντικότητας χαρακτηριστικών (Feature Importance) XGBoost

Υψηλής Σημασίας Χαρακτηριστικά

1. **CustServCalls** (Κλήσεις εξυπηρέτησης πελατών):
    - Το πιο σημαντικό χαρακτηριστικό για τον αλγόριθμο XGBoost είναι ο αριθμός των κλήσεων στην εξυπηρέτηση πελατών και αποτελεί κρίσιμο δείκτη για την πιθανότητα αποχώρησης πελατών. Πελάτες που επικοινωνούν συχνά είναι πιθανό να έχουν θέματα που δεν έχουν λυθεί.
2. **ContractRenewal** (Ανανέωση συμβολαίου):
    - Το δεύτερο πιο σημαντικό χαρακτηριστικό είναι η πληροφορία εάν ο πελάτης έχει ανανεώσει το συμβόλαιο του ή όχι. Η έλλειψη ανανέωσης του συμβολαίου είναι δείκτης που κάνει ξεκάθαρη την πρόθεση αποχώρησης των πελατών
3. **DataPlan** (Πρόγραμμα δεδομένων):
    - Επίσης σημαντικό χαρακτηριστικό έχουν οι πελάτες χωρίς πρόγραμμα δεδομένων μπορεί να μην εκμεταλλεύονται πλήρως τις υπηρεσίες και να θεωρούν ότι δεν έχουν επαρκή αξία από την εταιρεία.
4. **DayMins** (Λεπτά ημερήσιας χρήσης):
    - Στην μέση του δείκτη σημαντικότητας βρίσκεται ενας άλλος παράγοντας που χρειάζεται να λάβει υπόψιν της η εταιρεία που σχετίζεται με την υψηλή χρήση ημερήσιων λεπτών. Μπορεί να δείχνει ενεργούς χρήστες που ενδεχομένως αναμένουν καλύτερες προσφορές.
5. **RoamMins** (Λεπτά σε περιαγωγή):
    - Στην μέση από άποψη σημαντικότητας παρατηρούμε ότι βρίσκονται οι πελάτες με υψηλή χρήση περιαγωγής μπορεί να επηρεαστούν από υψηλές χρεώσεις.
6. **MonthlyCharge** (Μηνιαία χρέωση):
    - Άλλος ένας δείκτης μεσαίας σημαντικότητας είναι οι πελάτες με υψηλούς λογαριασμούς, αυτοί οι πελάτες είναι πιθανότερο να αποχωρήσουν λόγω κόστους.
7. **DataUsage** (Χρήση δεδομένων):
    - Χαμηλής σημαντικότητας δείκτης είναι η χρήση δεδομένων που σχετίζεται με τις ανάγκες του πελάτη, αλλά λιγότερο σημαντική από το αν έχουν πρόγραμμα - πακέτο δεδομένων.
8. **OverageFee** (Χρέωση υπέρβασης):

- Επίσης χαμηλή σημαντικότητα έχουν οι πελάτες με υψηλές υπερβάσεις που ενδέχεται να μην είναι ικανοποιημένοι, αλλά το χαρακτηριστικό αυτό δεν επηρεάζει ιδιαίτερα την απόφαση αλλαγή παρόχου.

9. **DayCalls** (Ημερήσιες κλήσεις):
    - Μικρότερη σημαντικότητα έχει ο αριθμός των κλήσεων που από μόνος του αυτός ο δείκτης δεν είναι ικανός για να αποχωρήσει ο πελάτης.
10. **AccountWeeks** (Εβδομάδες λογαριασμού):
    - Την μικρότερη σημαντικότητα εχει η διάρκεια που ένας πελάτης είναι με την εταιρεία καθώς δεν συνδέεται άμεσα με το churn.

Συμπεράσματα

1. **Κρίσιμα Σημεία:**
    - Το **CustServCalls** και το **ContractRenewal** είναι οι βασικοί δείκτες που επηρεάζουν την πιθανότητα αποχώρησης.
    - Η χρήση δεδομένων και οι περιαγωγής έχουν σχετική σημασία, αλλά δεν παίζουν καταλυτικό ρόλο στην λήψη της απόφασης αποχώρησης του πελάτη .
2. **Προτάσεις Μάρκετινγκ:**
    - Βελτίωση της εμπειρίας εξυπηρέτησης πελατών, με έμφαση στη γρήγορη και αποτελεσματική επίλυση προβλημάτων.
    - Δημιουργία κινήτρων για ανανέωση συμβολαίων, όπως ειδικές εκπτώσεις ή επιπλέον υπηρεσίες.
    - Παροχή ελκυστικών προγραμμάτων δεδομένων που καλύπτουν τις ανάγκες των πελατών.
3. **Διατήρηση Πελατών:**
    - Στοχευμένες καμπάνιες για πελάτες που έχουν συχνές κλήσεις εξυπηρέτησης ή υψηλές μηνιαίες χρεώσεις.
    - Ανάλυση της χρήσης περιαγωγής για να προταθούν ειδικά σχέδια με χαμηλότερο κόστος.

5.2.3 Αλγόριθμος Random Forest

Παρακάτω τα αποτελέσματα του αλγορίθμου **Random Forest** που θα χρειαστεί να αναλύσουμε για να κατανοήσουμε περισσότερο πως λειτουργεί ο αλγόριθμος πάνω στα δεδομένα που του δώσαμε.

Random Forest							
	Data	Acc	FP	TP	FN	TN	RunTimeSec
0	Train	0.9644	0.0028	0.1068	0.0328	0.8576	64.41508269309998
1	Test	0.9207683073229291	0.007202881152460984	0.08883553421368548	0.07202881152460984	0.8319327731092437	0.07111709117889405

Πίνακας 12: Αποτελέσματα από αλγορίθμου Random Forest

Ο πίνακας παρέχει πληροφορίες για την απόδοση του Random Forest σε δύο σετ δεδομένων: το **Training** και το **Testing**. Ακολουθεί αναλυτική περιγραφή:

Training Data:

- **Accuracy (Acc):** 0.9644
    - Πολύ υψηλή ακρίβεια, δείχνει ότι το μοντέλο προβλέπει σωστά το 96.44% των περιπτώσεων στο εκπαιδευτικό σύνολο.
- **False Positives (FP):** 0.0028
    - Πολύ χαμηλός αριθμός ψευδών θετικών (περίπου 0.28%), κάτι που σημαίνει ότι το μοντέλο σπάνια προβλέπει εσφαλμένα ότι ένας πελάτης θα αποχωρήσει.
    -
- **True Positives (TP):** 0.1068
    - Το 10.68% των δειγμάτων αναγνωρίζονται σωστά ως αποχωρούντες πελάτες.
- **False Negatives (FN):** 0.0328
    - Μόνο 3.28% των δειγμάτων που θα αποχωρούσαν δεν αναγνωρίστηκαν σωστά.
- **True Negatives (TN):** 0.8576
    - Το 85.76% των δειγμάτων που δεν αποχωρούν αναγνωρίστηκαν σωστά.
- **Runtime (RunTimeSec):** 64.42 δευτερόλεπτα
    - Το μοντέλο χρειάστηκε αρκετό χρόνο για την εκπαίδευση, λόγω της πολυπλοκότητας του Random Forest.

Testing Data:

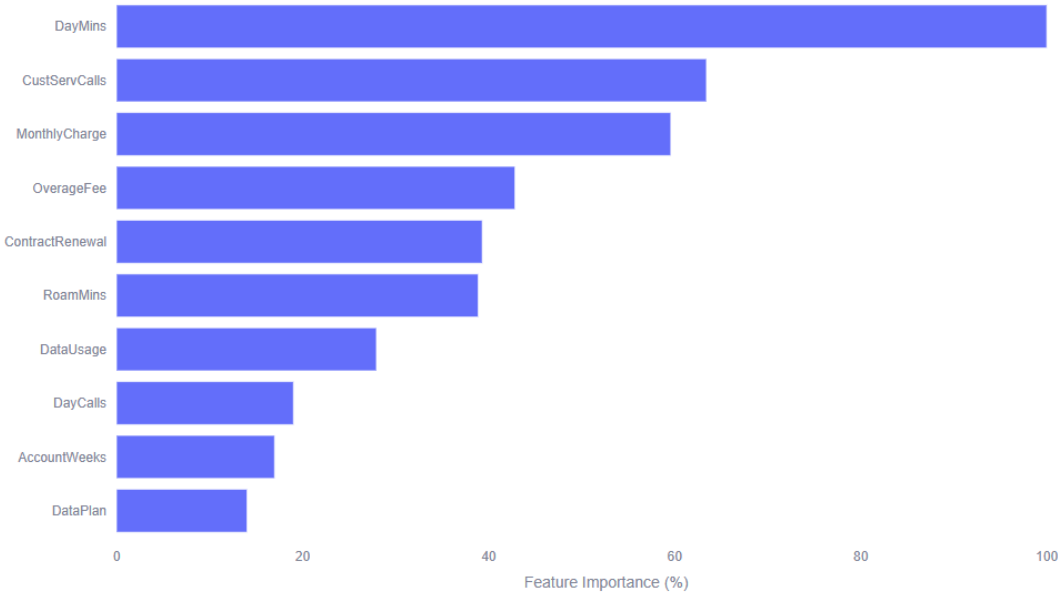
- **Accuracy (Acc):** 0.9208
    - Πολύ καλή ακρίβεια, με ελαφρώς χαμηλότερη επίδοση στο testing (92.08%), κάτι που δείχνει ότι το μοντέλο γενικεύει καλά.

- **False Positives (FP):** 0.0072
    - Ο αριθμός των ψευδών θετικών είναι χαμηλός (περίπου 0.72%), παραμένοντας σε καλά επίπεδα.
- **True Positives (TP):** 0.0888
    - Το 8.88% των δειγμάτων αναγνωρίζονται σωστά ως αποχωρούντες πελάτες.
- **False Negatives (FN):** 0.0720
    - Το 7.2% των δειγμάτων που θα αποχωρούσαν δεν αναγνωρίστηκαν σωστά, δείχνοντας περιθώριο βελτίωσης.
- **True Negatives (TN):** 0.8319
    - Το 83.19% των δειγμάτων που δεν αποχωρούν αναγνωρίστηκαν σωστά.
- **Runtime (RunTimeSec):** 0.0711 δευτερόλεπτα
    - Πολύ χαμηλός χρόνος για την πρόβλεψη, κάτι που καθιστά το Random Forest αποδοτικό για γρήγορες προβλέψεις.

Συμπεράσματα:

1. **Υψηλή Απόδοση:**
    - Το Random Forest αποδίδει εξαιρετικά τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα δοκιμής, με μικρή μείωση στην ακρίβεια στο testing, δείχνοντας ότι το μοντέλο δεν υπερπροσαρμόστηκε (overfitting).
2. **Αξιοπιστία Προβλέψεων:**
    - Ο χαμηλός αριθμός FP και FN δείχνει ότι το μοντέλο είναι ικανό να αναγνωρίσει με αξιοπιστία τόσο τους πελάτες που αποχωρούν όσο και εκείνους που παραμένουν.
3. **Χρόνος Εκτέλεσης:**
    - Παρότι το training απαιτεί σημαντικό χρόνο λόγω της φύσης του Random Forest, οι προβλέψεις στα δεδομένα testing εκτελούνται πολύ γρήγορα.

Το διάγραμμα παρουσιάζει τη σημασία των χαρακτηριστικών όπως προέκυψε από την εκπαίδευση του μοντέλου **Random Forest**.



Εικόνα 11 : Ραβδόγραμμα σημαντικότητας χαρακτηριστικών (Feature Importance) Random Forest

Υψηλής Σημασίας Χαρακτηριστικά

1. **DayMins (Χρόνος Ημερήσιων Κλήσεων)**:
    - Το σημαντικότερο χαρακτηριστικό στο μοντέλο είναι οι πελάτες με υψηλή διάρκεια ημερήσιων κλήσεων έχουν μεγαλύτερη πιθανότητα να αποχωρήσουν. Ίσως συνδέεται με αυξημένα κόστη.
2. **CustServCalls (Αριθμός Κλήσεων στην Εξυπηρέτηση Πελατών)**:
    - Το δεύτερο πιο σημαντικό χαρακτηριστικό είναι οι πελάτες που επικοινωνούν συχνά με την εξυπηρέτηση πελάτων πιθανόν αντιμετωπίζουν προβλήματα ή είναι δυσαρεστημένοι.
3. **MonthlyCharge (Μηνιαία Χρέωση)**:
    - Το τρίτο πιο σημαντικό χαρακτηριστικό είναι οι υψηλές μηνιαίες χρεώσεις συνδέονται με αυξημένο κίνδυνο αποχώρησης.

Μεσαίας Σημασίας Χαρακτηριστικά

4. **OverageFee (Χρεώσεις Υπέρβασης)**:
    - Οι υψηλές χρεώσεις υπέρβασης επηρεάζουν αρνητικά τη διατήρηση πελατών αλλά όχι σε σημαντικό βαθμό για να χακτηριστεί σημαντικό.
5. **ContractRenewal (Ανανέωση Συμβολαίου)**:
    - Η μη ανανέωση του συμβολαίου χαρακτηρίζεται ως μεσαίας σημαντικότητας και συσχετίζεται άμεσα με το churn.

6. **RoamMins (Χρόνος Περιφερειακών Κλήσεων)**:
    - Το μοντέλο μας δείχνει ότι οι πελάτες που κάνουν συχνές περιφερειακές κλήσεις ενδέχεται να είναι πιο απαιτητικοί ή να αναζητούν πιο ανταγωνιστικές τιμές.

Χαμηλής Σημασίας Χαρακτηριστικά

7. **DataUsage (Χρήση Δεδομένων)**:
    - Η χρήση δεδομένων έχει μικρότερη επίδραση στην πρόβλεψη αποχώρησης.
8. **DayCalls (Αριθμός Ημερήσιων Κλήσεων)**:
    - Η συχνότητα των ημερήσιων κλήσεων έχει περιορισμένη σημασία για την πρόβλεψη.
9. **AccountWeeks (Διάρκεια Συμβολαίου σε Εβδομάδες)**:
    - Το πόσο καιρό είναι ένας πελάτης με την εταιρεία έχει μικρή σημασία.
10. **DataPlan (Πακέτο Δεδομένων)**:
    - Η ύπαρξη ή μη πακέτου δεδομένων επηρεάζει ελάχιστα την πρόβλεψη.

Συμπεράσματα

- Τα κορυφαία χαρακτηριστικά (DayMins, CustServCalls, MonthlyCharge) δείχνουν τη σημαντικότητα της παρακολούθησης της χρήσης και των παραπόνων πελατών.
- Οι μεσαίας σημασίας μεταβλητές υποδεικνύουν ανάγκη για στρατηγική διαχείριση των χρεώσεων και των ανανεώσεων.
- Τα χαμηλής σημασίας χαρακτηριστικά δεν πρέπει να παραμεληθούν αλλά απαιτούν λιγότερη προτεραιότητα.

Διατήρηση Πελατών:

- Το τμήμα Μάρκετινγκ χρειάζεται να δημιουργήσει στοχευμένες προσφορές για πελάτες με υψηλή διάρκεια ημερήσιων κλήσεων και να προσαρμόσει την τιμολογιακής Πολιτικής με σκοπό να μειώσει τις χρεώσεις μέσω προσφορών για να διατηρήσει πελάτες με υψηλή μηνιαία χρέωση.
- Και βέβαια η εταιρεία θα πρέπει να Βελτιώσει την Εξυπηρέτηση της μέσω της ανάλυσης παραπόνων για την μείωση της δυσαρέσκειας των πελάτων.

5.2.4 Αλγόριθμος Decision Tree

Decision Tree							
A/A	Data	Acc	FP	TP	FN	TN	RunTimeSec
0	Train	0.926	0.0096	0.0752	0.0644	0.8508	0.4740748405456543
1	Test	0.9039615846338536	0.006002400960384154	0.07082833133253301	0.09003601440576231	0.8331332533013205	0.0001673698425292969

Πίνακας 12: Αποτελέσματα αλγορίθμου Decision Tree

1. **Acc (Accuracy)**:
    - Αντιπροσωπεύει την ακρίβεια του μοντέλου, δηλαδή το ποσοστό των σωστών προβλέψεων (TP και TN) επί του συνόλου.
    - **Train**: 92.6%
    - **Test**: 90.4% (ελαφρώς χαμηλότερη, ένδειξη ότι το μοντέλο δεν έχει overfitting).
2. **FP (False Positives)**:
    - Το ποσοστό των περιπτώσεων που το μοντέλο προέβλεψε λανθασμένα churn, ενώ δεν ισχύει.
    - **Train**: 0.96%
    - **Test**: 0.6% (μειωμένο σε σχέση με το train, δείχνοντας καλή γενίκευση).
3. **TP (True Positives)**:
    - Το ποσοστό των περιπτώσεων όπου το μοντέλο προέβλεψε σωστά churn.
    - **Train**: 7.52%
    - **Test**: 7.08% (μικρή διαφορά, καλή συνέπεια μεταξύ train και test).
4. **FN (False Negatives)**:
    - Το ποσοστό των περιπτώσεων που το μοντέλο απέτυχε να εντοπίσει churn.
    - **Train**: 6.44%
    - **Test**: 9% (ελαφρώς αυξημένο, κάτι που δείχνει ότι μερικές περιπτώσεις δεν γενικεύτηκαν σωστά).
5. **TN (True Negatives)**:
    - Το ποσοστό των περιπτώσεων όπου το μοντέλο προέβλεψε σωστά ότι δεν υπάρχει churn.
    - **Train**: 85.08%

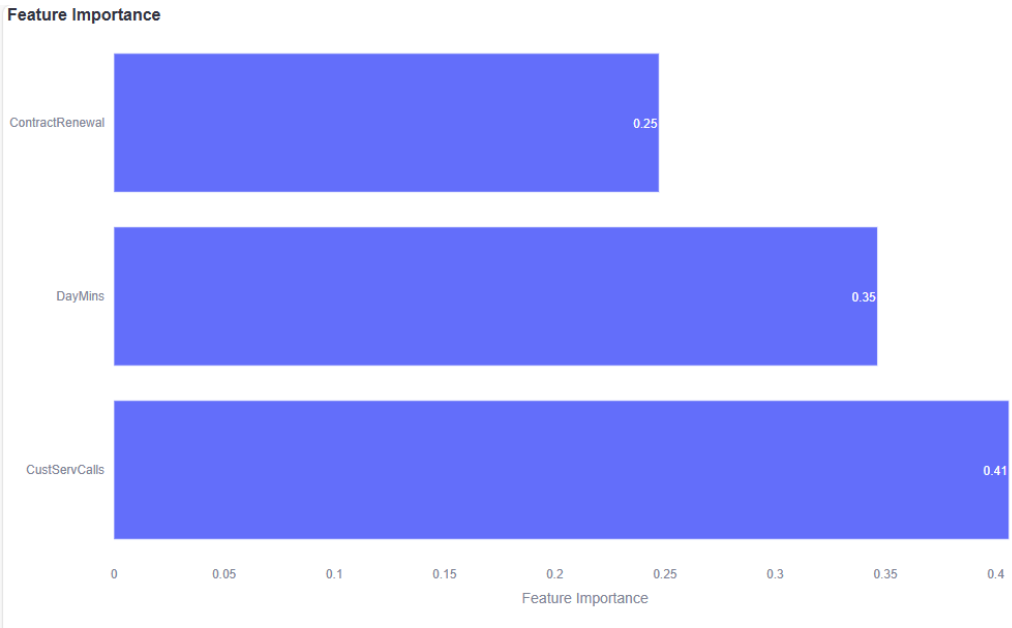
- **Test**: 83.3% (παραμένει αρκετά υψηλό).

6. **RunTimeSec**:
    - Χρόνος που απαιτείται για την εκπαίδευση ή τη δοκιμή.
    - **Train**: 0.47 δευτερόλεπτα.
    - **Test**: 0.000167 δευτερόλεπτα (εξαιρετικά γρήγορη πρόβλεψη).

Συμπεράσματα

- Το Decision Tree εμφανίζει ισορροπημένη απόδοση, με υψηλή ακρίβεια και στα δύο σύνολα δεδομένων.
- Οι χαμηλές τιμές False Positives (FP) και False Negatives (FN) υποδεικνύουν ένα μοντέλο που κάνει ελάχιστα λάθη, αλλά στο test set παρατηρείται μια μικρή αύξηση στα FN.
- Ο χρόνος εκτέλεσης είναι εξαιρετικά γρήγορος, ειδικά στο test set, καθιστώντας το μοντέλο αποδοτικό για άμεσες προβλέψεις.

Το παρακάτω διάγραμμα παρουσιάζει τη σημασία των χαρακτηριστικών όπως προέκυψε από την εκπαίδευση του **Decision Tree** μοντέλου



Εικόνα 12 : Ραβδόγραμμα σημαντικότητας χαρακτηριστικών (Feature Importance) Decision Tree

Υψηλής Σημασίας Χαρακτηριστικά

1. **CustServCalls (0.41)**
    - Το πιο σημαντικό χαρακτηριστικό, ο αριθμός των κλήσεων στην εξυπηρέτηση πελατών σχετίζεται άμεσα με την πιθανότητα αποχώρησης. Πελάτες που επικοινωνούν συχνά πιθανόν να αντιμετωπίζουν προβλήματα ή να μην είναι ικανοποιημένοι από τις υπηρεσίες.
2. **DayMins (0.35)**
    - Το δεύτερο πιο σημαντικό χαρακτηριστικό, η χρήση των λεπτών κλήσεων κατά τη διάρκεια της ημέρας είναι ένας σημαντικός παράγοντας που επηρεάζει το churn. Οι πελάτες με υψηλή χρήση ενδέχεται να είναι πιο απαιτητικοί.
3. **ContractRenewal (0.25)**
    - Το τρίτο σημαντικότερο χαρακτηριστικό, η ανανέωση συμβολαίου έχει ισχυρή επίδραση στην αποχώρηση. Οι πελάτες που δεν ανανεώνουν συμβόλαια πιθανόν να είναι πιο κοντά στο churn.

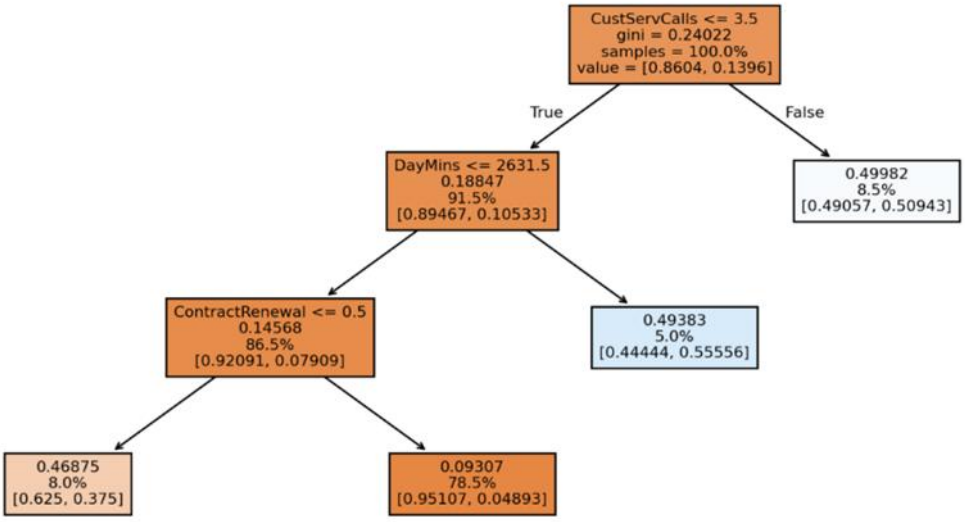
Συμπεράσματα

Το διάγραμμα υποδεικνύει ότι οι περισσότερες δράσεις θα πρέπει να επικεντρωθούν στη βελτίωση της εμπειρίας πελατών, ειδικά για αυτούς που επικοινωνούν συχνά με την εξυπηρέτηση, χρησιμοποιούν πολλά λεπτά κλήσεων κατά τη διάρκεια της ημέρας και είναι κοντά στη λήξη του συμβολαίου τους.

Η προσαρμογή προγραμμάτων και η στρατηγική επικοινωνίας με βάση αυτά τα χαρακτηριστικά μπορεί να συμβάλει σημαντικά στη μείωση του churn.

Διάγραμμα Δέντρου Αποφάσεων (Decision Tree Depth 3)

Το διάγραμμα δείχνει την ιεραρχία των διακλαδώσεων στο Decision Tree, το οποίο επιλέγει τα χαρακτηριστικά που διαχωρίζουν καλύτερα τους πελάτες με βάση την πιθανότητα churn. Το **Decision Tree** εξηγεί πώς το μοντέλο καταλήγει στις προβλέψεις, υποδεικνύοντας συγκεκριμένα μονοπάτια και υποομάδες πελατών.



Εικόνα 13 : Δέντρο Απόφασης (decision tree) που απεικονίζει τη διαδικασία πρόβλεψης

1. **CustServCalls ως Πρωτεύον Κριτήριο**
    - Το χαρακτηριστικό **CustServCalls** έχει τη μεγαλύτερη σημασία. Οι πελάτες που κάνουν περισσότερες από 3.5 κλήσεις είναι πιθανότερο να αποχωρήσουν θα χρειαστεί να αναλυθούν οι συχνές αιτίες κλήσεων και να εφαρμοστούν στοχευμένες λύσεις για την αποφυγή Churn.

2. **DayMins ως Δευτερεύον Κριτήριο**
    - Η υψηλή χρήση λεπτών (DayMins) σχετίζεται με μεγαλύτερο churn θα χρειαστεί να υπάρξει παροχή εξατομικευμένων πακέτων με βάση τη χρήση του κάθε πελάτη.
3. **ContractRenewal ως Κρίσιμη Μεταβλητή**
    - Οι πελάτες που δεν ανανεώνουν συμβόλαια έχουν δραματικά αυξημένο κίνδυνο αποχώρησης, συστήνεται η προώθηση κινήτρων για ανανέωση, όπως εκπτώσεις ή πρόσθετες υπηρεσίες.

Επόμενα Βήματα

- Στοχευμένες καμπάνιες για πελάτες με >3.5 κλήσεις και χαμηλή χρήση λεπτών.
- Προσφορές για πελάτες που δεν ανανεώνουν συμβόλαια.
- Επικοινωνία με πελάτες υψηλής χρήσης για ενίσχυση της εμπιστοσύνης και μείωση του churn.

Ανάλυση και Σύγκριση Μοντέλων Μηχανικής Μάθησης για την Πρόβλεψη Churn

XGBoost

- **Ακρίβεια (Test):** ~92%.
- **Σημασία Χαρακτηριστικών:**
    - Το σημαντικότερο χαρακτηριστικό είναι το **CustServCalls**, υποδεικνύοντας ότι οι συχνές κλήσεις στο τμήμα εξυπηρέτησης πελατών συνδέονται έντονα με την πιθανότητα αποχώρησης.
    - Εξίσου σημαντικό είναι το **ContractRenewal**, που δείχνει ότι η ανανέωση συμβολαίου μειώνει σημαντικά το churn.
- **Χρόνος Εκτέλεσης:**
    - Εξαιρετικά γρήγορος τόσο για την εκπαίδευση όσο και για την πρόβλεψη.
- **Πλεονεκτήματα:**
    - Υψηλή ακρίβεια και δυνατότητα fine-tuning για περαιτέρω βελτίωση.
    - Εξαιρετική απόδοση σε μη ισορροπημένα δεδομένα.
- **Μειονεκτήματα:**
    - Σχετικά δύσκολη ερμηνεία των αποτελεσμάτων για μη τεχνικό κοινό.

Random Forest

- **Ακρίβεια (Test):** ~92%.
- **Σημασία Χαρακτηριστικών:**
    - Το **DayMins** είναι το σημαντικότερο χαρακτηριστικό, με το **CustServCalls** και το **MonthlyCharge** να ακολουθούν, αναδεικνύοντας έναν ισορροπημένο καταμερισμό σημασίας στα χαρακτηριστικά.
- **Χρόνος Εκτέλεσης:**
    - Αργότερος από το XGBoost στην εκπαίδευση, αλλά πολύ γρήγορος στις προβλέψεις.
- **Πλεονεκτήματα:**
    - Ισορροπημένη σημασία χαρακτηριστικών.
    - Αποτελεσματικό σε δεδομένα με πολύπλοκες, μη γραμμικές σχέσεις.

- **Μειονεκτήματα:**
    - Υψηλότερος χρόνος εκπαίδευσης.
    - Ελαφρώς μικρότερη ευκολία ερμηνείας συγκριτικά με το Decision Tree.

Decision Tree

- **Ακρίβεια (Test):** ~90%.
- **Σημασία Χαρακτηριστικών:**
    - Δίνει τη μεγαλύτερη σημασία στο **CustServCalls**, ακολουθούμενο από το **DayMins** και το **ContractRenewal**, εστιάζοντας κυρίως στα κορυφαία χαρακτηριστικά.
- **Χρόνος Εκτέλεσης:**
    - Εξαιρετικά γρήγορο τόσο στην εκπαίδευση όσο και στις προβλέψεις.
- **Πλεονεκτήματα:**
    - Απλό στην κατανόηση και εύκολο στην παρουσίαση προς μη τεχνικό κοινό.
- **Μειονεκτήματα:**
    - Ευάλωτο στο overfitting χωρίς περιορισμούς βάθους.
    - Περιορισμένη δυνατότητα κατανόησης συσχετίσεων μεταξύ πολλών χαρακτηριστικών.

5.2.5 Σύγκριση Αποτελεσμάτων

Κριτήριο	**XGBoost**	**Random Forest**	**Decision Tree**
Ακρίβεια (Test)	~92%	~92%	~90%
Χρόνος Εκτέλεσης	Πολύ γρήγορος	Μέτριος	Εξαιρετικά γρήγορος
Σημασία Χαρακτηριστικών	Εξαιρετικά ακριβής	Ισορροπημένη σημασία	Εστίαση στα κορυφαία
Χρήση σε Μάρκετινγκ	Προσαρμοσμένες καμπάνιες	Ανάλυση μεσαίας σημασίας παραγόντων	Εύκολο στην παρουσίαση
Πολυπλοκότητα	Υψηλή	Μέτρια	Χαμηλή

Πίνακας 13 : Σύγκριση Αποτελεσμάτων Αλγορίθμων

Προτάσεις για το Μάρκετινγκ

Εστίαση σε Συχνές Κλήσεις στο Τμήμα Εξυπηρέτησης Πελατών:

- Θα πρέπει να βελτιώσει την εμπειρία εξυπηρέτησης πελατών για να μειώσει τη δυσαρέσκεια.
- Να αναλύσει τους λόγους των κλήσεων και να εφαρμόσει στοχευμένες λύσεις.

Παροχή Ελκυστικών Πακέτων:

- Χρειάζεται να προσφέρετε ειδικά προγράμματα με περισσότερα λεπτά ομιλίας για τους πελάτες με υψηλή χρήση (**DayMins**).
- Να δημιουργήσει κίνητρα για την ανανέωση συμβολαίων (**ContractRenewal**).

Στρατηγικές για Υψηλές Μηνιαίες Χρεώσεις:

- Να παρέχει οικονομικά πακέτα για πελάτες με υψηλό **MonthlyCharge**.
- Να ενισχύσει της εμπιστοσύνη τους με προνόμια και εκπτώσεις.

Χρήση Αλγορίθμων:

- Το **XGBoost** είναι ιδανικό για ακριβείς προβλέψεις και στοχευμένες καμπάνιες.
- Το **Random Forest** μπορεί να εντοπίσει μεσαίας σημασίας παράγοντες για ευρύτερες στρατηγικές.
- Το **Decision Tree** είναι εξαιρετικό για αρχική ανάλυση και επικοινωνία αποτελεσμάτων.

- Το **CustServCalls**, το **DayMins** και το **ContractRenewal** είναι τα κρισιμότερα χαρακτηριστικά που επηρεάζουν το churn.
- Οι ενέργειες μάρκετινγκ θα πρέπει να εστιάσουν στη βελτίωση της εξυπηρέτησης, την παροχή προσαρμοσμένων πακέτων και τη μείωση των παραπόνων.
- Με βάση τα χαρακτηριστικά κάθε αλγορίθμου, η εταιρεία μπορεί να επιλέξει την κατάλληλη προσέγγιση για να μειώσει το churn και να βελτιώσει τη δέσμευση των πελατών.

6. Συμπεράσματα

Η παρούσα διπλωματική εργασία επικεντρώθηκε στην πρόβλεψη και διαχείριση της συμπεριφοράς πελατών στην τηλεπικοινωνιακή βιομηχανία με τη χρήση τεχνικών μηχανικής μάθησης (Machine Learning) και ανάλυσης δεδομένων. Η ανάλυση στόχευσε κυρίως στη μείωση της αποχώρησης πελατών (customer churn), την αύξηση της ικανοποίησης και τη διατήρηση της πιστότητας πελατών.

Επιλέχθηκε το να στοχεύσουμε στο churn διότι είναι το στοίχημα που πρέπει να κερδίσουν οι εταιρείες τηλεπικοινωνιών για να του οδηγήσει στο μέλλον.

Καθώς το churn μειώνει το έξοδο αναζήτησης πελάτων που είναι αυξημένο καθώς για κάθε νέο πελάτη η δαπάνη είναι 10πλάσια συγκριτικά με το να διατηρήσει στο δίκτυο της έναν υπάρχων.

Κατανόηση της Σημασίας του Customer Churn

Η αποχώρηση πελατών αποτελεί έναν από τους σημαντικότερους παράγοντες που επηρεάζουν την κερδοφορία των τηλεπικοινωνιακών εταιρειών. Το φαινόμενο αυτό σχετίζεται με την ποιότητα υπηρεσιών, το κόστος διατήρησης πελατών και την αποτελεσματικότητα των στρατηγικών μάρκετινγκ. Βασικά χαρακτηριστικά που επηρεάζουν την πιθανότητα αποχώρησης περιλαμβάνουν:

- **Υψηλή ημερήσια χρήση (Total_day_minutes)**, που συνδέεται με δυσαρέσκεια λόγω κόστους.
- **Συχνές κλήσεις στο τμήμα εξυπηρέτησης πελατών (CustServCalls)**, που υποδηλώνουν πιθανή δυσαρέσκεια ή μη επίλυση προβλημάτων.
- **Ανανέωση συμβολαίων (ContractRenewal)**, όπου η μη ανανέωση αυξάνει την πιθανότητα αποχώρησης.

Εφαρμογή Τεχνικών Μηχανικής Μάθησης

Η εργασία αξιοποίησε αλγορίθμους όπως XGBoost, Random Forest και Decision Tree για την ανάλυση δεδομένων πελατών. Κάθε ο κάθε αλγόριθμος προσέφερε διαφορετικά πλεονεκτήματα:

- **XGBoost**: Υψηλή ακρίβεια (~92%) και δυνατότητα λεπτομερούς ανάλυσης χαρακτηριστικών. Κατάλληλο για προσαρμοσμένες καμπάνιες.
- **Random Forest**: Ισορροπημένη σημασία χαρακτηριστικών με υψηλή ακρίβεια (~92%). Ιδανικό για στρατηγικές ευρύτερης εφαρμογής.
- **Decision Tree**: Απλό και γρήγορο, με ακρίβεια ~90%. Χρήσιμο για οπτικοποιήσεις και αρχική ανάλυση.

Η σύγκριση των αποτελεσμάτων ανέδειξε την ανάγκη για συνδυαστική χρήση των αλγορίθμων για βέλτιστα αποτελέσματα.

Ανάλυση Δεδομένων και Εύρεση Ευρημάτων

Τα δεδομένα που χρησιμοποιήθηκαν περιλάμβαναν βασικά χαρακτηριστικά, όπως:

- **Ημερήσια διάρκεια κλήσεων (DayMins)**: Βρέθηκε ότι πελάτες με υψηλή χρήση είναι πιθανότερο να αποχωρήσουν.
- **Μηνιαίες χρεώσεις (MonthlyCharge)**: Οι υψηλές χρεώσεις συσχετίζονται με δυσαρέσκεια.
- **Χρεώσεις υπέρβασης (OverageFee)**: Οι πελάτες που υπερβαίνουν τα όρια του προγράμματός τους είναι ευάλωτοι στο churn.

Η Pearson ανάλυση συσχέτισης έδειξε ισχυρή σύνδεση μεταξύ των CustServCalls και του churn, αναδεικνύοντας την εξυπηρέτηση πελατών ως βασικό παράγοντα διατήρησης.

Προτάσεις Στρατηγικών Μάρκετινγκ

Με βάση τα ευρήματα, προτείνονται οι εξής στρατηγικές για τη μείωση του churn:

1. **Βελτίωση Εξυπηρέτησης Πελατών:**
    - Ανάλυση και επίλυση των λόγων συχνών κλήσεων στο τμήμα εξυπηρέτησης.
    - Ενίσχυση της εμπειρίας πελάτη μέσω γρήγορων λύσεων και εξειδικευμένων πακέτων.
2. **Δημιουργία Προσαρμοσμένων Πακέτων:**
    - Παροχή προγραμμάτων με περισσότερα λεπτά και μειωμένες χρεώσεις για πελάτες με υψηλή χρήση.
    - Προσφορές για πελάτες με υψηλές χρεώσεις περιαγωγής ή υπερβάσεων.
3. **Ενθάρρυνση Ανανέωσης Συμβολαίων:**
    - Παροχή κινήτρων για ανανέωση, όπως εκπτώσεις ή πρόσθετες υπηρεσίες.
4. **Εφαρμογή Loyalty Programs:**
    - Ανάπτυξη προγραμμάτων επιβράβευσης για πιστούς πελάτες.
    - Στόχευση πελατών με συχνή χρήση voicemail ή διεθνών κλήσεων.

Προκλήσεις και Περιορισμοί

Η εργασία ανέδειξε προκλήσεις στη διαχείριση δεδομένων και στην εφαρμογή τεχνικών μηχανικής μάθησης, όπως:

- **Ανισορροπία δεδομένων (data imbalance):** Η χαμηλή συχνότητα του churn σε σχέση με τους πελάτες που παραμένουν απαιτεί εξειδικευμένες μεθόδους αντιμετώπισης.
- **Διαχείριση ευαίσθητων δεδομένων:** Η χρήση προσωπικών δεδομένων απαιτεί αυστηρή συμμόρφωση με κανονισμούς, όπως ο GDPR.

7. Μελλοντικές Κατευθύνσεις

Μελλοντικές Κατευθύνσεις

Η ανάλυση πελατειακής συμπεριφοράς και η πρόβλεψη αποχώρησης πελατών (churn) προσφέρουν σημαντικές δυνατότητες βελτίωσης με τη χρήση σύγχρονων τεχνολογιών και την ενσωμάτωση νέων προσεγγίσεων. Στο πλαίσιο αυτό, η **ανάλυση συναισθημάτων (sentiment analysis)** αποτελεί μια πολλά υποσχόμενη τεχνική που μπορεί να ενισχύσει την ακρίβεια και τη χρησιμότητα των προγνωστικών συστημάτων, προσφέροντας στις εταιρείες τηλεπικοινωνιών βαθύτερη κατανόηση της εμπειρίας των πελατών τους.

Ενσωμάτωση της Sentiment Analysis

Η sentiment analysis, ως μια μέθοδος ανάλυσης της συναισθηματικής διάθεσης που εκφράζεται μέσω κειμένων ή δεδομένων επικοινωνίας, μπορεί να ενισχύσει τη στρατηγική πρόβλεψης και διαχείρισης πελατών με τους εξής τρόπους:

1. **Εξαγωγή Συναισθημάτων από Επικοινωνίες Πελατών**
    - Η ανάλυση των δεδομένων από κλήσεις εξυπηρέτησης πελατών, email, και αλληλεπιδράσεις στα κοινωνικά δίκτυα μπορεί να αποκαλύψει σημάδια δυσαρέσκειας ή ικανοποίησης. Αυτές οι πληροφορίες μπορούν να ενσωματωθούν στα προγνωστικά μοντέλα, βελτιώνοντας την ακρίβεια πρόβλεψης του churn.
2. **Προσαρμογή Στρατηγικών Επικοινωνίας**
    - Οι εταιρείες μπορούν να αναπτύξουν στοχευμένες καμπάνιες επικοινωνίας βασισμένες σε πραγματικό χρόνο συναισθηματικά δεδομένα. Για παράδειγμα, η αρνητική διάθεση που ανιχνεύεται σε σχόλια πελατών μπορεί να ενεργοποιήσει εξατομικευμένες προσφορές ή παρεμβάσεις.
3. **Αναγνώριση Προβλημάτων σε Πραγματικό Χρόνο**
    - Μέσω της sentiment analysis, οι εταιρείες μπορούν να εντοπίσουν δυσαρέσκεια προτού αυτή κλιμακωθεί σε αποχώρηση. Για παράδειγμα, οι αρνητικές αξιολογήσεις ή η αλλαγή στον τόνο των αλληλεπιδράσεων μπορεί να αποτελούν πρώιμα προειδοποιητικά σημάδια.

Τεχνολογικές Βελτιώσεις και Ενσωμάτωση AI

Η sentiment analysis εφόσον συνδυαστεί με άλλες τεχνικές μηχανικής μάθησης, όπως το XGBoost, το Random Forest ή το Decision Tree θα ενισχύσει την πρόγνωση των συστημάτων.

Με αποτέλεσμα να:

- Εντοπίζονται πελάτες που εκφράζουν θετικά συναισθήματα για στοχευμένες επιβραβεύσεις που θα φέρουν υπεραξία στην εταιρεία
- Ανίχνευση δυσαρέσκειας των πελατών για την άμεση προσαρμογή των υπηρεσιών ώστε να η εταιρεία να αποφύγει την αποχώρηση τους.
- Εξαγωγή συναισθημάτων σχετικά με ανταγωνιστικές εταιρείες από τα κοινωνικά δίκτυα και να προωθήσει στοχευμένες καμπάνιες που σκοπό θα έχουν να φέρουν πελάτες με μειωμένο κόστος αναζήτησης .

Προοπτικές και Προκλήσεις

Η ενσωμάτωση της sentiment analysis παρουσιάζει προκλήσεις, όπως η εξασφάλιση ποιοτικών δεδομένων και η διαχείριση της πολυπλοκότητας της γλώσσας, όπως ο σαρκασμός ή η ειρωνεία. Ωστόσο, η συνεχής εξέλιξη των μοντέλων NLP και των υπολογιστικών πόρων προσφέρει την ευκαιρία για υπέρβαση αυτών των εμποδίων.

Με τον συνδυασμό της sentiment analysis και τεχνικών μηχανικής μάθησης μπορεί να δοθεί μια νέα διάσταση στην πρόβλεψη του churn, καθιστώντας τα προγνωστικά συστήματα πιο ευέλικτα και αποτελεσματικά, ενώ ταυτόχρονα να ενισχύσει τη δέσμευση και την ικανοποίηση των πελατών. Η ανάλυση πελατειακής συμπεριφοράς αποτελεί δυναμικό πεδίο, με δυνατότητες βελτίωσης μέσω ανάλυση δεδομένων (**real-time)** σε πραγματικό χρόνο για δυναμική προσαρμογή στρατηγικών με την χρήση νεότερων τεχνικών, όπως το deep learning, για ακριβέστερες προβλέψεις. Δημιουργία ευφυών συστημάτων που θα συνδέουν δεδομένα πελατών με προτάσεις στρατηγικών μέσω διασύνδεσης CRM συστημάτων με AI**.**

Συνολικά η εργασία απέδειξε ότι οι τεχνικές μηχανικής μάθησης μπορούν να προσφέρουν σημαντικές λύσεις για την πρόβλεψη και μείωση του churn, ενισχύοντας τη βιωσιμότητα και την ανταγωνιστικότητα των τηλεπικοινωνιακών επιχειρήσεων. Οι προτεινόμενες στρατηγικές, σε συνδυασμό με τη συνεχή επένδυση σε δεδομένα και τεχνολογία, μπορούν να οδηγήσουν σε αυξημένη ικανοποίηση και διατήρηση πελατών, συμβάλλοντας στη μακροχρόνια επιτυχία των εταιρειών.

Βιβλιογραφία

[1] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, 785–794. https://doi.org/10.1145/2939672.2939785

[2] Ranjan, S., & Sood, S. (2020). Sentiment analysis based telecom churn prediction. *Journal of Web Engineering & Technology, 7*(1), 6–12. https://doi.org/10.5281/zenodo.341670496

[3] Lee, E.-B., Kim, J., & Lee, S.-G. (2017). Predicting customer churn in mobile industry using data mining technology. *Industrial Management & Data Systems, 117*(1), 90–109. https://doi.org/10.1108/IMDS-12-2015-0509

[4] Routray, S. K. (2021). Marketing strategy through machine learning techniques: A case study at telecom industry. *International Journal of Innovation Engineering and Science Research, 5*(3), 21–29. https://doi.org/10.5281/zenodo.4962021

[5] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://www.deeplearningbook.org/

[6] Alizadeh, H., Nazarpour Kashani, H., Jalali Filshoor, M., & Pourjabari Khameneh, A. (2023). Evaluation of consumer behavior prediction based on artificial intelligence in marketing. *Conference Paper*, November 2023. Retrieved from https://www.researchgate.net/publication/375596970

[7] Etim, G. S., Etuk, I. U., James, E. E., & Ekpe, S. (2021). Effect of relationship marketing on customer retention in the telecommunications industry. *British Journal of Management and Marketing Studies, 4*(4), 68–81. https://doi.org/10.52589/BJMMS-YIPPKDVZ

[8] Zulaikha, S., Mohamed, H., Kurniawati, M., Rusgianto, S., & Rusmita, S. A. (2021). Customer predictive analytics using artificial intelligence. *The Singapore Economic Review*. https://doi.org/10.1142/S0217590820480021

[9] Ameen, N., Tarhini, A., Reppel, A., & Anand, A. (2021). Customer experiences in the age of artificial intelligence. *Computers in Human Behavior, 114*, 106548. https://doi.org/10.1016/j.chb.2020.106548

[10] Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data, 6*(28). https://doi.org/10.1186/s40537-019-0191-6

[11] Alshamari, M. A. (2023). Evaluating user satisfaction using deep-learning-based sentiment analysis for social media data in Saudi Arabia's telecommunication sector. *Computers, 12*(170). https://doi.org/10.3390/computers12090170

[12] Kumar, A., & Singh, R. (2022). TelcoSentiment: Sentiment Analysis on Mobile Telecommunication Services. *Journal of Positive School Psychology*, 6(3), 1234-1245.

[13] Πιζάνιας, Δ. Ε. (2018). Εξόρυξη άποψης και ανάλυση συναισθήματος σε μέσα κοινωνικής δικτύωσης. *Μεταπτυχιακή Διπλωματική Εργασία, Πανεπιστήμιο Πειραιώς, Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης*.

[14]Βελισσαρίου, Α., & Πράπας, Δ. (2010). *Επιχειρηματική Ευφυΐα και Εξόρυξη Δεδομένων*. Εκδόσεις Κλειδάριθμος.

[15] Σταλίδης, Γ., & Καρδαράς, Δ. (2015). *Διαχείριση Δεδομένων και Επιχειρηματική Ευφυΐα*. ΣΕΑΒ.

[16] https://www.interviewquery.com/p/customer-churn-datasets#advanced-customer-churn-datasets-projects

[17] https://www.interviewquery.com/p/free-datasets

[18] https://www.kaggle.com/datasets/mehmetsabrikunt/internet-service-churn

[19] https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets