ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΠΛΗΡΟΦΟΡΙΚΗ

Διπλωματική Εργασία

ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΚΑΙ ΜΑΡΚΕΤΙΝΓΚ – AI DRIVEN MARKETING

του

Μπομπότα Βασίλειου

Υποβλήθηκε ως προαπαιτούμενο για την απόκτηση του Μεταπτυχιακού Διπλώματος ειδίκευσης στην Εφαρμοσμένη Πληροφορική

Επιβλέπων Καθηγήτρια
κα. Μαρία Βλαχοπούλου
mavla@uom.edu.gr

Θεσσαλονίκη, Ιανουάριος 2024

ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΚΑΙ ΜΑΡΚΕΤΙΝΓΚ – AI DRIVEN MARKETING

Μπομπότας Βασίλειος του Αντωνίου

Διπλωματική Εργασία

υποβαλλόμενη για τη μερική εκπλήρωση των απαιτήσεων του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΤΙΤΛΟΥ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΠΛΗΡΟΦΟΡΙΚΗ

Επιβλέπων Καθηγήτρια
κα. Μαρία Βλαχοπούλου

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 29/02/2024

Βλαχοπούλου Μαρία	Φούσκας Κωνσταντίνος	Κίτσιος Φώτιος
Καθηγήτρια	Αναπληρωτής Καθηγητής	Καθηγητής
…….	……	…….

Μπομπότας Βασίλειος

……

Ευχαριστίες

Με την εκπόνηση της παρούσας διπλωματικής εργασίας ολοκληρώνεται η φοίτησή μου στο ΠΜΣ στην Εφαρμοσμένη Πληροφορική.

Αρχικά, θα ήθελα να απονείμω τις ιδιαίτερες ευχαριστίες μου στην επιβλέπων καθηγήτρια αυτής της διπλωματικής εργασίας, Καθηγήτρια κα. Βλαχοπούλου Μάρω, για την ευκαιρία που μου έδωσε να εντρυφήσω σε ένα σημαντικό θέμα, το οποίο με προκαλεί ιδιαίτερο ενδιαφέρον και μου αρέσει, καθώς και για την καθοδήγηση και τη συνεχή συνδρομή της, καθ' όλη τη διάρκεια της συγγραφής.

Κλείνοντας, θα ήθελα να εκφράσω τις ιδιαίτερες ευχαριστίες μου στην οικογένειά μου για την στήριξη που με παρέχει όλα αυτά τα χρόνια των σπουδών μου.

Περίληψη

Η εφαρμογή καινοτόμων πρακτικών για την άσκηση αποτελεσματικού μάρκετινγκ κρίνεται ως προαπαιτούμενο για τις σύγχρονες επιχειρήσεις. Οι ολοένα και αυξανόμενες απαιτήσεις που επιβάλει ο έντονος ανταγωνισμός, ωθεί τις επιχειρήσεις σήμερα στην εφαρμογή καινοτόμων πρακτικών και εφαρμογών με σκοπό την απόκτηση ανταγωνιστικού πλεονεκτήματος, το οποίο θα τις ξεχωρίσει και θα τις αναδείξει από τις υπόλοιπες επιχειρήσεις. Στον σύγχρονο κόσμο, οι επιχειρήσεις οφείλουν να παίρνουν αποφάσεις,, μετατρέποντας την πληροφορία που λαμβάνουν καθημερινά σε πολύτιμη γνώση. Συνεπώς, η πληροφορία και ο μετασχηματισμός της σε γνώση, αποτελούν πλέον προτεραιότητα για την σύγχρονη επιχείρηση και το κενό αυτό έρχεται να καλύψει το πεδίο της Τεχνητή Νοημοσύνης.

Η παρούσα διπλωματική εργασία αναλύει την εφαρμογή της Τεχνητής Νοημοσύνης στον τομέα του Μάρκετινγκ. Κεντρικό σημείο εξέτασης είναι η εφαρμογή της Μηχανικής Μάθησης στην πρόβλεψη των συμπεριφορών των καταναλωτών και την προσαρμογή στοχευμένων στρατηγικών μάρκετινγκ. Με εστίαση σε πρακτικά παραδείγματα, η έρευνα εξετάζει τον τρόπο με τον οποίο η Τεχνητή Νοημοσύνη, χρησιμοποιώντας αλγόριθμους, όπως το Τυχαίο Δάσος, οι Κ-Πλησιέστεροι γείτονες, οι Μηχανές Διανυσμάτων Υποστήριξης και τα Βαθιά Νευρωνικά Δίκτυα μπορούν να συμβάλουν στην πρόβλεψη και προσαρμογή των καμπανιών μάρκετινγκ.

Στην παρούσα διπλωματική εργασία αναπτύχθηκαν μοντέλα πρόβλεψης, χρησιμοποιώντας τους παραπάνω τέσσερις αλγορίθμους μηχανικής μάθησης για την υλοποίηση του συγκεκριμένου προβλήματος ταξινόμησης. Τα αποτελέσματα μας οδήγησαν στην λήψη του μοντέλου με την μεγαλύτερη ακρίβεια πρόβλεψης**.** Συνοψίζοντας, η παρούσα έρευνα αποκαλύπτει τη σημασία της ενσωμάτωσης της Τεχνητής Νοημοσύνης στις στρατηγικές μάρκετινγκ, προκειμένου να επιτευχθεί αποτελεσματική προσέγγιση και προσαρμογή των επιχειρήσεων στις ανάγκες της σύγχρονης αγοράς για την εφαρμογή αποτελεσματικότερου και πιο στοχευμένου μάρκετινγκ.

Λέξεις Κλειδιά: Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Ψηφιακό Μάρκετινγκ, Στατιστική Ανάλυση

ABSTRACT

The implementation of innovative practices for effective marketing is considered a prerequisite for modern businesses. The escalating demands imposed by intense competition drive businesses today to adopt innovative practices and applications to gain a competitive advantage that sets them apart. In the contemporary world, businesses must make decisions, transforming the information they receive daily into valuable knowledge. Therefore, information and its transformation into knowledge are now a priority for modern enterprises, and this gap is being addressed by the field of Artificial Intelligence.

This thesis analyzes the application of Artificial Intelligence in the field of Marketing. The central point of examination is the implementation of Machine Learning in predicting consumer behaviors and adapting targeted marketing strategies. Focusing on practical examples, the research explores how Artificial Intelligence, using algorithms such as Random Forest, k-Nearest Neighbors, Support Vector Machines, and Deep Neural Networks, can contribute to the prediction and adaptation of marketing campaigns.

In this thesis, prediction models were developed using the aforementioned four machine learning algorithms for the implementation of the specific classification problem. The results led to the selection of the model with the highest prediction accuracy. In summary, this research reveals the significance of integrating Artificial Intelligence into marketing strategies to achieve an effective approach and adaptation of businesses to the needs of the modern market for the implementation of more efficient and targeted marketing

Keywords: Artificial Intelligence, Machine Learning, Digital Marketing, Statistical Analysis.

Περιεχόμενα

Κατάλογος Εικόνων

Κατάλογος Πινάκων

Εισαγωγή

1.1 Πρόβλημα

Στη σύγχρονη εποχή της τεχνολογικής εξέλιξης, η Τεχνητή Νοημοσύνη (ΤΝ) αναδύεται ως καθοριστικός παράγοντας του παγκόσμιου τοπίου της τεχνολογίας. Αυτή η επιστημονική προσέγγιση αντλεί γνώση από διάφορους τομείς, όπως οι υπολογιστές, η λογική, η βιολογία, η ψυχολογία κι η φιλοσοφία, με στόχο την ανάπτυξη ευφυών συστημάτων. Μέσω των προηγμένων υπολογιστικών τεχνικών, η Τεχνητή Νοημοσύνη επιδιώκει να αναπαράγει και να βελτιώνει τις ανθρώπινες λειτουργίες, όπως η μάθηση, η κρίση και η λήψη αποφάσεων.

Κεντρική σημασία έχει η εφαρμογή της Τεχνητής Νοημοσύνης στη μηχανική μάθηση. Η μηχανική μάθηση αποτελεί ένα πεδίο της επιστήμης των υπολογιστών που ξεχωρίζει από τις παραδοσιακές υπολογιστικές προσεγγίσεις. Στον κόσμο της παραδοσιακής επιστήμης των υπολογιστών, οι αλγόριθμοι είναι καθορισμένες σειρές εντολών που προγραμματίζονται για να εκτελέσουν συγκεκριμένες εργασίες. Αντίθετα, οι αλγόριθμοι της μηχανικής μάθησης επιτρέπουν στους υπολογιστές να εκπαιδεύονται με δεδομένα εισόδου, χρησιμοποιώντας στατιστική ανάλυση για την πρόβλεψη τιμών μέσα σε ένα συγκεκριμένο εύρος.

Η συνεχής εξέλιξη της μηχανικής μάθησης ανοίγει νέους ορίζοντες σε διάφορους τομείς, εκ των οποίων το μάρκετινγκ αναδεικνύεται ως ένας κρίσιμος τομέας εφαρμογής τους. Με την χρήση των μοντέλων της Μηχανικής Μάθησης, τα καθημερινά συλλεγόμενα δεδομένα μπορούν να μετατραπούν σε στρατηγικά εργαλεία προώθησης και εμπορικών προσφορών. Η ενσωμάτωση της Τεχνητής Νοημοσύνης στο μάρκετινγκ είναι ένα πολύτιμο εργαλείο για τις επιχειρήσεις. Με την ανάλυση δεδομένων, μπορούν να λαμβάνουν ενημερωμένες αποφάσεις βασισμένες σε προβλέψεις για τις ανάγκες των πελατών, βελτιώνοντας τη συνολική εμπειρία του χρήστη και μεγιστοποιώντας τις πιθανότητες μετατροπής σε πωλήσεις. Στην ουσία, η Τεχνητή Νοημοσύνη αναδεικνύεται ως κρίσιμο εργαλείο για την ανάπτυξη διαδικτυακών στρατηγικών, επιτρέποντας τις επιχειρήσεις να εξελίσσονται και να προσαρμόζονται τις συνεχώς μεταβαλλόμενες συνθήκες της αγοράς εργασίας.

Πλέον, οι επιχειρήσεις αναγνωρίζουν τη σημασία της Τεχνητής Νοημοσύνης στον τομέα του μάρκετινγκ και αποφασίζουν να υιοθετήσουν ψηφιακές μετασχηματιστικές πρακτικές. Ο βασικός τους στόχος είναι η αύξηση της παραγωγικότητας και η αναβάθμιση του περιεχομένου που προσφέρουν. Το Ψηφιακό Μάρκετινγκ αναγνωρίζει και εκμεταλλεύεται τις δυνατότητες που παρέχει η Τεχνητή Νοημοσύνη, επιτρέποντας τη συλλογή και ανάλυση δεδομένων, με σκοπό την αποτελεσματική ενσωμάτωσή τους στη διαδικασία λήψης αποφάσεων.

Για να επιτευχθούν οι προηγούμενες διαδικασίες στον τομέα του μάρκετινγκ, τα συστήματα τεχνητής νοημοσύνης πρέπει να υποστούν «εκπαίδευση» μέσω δεδομένων που προέρχονται από εμπορικές δραστηριότητες, όπως η ανάλυση συμπεριφοράς καταναλωτών, η εκτίμηση αποτελεσματικότητας προωθητικών ενεργειών και η προσαρμογή στρατηγικών πωλήσεων. Μέσω των πληροφοριών αυτών, τα συστήματα μπορούν να αντλήσουν γνώση από παρόμοιες εμπορικές καταστάσεις, αναδεικνύοντας συσχετίσεις μεταξύ των χαρακτηριστικών.

Ανάλογα με την κατηγορία των δεδομένων, χρησιμοποιούνται και διάφορες τεχνικές τεχνητής νοημοσύνης, συμπεριλαμβανομένης της Μηχανικής Μάθησης. Στον τομέα του μάρκετινγκ, η Μηχανική Μάθηση επικεντρώνεται κυρίως σε προβλήματα που αφορούν δομημένα δεδομένα, τις αναλύσεις δεδομένων πελατών, μετρήσεις απόδοσης εκστρατειών, κτλ. Στο πλαίσιο αυτό, ο στόχος είναι, με βάση την εκπαίδευση από υπάρχοντα δεδομένα, να προβλεφθεί, εάν ο πελάτης ενδέχεται να εκδηλώσει ενδιαφέρον για την αγορά τους προϊόντας.

Με σκοπό την καλύτερη κατανόηση του προβλήματος πήραμε ένα σύνολο δεδομένων από την πλατφόρμα (Kaggle). Εξετάσαμε τους παράγοντες που συμβάλουν στην αποτελεσματική προώθηση των προϊόντων και χρήσης του μάρκετινγκ αναπτύσσοντας μοντέλα μηχανικής μάθησης, αναπαριστώντας τα αποτελέσματα οπτικά σε διαγράμματα. Λάβαμε υπόψη τους σημαντικότερους αλγορίθμους μηχανικής μάθησης στον τομέα της κατηγοριοποίησης. Εν συνεχεία, συγκρίναμε τα μοντέλα αυτά με βάση τα αποτελέσματά τους, επιλέγοντας το μοντέλο με την μεγαλύτερη ικανότητα πρόβλεψης.

1.2 Σκοπός

Ο σκοπός της συγκεκριμένης διπλωματικής εργασίας είναι η δημιουργία και βελτιστοποίηση τεσσάρων μοντέλων μηχανικής μάθησης, χρησιμοποιώντας μια βάση δεδομένων, όπου ανακτήθηκε από την πλατφόρμα Kaggle. Ο στόχος είναι η επιλογή του βέλτιστου μοντέλου μέσω της εφαρμογής των αλγορίθμων των επιλεγμένων αξιολογητών στα αποτελέσματά τους. Για να επιτευχθούν αυτά, επιδιώκεται η εύρεση των καλύτερων παραμέτρων και ανεξάρτητων μεταβλητών, χρησιμοποιώντας την οπτικοποίηση του σε διαγράμματα.

1.3 Περιεχόμενα Μελέτης

Αρχικά, στο Κεφάλαιο 2 γίνεται λόγος στις έννοιες και όρους, οι οποίοι σχετίζονται με την Τεχνητή Νοημοσύνη και το Ψηφιακό Μάρκετινγκ. Συγκεκριμένα, αναλύονται όροι, όπως η Μηχανική Μάθηση, με σκοπό την κατανόηση του αναγνώστη των εννοιών που θα μας απασχολήσουν στο συγκεκριμένο επιστημονικό πεδίο. Περιγράφονται οι κύριο αλγόριθμοι Μηχανικής Μάθησης με τους οποίους θα αναπτυχθούν τα μοντέλα της Μηχανικής Μάθησης. Γίνεται εκτενής αναφορά στις μετρικές που θα χρησιμοποιηθούν με βάση τις οποίες θα αξιολογήσουμε τα μοντέλα, σύμφωνα με την εγκυρότητα και αποτελεσματικότητά τους. Τέλος, αναφερόμαστε στις εφαρμογές που έχει η Τεχνητή Νοημοσύνη σήμερα στο Ψηφιακό Μάρκετινγκ.

Στο Κεφάλαιο 3 γίνεται αναφορά στα βήματα που εκτελούνται για την υλοποίηση των μοντέλων. Αρχικά, γίνεται λήψη της βάσης δεδομένων με τα χαρακτηριστικά πελατών από την πλατφόρμα της Kaggle. Εν συνεχεία, προχωράμε στην προεπεξεργασία των δεδομένων, λαμβάνοντας υπόψη διάφορες παραμέτρους. Λαμβάνουμε υπόψη τις μεταβλητές με την μεγαλύτερη σημαντικότητα, εκτελώντας αναλύσεις δεδομένων και εφαρμόζοντας διάφορες τεχνικές. Μεταβαίνουμε στην στατιστική ανάλυση των δεδομένων και οπτικοποίησης τους, αναπαριστώντας τα οπτικά μέσω διαγραμμάτων.

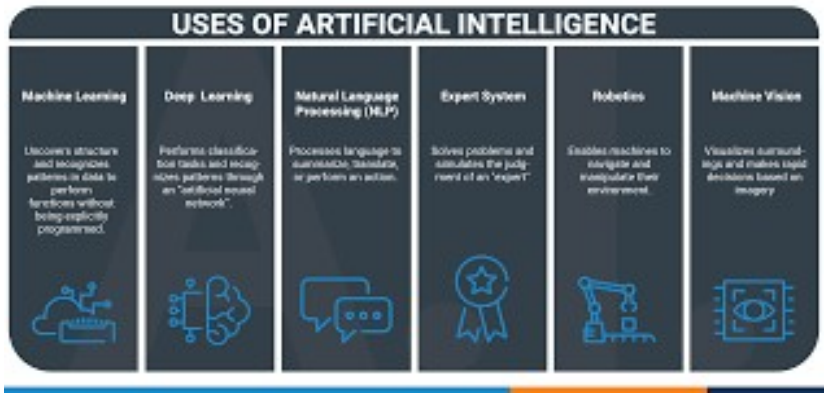
Στο Κεφάλαιο 4 αναφέρονται οι κύριοι αλγόριθμοι για προβλήματα ταξινόμησης. Συγκεκριμένα, αναπτύσσονται τέσσερις αλγόριθμοι Μηχανικής Μάθησης, οι οποίοι είναι τα Τυχαία Δάση, οι Μηχανές Διανυσμάτων Στήριξης, οι Κ-Πλησιέστεροι Γείτονες και τα Βαθιά Νευρωνικά Δίκτυα.

Στο Κεφάλαιο 5 αναφέρονται τα συμπεράσματα τα οποία προέκυψαν από την ανάπτυξη και των παραπάνω μοντέλων και επιλέγεται το μοντέλο με την μεγαλύτερη ακρίβεια.

2. Βιβλιογραφική Επισκόπηση

2.1 Τεχνητή Νοημοσύνη

Η τεχνητή νοημοσύνη (ΤΝ) πρωτοεμφανίστηκε ως μια φιλόδοξη ιδέα το 1945, με την πρόταση του Vannevar Bush, για την δημιουργία τους συστήματος που θα ενίσχυε την ανθρώπινη γνώση. Μετέπειτα, το 1950, ο Alan Turing διατύπωσε το ερώτημα της δυνατότητας προσομοίωσης του ανθρώπινου νοητικού επιπέδου από τις μηχανές. Στη συνέχεια, αναπτύχθηκαν και άλλοι κλασικοί ορισμοί, όπως του Marvin Lee Minsky, ο οποίος περιέγραψε την Τεχνητή Νοημοσύνη ως την επιστήμη, την οποία καθιστά τις μηχανές ικανές να πραγματοποιήσουν εργασίες που απαιτούν νοημοσύνη, όπως ο άνθρωπος. Τέλος, ο ορισμός του John McCarthy, το 1955, περιγράφει την Τεχνητή Νοημοσύνη ως τη δυνατότητα μιας μηχανής να εκδηλώνει ευφυή συμπεριφορά, όπως ακριβώς ο άνθρωπος.



Εικόνα 2.1: Εφαρμογές της Τεχνητής Νοημοσύνης

Η εξέλιξη της Τεχνητής Νοημοσύνης φαίνεται στη συνεχή πρόοδο των αλγορίθμων και των μοντέλων της, από τα πρώτα συστήματα ευφυούς υπολογιστικής μέχρι τις προηγμένες προσεγγίσεις βαθιάς μάθησης. Η Τεχνητή Νοημοσύνη εντάσσεται στην επιστήμη των δεδομένων, απαιτώντας μεγάλο όγκο δεδομένων. Ωστόσο, αν και εξαρτάται από τα δεδομένα, χρησιμοποιεί εξειδικευμένους αλγορίθμους για την ανάλυσή τους. Η έρευνα στον τομέα αυτό διατηρείται εδώ και πέντε δεκαετίες, με το ενδιαφέρον να εντείνεται όλο και περισσότερο τα τελευταία χρόνια. Συνεπώς, η Τεχνητή Νοημοσύνη αναδεικνύεται πλέον ως ο σημαντικότερος παράγοντα της τεχνολογικής προόδου, χρησιμοποιώντας υπολογιστές για την προσομοίωση και την εκμάθηση ανθρώπινων ευφυών συμπεριφορών.

Η τεχνητή νοημοσύνη εγκαθιδρύεται σταδιακά ως σημαντικό εργαλείο στον χώρο της εργασίας, αναλαμβάνοντας καθήκοντα, τα οποία είναι ρουτίνας καθώς κι χρονοβόρα. Αυτό συμβάλει στη βελτιστοποίηση των λειτουργιών και στην κατανομή του χρόνου

και των πόρων της επιχείρησης με πιο στρατηγικό τρόπο. Η τεχνητή νοημοσύνη μπορεί να διεκπεραιώνει εργασίες, όπως ανάλυση δεδομένων, εξυπηρέτηση πελατών ή διαχείριση διοικητικών καθηκόντων, επιτρέποντας τους ανθρώπους να αφοσιωθούν σε αποφάσεις υψηλότερου επιπέδου, στην επίλυση προβλημάτων και στην καινοτομία, χαρακτηριστικά που οι μηχανές δεν διαθέτουν ακόμα.

Οι κύριες εφαρμογές της Τεχνητής Νοημοσύνης συγκεντρώνονται στη μηχανική μάθηση, όπου στατιστικοί αλγόριθμοι προσομοιάζουν ανθρώπινες γνωστικές λειτουργίες. Σύμφωνα με τις τελευταίες έρευνες, προβλέπεται ότι οι εξελίξεις στον τομέα θα συμβάλουν σημαντικά σε οικονομικά και κοινωνικά οφέλη, επηρεάζοντας παράγοντες, όπως την υγεία, την εκπαίδευση, τη γεωργία και την έρευνα. Παρά τα σημαντικά της οφέλη, η ανάπτυξη της Τεχνητής Νοημοσύνης συνοδεύεται και από προκλήσεις, όπως η αλγοριθμική διάκριση, η έλλειψη διαφάνειας και η ανάγκη για ηθική αντιμετώπιση των προβλημάτων. Συνεπώς, απαιτείται σοβαρή σκέψη και δράση προκειμένου να διασφαλιστούν τα οφέλη και να αντιμετωπιστούν αποτελεσματικά οι προκλήσεις.

Γενικότερα, στον ευρύ κόσμο της Τεχνητής Νοημοσύνης αναδεικνύονται έξι σημαντικοί υποκλάδοι, καθένας από τους οποίους, αποτελεί ξεχωριστό επιστημονικό πεδίο:

1. **Μηχανική Μάθηση.** Η Μηχανική Μάθηση αναπτύσσει την τεχνική που επιτρέπει τους υπολογιστές να μαθαίνουν χωρίς προγραμματισμό. Χρησιμοποιεί αλγόριθμους για να αναγνωρίζει προτεραιότητες και προτείνει βελτιστοποιημένες λύσεις.
2. **Νευρωνικά Δίκτυα**. Τα Νευρωνικά Δίκτυα αναφέρονται σε ένα σύνολο αλγορίθμων που μιμούνται τη διαδικασία λειτουργίας του ανθρώπινου εγκεφάλου.
3. **Ανάλυση Δεδομένων**. Αναλύει τα δεδομένα με τη χρήση προηγμένων αλγορίθμων, επιτρέποντας την ανακάλυψη προτύπων, την εξαγωγή συμπερασμάτων και τη λήψη αποφάσεων βασισμένων σε δεδομένα.
4. **Ρομποτική**. Στο πλαίσιο αυτό, διερευνάται ο σχεδιασμός, η παραγωγή, η λειτουργία και η χρήση ρομπότ. Η τεχνολογία αυτή αναζητεί τρόπους βελτίωσης της αυτονομίας και της απόδοσης ρομπότ, ενσωματώνοντας πτυχές της Τεχνητής Νοημοσύνης για να επιτύχει εξελιγμένες λειτουργίες.
5. **Συστήματα Εμπειρογνωμόνων**. Αναφέρονται σε συστήματα υπολογιστή που μιμούνται τη διαδικασία λήψης αποφάσεων του ανθρώπινου εμπειρογνώμονα.
6. **Ασαφής Λογική**. Αντιπροσωπεύει μια τεχνική στην οποία αναπαρίστανται και τροποποιούνται αβέβαιες πληροφορίες, μετρώντας το βαθμό στον οποίο μια υπόθεση είναι σωστή.

Εικόνα 2.2: Κύριες υποκατηγορίες της Τεχνητής Νοημοσύνης

2.2 Μηχανική Μάθηση

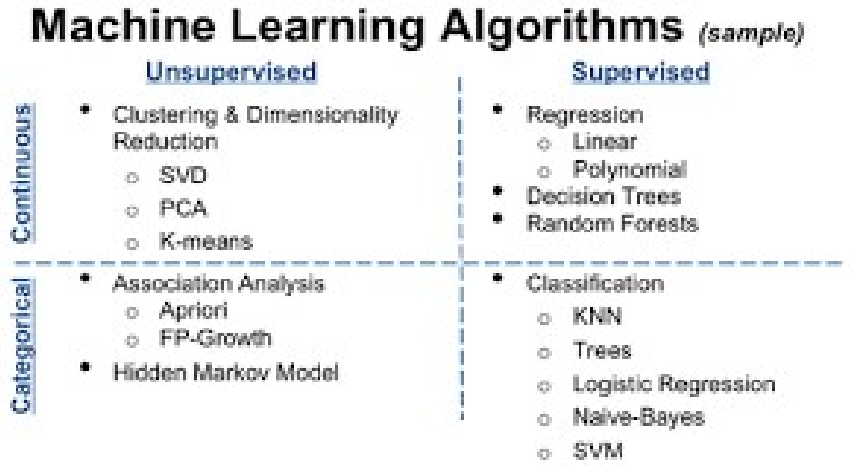
Η Μηχανική Μάθηση (ΜΜ), ως ο πιο κοινός υποτομέας της Τεχνητής Νοημοσύνης αποτελεί ένα υπό πεδίο της επιστήμης των υπολογιστών που στοχεύει κυρίως στο να επιτρέπει τους υπολογιστές να «μαθαίνουν» χωρίς να χρειάζεται να προγραμματιστούν απευθείας, ανοίγοντας νέους ορίζοντες στη διαδικασία εκμάθησης των αλγορίθμων. Οι υπολογιστές «μαθαίνουν» βελτιώνοντας την απόδοσή τους σε εργασίες μέσα από τη «εμπειρία», που συνήθως σημαίνει εκπαίδευση σε δεδομένα. Επικεντρώνεται κυρίως σε πρακτικούς στόχους και εφαρμογές, ιδιαίτερα στην πρόβλεψη και τη βελτιστοποίηση. Εφαρμόζοντας αλγόριθμους βασισμένους σε στατιστικά, η μηχανική μάθηση σχεδιάζει να «μάθει» από μεγάλες βάσεις δεδομένων, ανακατασκευάζοντας τις ανθρώπινες γνωστικές ικανότητες.

Η μηχανική μάθηση επομένως αποτελεί ένα εξαιρετικά δυναμικό πεδίο που επιτρέπει την εξέλιξη των αλγορίθμων μέσω επαναλαμβανόμενης εκπαίδευσης. Το ανακατασκευαστικό χαρακτηριστικό της διαδικασίας σημαίνει ότι τα αποτελέσματα βελτιώνονται συνεχώς, σύμφωνα με τον όγκο της εκπαίδευσης και την εμπειρία του αλγορίθμου. Έτσι, η μηχανική μάθηση ανοίγει νέους δρόμους για την ανάπτυξη τεχνολογιών που επιλύουν προβλήματα σε ποικίλες καταστάσεις, αποτελώντας θεμέλιο λίθο για την εξέλιξη της τεχνητής νοημοσύνης.

Η Μηχανική Μάθηση συνδέεται στενά με την υπολογιστική στατιστική. Έχει στενούς δεσμούς με τη μαθηματική βελτιστοποίηση, παρέχοντας μεθόδους, θεωρία και εφαρμογές. Εφαρμόζεται σε υπολογιστικές εργασίες, όπου ο σχεδιασμός και ο ρητός προγραμματισμός είναι δύσκολος. Ενίοτε μπορεί να συγχέεται με την εξόρυξη δεδομένων, αλλά αποσκοπεί κυρίως στην εξερεύνηση και ανάλυση δεδομένων. Στον

τομέα της ανάλυσης δεδομένων, χρησιμοποιείται για τη δημιουργία πολύπλοκων μοντέλων και αλγορίθμων που οδηγούν σε προβλέψεις (Tan et.al., 2016).

Η ποιότητα των δεδομένων αποτελεί καίριο παράγοντα στην απόδοση των μοντέλων Μηχανικής Μάθησης. Ανάλογα με το είδος του προβλήματος, η Μηχανική Μάθηση επιλέγει την κατηγορία μάθησης που ταιριάζει καλύτερα. Κατηγοριοποιείται σε τρεις βασικές κατηγορίες, ανάλογα με τον τύπο ανάδρασης, που αποτελεί κρίσιμο παράγοντα. Αυτές οι κατηγορίες είναι η επιβλεπόμενη μάθηση, η μη-επιβλεπόμενη μάθηση, η ημί-επιβεπόμενη μάθηση και η ενισχυτική μάθηση.



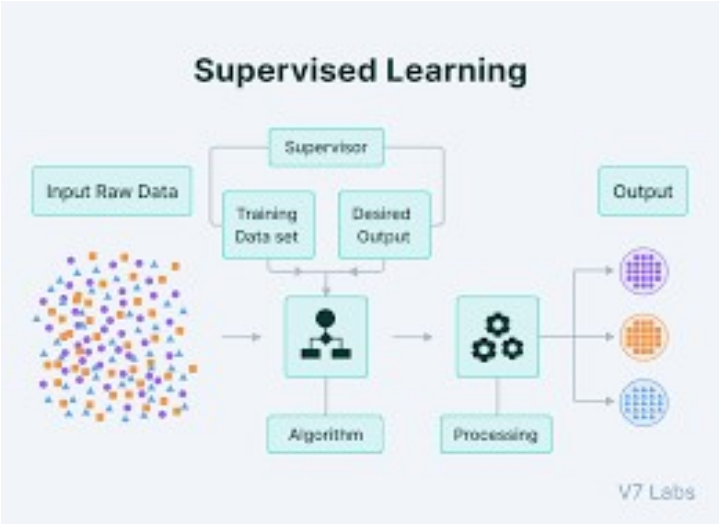
Εικόνα 2.3: Πίνακας κατηγοριών αλγορίθμων Μηχανικής Μάθησης

2.2.1 Επιβλεπόμενη Μάθηση

Η επιβλεπόμενη μάθηση αντιπροσωπεύει μια σημαντική κατηγορία τεχνικών Μηχανικής Μάθησης, όπου ο υπολογιστής αναπτύσσει γενικούς κανόνες ή «μοντέλα» από δεδομένα εκπαίδευσης. Κατά τη διάρκεια αυτής της διαδικασίας, το μοντέλο μαθαίνει να αντιστοιχεί εισόδους σε συγκεκριμένες εξόδους, προκειμένου να είναι σε θέση να προβλέπει τις εξόδους για νέα άγνωστα δεδομένα. Η πρόβλεψη είναι το κύριο ενδιαφέρον της επιβλεπόμενης μάθησης.

Στην επιβλεπόμενη μάθηση, χρησιμοποιούνται ετικετοποιημένα δεδομένα για την εκπαίδευση του μοντέλου, με σκοπό την πρόβλεψη του τύπου ή της τιμής νέων δεδομένων. Τα κύρια καθήκοντα, τα οποία αναλαμβάνει η επιβλεπόμενη μάθηση περιλαμβάνουν τη «κατηγοριοποίηση» και τη «παλινδρόμηση». Συνήθως, οι ερευνητές ενδιαφέρονται λιγότερο για το να ανακαλύψουν την «αληθινή» σύνδεση μεταξύ των μεταβλητών, παρά για το να μάθουν μια συνάρτηση που μεγιστοποιεί την ακρίβεια της πρόβλεψης της εξόδου χρησιμοποιώντας την είσοδο.

Η προβλεπτική ακρίβεια πρέπει να αξιολογηθεί χρησιμοποιώντας ένα διαφορετικό σύνολο ελέγχου, καθώς χωρίς περιορισμούς στο μοντέλο, μπορεί να επιτευχθεί τέλεια ακρίβεια για το σύνολο εκπαίδευσης μέσω απομνημόνευσης. Οι ερευνητές συνήθως χωρίζουν περαιτέρω το σύνολο εκπαίδευσης σε ένα υποσύνολο εκπαίδευσης και ένα υποσύνολο επικύρωσης. Τα μοντέλα θα εκπαιδευτούν χρησιμοποιώντας το υποσύνολο εκπαίδευσης και θα ρυθμιστούν ή θα επιλεγούν χρησιμοποιώντας το υποσύνολο επικύρωσης. Το τελικό επιλεγμένο μοντέλο θα αξιολογηθεί στο σύνολο ελέγχου για να αξιολογηθεί η εκτός δείγματος απόδοσή του. (π.χ. Hartmann, Heitmann, Schamp, & Netzer, 2019).



Εικόνα 2.4: Λειτουργία επιβλεπόμενης μάθησης

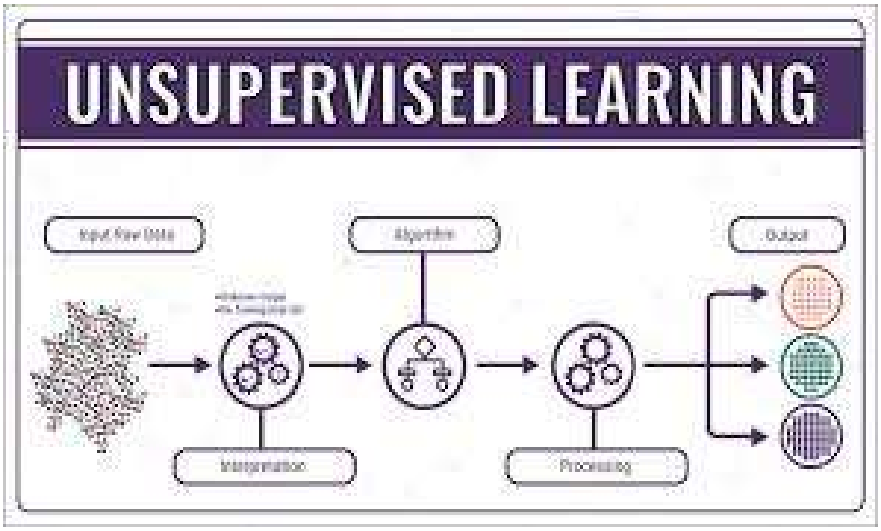
2.2.2 Μη Επιβλεπόμενη Μάθηση

Στις εργασίες της μη επιβλεπόμενης μάθησης, το σύνολο εκπαίδευσης περιλαμβάνει μόνο τις μεταβλητές εισόδου, ενώ οι μεταβλητές εξόδου είναι είτε απροσδιόριστες είτε άγνωστες. Ο στόχος είναι να βρεθούν κρυφά πρότυπα ή να εξαχθεί πληροφορία από τα δεδομένα.

Σε μια εργασία μείωσης διαστάσεων, τα δεδομένα υψηλής διάστασης μετασχηματίζονται σε χαμηλότερες διαστάσεις διατηρώντας την πληροφορία από τα αρχικά δεδομένα. Τα χαρακτηριστικά εξάγονται από τα δεδομένα εισόδου. Τα εξαγόμενα χαρακτηριστικά μεταφέρουν τη βασική πληροφορία των αρχικών δεδομένων και μπορούν να ερμηνευθούν ή να χρησιμοποιηθούν ως είσοδος για επόμενη ανάλυση. Αυτή η διαδικασία είναι καθοδηγούμενη από τα ίδια τα δεδομένα, επιτρέποντας στο σύστημα να ανακαλύψει μοτίβα, δομές και χαρακτηριστικά χωρίς προκαθορισμένες κατευθύνσεις. Συνεπώς, ένα από τα κύρια πλεονεκτήματα αυτής

της προσέγγισης είναι η ικανότητα ανακάλυψης κρυμμένων μοτίβων και δομών στα δεδομένα χωρίς προκαθορισμένες πληροφορίες.

Στο πλαίσιο της μη επιβλεπόμενης μάθησης, εφαρμόζονται διάφορες εργασίες που περιλαμβάνουν τη συσταδοποίηση, την εκτίμηση πυκνότητας, την εκμάθηση χαρακτηριστικών, την μείωση της διαστασιμότητας, την εύρεση κανόνων συσχέτισης, τον εντοπισμό ανωμαλιών. Η μη επιβλεπόμενη μάθηση προσφέρει ένα ευρύ φάσμα εφαρμογών και αποτελεί ένα ισχυρό εργαλείο για την εξαγωγή γνώσης από πολύπλοκα και μη δομημένα σύνολα δεδομένων.



Εικόνα 2.5: Λειτουργία μη επιβλεπόμενης μάθησης

2.2.3 Ημι-επιβλεπόμενη Μάθηση

Η γραμμή μεταξύ επιβλεπόμενης και μη επιβλεπόμενης μάθησης μπορεί να είναι ασαφής, οδηγώντας σε ημι-επιβλεπόμενες εργασίες μάθησης. Σε μια εργασία ημι-επιβλεπόμενης μάθησης, η έξοδος γνωρίζεται μόνο για ένα υποσύνολο των δεδομένων (Zhu, 2005).

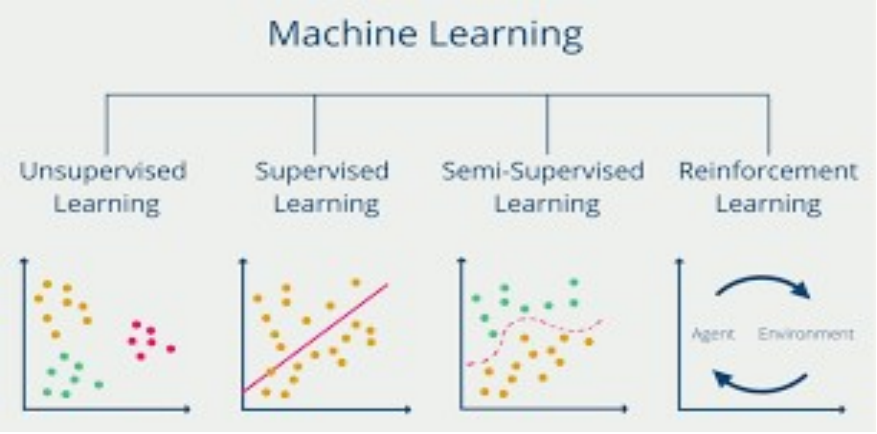
Στον χώρο της μηχανικής μάθησης, όπου η εκπαίδευση μοντέλων εξαρτάται σε μεγάλο βαθμό από την ποιότητα και την ποσότητα των δεδομένων, η ημι-επιβλεπόμενη μάθηση αναδεικνύεται ως ένα καίριο εργαλείο που επιτρέπει την αποδοτική χρήση μη επισημασμένων δεδομένων. Στην πραγματικότητα, τα ετικετοποιημένα και τα μη ετικετοποιημένα δεδομένα αναμιγνύονται στη διαδικασία μάθησης, με την τελευταία να αποτελεί συχνά την κύρια πηγή πληροφοριών λόγω της υπεροχής σε όγκο.

Υπό κανονικές συνθήκες, η ποσότητα των μη επισημασμένων δεδομένων υπερτερεί σημαντικά της ποσότητας των επισημασμένων δεδομένων, καθιστώντας την ιδέα της ημι-επιβλεπόμενης μάθησης ιδανική. Παρά την ιδανικότητά της, ωστόσο, δεν χρησιμοποιείται ευρέως σε πρακτικές εφαρμογές, με τους αλγορίθμους της να αποτελούν την εξαίρεση παρά τον κανόνα.

2.2.4.Ενισχυτική Μάθηση

Η ενισχυτική μάθηση αντιπροσωπεύει έναν σημαντικό τύπο αλγορίθμου μηχανικής μάθησης, που επιτρέπει σε λογισμικό πράκτορες και μηχανές να αξιολογούν αυτόματα τη βέλτιστη συμπεριφορά σε ένα συγκεκριμένο πλαίσιο ή περιβάλλον. Σε αυτόν τον τύπο μάθησης, η προσέγγιση καθορίζεται από το περιβάλλον, και ο στόχος είναι να ληφθούν μέτρα που θα αυξήσουν την ανταμοιβή ή θα ελαχιστοποιήσουν τον κίνδυνο με βάση τις πληροφορίες που προκύπτουν από την αλληλεπίδραση με το περιβάλλον. Αυτές οι εργασίες διατυπώνονται συχνά ως διαδικασία λήψης αποφάσεων Markov (MDP). Ο αλγόριθμος μάθησης πρέπει να καθορίσει τα μέτρα που θα λάβει τόσο για να μάθει τα χαρακτηριστικά του περιβάλλοντος όσο και για να διαμορφώσει την βέλτιστη πολιτική δράσης δεδομένων των καταστάσεων.

Το AlphaZero και το AlphaGo είναι παραδείγματα προηγμένων μοντέλων που χρησιμοποιούνται για την εκπαίδευση και τη λήψη αποφάσεων. Η Ενισχυτική μάθηση εκπροσωπεί έναν καθοριστικό πυλώνα στον χώρο της τεχνητής νοημοσύνης, προσφέροντας προηγμένες λύσεις σε προβλήματα λήψης αποφάσεων. Η Ενισχυτική μάθηση παρέχει ευελιξία και αποτελεσματική ανταπόκριση σε μεταβαλλόμενα περιβάλλοντα. Ωστόσο, η διαχείριση του trade-off μεταξύ εξερεύνησης και εκμάθησης αποτελεί σημαντική πρόκληση. Αυτή η προσέγγιση έχει ευρεία εφαρμογή σε πολλούς τομείς, όπως η ρομποτική, οι εργασίες αυτόνομης οδήγησης, η κατασκευή και διαχείριση της εφοδιαστικής αλυσίδας, κ.α.



Εικόνα 2.6: Κατηγορίες Μηχανικής Μάθησης

2.3 Προβλήματα Ενασχόλησης Μηχανικής Μάθησης

Η μηχανική μάθηση αντιμετωπίζει τέσσερα κεντρικά προβλήματα: την κατηγοριοποίηση, την παλινδρόμηση, τη συσταδοποίηση και την μείωση της διαστασιμότητας.

2.3.1 Πρόβλημα Κατηγοριοποίησης

Η κατηγοριοποίηση αποτελεί μια ευρέως χρησιμοποιούμενη μέθοδο επιβλεπόμενης μάθησης στον χώρο της μηχανικής μάθησης. Στην κατηγοριοποίηση, τα δεδομένα εκπαίδευσης περιλαμβάνουν παρατηρήσεις με αντίστοιχες κατηγορικές τιμές εξόδου. Ο στόχος είναι να εκπαιδευτεί το μοντέλο ώστε να κατηγοριοποιεί νέες παρατηρήσεις σε συγκεκριμένες κατηγορίες. Μαθηματικά, αναπαρίσταται μέσω μιας συνάρτησης (f), η οποία αντιστοιχεί τις μεταβλητές εισόδου (X) σε μεταβλητές εξόδου (Y) ως στόχο, ετικέτα ή κατηγορίες.

Η κατηγοριοποίηση εφαρμόζεται σε ποικίλα προβλήματα, ανεξαρτήτως εάν τα δεδομένα είναι δομημένα ή μη δομημένα. Για παράδειγμα, στο πρόβλημα ανίχνευσης spam σε υπηρεσίες ηλεκτρονικού ταχυδρομείου, όπου η εργασία είναι να διακρίνουμε μεταξύ «spam» και «non spam,» το πρόβλημα αυτό αποτελεί ένα κλασικό παράδειγμα κατηγοριοποίησης.

Στη συνέχεια, περιγράφονται συνοπτικά κάποια κοινά προβλήματα κατηγοριοποίησης. Αυτά τα προβλήματα περιλαμβάνουν διάφορες περιπτώσεις εφαρμογής της κατηγοριοποίησης, επισημαίνοντας την πολυπλοκότητα και την ευρεία εφαρμοστικότητα της τεχνικής στην επίλυση πρακτικών προβλημάτων.

Δυαδική Κατηγοριοποίηση. Η δυαδική κατηγοριοποίηση αναφέρεται σε καθήκοντα κατηγοριοποίησης με δύο ετικέτες κλάσης, της «αληθές και ψευδές» ή «ναι και όχι». Σε αυτά τα καθήκοντα, ένα σύνολο δεδομένων χωρίζεται σε δύο κατηγορίες, με τη μία να αντιπροσωπεύει την κανονική κατάσταση και την άλλη μια πιθανή ανωμαλία. Ένα παράδειγμα αποτελεί ο διαχωρισμός μεταξύ «spam» και «not spam» στα email.

Πολυκατηγορική Κατηγοριοποίηση. Η Κατηγοριοποίηση αυτή αναφέρεται σε εργασίες κατηγοριοποίησης που έχουν περισσότερες από δύο ετικέτες κλάσης. Σε αντίθεση με τη δυαδική κατηγοριοποίηση, όπου κάθε παράδειγμα ανήκει σε μία από τις δύο κατηγορίες, στην πολυκατηγορική κατηγοριοποίηση, τα παραδείγματα μπορούν να ανήκουν σε μία από τις πιθανές κατηγορίες. Για παράδειγμα, σε ένα σύνολο δεδομένων επιθέσεων δικτύου, η κατηγοριοποίηση μπορεί να περιλαμβάνει διάφορες ετικέτες, όπως είδη επιθέσεων (DoS, U2R, R2L, Probing). Κάθε παράδειγμα ταξινομείται σε μία από αυτές τις πολλαπλές κατηγορίες.

Πολυετική Κατηγοριοποίηση. Στην πολυετική κατηγοριοποίηση, ένα παράδειγμα συσχετίζεται με διάφορες κλάσεις ή ετικέτες. Αυτή η προσέγγιση είναι μια γενίκευση της πολυκατηγορικής κατηγοριοποίησης, όπου οι κατηγορίες έχουν ιεραρχική δομή και κάθε παράδειγμα μπορεί να ανήκει ταυτόχρονα σε περισσότερες από μία κατηγορία σε κάθε επίπεδο. Για παράδειγμα, στην πολυεπίπεδη ταξινόμηση κειμένου, μια είδηση μπορεί να αναταχθεί υπό «όνομα πόλης», «τεχνολογία», ή «τελευταία νέα» ταυτόχρονα.

2.3.2 Πρόβλημα Παλινδρόμησης

Σε αντίθεση με το πρόβλημα της κατηγοριοποίησης, στο πρόβλημα της παλινδρόμησης, το μοντέλο προβλέπει την έξοδο ως συνεχείς τιμές, δηλαδή οι έξοδοι είναι αριθμητικές τιμές. Το μοντέλο εκπαιδεύεται για να παράγει ένα συνεχές φάσμα τιμών, επιτρέποντας την πρόβλεψη αριθμητικών εξόδων για νέες παρατηρήσεις. Για παράδειγμα, ένα μοντέλο παλινδρόμησης μπορεί να προβλέψει τη θερμοκρασία μιας πόλης βάσει διάφορων χαρακτηριστικών, όπως η ώρα της ημέρας, η εποχή, και άλλοι παράγοντες.

Παρά τις διακριτικές διαφορές, παρατηρούνται επικαλύψεις μεταξύ των δύο τύπων αλγορίθμων μηχανικής μάθησης. Τα μοντέλα παλινδρόμησης, όπως η γραμμική, πολυωνυμική, Lasso και Ridge παλινδρόμηση, έχουν εφαρμογές σε πολλούς τομείς, τις οικονομικές προβλέψεις, την εκτίμηση κόστους, την ανάλυση τάσης, μάρκετινγκ, την εκτίμηση χρονοσειρών, κα. Για να κατανοήσουμε καλύτερα την παλινδρόμηση, εξετάσουμε σύντομα κάποιους γνωστούς τύπους αλγορίθμων παλινδρόμησης:

Γραμμική Παλινδρόμηση. Χρησιμοποιείται για να εξετάσει τη γραμμική σχέση μεταξύ μιας εξαρτώμενης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Η γραμμική παλινδρόμηση αποτελεί μια από της πιο δημοφιλείς τεχνικές μοντελοποίησης στον χώρο της μηχανικής μάθησης. Στο πλαίσιο αυτής της τεχνικής, η εξαρτώμενη μεταβλητή είναι συνεχής, ενώ οι ανεξάρτητες μεταβλητές μπορεί να είναι είτε συνεχείς είτε διακριτικές. Η γραμμική παλινδρόμηση δημιουργεί μια γραμμική σχέση μεταξύ της εξαρτώμενης μεταβλητής y και μιας ή περισσότερων ανεξάρτητων μεταβλητών χ. Η γραμμική παλινδρόμηση καθορίζεται από την εξίσωση:

$$Y=a+bX+e$$

Όπου α είναι η τομή, b η κλίση τους γραμμής και e αποτελεί τον όρο του σφάλματος.

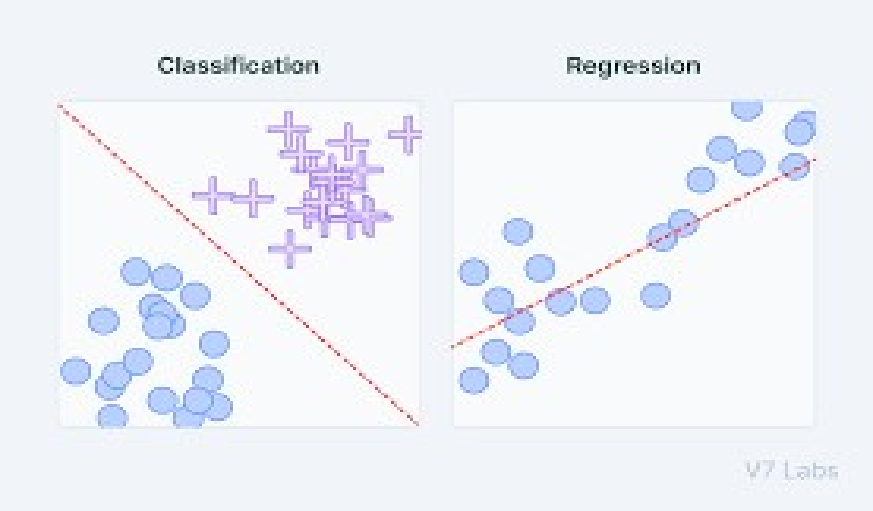
Πολυωνυμική Παλινδρόμηση. Διερευνά πολυωνυμικές σχέσεις μεταξύ των μεταβλητών. Η πολυωνυμική παλινδρόμηση αποτελεί μια σημαντική μορφή ανάλυσης παλινδρόμησης στη μηχανική μάθηση, διακρινόμενη από το γεγονός ότι η

σχέση μεταξύ της ανεξάρτητης μεταβλητής x και της εξαρτώμενης μεταβλητής y δεν είναι γραμμική, αλλά πολυωνυμική. Με άλλα λόγια, εάν η σχέση μεταξύ των μεταβλητών δεν περιγράφεται από μια απλή ευθεία γραμμή, αλλά έχει πολυωνυμικό χαρακτήρα, τότε η πολυωνυμική παλινδρόμηση είναι η κατάλληλη μέθοδος για την μοντελοποίηση τους. Η γενική εξίσωση για την πολυωνυμική παλινδρόμηση προκύπτει από την απλή γραμμική παλινδρόμηση (περίπτωση πολυωνυμικής παλινδρόμησης βαθμού 1) και δίνεται από τη σχέση:

$y=b0+b1x+b2x2+…+bnxn+e$

Οι εφαρμογές της πολυωνυμικής παλινδρόμησης είναι ευρείες, καθώς μπορεί να προσαρμοστεί σε ποικίλα σενάρια. Συχνά χρησιμοποιείται όταν η σχέση μεταξύ των μεταβλητών είναι περισσότερο πολύπλοκη από μια απλή ευθεία γραμμή, και στις περιπτώσεις όπου παρατηρούνται καμπυλότητες ή κορυφές στα δεδομένα.

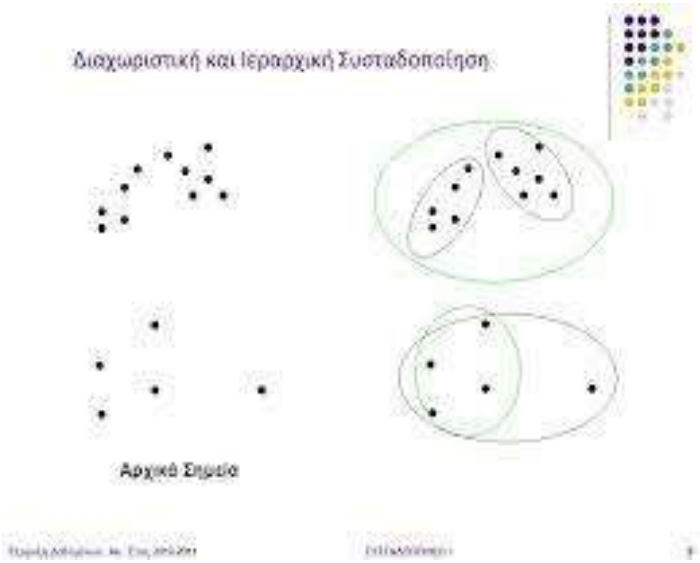
Lasso Ridge. Η παλινδρόμηση LASSO (Least Absolute Shrinkage and Selection Operator) και η παλινδρόμηση Ridge ανήκουν στην κατηγορία των ισχυρών τεχνικών μοντελοποίησης που εφαρμόζονται σε περιπτώσεις μεγάλου αριθμού χαρακτηριστικών. Είναι δυνατόν να χρησιμοποιηθούν για την αποφυγή της υπερεκπαίδευσης και την μείωση της πολυπλοκότητας του μοντέλου, καθιστώντας τις ιδιαίτερα χρήσιμες σε προβλήματα μεγάλης κλίμακας.



Εικόνα 2.7: Κατηγοριοποίηση vs Παλινδρόμηση

2.3.3 Πρόβλημα Συσταδοποίησης

Η συσταδοποίηση είναι μια διαδικασία που έχει ως στόχο την ομαδοποίηση των δειγμάτων σε ομάδες, γνωστές ως συστάδες, με βάση την ομοιότητά τους. Οι συσταδοποιητές, όπως ο αλγόριθμος k-means, επιτρέπουν τον εντοπισμό ομοιογενών ομάδων, ενισχύοντας την κατανόηση των δομικών πτυχών των δεδομένων. Αναδύεται ως ισχυρή τεχνική μη εποπτευόμενης μάθησης. Στόχος της είναι ο εντοπισμός και η ομαδοποίηση σχετικών σημείων δεδομένων σε μεγάλα σύνολα, χωρίς να περιορίζεται από προκαθορισμένα αποτελέσματα.



Εικόνα 2.8: Αλγόριθμος Κ-Πλησιέστεροι Γείτονες

Η εφαρμογή μεθόδων συσταδοποίησης βασίζεται στην ιδέα ότι τα δείγματα μπορούν να ομαδοποιηθούν μαζί βάσει της ομοιότητας τους, ενισχύοντας έτσι την κατανόηση των ενδεχόμενων δομών στα δεδομένα. Βασικός στόχος της συσταδοποίησης είναι να ανακαλύψει φυσικά πρότυπα και ομάδες παρόμοιων δειγμάτων, χωρίς να υποθέτει προκαταβολικά τις ιδιότητες που μπορεί να έχουν αυτές οι ομάδες. Η συσταδοποίηση είναι καίρια για την οργάνωση και την εξαγωγή πληροφοριών από πολύπλοκα σύνολα δεδομένων. Αναδεικνύεται ως ισχυρή εργαλειοθήκη για την απεικόνιση μοτίβων και τάσεων σε δεδομένα, αποσαφηνίζοντας δομές και επιτρέποντας την καλύτερη κατανόηση του περιεχομένου τους. Η συσταδοποίηση χρησιμοποιείται εκτενώς σε ποικίλα πεδία εφαρμογής. Στην κυβερνοασφάλεια, το ηλεκτρονικό εμπόριο, την επεξεργασία κινητών δεδομένων, την ανάλυση υγείας, το μοντέλο χρήστη και την ανάλυση συμπεριφοράς.

Μια απλή και ευρέως χρησιμοποιούμενη προσέγγιση είναι ο αλγόριθμος Συσταδοποίησης k-means. Σε αυτόν τον αλγόριθμο, ο αριθμός των συστάδων προκαθορίζεται από τον χρήστη και κάθε σύμπλεγμα αντιπροσωπεύεται από ένα

κέντρο συμπλέγματος. Η διαδικασία περιλαμβάνει την επαναληπτική ανάθεση των σημείων δεδομένων σε συμπλέγματα βάσει της ομοιότητάς τους με τα κέντρα, και την ενημέρωση της θέσης των κέντρων με βάση τη σύνθεση των ανατιθέμενων σημείων.

Παρά την αποτελεσματικότητα της Συσταδοποίησης, υπάρχουν προκλήσεις, όπως η επιλογή του κατάλληλου αριθμού συστάδων, που μπορεί να επηρεάσει σημαντικά τα αποτελέσματα. Επισημαίνεται ότι σε συνθήκες υψηλής διάστασης χαρακτηριστικών, η επιλογή αυτή γίνεται ακόμα πιο πολύπλοκη.

2.3.4 Μείωση της Διαστασιμότητας

Η διαδικασία αυτή γνωστή ως «Μείωση της Διαστατικότητας» είναι ζωτικής σημασίας. Καθώς αυξάνεται η διαστασιμότητα είναι απαραίτητο να μειώσουμε τις διαστάσεις του χώρου χαρακτηριστικών. Αυτό επιτυγχάνεται με την εξάλειψη περιττών ή συσχετισμένων χαρακτηριστικών. Σε αυτό το πλαίσιο, η μηχανική μάθηση διανύει έναν δρόμο εξέλιξης, βοηθώντας τους να ανακαλύψουμε την ουσία πίσω από τα δεδομένα, εκτός από την απλή παρατήρηση των χαρακτηριστικών τους.

Καθώς αυξάνουμε τον αριθμό των συλλεγμένων χαρακτηριστικών για ένα σύνολο δεδομένων, η διαστασιμότητα του διανύσματος χαρακτηριστικών και του αντίστοιχου χώρου χαρακτηριστικών αυξάνεται αντίστοιχα. Σε αυτό το σημείο, είναι σημαντικό να συνειδητοποιήσουμε πως η οπτικοποίηση όλων αυτών των διαστάσεων γίνεται αδύνατη, και η κατανόηση των σχέσεων μεταξύ τους ακόμη πιο πολύπλοκη. Η διαχείριση υψηλών διαστάσεων αποτελεί σημαντική πρόκληση για τους ερευνητές και προγραμματιστές, καθώς η μείωση της διαστατικότητας είναι απαραίτητη για την καλύτερη ερμηνεία, χαμηλότερες υπολογιστικές δαπάνες κι πρόληψη του overfitting.

Η χρήση τεχνικών, όπως η ανάλυση σε κύριες συνιστώσες (PCA) επιτρέπει τη συμπίεση των δεδομένων σε λίγες σημαντικές διαστάσεις, διευκολύνοντας την οπτικοποίηση και τον εντοπισμό μοντέλων.

Η διαδικασία της μείωσης της διαστατικότητας συχνά περιλαμβάνει την «Επιλογή Χαρακτηριστικών» και την «Εξαγωγή Χαρακτηριστικών». Η βασική διαφορά είναι ότι η «Επιλογή Χαρακτηριστικών» διατηρεί ένα υποσύνολο των αρχικών χαρακτηριστικών, ενώ η «Εξαγωγή Χαρακτηριστικών» δημιουργεί νέα.

Επιλογή Χαρακτηριστικών. Η «Επιλογή Χαρακτηριστικών» επικεντρώνεται στη διατήρηση του υποσυνόλου των αρχικών χαρακτηριστικών που παρέχουν σημαντική πληροφορία για το πρόβλημα. Αυτή η διαδικασία είναι ιδανική όταν υπάρχουν ήδη κατάλληλα χαρακτηριστικά. Στοχεύει στην επιλογή του υποσυνόλου μοναδικών

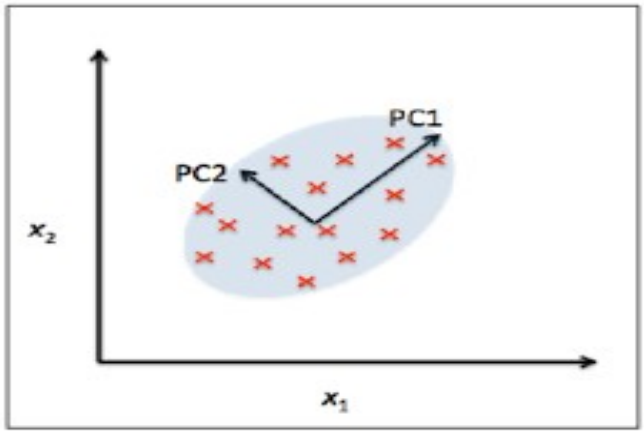
χαρακτηριστικών που συνεισφέρουν σημαντικά στην κατασκευή του μοντέλου, ενώ απορρίπτονται τα άσχετα ή λιγότερο σημαντικά.

Αυτή η διαδικασία ελαχιστοποιεί το πρόβλημα υπερπροσαρμογής, επιτρέπει την ταχύτερη εκπαίδευση των αλγορίθμων μηχανικής μάθησης, και αυξάνει την ακρίβεια του μοντέλου. Η «επιλογή χαρακτηριστικών» θεωρείται κεντρική έννοια στη μηχανική μάθηση, επηρεάζοντας σημαντικά την αποτελεσματικότητα και την αποδοτικότητα του επιθυμητού μοντέλου.

Εξαγωγή Χαρακτηριστικών. Η «Εξαγωγή Χαρακτηριστικών» δημιουργεί νέα χαρακτηριστικά που αντιπροσωπεύουν σύνθετες πτυχές των αρχικών δεδομένων. Αυτή η προσέγγιση είναι χρήσιμη όταν τα υπάρχοντα χαρακτηριστικά δεν είναι επαρκή για να αναδείξουν τις πληροφορίες. Στοχεύει στο να μειώσει τον αριθμό των χαρακτηριστικών σε ένα σύνολο δεδομένων, δημιουργώντας νέα χαρακτηριστικά από τα υπάρχοντα και, στη συνέχεια, απορρίπτοντας τα αρχικά χαρακτηριστικά.

Μια ευρέως χρησιμοποιούμενη τεχνική είναι η ανάλυση κύριων συνιστωσών (PCA), που μειώνει τις διαστάσεις των δεδομένων, δημιουργώντας νέα συνιστώσα από τα υπάρχοντα χαρακτηριστικά. Η PCA, μια από της δημοφιλέστερες τεχνικές εξαγωγής χαρακτηριστικών, αποτελεί ισχυρό εργαλείο για τη μείωση των διαστάσεων. Δημιουργεί νέα χαρακτηριστικά, συνδυάζοντας τα αρχικά, και επιτρέπει την εκπροσώπηση των δεδομένων με λιγότερες, αλλά πιο ενδιαφέρουσες, μεταβλητές.

Και οι δύο τεχνικές συμβάλλουν στη μείωση της διαστατικότητας, με την «Επιλογή Χαρακτηριστικών» να εστιάζει στη διατήρηση και επιλογή του υποσυνόλου, και την «Εξαγωγή Χαρακτηριστικών» να δημιουργεί νέα χαρακτηριστικά για την πλήρη αναπαράσταση των δεδομένων. Η κατανόηση της κατάλληλης τεχνικής για τη μείωση τους διαστατικότητας είναι καθοριστική για την επιτυχημένη ανάλυση και εφαρμογή μοντέλων μηχανικής μάθησης.



Εικόνα 2.9: Παράδειγμα εφαρμογής ανάλυσης κύριων συνιστωσών

2.3 Αλγόριθμοι Μηχανικής Μάθησης

2.3.1 Δέντρα Απόφασης

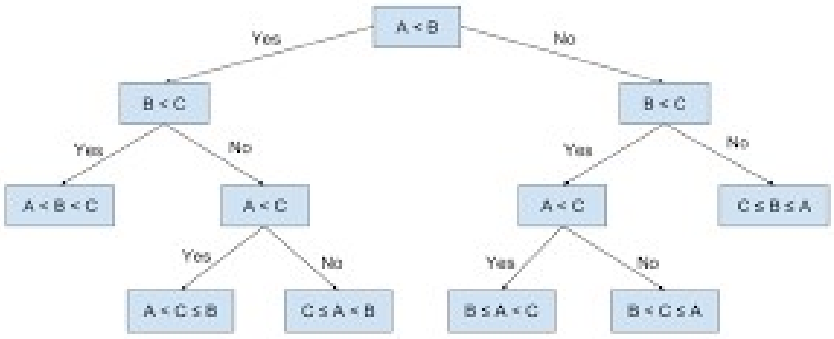
Τα δέντρα απόφασης αποτελούν ευρέως χρησιμοποιούμενη τεχνική στη μηχανική μάθηση, ειδικά σε προβλήματα κατηγοριοποίησης και παλινδρόμησης. Στην ουσία, αποσκοπούν στο να δημιουργήσουν μια ιεραρχία ερωτήσεων τύπου «εάν/αλλιώς» (if/else), όπου κάθε ερώτηση σχετίζεται με τα δεδομένα και οδηγεί σε μια απόφαση. Κάθε κόμβος απόφασης αναπαριστά μια συγκεκριμένη ερώτηση και τις πιθανές απαντήσεις, ενώ οι κόμβοι πιθανότητας αντιστοιχούν σε περιπτώσεις, όπου δεν είναι δυνατό να ληφθεί απόφαση με βεβαιότητα.

Παρόλο που η χρήση των δέντρων απόφασης είναι πολύ επινοητική, η πρόκληση παραμένει στο πώς να κατασκευάσουμε αποτελεσματικά αυτά τα δέντρα από τα δεδομένα που διαθέτουμε. Για τη δημιουργία του δέντρου απόφασης, ο αλγόριθμος εφαρμόζει έναν κανόνα διαίρεσης σε διαδοχικά μικρότερα τμήματα των δεδομένων, με κάθε τμήμα να είναι ένα κόμβος στο δέντρο. Ο κόμβος που αποτελείται από όλα τα δεδομένα είναι ο ριζικός κόμβος. Ο ριζικός κόμβος διαιρείται με βάση το BMI. Οι διαιρέσεις επιλέγονται για να ελαχιστοποιήσουν κάποιο μέτρο ανικανότητας του κόμβου (π.χ. ποικιλομορφία των κατηγοριών) ή ανομοιογένειας (π.χ. διακύμανση).

Η διαδικασία διαίρεσης επαναλαμβάνεται σε κάθε κλαδί του δέντρου μέχρι που περαιτέρω διαιρέσεις δεν οδηγούν σε περαιτέρω μείωση της ανικανότητας του κόμβου, ή μέχρι να επιτευχθεί κάποιο άλλο κριτήριο διακοπής (π.χ. καθορισμένο ελάχιστο πλήθος παρατηρήσεων σε τερματικούς κόμβους).

Οι αλγόριθμοι Δέντρων Αποφάσεων είναι αποτελεσματικοί σε ποικίλους τομείς, από την ανάλυση της συμπεριφοράς χρήστη έως την ανάλυση της κυβερνοασφάλειας. Η δομή τους είναι εύκολα ερμηνεύσιμη και παρέχει σαφείς αποφάσεις, καθιστώντας τα ιδανικά για εφαρμογές, όπου η διαφάνεια του μοντέλου είναι σημαντική. Ένα μειονέκτημα έγκειται στο γεγονός ότι τα δέντρα αποφάσεων μπορεί να είναι αρκετά ευαίσθητα σε μικρές διαταραχές στα δεδομένα.

Οι μέθοδοι συνόλου δέντρων αποτελούν ισχυρά εργαλεία στον χώρο της μηχανικής μάθησης και περιλαμβάνουν τα τυχαία δάση αποφάσεων και το Gradient Boosting. Και οι δύο αυτές προσεγγίσεις βασίζονται σε ένα σύνολο εκπαιδευμένων δέντρων αποφάσεων με σκοπό την πρόβλεψη μιας μεταβλητής απόκρισης. Παρόλο που αυτό το βασικό στοιχείο είναι κοινό, η διαφορά στον τρόπο δημιουργίας των δέντρων καθορίζει τις επιδόσεις και τα χαρακτηριστικά κάθε μεθόδου.



Εικόνα 2.10: Παράδειγμα Δέντρου Απόφασης

2.3.2 Τυχαία Δάση

Τα Τυχαία Δάση (Random Forests) αναδεικνύονται ως πρωτοπόρα τεχνική στον χώρο της μηχανικής μάθησης. Στον πυρήνα της αποτελεσματικότητας των Τυχαίων Δασών βρίσκεται η ιδέα της ποικιλομορφίας, συναρπάζοντας με την ικανότητά τους να προσφέρουν υψηλή ακρίβεια και έλεγχο σε προβλήματα ταξινόμησης και παλινδρόμησης. Αυτή η τεχνική ενσωματώνει πολλά δέντρα απόφασης σε ένα ενιαίο τυχαίο δάσος, προσφέροντας ακριβείς προβλέψεις με μειωμένο κίνδυνο υπερεκπαίδευσης. Ο κίνδυνος υπερεκπαίδευσης μειώνεται σημαντικά, καθώς τα δάση ενσωματώνουν πληροφορίες από πολλαπλά δέντρα, βελτιώνοντας τη γενίκευση σε νέα δείγματα.

Ο αλγόριθμος δημιουργεί εκατοντάδες ή χιλιάδες βαθιά δέντρα αποφάσεων. Μέσω της μεθόδου bootstrap, τα τυχαία δάση δημιουργούν αντίγραφα των δεδομένων, εκπαιδεύοντας ένα ξεχωριστό δέντρο σε κάθε αντίγραφο. Η συνολική πρόβλεψη προκύπτει από τον συνδυασμό πολλών δέντρων, ελαχιστοποιώντας τον κίνδυνο υπερεκπαίδευσης. Αν και κάθε ένα από αυτά τα δέντρα είναι πιθανόν να είναι υπερπροσαρμοσμένο, η συνδυαστική λειτουργία τους αντιμετωπίζει το πρόβλημα της υπερπροσαρμογής. Συνεπώς, το μοντέλο λαμβάνει την έξοδο πολλών δέντρων και διασφαλίζει ότι η τελική πρόβλεψη είναι σταθερή και αξιόπιστη.

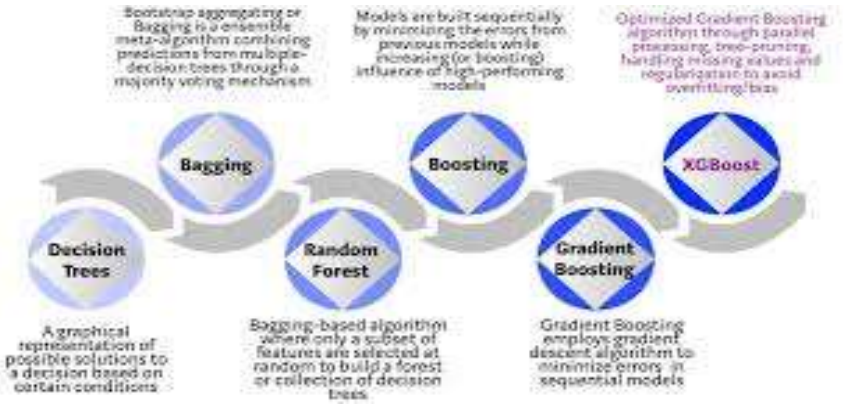
Bagging. Το Bagging (ή Bootstrap Aggregating) εφαρμόζει τον ίδιο βασικό αλγόριθμο σε κάθε ανατιθέμενο αντίγραφο των αρχικών εκπαιδευτικών δεδομένων και στη συνέχεια δημιουργεί μια τελική πρόβλεψη βασισμένη στα αποτελέσματα από τα αποτελεσματικά, παραμετροποιημένα μοντέλα. Εκπαιδεύει μοντέλα σε τυχαία υποσύνολα μεταβλητών/ χαρακτηριστικών αντί για μεταβλητές σε μια προσπάθεια να μειώσει τη συσχέτιση μεταξύ των μοντέλων σε ένα σύνολο.

Adaptive Boosting AdaBoost. Το Adaptive Boosting (AdaBoost) συνιστά μια εξελιγμένη προσέγγιση στη μετα-μάθηση, παρέχοντας αξιόπιστες και ακριβείς προβλέψεις, ενώ παράλληλα αντιμετωπίζει δυνητικά ζητήματα υπερεκπαίδευσης. Στην πράξη, το AdaBoost εξειδικεύεται στη βελτίωση της απόδοσης των δέντρων απόφασης, που είναι ο βασικός του εκτιμητής. Αυτό είναι εξαιρετικά χρήσιμο σε προβλήματα δυαδικής ταξινόμησης, αλλά η ευαισθησία του σε θορυβώδη δεδομένα και ακραίες τιμές μπορεί να αποτελέσει μια πρόκληση.

Extreme Gradient Boosting (XGBoost). Το Gradient Boosting, ανάλογα με τα Random Forests, είναι ένας προηγμένος αλγόριθμος μηχανικής μάθησης που κατασκευάζει ένα συνολικό μοντέλο βασισμένο σε μια σειρά από μοντέλα αποφάσεων, συχνά χρησιμοποιώντας δέντρα αποφάσεων. Στην ουσία, ο στόχος του Gradient Boosting είναι η ελαχιστοποίηση της συνάρτησης καθυστέρησης με τη χρήση της κλίσης, παραπλήσια με τον τρόπο που τα νευρωνικά δίκτυα βελτιστοποιούν τα βάρη τους.

Ο αλγόριθμος Gradient Boosting, όπως το XGBoost, μειώνει επαναληπτικά το σφάλμα ταξινόμησης προσθέτοντας περισσότερα δέντρα στο μοντέλο. Η στρατηγική αυτή στοχεύει στη βελτίωση της απόδοσης του μοντέλου, προσθέτοντας δέντρα που επικεντρώνονται στα σημεία, όπου το μοντέλο υστερεί. Το Extreme Gradient Boosting (XGBoost) αναδεικνύεται ως προηγμένη μορφή του Gradient Boosting, λαμβάνοντας υπόψη λεπτομερείς προσεγγίσεις για την εύρεση του καλύτερου μοντέλου.

Ένα από τα σημαντικά πλεονεκτήματα του XGBoost είναι η ταχύτητά του στην εκτέλεση και η ικανότητά του να αντιμετωπίζει αποτελεσματικά μεγάλα σύνολα δεδομένων. Παρόλα αυτά, πρέπει να σημειωθεί ότι το XGBoost είναι ευαίσθητο σε δεδομένα υψηλού θορύβου και ακραίες τιμές, προκαλώντας πιθανά προβλήματα υπερεκπαίδευσης.



Εικόνα 2.11: Η εξέλιξη του Xgboost μέσα από τα Δέντρα Απόφασης

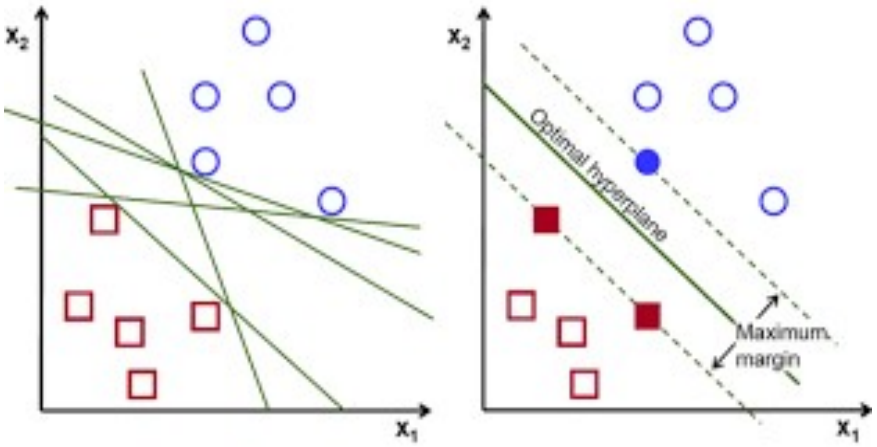
2.3.3 Μηχανές Διανυσμάτων Στήριξης

Οι μηχανές υποστήριξης (Support Vector Machines – SVM) και οι μέθοδοι πυρήνα συνθέτουν ένα ισχυρό πλαίσιο στον κόσμο της μηχανικής μάθησης. Ο αλγόριθμος SVM έχει ως βασικό στόχο τη δημιουργία ενός μοντέλου που να μπορεί να προβλέψει την εξαρτώμενη μεταβλητή από τις ανεξάρτητες. Η ύψιστη αποτελεσματικότητά του σε χώρους υψηλής διάστασης καθιστά τον SVM ένα ισχυρό εργαλείο στον τομέα της μηχανικής μάθησης, συνεισφέροντας έτσι σημαντικά στην επίλυση προβλημάτων κατηγοριοποίησης και παλινδρόμησης.

Η ιδιαιτερότητα των SVM βρίσκεται στον τρόπο που διαχωρίζουν τις κατηγορίες. Το SVM για κατηγοριοποίηση επιδιώκει να εντοπίσει τον βέλτιστο διαχωριστικό χώρο (υπερεπίπεδο) που θα διαχωρίσει αποτελεσματικά τις διάφορες κατηγορίες. Τα «διανύσματα υποστήριξης» παίζουν κρίσιμο ρόλο στον καθορισμό αυτού του υπερεπιπέδου και καθορίζουν ένα υπερεπίπεδο σε έναν χώρο χαρακτηριστικών. Συνεπώς, αναζητούν το υπερεπίπεδο που μεγιστοποιεί το περιθώριο μεταξύ διαφορετικών κατηγοριών, επιτρέποντας έτσι μια πιο αποτελεσματική διαχωριστική επιφάνεια. Η ευελιξία των SVM προκύπτει από τη δυνατότητά τους να αντιμετωπίζουν πολλαπλές εισόδους και να υποθέτουν πολύπλοκες μη γραμμικές συναρτήσεις.

Το κλειδί της επιτυχίας των SVM είναι η χρήση πυρήνων, επιτρέποντας την εργασία σε χώρους χαρακτηριστικών υψηλότερης διάστασης. Η βέλτιστη συνάρτηση πυρήνα επιλέγεται συνήθως από ένα σύνολο συνήθως χρησιμοποιούμενων συναρτήσεων πυρήνα που επιλέγονται μέσω διασταυρούμενης επικύρωσης. Ο πυρήνας επιτρέπει την αντιστοίχιση σε ένα γινόμενο σημείων στον χώρο χαρακτηριστικών, ενισχύοντας τη δυνατότητα αντιμετώπισης μη γραμμικών συναρτήσεων. Ένα εντυπωσιακό χαρακτηριστικό της SVM είναι η ικανότητά της να επιτυγχάνει ένα ισχυρό περιθώριο, δηλαδή μια μέγιστη απόσταση από τα πλησιέστερα σημεία εκπαίδευσης σε κάθε κλάση. Αυτό το περιθώριο συνδέεται άμεσα με την ακρίβεια της κατηγοριοποίησης, όπου το μεγαλύτερο περιθώριο συνήθως συνεπάγεται χαμηλότερο σφάλμα γενίκευσης.

Παρόλα αυτά, η απόδοση της SVM μπορεί να μειωθεί όταν το σύνολο δεδομένων περιέχει θόρυβο. Είναι πολύ σημαντικό να λαμβάνεται υπόψη η ποιότητα των δεδομένων κατά την εφαρμογή της SVM. Παρά την ακρίβεια των προβλέψεων, οι SVM αντιμετωπίζουν την πρόκληση του «μαύρου κουτιού», δηλαδή δεν παρέχουν λεπτομερείς μετρήσεις για τον συνδυασμό των προβλέψεων, καθιστώντας τη μέθοδο δυσνόητη για ερμηνεία. Δημοφιλείς συναρτήσεις πυρήνα περιλαμβάνουν το πολυωνυμικό πυρήνα, τον πυρήνα Gaussian και τον πυρήνα sigmoid.



Εικόνα 2.12: Παράδειγμα Κατηγοριοποίησης με SVM

2.3.4 Κ-Πλησιέστεροι Γείτονες

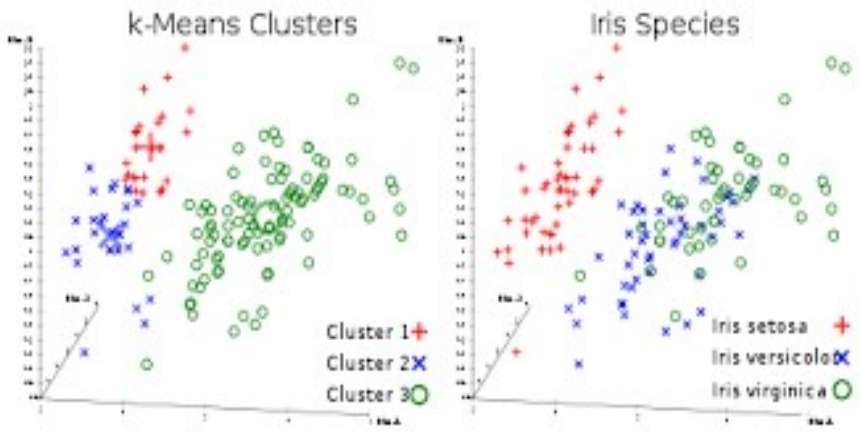
Η μάθηση βάσει παραδειγμάτων, ευρέως γνωστή και ως μάθηση k-κοντινότερων γειτόνων (kNN), αποτελεί μια απλή αλλά αποτελεσματική προσέγγιση στον χώρο της μηχανικής μάθησης. Σε αυτήν την προσέγγιση, η μάθηση εξαρτάται από την αποθήκευση όλων των υπαρχόντων ετικετοποιημένων δεδομένων (ή αλλιώς, των δεδομένων εκπαίδευσης) σε μια βάση δεδομένων.

Όταν παρατηρείται ένα νέο, μη-ταξινομημένο παράδειγμα, ο αλγόριθμος τοποθετεί το παράδειγμα αυτό σε ένα n-διάστατο χώρο χαρακτηριστικών βάσει των τιμών των χαρακτηριστικών του. Για κάθε σημείο δεδομένων στη βάση δεδομένων, υπολογίζουμε την απόστασή του από το νέο δείγμα δεδομένων, χρησιμοποιώντας μια μέθοδο μέτρησης απόστασης, όπως η Ευκλείδεια απόσταση. Στη συνέχεια, εντοπίζουμε τους k κοντινότερους γείτονές του.

Το clustering K-means είναι ένας από τους πιο απλούς αλγορίθμους ανεπτυγμένης μάθησης (unsupervised learning). Αρχικά, το clustering K-means επιλέγει τυχαία k κεντροειδείς (centroids), με κάθε κεντροειδές να καθορίζει μια συστάδα (δηλαδή, κάθε παρατήρηση ανατίθεται στο πιο κοντινό κεντροειδές). Παρά την απλότητά του, ο αλγόριθμος kNN αποδεικνύεται πολύ αποτελεσματικός στην πράξη και αποτελεί ένα χρήσιμο εργαλείο για προβλήματα κατηγοριοποίησης και παλινδρόμησης.

Ωστόσο, παρά την ανθεκτικότητά του, ο KNN αντιμετωπίζει ένα πρόβλημα, το οποίο είναι η επιλογή του βέλτιστου αριθμού γειτόνων (k). Αυτή η επιλογή κρίνεται κρίσιμη για τη σωστή λειτουργία του αλγορίθμου. Η επιλογή του k, επηρεάζει σημαντικά τα αποτελέσματα. Ένα πολύ μικρό k μπορεί να καθιστά το μοντέλο ευαίσθητο στο θόρυβο, ενώ ένα πολύ μεγάλο k μπορεί να απομακρύνει την ευελιξία του αλγορίθμου. Εφαρμόζονται διάφορες μέθοδοι για την επιλογή του κατάλληλου k.

Ο KNN μπορεί να χρησιμοποιηθεί σε ποικίλες εφαρμογές, και η επιλογή του εξαρτάται σταθερά από την ποιότητα των δεδομένων. Ενδείκνυται για περιπτώσεις που δεν είναι δυνατή η κατασκευή ενός γενικού μοντέλου ή όταν τα δεδομένα είναι δυσεπεξεργάστηκτα από τους αλγορίθμους.



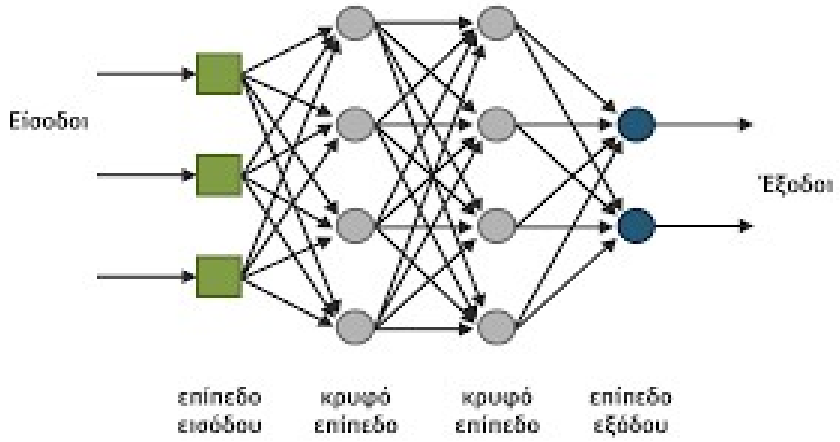
Εικόνα 2.13: Παράδειγμα Κατηγοριοποίησης με Κ-Πλησιέστερους Γείτονες

2.3.5 Νευρωνικά Δίκτυα

Τα νευρωνικά δίκτυα αντιπροσωπεύουν ένα σύστημα που προέκυψε από τη μελέτη των νευρώνων και των κυκλωμάτων στον εγκέφαλο. Αποτελούνται από νευρώνες και ακμές, όπου τα βάρη προσδίδονται σε κάθε ακμή, προκειμένου να καθοριστεί η ένταση των συνδέσεων. Κάθε νευρώνας χρησιμοποιεί μια συνάρτηση ενεργοποίησης για τον υπολογισμό της εξόδου από ένα εισαγόμενο σήμα, λαμβάνοντας υπόψη τα βάρη. Η πιο γνωστή συνάρτηση ενεργοποίησης αποτελεί η σιγμοειδής συνάρτηση.

Οι νευρώνες οργανώνονται σε επίπεδα, συμπεριλαμβανομένου του επιπέδου εισόδου, ενός ή περισσοτέρων κρυμμένων επιπέδων και του επιπέδου εξόδου. Τα κρυμμένα επίπεδα αποτελούν το επίπεδο αφαίρεσης που είναι απαραίτητο για τη μετάβαση από την είσοδο στην έξοδο. Ο αριθμός των κρυμμένων επιπέδων καθορίζει εάν το σύστημα είναι επιφανειακής μάθησης ή βαθιάς μάθησης, ενώ υπάρχει ισορροπία μεταξύ αυτού του αριθμού και του χρόνου εκπαίδευσης του μοντέλου. Το βασικότερο είδος νευρωνικού δικτύου γνωστό ως προωθημένο νευρωνικό δίκτυο απλώς μεταδίδει πληροφορία από το επίπεδο εισόδου στα κρυμμένα επίπεδα και, τελικά, στο επίπεδο εξόδου. Η κατάσταση του συστήματος δεν εξαρτάται από προηγούμενες καταστάσεις, αντιπροσωπεύοντας ένα ασυμπτωτικά μνημονικό σύστημα. Η βαθιά μάθηση ξεπερνά τη μηχανική μάθηση, ειδικά όταν εκπαιδεύεται σε μεγάλα σύνολα δεδομένων. Παρά τη γενική αυτή τάση, η απόδοση μπορεί να διαφέρει, εξαρτώμενη από τα χαρακτηριστικά των δεδομένων και το πειραματικό πλαίσιο.

Έχει εφαρμογές σε πολλούς τομείς, όπως η αναγνώριση εικόνας, η φωνητική αναγνώριση, και η αυτόματη μετάφραση. Η ικανότητά της να αντλεί συμπεράσματα από πολύπλοκα δεδομένα την καθιστά ιδανική για απαιτητικές εργασίες. Οι πιο κοινοί αλγόριθμοι βαθιάς μάθησης είναι: το Πολυεπίπεδο Πεπερσέπτρον (MLP), το Συνελικτικό Νευρωνικό Δίκτυο (CNN), το Μακροπρόθεσμο Αναδρομικό Νευρωνικό Δίκτυο (LSTM-RNN).



Εικόνα 2.14: Παράδειγμα εφαρμογής Νευρωνικού Δικτύου

Πολυεπίπεδο Πεπερσέπτρον. Στον κόσμο της βαθιάς μάθησης, η βασική αρχιτεκτονική που διακρίνεται είναι το Πολυεπίπεδο Πεπερσέπτρον (MLP). Είναι ένα πλήρως συνδεδεμένο νευρωνικό δίκτυο που συνδυάζει διάφορα επίπεδα επεξεργασίας για την ανάλυση και την εξαγωγή χαρακτηριστικών από τα δεδομένα. Το τυπικό MLP αποτελείται από τρία βασικά επίπεδα: το επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και το επίπεδο εξόδου. Κάθε κόμβος σε ένα επίπεδο συνδέεται με κάθε κόμβο στο επόμενο επίπεδο με συγκεκριμένα βάρη.

Συνελικτικό Νευρωνικό Δίκτυο. Η βασική δομή του CNN περιλαμβάνει συνελικτικά επίπεδα που αναγνωρίζουν περιοχές χαρακτηριστικών, επίπεδα συσσώρευσης που εκτελούν υποδειγματοληψία, και πλήρως συνδεδεμένα επίπεδα για την τελική επεξεργασία. Αν και το CNN έχει μεγαλύτερο υπολογιστικό φορτίο, η αυτόματη εξαγωγή σημαντικών χαρακτηριστικών το καθιστούν ισχυρότερο.

LSTM Νευρωνικό Δίκτυο. Το μοντέλο Μακρού-Σύντομου Χρόνου Μνήμης (LSTM) ανήκει στην κατηγορία των αναδρομικών νευρωνικών δικτύων (RNN) και αποτελεί ισχυρό εργαλείο στον χώρο τους βαθιάς μάθησης. Το LSTM διαθέτει συνδέσμους ανατροφοδότησης, σε αντίθεση με τα κανονικά νευρωνικά δίκτυα προώθησης. Η δομή του επιτρέπει την ανάλυση και εκμάθηση ακολουθιακών δεδομένων με μνήμη που μπορεί να διατηρεί πληροφορίες για μεγάλα χρονικά διαστήματα.

2.5 Αξιολόγηση Μοντέλων

Μετά τη δημιουργία και εκπαίδευση των μοντέλων, απαιτείται η αξιολόγησή τους χρησιμοποιώντας διάφορες μετρικές. Οι αλγόριθμοι μηχανικής μάθησης εκτιμώνται με βάση διάφορα κριτήρια, και στο τέλος πραγματοποιείται η επιλογή του βέλτιστου ή του πιο αποδοτικού αλγορίθμου, λαμβάνοντας υπόψη το δείγμα των δεδομένων.

Για την αξιολόγηση των μοντέλων παλινδρόμησης χρησιμοποιούμε συνήθως τις εξής τέσσερις μετρικές αξιολόγησης: το Μέσο Τετραγωνικό Σφάλμα (MAE), το Μέσο Απόλυτο Σφάλμα (MSE), το Ριζικό Μέσο Τετραγωνικό Σφάλμα (RMSE) και το τετραγωνικό R(R squared).

Μέσο Τετραγωνικό Σφάλμα (MSE). Το Μέσο Τετραγωνικό Σφάλμα είναι μια συνάρτηση κόστους που μετρά τη μέση τετραγωνική απόκλιση μεταξύ των πραγματικών και προβλεπόμενων τιμών. Υψηλό MSE υποδεικνύει μεγάλη απόκλιση και, συνεπώς, χαμηλή απόδοση του μοντέλου. Είναι μετρική υπολογισμού της διακύμανσης.

Μέσο Απόλυτο Σφάλμα (MAE). Το Μέσο Απόλυτο Σφάλμα αναπαριστά τον μέσο όρο της απόλυτης τιμής της διαφοράς μεταξύ των πραγματικών και των προβλεπόμενων τιμών στο σύνολο των δεδομένων. Είναι μετρική υπολογισμού μέσης τιμής.

Ριζικό Μέσο Τετραγωνικό Σφάλμα (RMSE). Το Ριζικό Μέσο Τετραγωνικό Σφάλμα αποτελεί την τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος και είναι μετρική υπολογισμού τυπικής απόκλισης.

Τετραγωνικό R (R squared). Το Τετραγωνικό R αναπαριστά το ποσοστό της διακύμανσης στην εξαρτημένη μεταβλητή, το οποίο μπορεί να κατανοηθεί από το μοντέλο της γραμμικής παλινδρόμησης. Οι τιμές αυτής της μετρικής είναι μικρότερες τους μονάδας.

Αντίθετα, σε προβλήματα δυαδικής ταξινόμησης με δύο κατηγορίες, χρησιμοποιούμε συνήθως τον «πίνακα σύγχυσης» και από αυτόν προκύπτουν διάφορες μετρήσεις απόδοσης.

2.5.1 Πίνακας Σύγχυσης

Ο πίνακας σύγχυσης αναδεικνύει την απόδοση του μοντέλου πρόβλεψης, παρέχοντας μια αναλυτική σύγκριση μεταξύ των πραγματικών και προβλεπόμενων τιμών. Απεικονίζει τον αριθμό των πραγματικών και των προβλεπόμενων κλάσεων, χωρίζοντας τα αποτελέσματα σε τέσσερις κατηγορίες: True Positive (TP), True Negative (TN), False Positive (FP) και False Negative (FN). Στο πλαίσιο της μηχανικής

μάθησης, ο πίνακας σύγχυσης λειτουργεί ως χρήσιμο εργαλείο για την αξιολόγηση της επίδοσης του αλγορίθμου ταξινόμησης. Αναδεικνύει σημαντικές μετρικές, όπως η ακρίβεια, η ανάκληση, το F1-score προσφέροντας μια συνολική εικόνα της απόδοσης του συστήματος.

Ο λόγος που ο πίνακας σύγχυσης είναι ιδιαίτερα χρήσιμος είναι ότι, σε αντίθεση με τους τύπους μετρήσεων ταξινόμησης, όπως η απλή ακρίβεια, ο πίνακας σύγχυσης δημιουργεί μια πιο ολοκληρωμένη εικόνα του τρόπου απόδοσης του μοντέλου. Μόνο η χρήση μιας μετρικής, όπως η ακρίβεια μπορεί να οδηγήσει σε μια κατάσταση, όπου το μοντέλο προσδιορίζει εντελώς και συνεχώς εσφαλμένα μια κατηγορία, αλλά περνά απαρατήρητο, επειδή κατά μέσο όρο η απόδοση είναι καλή. Εάν η πρόβλεψη συμπίπτει με το πραγματικό αποτέλεσμα, τότε υπάρχει σωστή κατηγοριοποίηση, δηλαδή όντως θετικό (TP-True Positive), ή όντως αρνητικό (TN-True Negative). Ενώ αν η πρόβλεψη δεν συμπίπτει με το πραγματικό αποτέλεσμα, τότε έχουμε λάθος κατηγοριοποίηση, δηλαδή λάθος πρόβλεψη (FP-False Positive), ή όντως αρνητικό (FN-False Negative).

2.5.2 Ακρίβεια και Ανάκληση (ή Ευαισθησία)

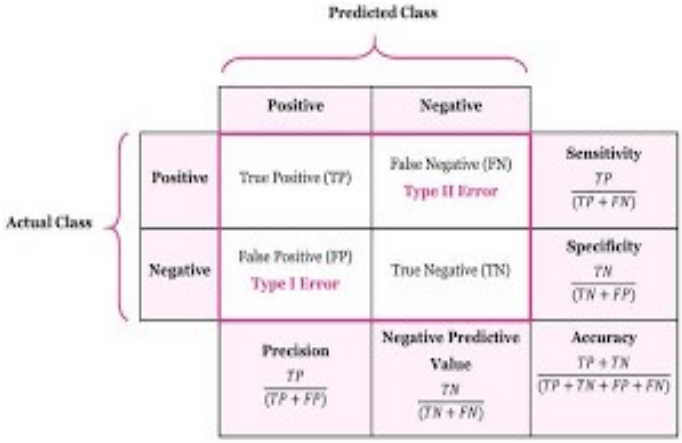
Η Ακρίβεια αντιστοιχεί στον λόγο των σωστά προβλεπόμενων θετικών τιμών προς τον συνολικό αριθμό των προβλεπόμενων θετικών τιμών.

Ακρίβεια = TP / (TP + FP)

Αντίστοιχα, η ανάκληση (ή αλλιώς ευαισθησία) ορίζεται ως το ποσοστό των σωστά προβλεπόμενων θετικών τιμών προς τον συνολικό αριθμό των θετικών τιμών στο σύνολο των δεδομένων.

Ανάκληση = TP /(TP + FN)

Για μια αποτελεσματική αξιολόγηση του μοντέλου, είναι ζωτικής σημασίας να εξεταστούν και οι δύο αυτοί δείκτες. Τα δύο αυτά ποσοστά είναι αντιστρόφως ανάλογα μεταξύ τους, δηλαδή όταν η ακρίβεια βελτιώνεται, το ποσοστό ευαισθησίας μειώνεται και αντίστροφα. Έχουν δημιουργηθεί διάφοροι αξιολογητές που βασίζονται στην ακρίβεια και την ευαισθησία του μοντέλου. Στην επόμενη ενότητα, θα αναλύσουμε τον αξιολογητή F1 score.



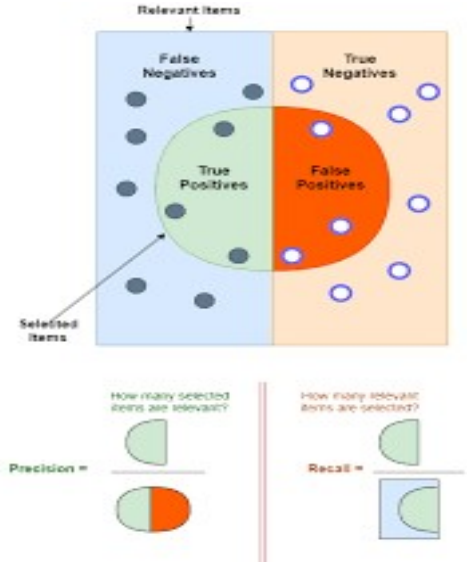
Εικόνα 2.15: Πίνακας Σύγχυσης (Confusion Matrix Metrics)

2.5.3 Αξιολογητής F1-Score

Το F1 σκορ είναι ο αρμονικός μέσος της ακρίβειας και της ανάκλησης. Σε μια στατιστική ανάλυση συστημάτων δυαδικής ταξινόμησης, το F1-σκορ είναι ένα μέτρο προγνωστικής απόδοσης. Υπολογίζεται από την ακρίβεια και την ανάκληση του τεστ.

Η υψηλότερη δυνατή τιμή του F1-σκορ είναι το 1.0, που υποδεικνύει τέλεια ακρίβεια και ανάκληση, και η χαμηλότερη δυνατή τιμή είναι το 0, εάν είτε η ακρίβεια είτε η ανάκληση είναι μηδέν.

F1-score = 2 * 1 / (1 / Ακρίβεια + 1 / Ανάκληση)

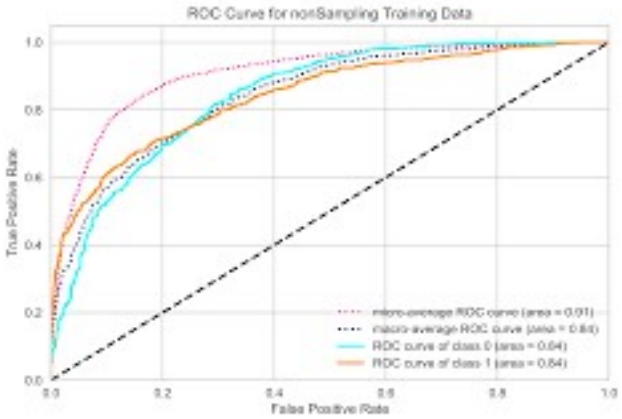


Εικόνα 2.16: F1-Score 2 * 1 / (1 / Ακρίβεια + 1 / Ανάκληση)

2.5.4 Καμπύλη Χαρακτηριστικών Λειτουργιάς Δείκτη (Καμπύλη ROC – AUC)

Η καμπύλη ROC (Receiver Operating Characteristics) αποτελεί μια εξαιρετική μέτρηση της απόδοσης των μοντέλων για τα προβλήματα ταξινόμησης. Παρουσιάζει τη σχέση μεταξύ του ποσοστού πραγματικά θετικών (TP) και του ποσοστού λανθασμένων προβλέψεων ως θετικά (FP) στο πλαίσιο διαγνωστικών δοκιμών. Η θέση κάθε σημείου στην πειραματική αυτή καμπύλη καθορίζεται από συγκεκριμένα TP και FP αποτελέσματα των δοκιμών που συνδέονται με ένα συγκεκριμένο διαχωριστικό όριο.

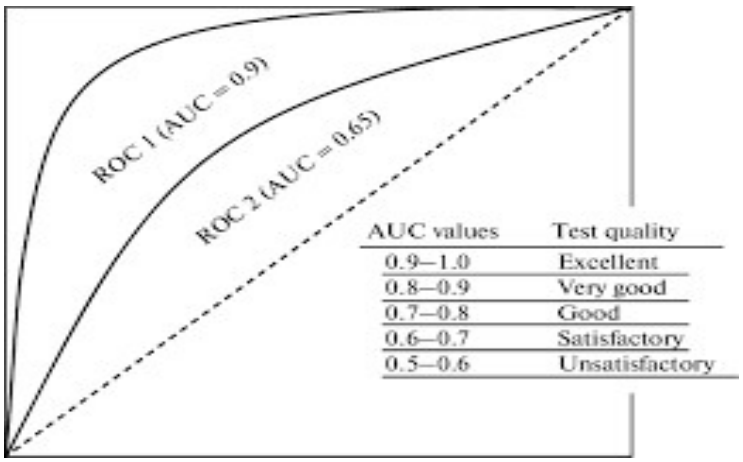
Η καμπύλη ROC καταγράφεται εντός ενός τετραγώνου, όπου οι τέσσερις γωνίες αντιστοιχούν στις άκρες των TP και FP ποσοστών (0 και 1), καθώς και στις συμπληρωματικές τους τιμές FN, TN. Η διαγώνια γραμμή διαχωρισμού αντιπροσωπεύει μια τυχαία πρόβλεψη, ενώ η γραμμή ανωτέρω αντιπροσωπεύει μια βελτιωμένη πρόβλεψη, με την τέλεια πρόβλεψη να παρουσιάζεται όταν η γραμμή δημιουργεί ορθή γωνία. Η πάνω αριστερή γωνία μιας καμπύλης ROC είναι η ιδανική περίπτωση με 100% των θετικών τιμών που ταξινομούνται σωστά (ανάκληση = 1) και 0% των θετικών τιμών που προβλέπονται εσφαλμένα στο 0 (FPR = 0). Καθώς είναι ιδανικό να μεγιστοποιούμε την ανάκληση ενώ ταυτόχρονα ελαχιστοποιούμε τον FPR, ένα μεγαλύτερο εμβαδόν κάτω από την καμπύλη ROC (AUC) είναι καλύτερο.



Εικόνα 2.17: Καμπύλη ROC (ROC Curve Classification Problem)

Το εμβαδόν κάτω από την καμπύλη (AUC – Area Under Curve) χρησιμοποιείται ως ένδειξη για το πόσο καλά διαχωρίζονται οι κατανομές. Όσο υψηλότερη είναι η AUC, τόσο καλύτερο το μοντέλο στην πρόβλεψη. Επομένως, όσο μεγαλύτερη η περιοχή κάτω από την καμπύλη, τόσο προτιμότερος θα είναι ο αντίστοιχος κατηγοριοποιητής. Η ελάχιστη τιμή που μπορεί να πάρει είναι 0.5, όταν οι δύο κατανομές συμπίπτουν, ενώ η μέγιστη τιμή είναι 1 όταν οι κατανομές δεν έχουν κοινά σημεία. (Stuart Russel & Peter Norvig, 2016). Συγκεκριμένα τιμές μεταξύ 0,50–0,70 μπορεί να υποδηλώνει κακή πρόβλεψη· 0,70–0,79 να υποδηλώνει αποδεκτή πρόβλεψη· 0,80–0,89 να

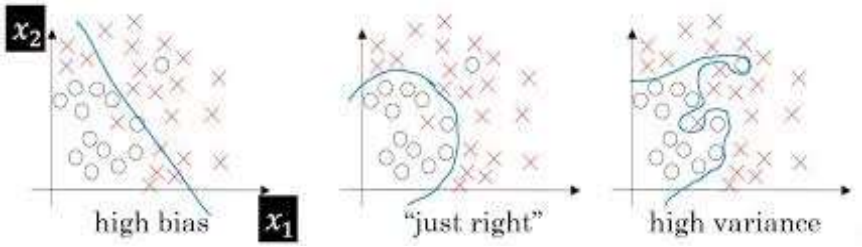
υποδηλώνει εξαιρετική πρόβλεψη· και > 0,90 να υποδηλώνει εξαιρετική πρόβλεψη (Hosmer & Lemeshow, 2013).



Εικόνα 2.18: Αξιολόγηση Καμπύλης ROC

2.6 Εκτίμηση Απόδοσης

Στη μηχανική μάθηση, ο στόχος ενός αλγορίθμου είναι να δημιουργήσει ένα μοντέλο που να περιγράφει με ακρίβεια τα παρατηρούμενα δεδομένα εκπαίδευσης και να είναι ικανό να γενικεύει σε νέα, ανεξάρτητα δεδομένα, αποφεύγοντας τα προβλήματα της υπερπροσαρμογής (overfitting) και της υποπροσαρμογής (underfitting).



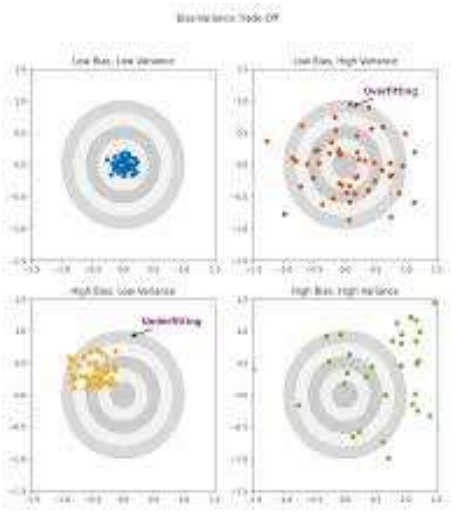
Εικόνα 2.19: Μετρικές εκτίμησης απόδοσης (Underfit vs Overfit)

Η υπερπροσαρμογή (overfit) συμβαίνει όταν το μοντέλο προσαρμόζεται υπερβολικά στα δεδομένα εκπαίδευσης, συμπεριλαμβάνοντας τυχαίες ή ασήμαντες αντιστοιχίσεις που δεν γενικεύονται στα νέα δεδομένα. Αντίθετα, η υποπροσαρμογή (underfit) συμβαίνει όταν το μοντέλο είναι υπερβολικά απλό, αδυνατώντας να αντιστοιχίσει την πολυπλοκότητα των πραγματικών δεδομένων, αφήνοντας περιθώρια για απώλεια ακρίβειας.

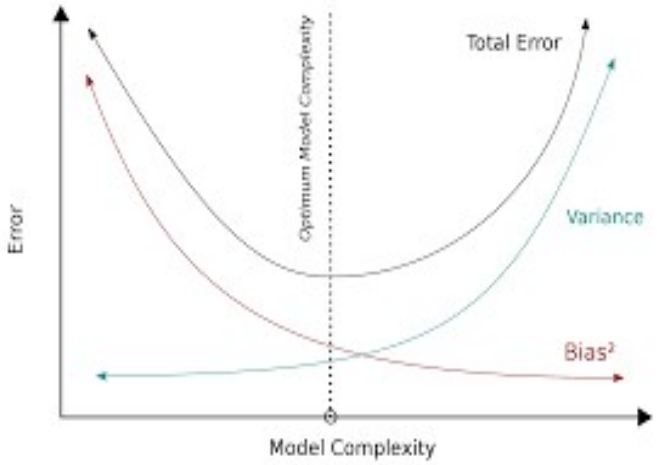
Η διαδικασία εκπαίδευσης ενός μοντέλου στη μηχανική μάθηση απαιτεί την εκτίμηση των παραμέτρων του, όπως οι συντελεστές και τα βάρη, βάσει των παρατηρούμενων δεδομένων στο σύνολο εκπαίδευσης. Αυτό επιτυγχάνεται ελαχιστοποιώντας μια συνάρτηση απώλειας, η οποία μετρά την απόσταση μεταξύ των προβλέψεων του μοντέλου και των παρατηρούμενων δεδομένων. Η συνάρτηση απώλειας λειτουργεί ως κριτήριο για την προσαρμογή του μοντέλου, με στόχο την παραγωγή προβλέψεων που είναι ταυτόχρονα κοντά στα δεδομένα εκπαίδευσης και ικανά να γενικεύουν σε νέα δεδομένα. Επομένως, η συνάρτηση απώλειας λειτουργεί ως οδηγός για τη βελτιστοποίηση των παραμέτρων του μοντέλου, επιτυγχάνοντας τον συμβιβασμό μεταξύ πολυπλοκότητας και γενίκευσης, προκειμένου το μοντέλο να επιτύχει υψηλή απόδοση τόσο στα δεδομένα εκπαίδευσης όσο και σε νέα, μη ορατά δεδομένα.

Ένα κρίσιμο ζήτημα στην μηχανική μάθηση είναι η σχέση μεταξύ προκατάληψης (bias) και διακύμανσης (variance). Η προκατάληψη (bias) είναι μια ένδειξη του μέσου σφάλματος του μοντέλου σε διάφορα σύνολα εκπαίδευσης. Αναφέρεται στην ανομοιότητα μεταξύ του μέσου των προβλεπόμενων τιμών και του πραγματικού

μέσου που προσπαθούμε να προβλέψουμε. Αντίθετα, η διακύμανση (variance) αντικατοπτρίζει την ευαισθησία του μοντέλου στο σύνολο εκπαίδευσης. Είναι η διασπορά των προβλεπόμενων τιμών γύρω από το μέσον τους. Υψηλή διακύμανση υποδεικνύει ότι το μοντέλο είναι πολύ ευαίσθητο στις αλλαγές στα δεδομένα εκπαίδευσης.



Εικόνα 2.20: Bias Variance Trade-off



Για την ελαχιστοποίηση του προβλεπόμενου σφάλματος, χρειάζεται ένας συμβιβασμός μεταξύ ελαχιστοποίησης της προκατάληψης (bias) και της διακύμανσης (variance). Αυξάνοντας την πολυπλοκότητα του μοντέλου, μειώνεται η προκατάληψη αλλά αυξάνεται η διακύμανση. Για την κατασκευή λιγότερο πολύπλοκων μοντέλων, χρησιμοποιούνται τεχνικές κανονικοποίησης, με τις πιο γνωστές τεχνικές κανονικοποίησης να αποτελούν η L1 και L2.

2.6.1 Σύνολο Ελέγχου

Η αρχή της επιλογής του μοντέλου στη μηχανική μάθηση αποτελεί κρίσιμο βήμα για την επίτευξη της καλύτερης δυνατής απόδοσης. Κατά τη διαδικασία αυτή, τα δεδομένα διαχωρίζονται σε τρία σύνολα: το σύνολο εκπαίδευσης, το σύνολο επικύρωσης, και το σύνολο δοκιμής.

Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκπαίδευση διάφορων μοντέλων. Κατά τη διάρκεια αυτής της φάσης, τα μοντέλα προσαρμόζονται στα δεδομένα και αποκτούν τη δυνατότητα να προβλέπουν αποτελέσματα.

Το σύνολο επικύρωσης χρησιμοποιείται για να επιλεγεί το βέλτιστο μοντέλο και να προσδιοριστούν οι βέλτιστες υπερπαράμετροι. Κατά την επιλογή, τα μοντέλα αξιολογούνται βάσει της επίδοσής τους στο σύνολο επικύρωσης, και εκείνο που επιλέγεται είναι το μοντέλο που παρουσιάζει την καλύτερη απόδοση.

Στη συνέχεια, το επιλεγμένο μοντέλο υποβάλλεται σε αξιολόγηση στο σύνολο δοκιμής. Αυτό το σύνολο επιτρέπει την αξιολόγηση του σφάλματος γενίκευσης ή σφάλματος δοκιμής, που αποτελεί το σφάλμα πρόβλεψης σε ένα ανεξάρτητο σύνολο δεδομένων που δεν χρησιμοποιήθηκε κατά τη διάρκεια της εκπαίδευσης.

Ένας τρόπος εσωτερικής επικύρωσης είναι η απλή μέθοδος της τυχαίας διαίρεσης ενός συνόλου δεδομένων σε ένα σύνολο εκπαίδευσης και ένα σύνολο ελέγχου, εφαρμόζοντας το μοντέλο στο σύνολο εκπαίδευσης και στη συνέχεια εφαρμόζοντας το μοντέλο στο σύνολο επικύρωσης.

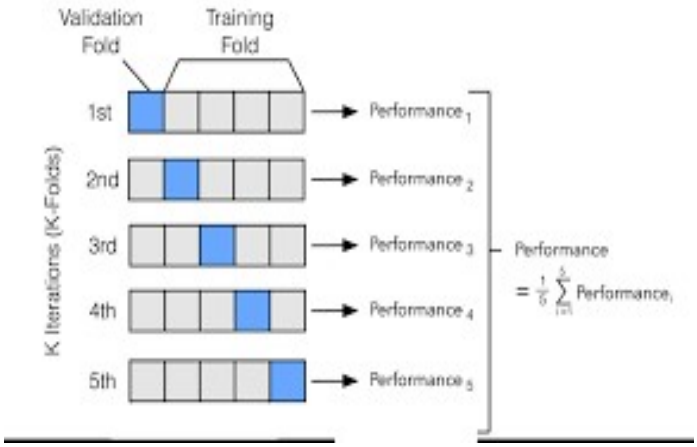
2.6.2 Διασταυρωμένη Επικύρωση

Η διασταυρούμενη επικύρωση είναι μια δημοφιλής τεχνική που επιτρέπει την αξιολόγηση της απόδοσης με βάση διαφορετικά υποσύνολα του συνόλου εκπαίδευσης. Στη διαδικασία αυτή, το σύνολο εκπαίδευσης διαιρείται σε k υποσύνολα, με τα k-1 υποσύνολα να χρησιμοποιούνται για την εκπαίδευση. Το υπόλοιπο υποσύνολο χρησιμοποιείται για την αξιολόγηση της απόδοσης του μοντέλου. Αυτή η διαδικασία επαναλαμβάνεται k φορές, και οι βαθμολογίες συνοψίζονται για κάθε σύνολο υπερπαραμέτρων προς αξιολόγηση.

Η διασταυρούμενη επικύρωση επιτρέπει την εξαγωγή πιο σταθερών και αξιόπιστων αποτελεσμάτων. Αποτελεί πολύτιμη εργαλειακή προσέγγιση, ιδίως όταν η δημιουργία ενός ξεχωριστού συνόλου επικύρωσης δεν είναι εφικτή λόγω του περιορισμένου μεγέθους του διαθέσιμου συνόλου δεδομένων. Με αυτόν τον τρόπο,

διασφαλίζεται ότι η επιλογή των υπερπαραμέτρων γίνεται με αξιοπιστία, ακόμη και σε περιβάλλοντα με περιορισμένο όγκο δεδομένων.

Η μέθοδος του Στατιστικού Συνόλου K-Fold είναι ένα αξιολογικό εργαλείο που χρησιμοποιείται για τον υπολογισμό και την αξιολόγηση της ικανότητας εκπαίδευσης ενός μοντέλου, καθώς και της γενικής του απόδοσης. Χρησιμοποιείται για σύγκριση των αποδόσεων διαφορετικών μοντέλων πρόβλεψης. Είναι μια μέθοδος υπολογιστικά απλή και μας δίνει άμεσα και κατανοητά αποτελέσματα. Η παράμετρος K (K-fold) λαμβάνει μια θετική ακέραια τιμή που καθορίζεται από τον χρήστη, με τις συνηθέστερες τιμές να είναι το k=5 και k=10.



Εικόνα 2.21: K-Fold Cross Validation

2.7 Εφαρμογές της Τεχνητής Νοημοσύνης στο Μάρκετινγκ

Οι επιχειρήσεις που επιδίδονται στον ψηφιακό μάρκετινγκ εκμεταλλεύονται εκτενώς την Τεχνητή Νοημοσύνη για τη συλλογή και ανάλυση μεγάλου όγκου δεδομένων. Η σημασία των πληροφοριών που αποκαλύπτονται από τα δεδομένα των χρηστών που κοινοποιούνται στο διαδίκτυο έχει ιδιαίτερο ενδιαφέρον για τους μάρκετερς τα τελευταία χρόνια. Ειδικότερα, η περαιτέρω κατανόηση του ψηφιακού αποτυπώματος των καταναλωτών και η ευρεία χρήση των διαδικτυακών υπηρεσιών μπορούν, μαζί με τη χρήση της τεχνητής νοημοσύνης, να βοηθήσουν στον σχεδιασμό εμπορικά επιτυχημένων προϊόντων και υπηρεσιών. Αυτή η πρακτική οδηγεί στη δημιουργία εκτεταμένης γνώσης, επιτρέποντας την αυτόματη λήψη αποφάσεων στον τομέα του μάρκετινγκ (Kühl et al., 2019).

Οι αλληλεπιδράσεις μεταξύ επιχειρήσεων και καταναλωτών γίνονται ολοένα πιο ατομικές και πανταχού παρόντες, δημιουργώντας εκτενή ψηφιοποιημένα ίχνη. Εξίσου, οι επιχειρήσεις αξιοποιούν έξυπνους πράκτορες και εικονικούς προσωπικούς βοηθούς για την εκτέλεση προηγμένων και ευφυών λειτουργιών. Αυτή η ολοκληρωμένη προσέγγιση επιτρέπει στις επιχειρήσεις να διαμορφώνουν ολοκληρωμένες στρατηγικές μάρκετινγκ, επιφέροντας σημαντική αύξηση στην αποτελεσματικότητα και την ανταγωνιστικότητά τους.

Η τεχνητή νοημοσύνη στο μάρκετινγκ κερδίζει επί του παρόντος σημαντικό έδαφος, λόγω της αυξανόμενης ισχύος υπολογιστών, του χαμηλότερου κόστους υπολογιστών, της διαθεσιμότητας των μεγάλων δεδομένων και της προόδου των αλγορίθμων και των μοντέλων μηχανικής μάθησης.

Ωστόσο, παρά το γεγονός ότι το ενδιαφέρον αυξάνεται γρήγορα, η χρήση μεθόδων μηχανικής μάθησης στο μάρκετινγκ βρίσκεται ακόμα σε πρώιμο στάδιο, και οι υπάρχουσες μελέτες είναι κάπως διασπασμένες. Μέχρι σήμερα, δεν φαίνεται να υπάρχει ενιαίο όραμα ή ενιαίο πλαίσιο για το πώς οι μέθοδοι μηχανικής μάθησης πρέπει να ενσωματώνονται στην έρευνα του μάρκετινγκ.

Οι εφαρμογές της τεχνητής νοημοσύνης στο μάρκετινγκ μπορούν να διαπερνούν δισεκατομμύρια σημεία δεδομένων στο διαδίκτυο και να περιγράφουν ακριβώς αυτό που χρειάζεται μια επιχείρηση. Οι εταιρείες επωφελούνται από την χρήση της Τεχνητής Νοημοσύνης αποτιμώντας τα μεγάλα δεδομένα σε πληροφορίες και γνώση, επιτρέποντας την ανάπτυξη πιο αποτελεσματικών στρατηγικών μάρκετινγκ και πωλήσεων, οι οποίες μεταφράζονται σε ανταγωνιστικό πλεονέκτημα (Paschen et al., 2020).

Τα μέσα κοινωνικής δικτύωσης και οι ηλεκτρονικές συσκευές έχουν αυξήσει δραματικά τις αλληλεπιδράσεις μεταξύ των επιχειρήσεων και των καταναλωτών, με τις πληροφορίες να κωδικοποιούνται σε πλούσιες μορφές πολυμέσων, όπως κείμενο, εικόνα και βίντεο. Είναι απαραίτητο για τις εταιρείες να κατανοούν τις δυνατότητες που προσφέρονται από τις προτιμήσεις των καταναλωτών και να αποκτούν πληροφόρηση σχετικά με τη θέση της μάρκας βασισμένες σε αυτό το πλούσιο πολυμεσικό περιεχόμενο.

Βασιζόμενη σε ιστορικά δεδομένα, η τεχνητή νοημοσύνη στο ψηφιακό μάρκετινγκ μπορεί να καθορίσει ποιο περιεχόμενο είναι πιθανότατα να επαναφέρει τους πελάτες στον ιστότοπο, ποια τιμή θα προκαλέσει τις περισσότερες μετατροπές σε πωλήσεις (ROI), πότε είναι η καλύτερη στιγμή για να δημοσιεύσει κάποιος, ποιος τίτλος θα προσελκύσει τη μεγαλύτερη προσοχή στο κοινό, κ.λ.π . Η τεχνητή νοημοσύνη μπορεί να αναγνωρίσει ποιοι πελάτες είναι πιθανότερο να ακυρώσουν μια συγκεκριμένη υπηρεσία και να αναλύσει τα χαρακτηριστικά εκείνα που προκαλούν αυτό το αποτέλεσμα.

Εφαρμογές του μάρκετινγκ, όπως η ψηφιακή αναζήτηση και διαφήμιση, η αλληλεπίδραση στα κοινωνικά μέσα, η καταγραφή μέσω κινητών, η online αγορά, αναπτύσσονται όλο και περισσότερο από επεκτάσιμους και έξυπνους αλγορίθμους, με τη βοήθεια τόσο μεγάλων τεχνολογικών εταιριών, όπως η Google και η Amazon.

Ως αποτέλεσμα αυτής της ανάλυσης, οι μάρκετερς μπορούν να σχεδιάσουν τις μελλοντικές τους καμπάνιες και να υιοθετήσουν πρακτικές που ενθαρρύνουν τους ανθρώπους να παραμείνουν. Συνεπάγεται επομένως ότι η τεχνητή νοημοσύνη έχει σημαντικά περισσότερες πιθανότητες να οδηγήσει μια επιχείρηση σε υψηλότερη απόδοση των επενδύσεων, μειώνοντας τα έξοδα της και βελτιώνοντας την αποτελεσματικότητα της.

Εικόνα 2.22: Εφαρμογές της Τεχνητής Νοημοσύνης στο Μάρκετινγκ

2.7.1 Προσωποποιημένο Μάρκετινγκ

Το μάρκετινγκ γίνεται όλο και πιο εξατομικευμένο. Οι λύσεις της τεχνητής νοημοσύνης δίνουν στους επαγγελματίες του μάρκετινγκ μια καλύτερη κατανόηση των πελατών τους, επιτρέποντάς τους να παραδίδουν το σωστό μήνυμα στον σωστό άνθρωπο τη σωστή στιγμή. Οι μάρκετερς μπορούν πλέον να επικεντρώνονται περισσότερο στον πελάτη και να ικανοποιούν τις ανάγκες του σε πραγματικό χρόνο. Χρησιμοποιώντας την τεχνητή νοημοσύνη, μπορούν γρήγορα να καθορίσουν ποιο περιεχόμενο να στοχεύσουν στους πελάτες και ποιο κανάλι να χρησιμοποιήσουν σε κάθε στιγμή, χάρη στα δεδομένα που συλλέγονται και παράγονται από τους αλγόριθμους της.

Η δημιουργία ενός πραγματικά συνεκτικού προφίλ καταναλωτή ενέχει τη συλλογή δεδομένων κατά τη διάρκεια κάθε αλληλεπίδρασης. Οι επαγγελματίες του μάρκετινγκ μπορούν να χρησιμοποιήσουν τις λύσεις της τεχνητής νοημοσύνης για να καλλιεργήσουν τις εκστρατείες μάρκετινγκ και να δημιουργήσουν υψηλά εξατομικευμένο περιεχόμενο, πηγαίνοντας ένα βήμα παραπέρα τα προφίλ αυτά. Οι μάρκετερς μπορούν να διασφαλίσουν ότι ασχολούνται με τις σωστές κατηγορίες καταναλωτών που είναι πιθανότατα να συμπεριφερθούν και να ανταποκριθούν θετικά στη διαφήμιση, καθώς αναπτύσσουν στοχευμένες στρατηγικές διαφήμισης. Οι επαγγελματίες του χώρου πετυχαίνουν αυτό εκμεταλλευόμενοι την ψηφιακή υπερνοημοσύνη των μοντέλων και των αλγορίθμων της τεχνητής νοημοσύνης.

Η τεχνητή νοημοσύνη τους βοηθά να εντοπίζουν γρήγορα ποιοι πελάτες είναι πιθανό να είναι πιο κατάλληλοι, να αναπτύσσουν καλύτερες τακτικές ενθάρρυνσης και να δημιουργούν σχετικό περιεχόμενο όταν ενσωματώνεται στα εργαλεία αυτοματισμού μάρκετινγκ. Τα δυναμικά emails με περιεχόμενο, ιδιαίτερα τα ατομικά προσωποποιημενα emails, είναι τα πιο αποτελεσματικά, διότι χρησιμοποιούν προσαρμοσμένα emails, ενώ ταυτόχρονα στοχεύουν σε αυτό που θέλουν να ακούσουν οι συνδρομητές. Οι στρατηγικές δυναμικού περιεχομένου εξασφαλίζουν ότι τα emails παραμένουν σχετικά με τους συνδρομητές, βασιζόμενα στις γεωγραφικές τους τοποθεσίες, τη ψυχογραφία τους, τα συμπεράσματα από τη συμπεριφορά τους και τις αντιλήψεις τους.

Παράδειγμα αποτελεί η LEGO, όπου το 2018, χρησιμοποίησε το Watson Ads Omni για να δημιουργήσει διαφημίσεις με διάδραση που χρησιμοποιούν την τεχνητή νοημοσύνη (Sweeney, 2018). Το σύστημα τεχνητής νοημοσύνης εκπαιδεύτηκε με τη γνώση ενός ευρέος φάσματος διαφορετικών προϊόντων LEGO, με διαφημίσεις προσαρμοσμένες στους καταναλωτές ανάλογα με τα συγκεκριμένα ενδιαφέροντά και τις ανάγκες τους. Το πλεονέκτημα μιας τέτοιας εφαρμογής είναι ότι η μάρκα μπορεί να έχει σημαντικές, ατομικές συζητήσεις με τους καταναλωτές κατά τη διάρκεια των

διαδρομών τους προς την αγορά. Πηγαίνοντας ένα βήμα παραπέρα στη δημιουργία εξατομικευμένου περιεχομένου, η NBA συνεργάστηκε με την εταιρεία WSC Sports, το 2015, για να προσφέρει στους οπαδούς σχεδόν άμεσα βίντεο με τα καλύτερα στιγμιότυπα από τα παιχνίδια μέσω των ιστότοπων της NBA. Χρησιμοποιώντας την τεχνητή νοημοσύνη, δημιουργήθηκαν πακέτα επιλογών για κάθε παίκτη σε ένα παιχνίδι, επιτρέποντας στην NBA να παρέχει εξατομικευμένα στιγμιότυπα σε ένα παγκόσμιο κοινό, π.χ. αποστέλλοντας βίντεο με αστέρια του NBA που είναι γεννημένα στην Αυστραλία σε Αυστραλούς θεατές που δεν καλύπτονταν ικανοποιητικά από το παραδοσιακό τηλεοπτικό υλικό (NBA, 2015).

2.7.2 Χαρτογράφηση Ταξιδιού Χρήστη

Η σύνθεση του συνολικού πλάνου του ταξιδιού του καταναλωτή είναι ιδιαιτέρως αξιόλογη για τις επιχειρήσεις, επιτρέποντάς τους να παρακολουθούν και να καθοδηγούν τον καταναλωτή, παρέχοντας τη σωστή πληροφορία, υπηρεσία και προώθηση στο κατάλληλο στάδιο και πλαίσιο. Εστιάζοντας σε ολόκληρο το ταξίδι, μπορεί να αποδειχθούν αποτελεσματικές οι περισσότερες πτυχές για τη δημιουργία διαρκούς συναισθηματικής σύνδεσης. Οι μέθοδοι deep learning και reinforcement learning μπορούν να βοηθήσουν τις επιχειρήσεις να κατανοήσουν και να διαχειριστούν ολόκληρο το ταξίδι απόφασης του χρήστη. Οι μέθοδοι μηχανικής μάθησης διαδραματίζουν καίριο ρόλο στη διαχείριση των αποφάσεων των καταναλωτών, παρέχοντας στους πελάτες έξυπνη, απλή και βολική υποστήριξη σε κάθε στάδιο του ταξιδιού τους.

Συνεπώς, συστήνεται η εφαρμογή μεθόδων μηχανικής μάθησης για να συνδράμουν στην χαρτογράφηση ολόκληρης της διαδρομής αγοράς του πελάτη, ειδικά στα αρχικά στάδια, με σκοπό την ανάπτυξη δυνατοτήτων υποστήριξης αποφάσεων που καλύπτουν όλες τις πτυχές των λειτουργιών του μάρκετινγκ και την εκτέλεση ολιστικής ανάλυσης της δομής της αγοράς, συμπεριλαμβανομένου του θέματος της θέσης της μάρκας και της ανταγωνιστικής ανάλυσης.

2.7.3 Συμμετοχή Χρήστη

Καθώς οι εταιρείες επικεντρώνονται στα ταξίδια απόφασης των πελατών, έξυπνοι πράκτορες βοηθούν στην αλληλεπίδραση με τους πελάτες, καθ' όλη τη διάρκεια διαδρομής του ταξιδιού, για τη βελτίωση της εμπειρίας του χρήστη.

Με την υποστήριξη προηγμένων μηχανισμών τεχνητής νοημοσύνης στο νέφος, εικονικοί βοηθοί όπως η Alexa ανταποκρίνονται στις φωνητικές ερωτήσεις των καταναλωτών για να παρέχουν πληροφορίες ή να πραγματοποιούν αγορές. Τα chatbots που ενεργοποιούνται από αλγόριθμους αναγνώρισης ομιλίας και

επεξεργασίας φυσικής γλώσσας αναλαμβάνουν όλο και περισσότερο την χειριστική εξυπηρέτηση πριν και μετά την αγορά. Σε όλο το ταξίδι του πελάτη, οι καινοτομίες που κινούνται από την τεχνητή νοημοσύνη ανασχηματίζουν γρήγορα τις πρακτικές της αλληλεπίδρασης.

Σήμερα, τα chatbots χρησιμοποιούνται στην εξυπηρέτηση πελατών για την αντιμετώπιση των περισσότερων απλών ερωτήσεων. Τα chatbots με χρήση της τεχνητής νοημοσύνης βελτιώνονται και γίνονται όλο και πιο έξυπνα συνεχώς με την πάροδο του χρόνου. Είναι ευέλικτα, προσαρμόσιμα και έξυπνα, προσφέροντας στους χρήστες μια πιο φυσική εμπειρία. Τα chatbots αποδεικνύονται ως πολύτιμη συμβολή για τις επιχειρήσεις, καθώς αποτελούν εξαιρετικά εργαλεία συλλογής δεδομένων που μειώνουν σημαντικά τις ανάγκες προσωπικού, αμβλύνοντας τα εμπόδια, μειώνοντας έτσι το κόστος εξυπηρέτησης πελατών, αλλά η επίδρασή τους στην ικανοποίηση του πελάτη μπορεί να είναι ποικίλη.

Σύμφωνα με έρευνα, πολλοί καταναλωτές προτιμούν ακόμη να επικοινωνούν με ανθρώπινους πράκτορες για πιο περίπλοκα αιτήματα. Στο Ηνωμένο Βασίλειο, σχεδόν το 50% προτιμά έναν άνθρωπο απέναντι σε ένα chatbot, και στις Ηνωμένες Πολιτείες, το 40% προτιμά έναν άνθρωπο απέναντι σε ένα chatbot (Elliott, 2018). Παρά τις προτιμήσεις των καταναλωτών για ανθρώπινες αλληλεπιδράσεις, η τεχνητή νοημοσύνη μπορεί ακόμα να παρέχει υποστήριξη στην εξυπηρέτηση πελατών. Για παράδειγμα, η τεχνητή νοημοσύνη μπορεί να δράσει για την ανάθεση πρακτόρων σε πελάτες. Αυτή η διαδικασία μπορεί να εξασφαλίσει ότι οι πελάτες συνδέονται με έναν πράκτορα που έχει την εμπειρογνωμοσύνη να αντιμετωπίσει τις ανάγκες τους. Ένα τέτοιο παράδειγμα αποτελεί η χρήση συστημάτων ταξινόμησης που χρησιμοποιούν επεξεργασία φυσικής γλώσσας για τον εντοπισμό των προβλημάτων που αναφέρουν οι πελάτες. Με τον καλύτερο συντονισμό των πρακτόρων με τους πελάτες, η τεχνητή νοημοσύνη μπορεί να διευκολύνει την αλληλεπίδραση και να διατηρήσει την αξία για τις εταιρείες.

2.7.4 Συστάσεις Χρήστη

Η προτεινόμενη σε συγκεκριμένους καταναλωτές σωστή προώθηση προϊόντων μπορεί σημαντικά να βελτιώσει την απόδοση του μάρκετινγκ. Η προβλεπτική ανάλυση, ως εφαρμογή της τεχνητής νοημοσύνης στο μάρκετινγκ, έχει τη δυνατότητα να απελευθερώσει μια ισχυρή έλξη σε όλες τις δραστηριότητες στο μάρκετινγκ. Η προβλεπτική ανάλυση που υποστηρίζεται από την τεχνητή νοημοσύνη μπορεί να πάρει υπάρχοντα δεδομένα και να αντλήσει τεράστια αξία από αυτά. Η προβλεπτική αξιολόγηση δυνητικών πελατών με την υποστήριξη της τεχνητής νοημοσύνης είναι μία από τις πιο δημοφιλείς εφαρμογές της τεχνητής νοημοσύνης στο μάρκετινγκ. Είναι μια καινοτόμος προσέγγιση για την ταξινόμηση και αξιολόγηση πιθανών

πελατών. Αναλύοντας τις πληροφορίες εκατομμυρίων καταναλωτών και προϊόντων για να αξιολογήσουν τη σχετικότητα, τα συστήματα προτάσεων είναι πλέον ένα ουσιώδες συστατικό στο μάρκετινγκ, αντιστοιχώντας αποτελεσματικά προϊόντα και καταναλωτές σε όλα τα ψηφιακά κανάλια.

Η λήψη αποφάσεων και η διαχείριση των πελατών είναι πλεονεκτήματα της προσέγγισης του μάρκετινγκ με τη χρήση της τεχνητής νοημοσύνης. Τα δεδομένα είναι κρίσιμα για τη βελτίωση των προτύπων του υλικού που συνιστάται στους πελάτες από αλγόριθμους μηχανικής μάθησης. Καθώς η τεχνητή νοημοσύνη μπορεί να βοηθήσει τις επιχειρήσεις να προβλέπουν τι θα αγοράσουν οι πελάτες, η χρήση της πρέπει να οδηγήσει σε σημαντικές βελτιώσεις στην ικανότητα πρόβλεψης. Ανάλογα με τα επίπεδα προβλεπτικής ακρίβειας, οι εταιρείες μπορεί ακόμα και να αλλάξουν σημαντικά τα μοντέλα επιχείρησής τους, παρέχοντας αγαθά και υπηρεσίες στους πελάτες συνεχώς βασισμένες σε δεδομένα και προβλέψεις για τις ανάγκες τους.

Η Lily AI αποτελεί ένα εργαλείο που βοηθά στη διαμόρφωση προϊόντων σε διαδικτυακές ρυθμίσεις. Συγκεκριμένα, η Lily AI επιτρέπει στους λιανοπωλητές μόδας να ενθαρρύνουν τους καταναλωτές να ολοκληρώσουν το look κατά την ολοκλήρωση της αγοράς. Η βαθιά κατανόηση των επιλογών των αγοραστών αναφορικά με όλες τις κατηγορίες ενδυμάτων από την Lily AI της επιτρέπει να δημιουργεί προτάσεις ενδυμασίας σε πραγματικό χρόνο που βοηθούν τους λιανοπωλητές να αυξήσουν το μέγεθος του καλαθιού κατά την ολοκλήρωση της αγοράς. Η τεχνολογία AI εφαρμόζεται επίσης στις στρατηγικές προϊόντων εντός του φυσικού καταστήματος.

2.7.5 Βελτιστοποίηση Λήψης Αποφάσεων

Η τεχνητή νοημοσύνη επιτρέπει στους ανθρώπους να αποκτήσουν καλύτερη κατανόηση και, ως αποτέλεσμα, να συμβάλουν σε καλύτερες αποφάσεις με την ανάλυση ποσοτικών και ποιοτικών δεδομένων.

Η τεχνητή νοημοσύνη μπορεί επίσης να εντοπίζει ποια προϊόντα να κατασκευάσει. Παράδειγμα αποτελεί η εταιρεία γρήγορης μόδας Choosy, η οποία αντλεί έμπνευση στον τομέα της μόδας σχεδόν αποκλειστικά από τις κορυφαίες τάσεις που εμφανίζονται στο Instagram, κυκλοφορώντας 10 στυλ κάθε εβδομάδα που οι πελάτες μπορούν να παραγγείλουν πριν προχωρήσουν στην παραγωγή (Pallant & Sands, 2018). Με το να δημιουργεί μόνο προϊόντα που οι πελάτες έχουν δεσμευτεί να αγοράσουν, η Choosy αποφεύγει την επισφάλεια των υπερβολικών αποθεμάτων και εκμεταλλεύεται τα οφέλη της μαζικής προσαρμογής ενώ μειώνει τον κίνδυνο.

2.7.6 Πολιτική Διαχείρισης Τιμών

Οι επιχειρήσεις χρησιμοποιούν δυναμικούς αλγορίθμους τιμολόγησης για να φτάσουν σε βέλτιστες τιμές για τα προϊόντα ή τις υπηρεσίες τους, προκειμένου να παραμείνουν ανταγωνιστικές και να αυξήσουν γρήγορα την κερδοφορία τους, μεγιστοποιώντας τις πωλήσεις τους. Αποτελεί μία από τις πιο κερδοφόρες εφαρμογές τεχνητής νοημοσύνης στον τομέα του μάρκετινγκ.

Η τιμολόγηση προϋποθέτει τον συνδυασμό πολλαπλών παραγόντων για τον καθορισμό της βέλτιστης τιμής και είναι μια δραστηριότητα υπολογιστικού χαρακτήρα. Κατά την ανάπτυξη της στρατηγικής τιμολόγησης, οι μάρκετερς αποφασίζουν πόσο να χρεώσουν τα προϊόντα και τις υπηρεσίες. Προσπαθούν να κατανοήσουν την ευαισθησία των καταναλωτών στις τιμές και χαρτογραφούν τις τιμές των ανταγωνιστών. Η τεχνολογία της Τεχνητής Νοημοσύνης μπορεί να βοηθήσει με διάφορους τρόπους, συμπεριλαμβανομένης της εκτίμησης της ελαστικότητας των τιμών των καταναλωτών, τη δυναμική τιμολόγηση (π.χ. άνοδος τιμών) καθώς και τον εντοπισμό ανωμαλιών στις τιμές (συμπεριλαμβανομένων σφαλμάτων τιμολόγησης, περιπτώσεων απάτης και μη κερδοφόρων πελατών). Η τεχνολογία AI επιτρέπει στους μάρκετερ να παρακολουθούν τις τάσεις της αγοράς και να καθορίζουν πιο ανταγωνιστικά σημεία ισορροπίας για να επηρεάσουν τους καταναλωτές στο σημείο της απόφασης (Arevalillo, 2019).

Ο αλγόριθμος "multiarmed bandit" βασισμένος σε τεχνητή νοημοσύνη μπορεί να προσαρμόζει δυναμικά την τιμή σε πραγματικό χρόνο (Misra κ.ά., 2019). Σε συνθήκες συχνής αλλαγής τιμολογιακού περιβάλλοντος, η εκτίμηση σε αλγόριθμους μηχανικής μάθησης μπορεί να προσαρμόζει γρήγορα τα σημεία ισορροπίας της τιμής για να ταιριάξουν με τις τιμές του ανταγωνιστή (Bauer & Jannach, 2018). Σύμφωνα με τον Dekimpe (2020), οι αλγόριθμοι τιμολόγησης "best response" ενσωματώνουν τις επιλογές του πελάτη, τις στρατηγικές του ανταγωνιστή και το δίκτυο προσφοράς με σκοπό τον βέλτιστο δυναμικό προσδιορισμό τιμολόγησης.

Στο πλαίσιο των ξενοδοχείων, η δυναμική τιμολόγηση μπορεί να επιτρέπει την αντιμετώπιση του προβλήματος της υποκατάληψης προσαρμόζοντας τις τιμές για να ισορροπήσει την προσφορά και τη ζήτηση και να μεγιστοποιήσει το κέρδος (O'Hear, 2017). Με σκοπό την υποστήριξη των αποφάσεων τιμολόγησης, η Airbnb χρησιμοποιεί τεχνητή νοημοσύνη για να βοηθήσει τους οικοδεσπότες να καταλήξουν σε αποφάσεις τιμολόγησης για την ιδιοκτησία τους. Συμμερίζεται το γεγονός ότι, η τιμολόγηση είναι ένα πολύπλοκο διαδικαστικό για τους οικοδεσπότες, λαμβάνοντας υπόψη παραδοσιακούς παράγοντες ζήτησης (Hill, 2015).

2.7.7 Πολιτική Διανομής και Προώθησης Προϊόντων

Η πρόσβαση στο προϊόν και η διαθεσιμότητα του είναι κρίσιμες συνιστώσες του μάρκετινγκ για την αυξημένη ικανοποίηση του πελάτη. Η τεχνητή νοημοσύνη επιτρέπει στους μάρκετερς να προβλέπουν και να βελτιστοποιούν τη διανομή, το απόθεμα, τις εκθέσεις καταστημάτων και τις διατάξεις καταστημάτων (τόσο φυσικών όσο και διαδικτυακών). Σήμερα, οι έμποροι μπορούν να χρησιμοποιούν σχέδια καταστρώματος πληροφοριών που ενημερώνονται από την τεχνητή νοημοσύνη, ή δυναμικά σχέδια που συνιστούν τον ιδανικό αριθμό και τη θέση του αποθέματος εντός και εκτός καταστήματος (McDowell, 2019).

Η διανομή του προϊόντος εξαρτάται από συσχετισμένες σχέσεις, λογιστική, διαχείριση αποθεμάτων, αποθήκευση και προβλήματα μεταφορών, τα οποία είναι σε μεγάλο βαθμό μηχανικά και επαναληπτικά. Η τεχνητή νοημοσύνη είναι η ιδανική λύση στη διαχείριση του τόπου παροχής μέσω της προσφοράς συνεργατικών ρομπότ για συσκευασία, drones για παράδοση, IoT για παρακολούθηση παραγγελιών και αναπλήρωση παραγγελιών (Huang & Rust, 2020).

Στη Walmart, μερικά καταστήματα άρχισαν να δοκιμάζουν αυτόνομα ρομπότ που σαρώνουν τα ράφια για τις θέσεις που χρειάζονται ανανέωση (McDowell, 2019). Το Prime Air της Amazon.com χρησιμοποιεί drones για την αυτοματοποίηση της αποστολής και παράδοσης. Η Domino’s Pizza πειραματίζεται με αυτόνομα αυτοκίνητα και ρομπότ παράδοσης για να παραδίδει πίτσα στην πόρτα του πελάτη. Ο οίκος μόδας Levi's χρησιμοποιεί αλγόριθμους για τη βελτιστοποίηση του τρόπου όπου τα προϊόντα διατίθενται στο κατάστημα και για τη βελτίωση της διαθεσιμότητας των μεγεθών. Η Nike χρησιμοποιεί γεωγραφικά και συμπεριφορικά δεδομένα από την εφαρμογή της για να ενημερώνεται σχετικά με τις προσφορές στο κατάστημα και χρησιμοποιεί αλγόριθμους συσταδοποίησης για να παρέχει συμβουλές σχετικά με τα αντικείμενα που πρέπει να εκτίθενται μαζί (McDowell, 2019).

2.7.8 Λειτουργίες Διαχείρισης Σχέσεων Πελατών

Οι λειτουργίες Διαχείρισης Σχέσεων Πελατών (CRM) επωφελήθηκαν μέσω της Τεχνητής Νοημοσύνης στη Διεπαφή Χρήστη (Seranmadevi και Kumar, 2019). Η Τεχνητή Νοημοσύνη και το Διαδίκτυο των Πραγμάτων (IoT) μετέτρεψαν τα παραδοσιακά καταστήματα λιανικής σε έξυπνα καταστήματα λιανικής. Τα έξυπνα καταστήματα λιανικής ανέβασαν την εμπειρία του πελάτη και την ευκολία των αγορών, καθώς και τη βελτίωση της αλυσίδας εφοδιαστικής (Sujata και συν., 2019).

Η διαχείριση των σχέσεων με τους πελάτες (CRM) περιλαμβάνει τις διαδικασίες και τα συστήματα που υποστηρίζουν μια στρατηγική που επιδιώκει τη δημιουργία

κερδοφόρων μακροπρόθεσμων σχέσεων με συγκεκριμένους πελάτες. Η σημασία του CRM έχει αυξηθεί δεδομένης της αυξανόμενης επίγνωσής μας ότι η κατάκτηση πελατών κοστίζει περισσότερο από τη διατήρηση υπαρχόντων πελατών (Ling & Yen, 2001). Με αυτή την άποψη, αρκετά αποτελέσματα της τεχνητής νοημοσύνης προσδοκούν να συμβάλουν στη βελτίωση των σχέσεων με τους πελάτες.

Εκμεταλλευόμενη τη δυνατότητα της τεχνητής νοημοσύνης να προβλέπει ποιοι πελάτες είναι πιο πιθανό να ανταποκριθούν σε διαφημιστικές καμπάνιες, το CRM επικεντρώνεται στη χρήση νέων τεχνολογιών και μεθόδων (Chatterjee et al., 2019). Έτσι, οι πρόσφατες εξελίξεις στην τεχνολογία ενισχύουν τη δυναμική του CRM μέσω της αποτελεσματικής χρήσης των διαθέσιμων δεδομένων κι της έντονης αλληλεπίδρασης με έναν τρόπο που προάγει τις σχέσεις με τους πελάτες (Bock et al., 2020, Kaplan και Haenlein, 2019) και τελικά επιτρέπει την εστίαση στον πελάτη (Latinovic & Chatterjee, 2019). Από στρατηγικής άποψης, η Τεχνητή Νοημοσύνη (ΤΝ) γίνεται όλο και πιο σημαντική στο μάρκετινγκ.

3. Στόχος της Έρευνας – Μεθοδολογία

3.1 Πηγή Δεδομένων

Αυτή η ερευνητική εργασία εμβαθύνει στο πολύπλοκο τοπίο της τμηματοποίηση των πελατών και των προτύπων συμπεριφοράς με πρωταρχικό στόχο τη βελτίωση των στρατηγικών μάρκετινγκ για βελτιωμένη ικανοποίηση και αφοσίωση των πελατών. Χρησιμοποιώντας προηγμένες τεχνικές ομαδοποίησης, αναλύουμε συστηματικά ένα ποικίλο σύνολο δεδομένων που περιλαμβάνει δημογραφικά χαρακτηριστικά, αγοραστικές συμπεριφορές και επίπεδα ικανοποίησης μιας πελατειακής βάσης.

Η προσαρμογή των προσεγγίσεων με βάση τα προσδιορισμένα τμήματα πελατών μπορεί να βελτιστοποιήσει την κατανομή των πόρων και να αυξήσει την αποτελεσματικότητα των προσπαθειών προώθησης. Τα ευρήματα υπογραμμίζουν τη σημασία του εξατομικευμένου μάρκετινγκ στο σημερινό δυναμικό καταναλωτικό τοπίο, δίνοντας έμφαση στις δυνατότητες αυξημένης ικανοποίησης των πελατών και μακροπρόθεσμης δέσμευσης.

Το dataset παρέχει μια συνολική εικόνα της συμπεριφοράς του πελάτη μέσα σε μια πλατφόρμα ηλεκτρονικού εμπορίου. Κάθε εγγραφή αντιπροσωπεύει και ένα ξεχωριστό πελάτη, στον οποίο αντιστοιχούν διάφορα δεδομένα αντίστοιχα (όπως το φύλο, η ηλικία, κτλ). Παρακάτω, παραθέτονται αναλυτικά αρκετές πληροφορίες για τις ανεξάρτητες μεταβλητές:

Όνομα μεταβλητής : Ταυτότητα Πελάτη (ID)
Τύπος : Αριθμητικός
Περιγραφή: Μια μοναδική ταυτοποίηση που ανατίθεται σε κάθε πελάτη, εξασφαλίζοντας τη διάκριση σε όλο το dataset.

Όνομα μεταβλητής : Φύλο (Gender)
Τύπος : Κατηγορικό (Άνδρας, Γυναίκα)
Περιγραφή : Καθορίζει το φύλο του πελάτη, επιτρέποντας αναλύσεις βασισμένες στο φύλο.

Όνομα μεταβλητής : Ηλικία (Age)
Τύπος : Αριθμητικός
Περιγραφή: Αναπαριστά την ηλικία του πελάτη, επιτρέποντας εισόδους ειδικές για ομάδες ηλικίας.

Όνομα μεταβλητής : Πόλη (City)
Τύπος : Κατηγορικός (Ονόματα Πόλεων)

Περιγραφή : Υποδεικνύει την πόλη κατοικίας κάθε πελάτη, παρέχοντας γεωγραφικά εισαγωγικά.

Όνομα μεταβλητής : Τύπος Συνδρομής (Membership Type)
Τύπος: Κατηγορικός (Χρυσό, Ασημένιο, Χάλκινο)
Περιγραφή : Αναγνωρίζει τον τύπο συνδρομής που διατηρεί ο πελάτης επηρεάζοντας τα προνόμια και τα ευεργετήματα.

Όνομα μεταβλητής : Συνολική Δαπάνη (Total Spend)
Τύπος : Αριθμητικός
Περιγραφή : Καταγράφει το συνολικό χρηματικό ποσό που ξοδεύει ο πελάτης στην πλατφόρμα ηλεκτρονικού εμπορίου.

Όνομα μεταβλητής : Αντικείμενα Αγορασμένα (Items Purchased)
Τύπος : Αριθμητικός
Περιγραφή : Καταμετρά το συνολικό αριθμό αντικειμένων που αγοράζει ο πελάτης.

Όνομα μεταβλητής : Μέση Βαθμολογία (Average Rating)
Τύπος : Αριθμητικός (0 έως 5, με δεκαδικά)
Περιγραφή : Αντιπροσωπεύει τη μέση βαθμολογία που δίνεται από τον πελάτη για τα αγορασμένα αντικείμενα, μετρώντας την ικανοποίηση.

Όνομα μεταβλητής : Εφαρμογή Έκπτωσης (Discount Applied)
Τύπος : Λογικό (Αληθές, Ψευδές)
Περιγραφή : Υποδεικνύει ότι εφαρμόστηκε έκπτωση στην αγορά του πελάτη, επηρεάζοντας τη συμπεριφορά αγοράς.

Όνομα μεταβλητής : Ημέρες Από την Τελευταία Αγορά (Days Since Last Purchase)
Τύπος : Αριθμητικός
Περιγραφή : Αντικατοπτρίζει τον αριθμό των ημερών που έχουν περάσει από την τελευταία αγορά του πελάτη, βοηθώντας στην ανάλυση διατήρησης.

Όνομα μεταβλητής : Επίπεδο Ικανοποίησης (Satisfaction Level)
Τύπος: Κατηγορικός (Ικανοποιημένος, Ουδέτερος, Ανικανοποίητος)
Περιγραφή : Καταγράφει το συνολικό επίπεδο ικανοποίησης του πελάτη, παρέχοντας ένα υποκειμενικό μέτρο της εμπειρίας του.

3.2 Βήματα Υλοποίησης

Η υλοποίηση αυτής της έρευνας περιλαμβάνει μια συστηματική και μεθοδική προσέγγιση για τη μόχλευση προηγμένων τεχνικών ανάλυσης δεδομένων και την

εξαγωγή σημαντικών γνώσεων από το σύνολο των δεδομένων. Οι ακόλουθες παράγραφοι αναλύουν τα βήματα που θα εφαρμοστούν στη διαδικασία υλοποίησης.

3.2.1 Προεπεξεργασία Δεδομένων

Το ταξίδι ξεκίνησε με μια ολοκληρωμένη φάση προεπεξεργασίας δεδομένων. Αυτό περιλάμβανε χειρισμό τιμών που λείπουν, κωδικοποίηση κατηγορικών μεταβλητών και τυποποίηση αριθμητικών χαρακτηριστικών για να διασφαλιστεί ένα ομοιόμορφο και καθαρό σύνολο δεδομένων. Επιπλέον, εφαρμόστηκε τυποποίηση σε αριθμητικά χαρακτηριστικά για να τα φέρει σε μια κοινή κλίμακα, επιτρέποντας την εφαρμογή αλγορίθμων ομαδοποίησης.

3.2.1.1 Έλεγχος Κενών Τιμών

Ο έλεγχος για κενές τιμές (NA) αποτελεί ένα σημαντικό βήμα κατά την ανάλυση δεδομένων και την επεξεργασία πληροφοριών. Οι απουσιάζουσες τιμές μπορούν να επηρεάσουν σοβαρά την ακρίβεια και τη συνέπεια των αποτελεσμάτων. Κατά τη διάρκεια του ελέγχου για NA, αναζητούμε και αναγνωρίζουμε τυχόν κενές ή μη διαθέσιμες τιμές στον πίνακα δεδομένων μας.

Η διαδικασία αυτή απαιτεί τη χρήση ειδικών μεθόδων και συναρτήσεων, καθώς και την εφαρμογή προσεκτικών στρατηγικών για την αντιμετώπιση των NA, όπως αντικατάσταση με μέσους όρους, διαγραφή συγκεκριμένων γραμμών ή στηλών, ή ακόμη και ενδεχομένως τη χρήση προηγμένων αλγορίθμων πρόβλεψης για την αντικατάσταση των απουσιαζουσών τιμών. Ένας συστηματικός έλεγχος για NA συμβάλλει στην εξασφάλιση της αξιοπιστίας και της εγκυρότητας των αναλύσεων μας, επιτρέποντάς μας να λαμβάνουμε αποφάσεις βασισμένες σε πλήρη και ακριβή δεδομένα.

```
cleaned_df = remove_empty_rows(df)
df=cleaned_df
display_data(df)
df.isnull().sum()
```

```
df.isnull().sum()
```

```
Customer ID                  0
Gender                       0
Age                          0
City                         0
Membership Type              0
Total Spend                  0
Items Purchased              0
Average Rating               0
Discount Applied             0
Days Since Last Purchase     0
Satisfaction Level           0
dtype: int64
```

Πίνακας 3.1: Έλεγχος Κενών-Απουσιαζουσών Τιμών

3.2.1.2 Αντικατάσταση των Ονομάτων των Πόλεων με Συντομογραφίες

Η διαδικασία της αντικατάστασης των ονομάτων πόλεων με συντομογραφίες είναι μια κοινή πρακτική στην οπτικοποίηση δεδομένων, ιδιαίτερα όταν δημιουργούμε γραφήματα με περιορισμένο χώρο για ετικέτες. Αυτή η τεχνική βοηθά στον εξ ορθολογισμό και την απομάκρυνση των οπτικών αναπαραστάσεων, καθιστώντας τις πιο συνοπτικές και οπτικά ελκυστικές. Αντί να εμφανίζονται τα πλήρη ονόματα των πόλεων, τα οποία μπορεί να είναι μεγάλα και μπορεί να καταστήσουν την πλοκή, η χρήση συντομεύσεων παρέχει μια συμπαγή αλλά ενημερωτική αναπαράσταση γεωγραφικών τοποθεσιών.

Χρησιμοποιώντας συντομογραφίες πόλεων, οι αναλυτές δεδομένων και οι ερευνητές μπορούν να μεταφέρουν πληροφορίες πιο αποτελεσματικά χωρίς να θυσιάζουν τη σαφήνεια. Επιπλέον, η χρήση συντομογραφιών σε πόλεις συμβάλλει σε μια καθαρότερη και πιο επαγγελματική εμφάνιση, ενισχύοντας τη συνολική οπτική επίδραση της παρουσίασης δεδομένων.

```
df['City'] = df['City'].map({ 'New York': 'NY', 'Los Angeles':
'LA', 'Chicago': 'CH', 'San Francisco': 'SF', 'Miami': 'MI'})
display_data(df)
```

3.2.1.3 Κωδικοποίηση Ετικέτας με Αριθμητική Σειρά

Η κωδικοποίηση ετικετών με αριθμητική σειρά είναι μια μέθοδος που χρησιμοποιείται στην προεπεξεργασία δεδομένων για τη μετατροπή κατηγορικών μεταβλητών σε αριθμητική μορφή, διατηρώντας παράλληλα την εγγενή σειρά ή ιεραρχία μεταξύ διαφορετικών κατηγοριών. Σε αυτήν την τεχνική, σε κάθε μοναδική κατηγορία αποδίδεται μια αριθμητική ετικέτα με βάση τη φυσική της σειρά. Αυτή η προσέγγιση είναι ιδιαίτερα χρήσιμη όταν έχουμε να κάνουμε με τακτικά δεδομένα, όπου οι κατηγορίες έχουν μια σημαντική αλληλουχία ή κατάταξη.

Η διαδικασία κωδικοποίησης ετικετών περιλαμβάνει την αντιστοίχιση ακεραίων σε κατηγορίες με τρόπο που να αντικατοπτρίζει τη φυσική τους σειρά. Αυτή η αριθμητική αναπαράσταση διευκολύνει την εφαρμογή αλγορίθμων μηχανικής μάθησης, καθώς πολλά μοντέλα απαιτούν αριθμητική εισαγωγή.

Η κωδικοποίηση ετικετών με αριθμητική σειρά χρησιμοποιείται συνήθως σε σενάρια, όπου τα κατηγορικά χαρακτηριστικά παρουσιάζουν σαφή ιεραρχία, όπως τα επίπεδα εκπαίδευσης (π.χ. γυμνάσιο, πτυχίο, μεταπτυχιακό). Αντιπροσωπεύοντας αυτές τις κατηγορίες με ταξινομημένες αριθμητικές ετικέτες, επιτρέπει στα μοντέλα μηχανικής εκμάθησης να κατανοούν καλύτερα και να χρησιμοποιούν την εγγενή δομή των δεδομένων, οδηγώντας σε πιο ακριβείς και ουσιαστικές προβλέψεις.

```
FindUniques (df,'Satisfaction Level')
FindUniques(df,'Membership Type')
df['Satisfaction Level'] = df['Satisfaction Level'].map({
'Unsatisfied': 1, 'Neutral': 2, 'Satisfied': 3})
df['Membership Type'] = df['Membership Type'].map({ 'Bronze': 1,
'Silver': 2, 'Gold': 3})
display_data(df)
```

3.2.1.4 Κωδικοποίηση Ετικέτας για Σωστό / Λάθος

Η διαδικασία αντικατάστασης των τιμών 1 και 0 με τις λογικές τιμές True και False αποτελεί συνήθη πρακτική στον τομέα της επεξεργασίας δεδομένων και της μηχανικής μάθησης. Συνήθως, αυτό γίνεται για να καθιστά πιο κατανοητές και ερμηνεύσιμες τις τιμές αυτές, καθιστώντας τον κώδικα πιο ευανάγνωστο και τις διαδικασίες αντιληπτές.

```
df=replace_bool_with_numbers(df)
display_data(df)
```

3.2.1.5 Κωδικοποίηση Ετικέτας με One Hot Encoding

Η One Hot Encoding είναι μια δημοφιλής τεχνική που χρησιμοποιείται στην προεπεξεργασία δεδομένων, ειδικά στο πλαίσιο της μηχανικής μάθησης, για τη μετατροπή κατηγορικών μεταβλητών σε μορφή κατάλληλη για αριθμητική ανάλυση.

Σε αυτή τη μέθοδο, κάθε μοναδική κατηγορία μετατρέπεται σε ένα δυαδικό διάνυσμα, όπου κάθε κατηγορία αντιπροσωπεύεται από μια στήλη και μια δυαδική τιμή (0 ή 1), υποδηλώνοντας την παρουσία ή την απουσία αυτής της κατηγορίας. Ουσιαστικά, δημιουργεί μια «εικονική μεταβλητή» για κάθε κατηγορία, σχηματίζοντας μια δυαδική μήτρα που μπορεί εύκολα να επεξεργαστεί από αλγόριθμους μηχανικής μάθησης.

Το πρωταρχικό πλεονέκτημα του One Hot Encoding είναι η ικανότητά του να αντιμετωπίζει το ζήτημα της τακτικότητας ή της ιεραρχίας εντός κατηγορικών μεταβλητών. Σε αντίθεση με το Label Encoding, το One Hot Encoding δεν επιβάλλει καμία αριθμητική σειρά στις κατηγορίες, καθιστώντας το κατάλληλο για σενάρια όπου δεν υπάρχει εγγενής κατάταξη ή ακολουθία μεταξύ των διαφορετικών κλάσεων. Αυτή η προσέγγιση είναι ιδιαίτερα χρήσιμη όταν πρόκειται για ονομαστικά δεδομένα, όπως χρώματα ή χώρες, όπου δεν υπάρχει φυσική σειρά.

Ενώ η One Hot Encoding αυξάνει τη διάσταση του συνόλου δεδομένων, εμποδίζει το μοντέλο να παρερμηνεύσει τυχόν τεχνητές σχέσεις μεταξύ κατηγοριών. Κάθε κατηγορία αντιμετωπίζεται ανεξάρτητα, επιτρέποντας στον αλγόριθμο να αναγνωρίζει και να εξετάζει την ιδιαιτερότητα κάθε τάξης κατά τη διάρκεια της εκπαίδευσης. Συνολικά, το One Hot Encoding είναι ένα πολύτιμο εργαλείο για το χειρισμό κατηγορικών μεταβλητών με τρόπο που βελτιώνει την απόδοση των μοντέλων μηχανικής εκμάθησης, διασφαλίζοντας ότι μπορούν να μάθουν αποτελεσματικά και να κάνουν προβλέψεις για διαφορετικά και μη συνηθισμένα κατηγορικά δεδομένα.

3.2.1.6 Αφαίρεση Στηλών

Η απόρριψη περιττών στηλών, όπως τα αναγνωριστικά, είναι ένα κρίσιμο βήμα στη γραμμή προεπεξεργασίας δεδομένων. Συχνά, τα σύνολα δεδομένων περιλαμβάνουν στήλες αναγνώρισης που δεν εξυπηρετούν κανένα αναλυτικό σκοπό ή συνεισφέρουν σημαντικές πληροφορίες στην ανάλυση. Αυτές οι στήλες, που συνήθως αποτελούνται από μοναδικά αναγνωριστικά ή αριθμητικούς κωδικούς, χρησιμοποιούνται συχνά για τη διαχείριση της βάσης δεδομένων, αλλά ενδέχεται να μην παρέχουν πολύτιμες πληροφορίες στο πλαίσιο της ανάλυσης δεδομένων ή της μηχανικής μάθησης.

Η αφαίρεση αυτών των περιττών στηλών είναι επωφελής για διάφορους λόγους. Πρώτον, μειώνει τη διάσταση του συνόλου δεδομένων, καθιστώντας το πιο διαχειρίσιμο και βελτιώνει την υπολογιστική απόδοση. Με την εξάλειψη των στηλών που δεν συμβάλλουν στους αναλυτικούς στόχους, οι επιστήμονες δεδομένων μπορούν να εξορθολογήσουν την ανάλυσή τους, να εστιάσουν σε σχετικά χαρακτηριστικά και να αποφύγουν τον πιθανό θόρυβο που εισάγεται από άσχετα αναγνωριστικά. Δεύτερον, η απόθεση άχρηστων στηλών ενισχύει την ερμηνευτικότητα και τη σαφήνεια του συνόλου δεδομένων. Οι περιττές στήλες μπορεί να γεμίσουν το σύνολο δεδομένων, καθιστώντας πιο δύσκολο να διακρίνουμε ουσιαστικά μοτίβα ή σχέσεις. Με την απόρριψη αυτών των εξωγενών χαρακτηριστικών, οι αναλυτές μπορούν να δημιουργήσουν ένα καθαρότερο και πιο εστιασμένο σύνολο δεδομένων, διευκολύνοντας μια πιο ακριβή και διορατική ανάλυση.

```
columns_to_drop = ['Customer ID']
df= drop_columns(df, columns_to_drop)
display_data(df)
```

3.2.1.7 Εύρεση Ακραίων Τιμών

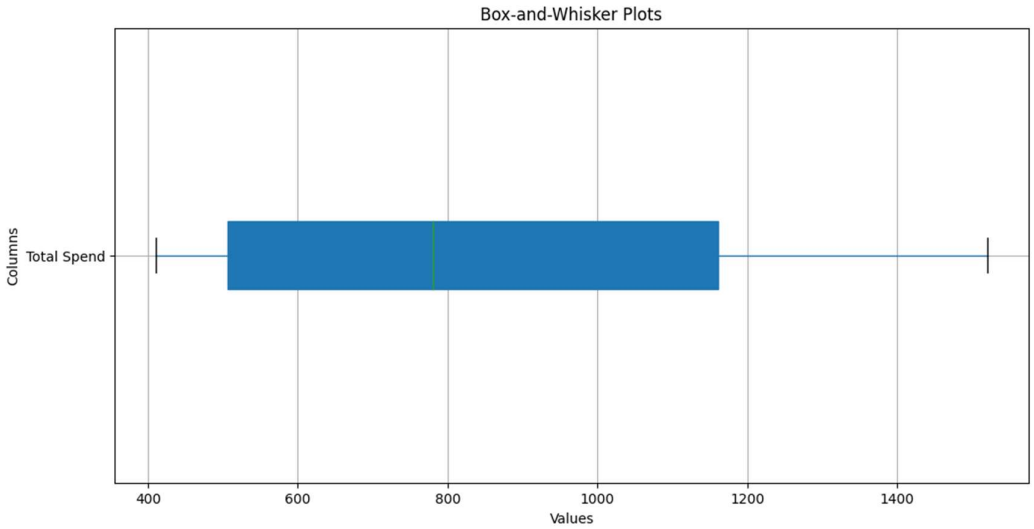
Ο εντοπισμός των ακραίων τιμών είναι ένα κρίσιμο βήμα στη διαδικασία διερευνητικής ανάλυσης δεδομένων, με στόχο τον εντοπισμό παρατηρήσεων που αποκλίνουν σημαντικά από το μεγαλύτερο μέρος του συνόλου δεδομένων. Τα ακραία σημεία είναι σημεία δεδομένων που εμφανίζουν ακραίες τιμές σε σύγκριση με την υπόλοιπη κατανομή και μπορούν ενδεχομένως να παραμορφώσουν τις στατιστικές αναλύσεις ή τα μοντέλα μηχανικής μάθησης. Διάφορες τεχνικές και στατιστικές μέθοδοι χρησιμοποιούνται για την αποκάλυψη ακραίων στοιχείων, επιτρέποντας στους αναλυτές δεδομένων και τους επιστήμονες να αποκτήσουν γνώσεις για ασυνήθιστα μοτίβα ή σφάλματα στα δεδομένα.

Μια κοινή προσέγγιση για την ανίχνευση ακραίων σημείων περιλαμβάνει τη χρήση στατιστικών μέτρων όπως το τεταρτημόριο εύρος (IQR). Το IQR βοηθά στον καθορισμό του εύρους εντός του οποίου συγκεντρώνεται το κεντρικό τμήμα των δεδομένων. Οι παρατηρήσεις που βρίσκονται εκτός ενός συγκεκριμένου εύρους πέρα από το IQR θεωρούνται πιθανές ακραίες τιμές. Τα εργαλεία οπτικοποίησης, όπως οι γραφικές παραστάσεις πλαισίου ή οι γραφικές παραστάσεις, είναι επίσης καθοριστικής σημασίας για την επισήμανση σημείων δεδομένων που απέχουν πολύ από την τυπική κατανομή.

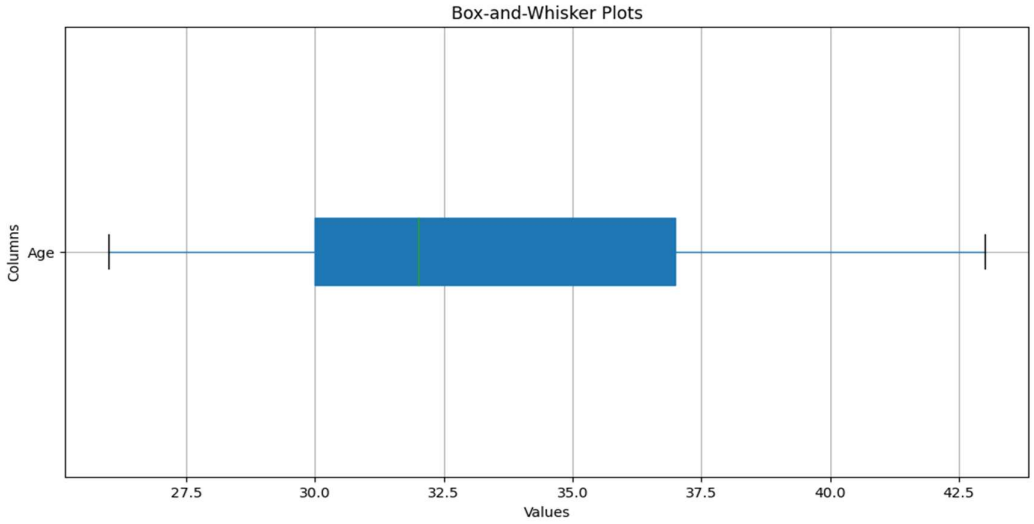
Η εύρεση ακραίων τιμών είναι απαραίτητη για τη διασφάλιση της ευρωστίας και της αξιοπιστίας των αναλύσεων δεδομένων. Εντοπίζοντας και αντιμετωπίζοντας τα

ακραία σημεία, οι αναλυτές μπορούν να λάβουν τεκμηριωμένες αποφάσεις σχετικά με το εάν θα εξαιρεθούν ή θα μετασχηματίσουν αυτά τα σημεία δεδομένων, οδηγώντας τελικά σε πιο ακριβείς και ουσιαστικές ερμηνείες των υποκείμενων προτύπων μέσα στο σύνολο δεδομένων.

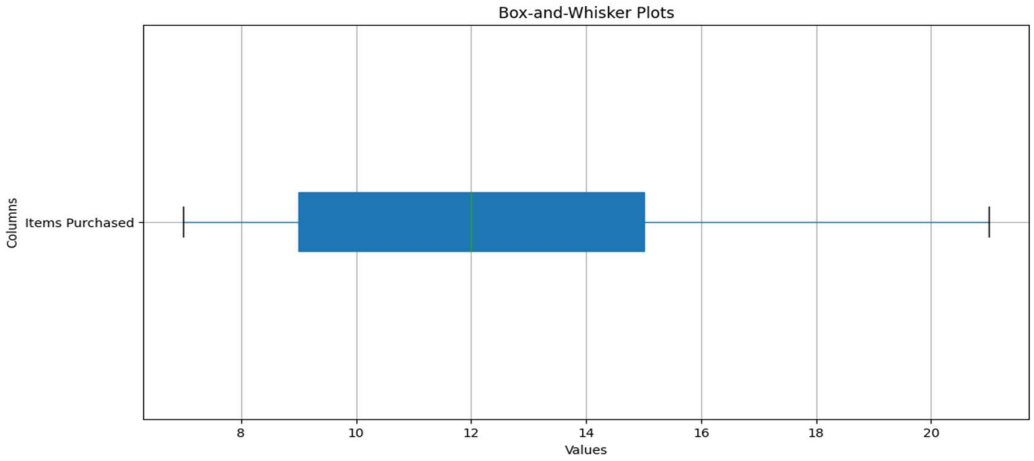
```
display_box_whisker(df,['Total Spend'])
display_box_whisker(df,['Age'])
display_box_whisker(df,['Items Purchased'])
display_box_whisker(df,['Average Rating'])
display_box_whisker(df,['Days Since Last Purchase'])
```



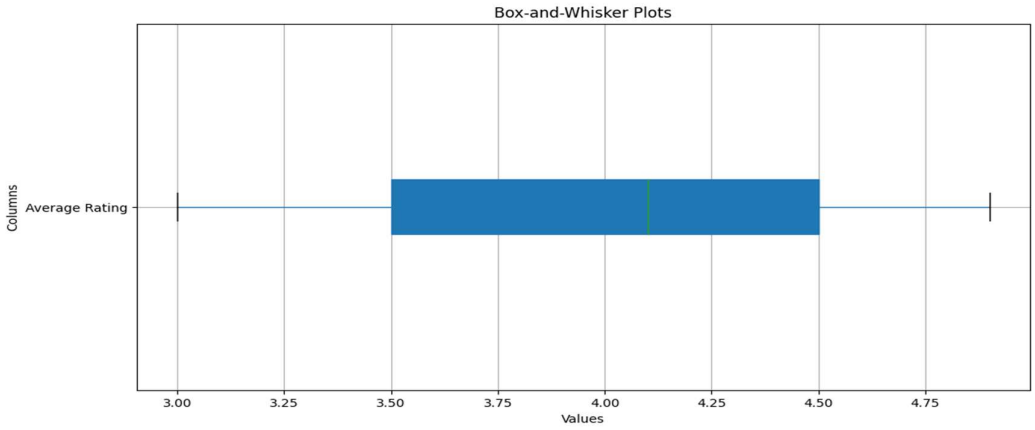
Πίνακας 3.2: Έλεγχος Ακραίων Τιμών Total Spend



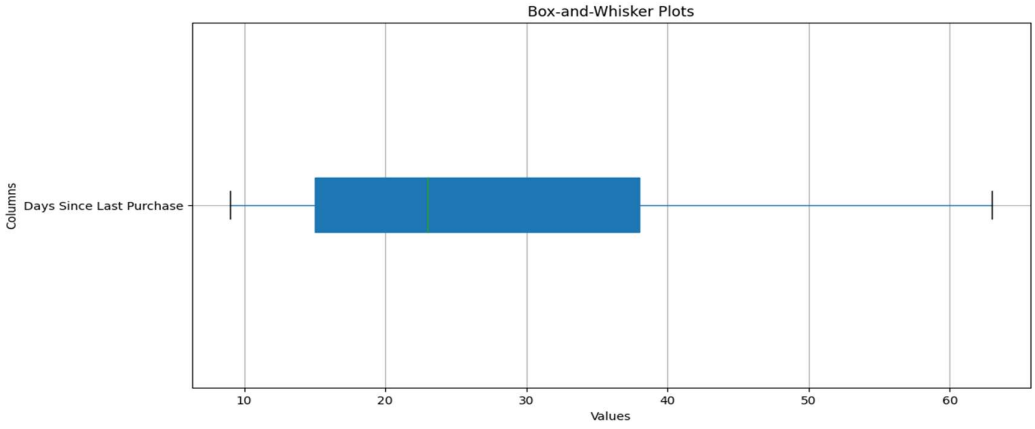
Πίνακας 3.3: Έλεγχος Ακραίων Τιμών Age



Πίνακας 3.4: Έλεγχος Ακραίων Τιμών Items Purchased



Πίνακας 3.5: Έλεγχος Ακραίων Τιμών Average Rating



Πίνακας 3.6: Έλεγχος Ακραίων Τιμών Days Since Last Purchase

3.2.1.8 Δημιουργία Churn Στήλης

Η δημιουργία της στήλης Churn σε ένα DataFrame χρησιμοποιείται για τον προσδιορισμό πελατών που ενδέχεται να διακόψουν τη χρήση ή τη συνεργασία με ένα προϊόν ή υπηρεσία. Συνήθως, οι πελάτες που έχουν χαμηλή μέση αξιολόγηση ή που έχουν παρέλθει πολλές ημέρες από την τελευταία τους αγορά θεωρούνται ως πιθανοί υποψήφιοι για διακοπή. Η διαδικασία αυτή ενισχύει τη δυνατότητα πρόβλεψης του χρόνου που ένας πελάτης θα μπορούσε να εγκαταλείψει την υπηρεσία ή το προϊόν.

Η στήλη Churn που δημιουργείται με βάση αυτά τα στατιστικά κριτήρια παρέχει μια χρήσιμη μετρική για την ανίχνευση των πελατών που ενδέχεται να χάσουν το ενδιαφέρον τους ή να αποχωρήσουν. Αυτή η προσέγγιση επιτρέπει στις επιχειρήσεις να προβλέπουν την πιθανή μείωση της πελατείας και να λαμβάνουν προληπτικά μέτρα για τη διατήρηση της πελατείας και τη βελτίωση της ποιότητας των υπηρεσιών τους. Αυτή η ανάλυση καθιστά δυνατή την καλύτερη κατανόηση της συμπεριφοράς των πελατών και την ανάληψη δράσης προς όφελος της μακροπρόθεσμης επιχειρηματικής επιτυχίας.

```
# Ορισμός κατωφλίου για το Average Rating
rating_threshold = 3.8
# Ορισμός κατωφλίου για τον αριθμό των ημερών από την τελευταία
αγορά
days_since_last_purchase_threshold = 35
# Δημιουργία της στήλης 'Churn' βάσει των κατωφλίων
df['Churn'] = ((df['Average Rating'] < rating_threshold) &
(df['Days Since Last Purchase'] >
days_since_last_purchase_threshold)).astype(int)
# Προσθήκη τυχαίων γραμμών με 'Churn' τιμή 1
num_random_churn_1 = 25  # Adjust the number of random rows with
Churn value of 1
random_rows_churn_1 = np.random.choice(df.index,
size=num_random_churn_1, replace=False)
df.loc[random_rows_churn_1, 'Churn'] = 1

# Identify at least 10 random rows with Membership Type 'Bronze'
bronze_rows = df[df['Membership Type'] == 'Bronze'].sample(15,
random_state=42)
# Set Churn Value to 1 for the selected random rows
bronze_rows['Churn'] = 1
# Update the original DataFrame with the modified rows
df.update(bronze_rows)

Satisfaction_rows = df[df['Total Spend'] < 700].sample(10,
random_state=42)
```

```
# Set Churn Value to 1 for the selected random rows
Satisfaction_rows['Churn'] = 1
# Update the original DataFrame with the modified rows
df.update(Satisfaction_rows)

# Μετρητής των τιμών 'Churn'
churn_counts = df['Churn'].value_counts()
# Εκτύπωση των μετρήσεων 'Churn'
print("\nChurn Counts:")
print(churn_counts)
# Εμφάνιση των δεδομένων του DataFrame
display_data(df)
```

3.2.2 Ανάλυση Δεδομένων με διάφορες Τεχνικές

3.2.2.1 Satisfaction Level vs Membership Type

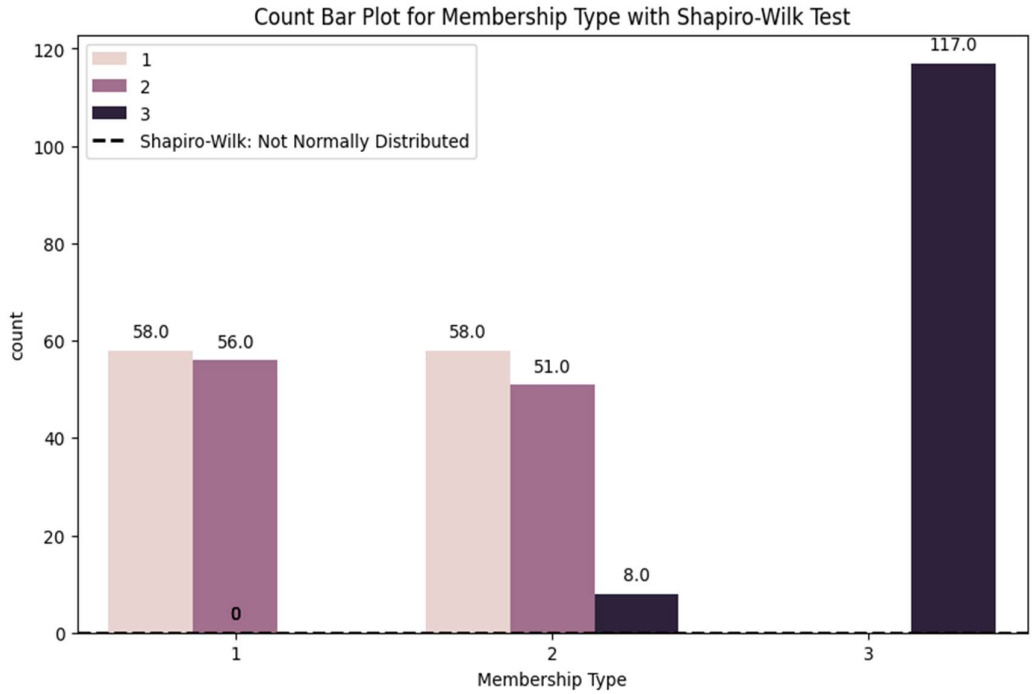
Το επίπεδο ικανοποίησης εναντίον του τύπου μέλους είναι μια γραφική αναπαράσταση που αποσκοπεί στη διερεύνηση της σχέσης μεταξύ των επιπέδων ικανοποίησης των πελατών και των διαφορετικών τύπων συμμετοχής. Αυτός ο τύπος απεικόνισης είναι ιδιαίτερα χρήσιμος για την κατανόηση του τρόπου με τον οποίο διαφορετικές κατηγορίες συμμετοχής επηρεάζουν την ικανοποίηση του πελάτη μέσα σε ένα σύνολο δεδομένων. Συνήθως, μια γραφική παράσταση ζεύγους περιλαμβάνει διαγράμματα σκέδασης για κάθε συνδυασμό των μεταβλητών, παρέχοντας πληροφορίες για πιθανά πρότυπα, τάσεις ή συσχετισμούς μεταξύ των δύο παραγόντων.

Στο πλαίσιο της ανάλυσης των πελατών, η σχεδίαση των επιπέδων ικανοποίησης έναντι των διαφόρων τύπων μελών επιτρέπει στους αναλυτές να αξιολογούν οπτικά εάν ορισμένες κατηγορίες συμμετοχής τείνουν να έχουν υψηλότερα ή χαμηλότερα επίπεδα ικανοποίησης. Τα διαγράμματα στην γραφική παράσταση ζεύγους μπορούν να αποκαλύψουν συστάδες ή τάσεις που μπορεί να υποδηλώνουν ισχυρή ή ασθενή συσχέτιση μεταξύ αυτών των μεταβλητών. Επιπλέον, η γραφική αναπαράσταση ζεύγους μπορεί να ενσωματώσει πρόσθετα οπτικά στοιχεία όπως το χρώμα ή το σχήμα για να διαφοροποιήσει τους διάφορους τύπους μελών, καθιστώντας ευκολότερη την ερμηνεία των δεδομένων.

Η ανάλυση του επιπέδου ικανοποίησης εναντίον του τύπου μέλους του ζευγαριού μπορεί να συμβάλει στη λήψη αποφάσεων που βασίζονται σε δεδομένα σχετικά με τα προγράμματα εμπλοκής και αφοσίωσης των πελατών. Για παράδειγμα, εάν ορισμένοι τύποι συμμετοχής δείχνουν σταθερά υψηλότερα επίπεδα ικανοποίησης, οι επιχειρήσεις μπορούν να προσαρμόσουν τις στρατηγικές τους για να ενισχύσουν

τις εμπειρίες των πελατών για αυτές τις συγκεκριμένες κατηγορίες. Από την άλλη πλευρά, ο εντοπισμός τυχόν προτύπων δυσαρέσκειας εντός ορισμένων μελών μπορεί να προκαλέσει στοχοθετημένες παρεμβάσεις για την αντιμετώπιση τους.

```
#sns.pairplot(df,x_vars=['Membership Type'],y_vars=["Satisfaction Level"],height=4,diag_kind='bar')
count_bar_plot_with_shapiro(df,"Membership Type",'Satisfaction Level')
```



Η γραφική παράσταση «Επίπεδο Ικανοποίησης έναντι Τύπου Μέλους» αποκαλύπτει ένα ξεχωριστό μοτίβο στη σχέση μεταξύ των επιπέδων ικανοποίησης των πελατών και των αντίστοιχων τύπων μέλους. Ειδικότερα, όταν ο τύπος μέλους κατηγοριοποιείται ως 3, τα επίπεδα ικανοποίησης ευθυγραμμίζονται σταθερά με την υψηλότερη βαθμολογία, που υποδηλώνεται ως 3. Αυτή η παρατήρηση υποδηλώνει μια ισχυρή θετική συσχέτιση μεταξύ της κατηγορίας μελών κορυφαίας βαθμίδας και της αυξημένης ικανοποίησης πελατών. Οι πελάτες που κατέχουν τύπο συνδρομής 3 εκφράζουν σταθερά τα υψηλότερα επίπεδα ικανοποίησης, υποδεικνύοντας την πιθανή επίδραση των χαρακτηριστικών ή των πλεονεκτημάτων της premium συνδρομής στη συνολική ικανοποίηση.

Αντίθετα, όταν οι τύποι συνδρομής κατηγοριοποιούνται ως 2 ή 1, τα επίπεδα ικανοποίησης εμπίπτουν κυρίως στις χαμηλότερες κατηγορίες, συγκεκριμένα 1 ή 2. Αυτό το μοτίβο υποδηλώνει ότι οι πελάτες με συνδρομές χαμηλότερου επιπέδου τείνουν να εκφράζουν μέτρια έως χαμηλότερα επίπεδα ικανοποίησης. Η διχοτόμηση στα επίπεδα ικανοποίησης με βάση τους τύπους μελών υπογραμμίζει τη σημασία των

διακρίσεων στα επίπεδα συνδρομής στη διαμόρφωση των συναισθημάτων των πελατών. Αυτές οι πληροφορίες είναι ζωτικής σημασίας για τις επιχειρήσεις που στοχεύουν να προσαρμόσουν τις στρατηγικές τους, καθώς υπογραμμίζουν τον αντίκτυπο των βαθμίδων συμμετοχής στη συνολική ικανοποίηση των πελατών και προτείνουν τομείς για πιθανές βελτιώσεις ή βελτιώσεις στις προσφορές μελών χαμηλότερης βαθμίδας.

Τύπος μέλους 3:
Όταν ο Τύπος μέλους κατηγοριοποιείται ως 3, τα Επίπεδα Ικανοποίησης ευθυγραμμίζονται με συνέπεια με την υψηλότερη βαθμολογία, που υποδηλώνεται ως 3. Αυτό υποδηλώνει μια ισχυρή θετική συσχέτιση μεταξύ της κατηγορίας μελών κορυφαίας κατηγορίας και της αυξημένης ικανοποίησης των πελατών. Οι πελάτες με Τύπο μέλους 3 εκφράζουν με συνέπεια τα υψηλότερα επίπεδα ικανοποίησης.

Τύποι μελών 2 ή 1:
Αντίθετα, όταν οι τύποι μελών κατηγοριοποιούνται ως 2 ή 1, τα επίπεδα ικανοποίησης εμπίπτουν κυρίως στις χαμηλότερες κατηγορίες, συγκεκριμένα 1 ή 2. Αυτό το μοτίβο υποδηλώνει ότι οι πελάτες με συνδρομές χαμηλότερου επιπέδου τείνουν να εκφράζουν μέτρια έως χαμηλότερα επίπεδα ικανοποίησης. Η διχοτόμηση στα επίπεδα ικανοποίησης που βασίζονται σε τύπους συμμετοχής υπογραμμίζει τη σημασία των διακρίσεων των βαθμίδων συνδρομής στη διαμόρφωση των συναισθημάτων των πελατών.

3.2.2.2 Satisfaction Level vs Total Spend

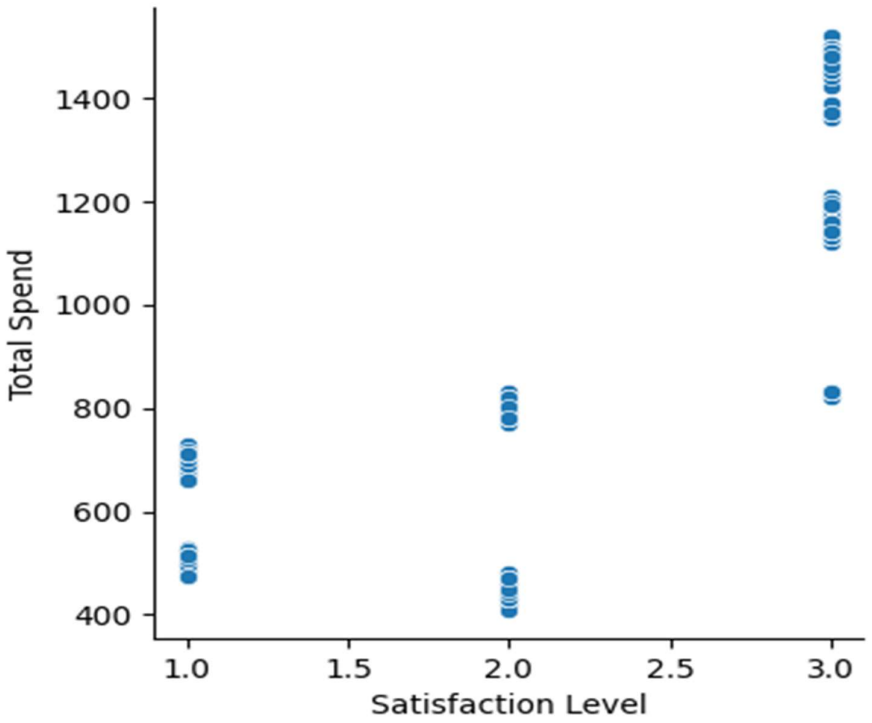
Η γραφική παράσταση του ζεύγους Επίπεδο ικανοποίησης έναντι Συνολικών Δαπανών είναι μια οπτική αναπαράσταση που διερευνά τη πιθανή σχέση μεταξύ των επιπέδων ικανοποίησης πελατών και του συνολικού ποσού που δαπανήθηκε από τους πελάτες. Αυτός ο τύπος γραφικής παράστασης ζεύγους είναι καθοριστικός για την αποκάλυψη προτύπων ή τάσεων που μπορεί να υπάρχουν μεταξύ αυτών των δύο βασικών μεταβλητών, παρέχοντας πολύτιμες γνώσεις σχετικά με τη συσχέτιση μεταξύ της ικανοποίησης των πελατών και της χρηματικής συνεισφοράς τους σε μια επιχείρηση.

Σε αυτό το γράφημα ζεύγους, κάθε σημείο στο διάγραμμα διασποράς αντιπροσωπεύει έναν πελάτη, με τον άξονα x να απεικονίζει το επίπεδο ικανοποίησης και τον άξονα y να αντιπροσωπεύει τη συνολική δαπάνη. Παρατηρώντας την κατανομή των πόντων και τη γενική τάση του διαγράμματος διασποράς, οι αναλυτές μπορούν να συμπεράνουν εάν υπάρχει θετική, αρνητική ή ουδέτερη συσχέτιση μεταξύ της ικανοποίησης των πελατών και της συμπεριφοράς των δαπανών τους. Μια θετική συσχέτιση θα σήμαινε ότι καθώς αυξάνονται τα επίπεδα ικανοποίησης, οι

συνολικές δαπάνες τείνουν να αυξάνονται επίσης, ενώ μια αρνητική συσχέτιση θα υποδηλώνει το αντίθετο.

Η ερμηνεία της γραφικής παράστασης του ζεύγους Επιπέδου Ικανοποίησης έναντι Συνολικών Δαπανών επιτρέπει στις επιχειρήσεις να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με τη δέσμευση πελατών και τις στρατηγικές μάρκετινγκ. Για παράδειγμα, εάν παρατηρηθεί μια σαφής θετική συσχέτιση, μπορεί να υποδηλώνει ότι η επένδυση σε πρωτοβουλίες για την ενίσχυση της ικανοποίησης των πελατών θα μπορούσε ενδεχομένως να οδηγήσει σε αύξηση των δαπανών. Αντίθετα, η έλλειψη συσχέτισης μπορεί να ωθήσει τις επιχειρήσεις να επαναξιολογήσουν τις στρατηγικές ικανοποίησης των πελατών τους ή να διερευνήσουν άλλους παράγοντες που επηρεάζουν τη συμπεριφορά των δαπανών των πελατών. Αυτή η οπτική εξερεύνηση της ικανοποίησης των πελατών και των προτύπων δαπανών είναι ζωτικής σημασίας για την επινόηση στοχευμένων προσεγγίσεων για τη βελτίωση των συνολικών σχέσεων με τους πελάτες και την προώθηση της επιχειρηματικής ανάπτυξης.

```
sns.pairplot(df,x_vars=['Satisfaction Level'],y_vars=["Total Spend"],height=4)
```



Πίνακας 3.8: Συσχέτιση μεταξύ Satisfaction Level & Total Spend

Η πλοκή Επιπέδου Ικανοποίησης έναντι Συνολικών Δαπανών αποκαλύπτει ενδιαφέρουσες πληροφορίες σχετικά με τη σχέση μεταξύ των επιπέδων ικανοποίησης των πελατών και των αντίστοιχων συνολικών δαπανών τους. Συγκεκριμένα, όταν το επίπεδο ικανοποίησης βαθμολογείται ως 3, υπάρχει μια ευδιάκριτη και προφανώς θετική συσχέτιση με υψηλότερη συνολική δαπάνη. Αυτή η παρατήρηση υποδηλώνει ότι οι πελάτες που εκφράζουν την υψηλότερη ικανοποίηση τείνουν να επιδεικνύουν πιο ουσιαστικές δαπάνες, επιδεικνύοντας μια θετική σύνδεση μεταξύ της ικανοποίησής τους και της προθυμίας τους να επενδύσουν περισσότερα στα προϊόντα ή τις υπηρεσίες.

Ωστόσο, η γραφική παράσταση δείχνει επίσης ότι η σχέση είναι λιγότερο σαφής για τους πελάτες με χαμηλότερα επίπεδα ικανοποίησης, ιδιαίτερα αυτούς που βαθμολογούνται με 2 και 1. Τα σημεία διασποράς για αυτά τα επίπεδα ικανοποίησης δεν δείχνουν μια σταθερή και ισχυρή θετική τάση με τις συνολικές δαπάνες. Αυτό υποδηλώνει ότι η σχέση μεταξύ ικανοποίησης και δαπανών γίνεται λιγότερο εμφανής καθώς μειώνονται τα επίπεδα ικανοποίησης. Είναι σημαντικό για τις επιχειρήσεις να δώσουν προσοχή σε αυτή τη διάκριση, καθώς υπονοεί ότι η ενίσχυση της ικανοποίησης από ένα χαμηλότερο επίπεδο μπορεί να μην μεταφραστεί απαραίτητα σε αναλογική αύξηση των δαπανών. Η κατανόηση αυτών των αποχρώσεων επιτρέπει στις επιχειρήσεις να προσαρμόσουν τις στρατηγικές τους με βάση τον ποικίλο αντίκτυπο των επιπέδων ικανοποίησης στη συμπεριφορά των δαπανών των πελατών.

Συνοπτικά:

1. **Σαφής θετική συσχέτιση στο επίπεδο ικανοποίησης 3:**
    - Παρατηρείται μια διακριτή και θετική σχέση μεταξύ του επιπέδου ικανοποίησης των πελατών 3 και των συνολικών δαπανών τους.
    - Οι πελάτες που εκφράζουν την υψηλότερη ικανοποίηση (επίπεδο 3) τείνουν να εμφανίζουν σημαντικά υψηλότερη συνολική δαπάνη.
    - Αυτή η θετική συσχέτιση υποδηλώνει ότι οι πελάτες με μεγαλύτερη ικανοποίηση είναι πιο διατεθειμένοι να κάνουν μεγαλύτερες δαπάνες.
2. **Λιγότερο σαφής σχέση για τα επίπεδα ικανοποίησης 2 και 1:**
    - Αντίθετα, η σχέση μεταξύ των επιπέδων ικανοποίησης 2 και 1 και των συνολικών δαπανών δεν είναι τόσο σαφώς καθορισμένη.
    - Τα σημεία διασποράς για αυτά τα επίπεδα ικανοποίησης δεν εμφανίζουν σταθερή και ισχυρή θετική τάση με τις συνολικές δαπάνες.
    - Η γραφική παράσταση υποδηλώνει ότι η συσχέτιση μεταξύ ικανοποίησης και δαπανών γίνεται λιγότερο εμφανής ή συνεπής καθώς μειώνονται τα επίπεδα ικανοποίησης.
    - Η ενίσχυση της ικανοποίησης από τα χαμηλότερα επίπεδα μπορεί να μην οδηγήσει απαραίτητα σε αναλογική αύξηση των δαπανών.

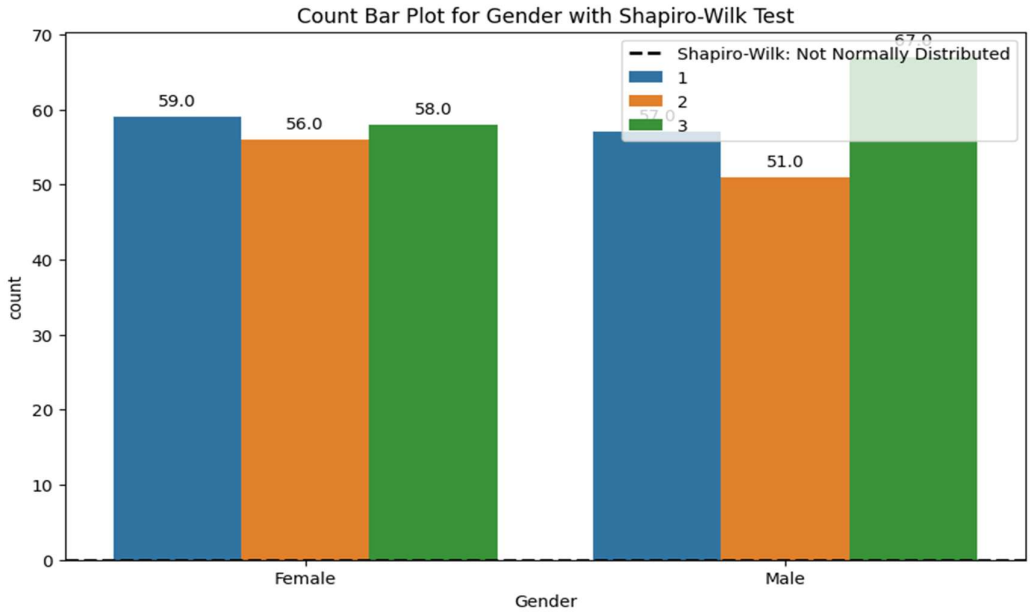
3.2.2.3 Satisfaction Level vs Gender

Η γραφική παράσταση του ζεύγους Επίπεδο Ικανοποίησης έναντι Φύλου είναι μια οπτική εξερεύνηση της σχέσης μεταξύ ικανοποίησης πελατών και φύλου μέσα σε ένα σύνολο δεδομένων. Αυτός ο τύπος γραφικής παράστασης ζεύγους δίνει τη δυνατότητα στους αναλυτές να εξετάσουν πιθανά πρότυπα ή διαφορές στα επίπεδα ικανοποίησης με βάση το φύλο των πελατών.

Μελετώντας την κατανομή και τη συνολική τάση της γραφικής παράστασης διασποράς, μπορούν να αποκτηθούν πληροφορίες για το εάν υπάρχουν αξιοσημείωτες διακρίσεις στα επίπεδα ικανοποίησης μεταξύ των διαφορετικών φύλων. Οι οπτικές ενδείξεις, όπως η ομαδοποίηση, μπορεί να υποδηλώνουν μοτίβα στην ικανοποίηση των πελατών που σχετίζονται με το φύλο. Αυτή η ανάλυση μπορεί να είναι ιδιαίτερα πολύτιμη για τις επιχειρήσεις που επιδιώκουν να κατανοήσουν και να αντιμετωπίσουν πιθανές προτιμήσεις ή ανησυχίες που επηρεάζουν την ικανοποίηση του πελάτη.

Η ανάλυση της γραφικής παράστασης του ζεύγους Επίπεδο Ικανοποίησης έναντι Φύλου επιτρέπει στις επιχειρήσεις να προσαρμόσουν τις στρατηγικές τους με βάση τις γνώσεις που σχετίζονται με το φύλο. Για παράδειγμα, εάν υπάρχουν ευδιάκριτες διαφορές στα επίπεδα ικανοποίησης μεταξύ των φύλων, στοχευμένες εκστρατείες μάρκετινγκ ή πρωτοβουλίες εξυπηρέτησης πελατών μπορούν να σχεδιαστούν για να ανταποκρίνονται στις συγκεκριμένες ανάγκες ή προτιμήσεις κάθε ομάδας φύλου. Τελικά, αυτή η οπτική αναπαράσταση βοηθά τις επιχειρήσεις να λαμβάνουν αποφάσεις βάσει δεδομένων για να βελτιώσουν την ικανοποίηση των πελατών και να προωθήσουν μια πιο περιεκτική και πελατοκεντρική προσέγγιση.

```
#sns.pairplot(df,x_vars=['Satisfaction
Level'],y_vars=["Gender"],height=4)
count_bar_plot_with_shapiro(df,"Gender",'Satisfaction Level')
```



Η πλοκή Επίπεδο Ικανοποίησης έναντι Φύλου αποκαλύπτει ενδιαφέροντα μοτίβα στην κατανομή των επιπέδων ικανοποίησης των πελατών μεταξύ διαφορετικών φύλων. Συγκεκριμένα, δεν υπάρχει στατιστικά σημαντική απόκλιση στην κατανομή των επιπέδων ικανοποίησης μεταξύ ανδρών και γυναικών. Και τα δύο φύλα δείχνουν μια σχετικά ομοιόμορφη κατανομή μεταξύ των επιπέδων ικανοποίησης, υποδηλώνοντας συγκρίσιμη κατανομή ικανοποίησης μεταξύ ανδρών και γυναικών πελατών.

Ωστόσο, προκύπτει μια λεπτή διάκριση, καθώς οι άνδρες εμφανίζουν ελαφρώς υψηλότερο Επιπέδου Ικανοποίησης 3. Ενώ η συνολική κατανομή παραμένει σχετικά ομοιόμορφη, τα δεδομένα υποδηλώνουν ότι ένα ελαφρώς μεγαλύτερο ποσοστό ανδρών εκφράζει το υψηλότερο επίπεδο ικανοποίησης σε σύγκριση με τις γυναίκες. Αυτή η διαφοροποιημένη διαφορά θα μπορούσε να είναι ενδεικτική των ποικίλων παραγόντων που επηρεάζουν την ικανοποίηση μεταξύ των φύλων και οι επιχειρήσεις μπορεί να θεωρήσουν ωφέλιμο να διερευνήσουν τους βαθύτερους λόγους πίσω από αυτή τη μικρή διαφοροποίηση.

Η κατανόηση αυτών των προτύπων που σχετίζονται με το φύλο στα επίπεδα ικανοποίησης μπορεί να ενημερώσει τις επιχειρήσεις στην προσαρμογή των στρατηγικών τους και των προσπαθειών δέσμευσης πελατών. Ενώ η συνολική διαφορά ικανοποίησης είναι ομοιόμορφη, η αναγνώριση των λεπτών παραλλαγών επιτρέπει πιο στοχευμένες προσεγγίσεις για την αντιμετώπιση των συγκεκριμένων αναγκών και προτιμήσεων διαφορετικών τμημάτων πελατών με βάση το φύλο.

Συνοπτικά:

1. **Ισότιμη κατανομή των επιπέδων ικανοποίησης:**
    - Η γραφική παράσταση Επίπεδο Ικανοποίησης έναντι Φύλου δεν αποκαλύπτει στατιστικά σημαντική απόκλιση στην κατανομή των επιπέδων ικανοποίησης μεταξύ ανδρών και γυναικών.
    - Και τα δύο φύλα παρουσιάζουν μια σχετικά ομοιόμορφη κατανομή μεταξύ των επιπέδων ικανοποίησης, υποδηλώνοντας ένα συγκρίσιμο επίπεδο ικανοποίησης.
2. **Μικρή ανισότητα μεταξύ των φύλων στο επίπεδο ικανοποίησης 3:**
    - Εμφανίζεται μια λεπτή διάκριση καθώς οι άνδρες εμφανίζουν ελαφρώς υψηλότερο Επιπέδου Ικανοποίησης 3.
    - Ενώ η συνολική κατανομή παραμένει σχετικά ομοιόμορφη, ένα ελαφρώς μεγαλύτερο ποσοστό ανδρών εκφράζει το υψηλότερο επίπεδο ικανοποίησης σε σύγκριση με τις γυναίκες.
    - Αυτή η διαφοροποιημένη διαφορά υποδηλώνει πιθανές διακυμάνσεις στους παράγοντες που επηρεάζουν την ικανοποίηση μεταξύ των φύλων, κάτι που δικαιολογεί περαιτέρω διερεύνηση.

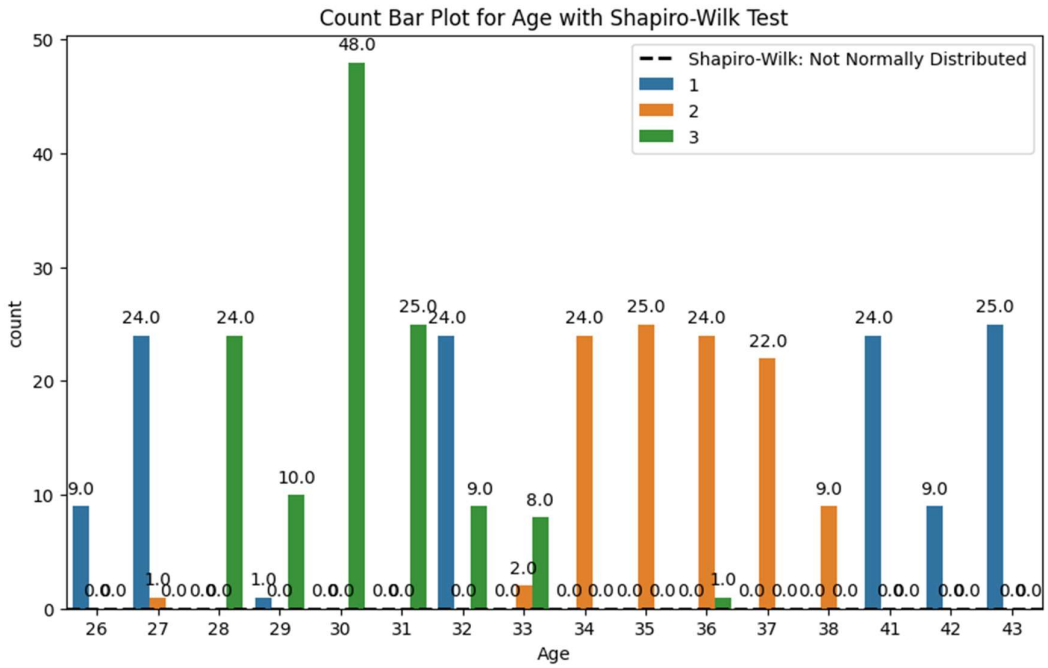
3.2.2.4 Satisfaction Level vs Age pair

Η πλοκή του ζεύγους Επίπεδο Ικανοποίησης έναντι Ηλικίας είναι μια οπτική εξερεύνηση της πιθανής σχέσης μεταξύ ικανοποίησης πελατών και ηλικίας. Αυτός ο τύπος γραφικής παράστασης ζεύγους επιτρέπει στους αναλυτές να διερευνήσουν εάν υπάρχουν ευδιάκριτα πρότυπα ή τάσεις στα επίπεδα ικανοποίησης με βάση διαφορετικές ηλικιακές ομάδες. Η εξέταση της κατανομής και του συνολικού σχεδίου στο διάγραμμα διασποράς παρέχει πληροφορίες για πιθανούς συσχετισμούς ή διακυμάνσεις στα επίπεδα ικανοποίησης σε διαφορετικές ηλικιακές ομάδες. Μοτίβα όπως συμπλέγματα ή τάσεις μπορεί να εμφανιστούν, υποδηλώνοντας πώς η ηλικία μπορεί να επηρεάσει την ικανοποίηση των πελατών. Αυτή η ανάλυση είναι πολύτιμη για επιχειρήσεις που στοχεύουν να κατανοήσουν τις διαφορετικές προτιμήσεις και προσδοκίες των πελατών σε διαφορετικά στάδια ζωής.

Η ερμηνεία της γραφικής παράστασης του ζεύγους Επίπεδο ικανοποίησης έναντι ηλικίας μπορεί να καθοδηγήσει τις επιχειρήσεις στην προσαρμογή των στρατηγικών τους σε συγκεκριμένα δημογραφικά στοιχεία ηλικίας. Για παράδειγμα, εάν ορισμένες ηλικιακές ομάδες παρουσιάζουν σταθερά υψηλότερα επίπεδα ικανοποίησης, οι στοχευμένες καμπάνιες μάρκετινγκ ή οι προσφορές προϊόντων μπορούν να σχεδιαστούν, ώστε να έχουν απήχηση σε αυτά τα δημογραφικά στοιχεία. Αντίθετα, ο εντοπισμός τυχόν προτύπων δυσαρέσκειας που σχετίζονται με την ηλικία μπορεί να ωθήσει τις επιχειρήσεις να αντιμετωπίσουν συγκεκριμένες ανησυχίες ή να

προσαρμόσουν τις υπηρεσίες τους, ώστε να ανταποκρίνονται καλύτερα στις προσδοκίες των πελατών σε αυτές τις ηλικιακές κατηγορίες. Συνολικά, αυτή η οπτική εξερεύνηση βοηθά τις επιχειρήσεις να λαμβάνουν αποφάσεις βάσει δεδομένων για να βελτιώσουν την ικανοποίηση των πελατών και να βελτιώσουν τη συνολική εμπειρία του πελάτη.

```
count_bar_plot_with_shapiro(df,"Age",'Satisfaction Level')
```



Η πλοκή Επίπεδο Ικανοποίησης εναντίον Ηλικίας προσφέρει ενδιαφέρουσες ιδέες για τη σχέση μεταξύ της ηλικίας των πελατών και των εκφραζόμενων επιπέδων ικανοποίησής τους. Συγκεκριμένα, υπάρχει μια ευδιάκριτη συγκέντρωση πελατών με Επίπεδο Ικανοποίησης 2 μεταξύ 33 και 38 ετών. Αυτή η συγκέντρωση υποδηλώνει ότι οι πελάτες που εμπίπτουν σε αυτό το ηλικιακό κλιμάκιο είναι πιο πιθανό να εκφράσουν μέτρια ικανοποίηση. Το συγκεκριμένο ηλικιακό εύρος από 33 έως 38 μπορεί να αντιπροσωπεύει ένα κρίσιμο δημογραφικό στοιχείο όπου ορισμένοι παράγοντες επηρεάζουν διαφορετικά τα επίπεδα ικανοποίησης σε σύγκριση με άλλες ηλικιακές ομάδες.

Από την άλλη πλευρά, οι πελάτες με Επίπεδο Ικανοποίησης 3 παρουσιάζουν διαφορετική ηλικιακή κατανομή. Η πλειονότητα αυτών των πελατών συγκεντρώνεται στο εύρος ηλικιών από 28 έως 33. Αυτή η συγκέντρωση σημαίνει ότι οι πελάτες εντός αυτής της ηλικιακής κατηγορίας είναι πιο επιρρεπείς στο να εκφράσουν το υψηλότερο επίπεδο ικανοποίησης. Είναι ενδιαφέρον ότι το Επίπεδο Ικανοποίησης 3 φαίνεται να κατανέμεται πιο ομοιόμορφα σε αυτό το εύρος ηλικιών, υποδηλώνοντας

ένα σταθερό μοτίβο υψηλών επιπέδων ικανοποίησης μεταξύ πελατών ηλικίας 28 έως 33 ετών.

Η κατανόηση αυτών των προτύπων ανάλογα με την ηλικία στα επίπεδα ικανοποίησης μπορεί να είναι ζωτικής σημασίας για τις επιχειρήσεις. Τους επιτρέπει να προσαρμόζουν τις στρατηγικές, τις καμπάνιες μάρκετινγκ και τις προσπάθειες αφοσίωσης των πελατών τους για να ανταποκρίνονται στις συγκεκριμένες ανάγκες και προτιμήσεις διαφορετικών ηλικιακών ομάδων, διασφαλίζοντας μια πιο στοχευμένη και αποτελεσματική προσέγγιση για την ενίσχυση της συνολικής ικανοποίησης των πελατών.

Συνοπτικά:

1. **Συγκέντρωση ικανοποίησης Επίπεδο 2 (Ηλικιακή ομάδα 33-38):**
    - Η γραφική παράσταση Επιπέδου Ικανοποίησης έναντι Ηλικίας αποκαλύπτει μια ευδιάκριτη συγκέντρωση πελατών με Επίπεδο Ικανοποίησης 2 μεταξύ 33 και 38 ετών.
    - Αυτή η συγκέντρωση υποδηλώνει ότι οι πελάτες αυτής της ηλικιακής κατηγορίας είναι πιο πιθανό να εκφράσουν μέτρια ικανοποίηση.
    - Το συγκεκριμένο εύρος ηλικιών από 33 έως 38 μπορεί να αντιπροσωπεύει ένα κρίσιμο δημογραφικό στοιχείο με μοναδικούς παράγοντες που επηρεάζουν τα επίπεδα ικανοποίησης.
2. **Συγκέντρωση ικανοποίησης Επίπεδο 3 (Ηλικιακή ομάδα 28-33):**
    - Οι πελάτες με επίπεδο ικανοποίησης 3 παρουσιάζουν διαφορετική ηλικιακή κατανομή, με την πλειοψηφία να συγκεντρώνεται στην ηλικιακή περιοχή 28 έως 33 ετών.
    - Αυτή η συγκέντρωση υποδηλώνει ότι οι πελάτες εντός αυτής της ηλικιακής κατηγορίας είναι πιο επιρρεπείς στο να εκφράσουν το υψηλότερο επίπεδο ικανοποίησης.
    - Το επίπεδο ικανοποίησης 3 φαίνεται να κατανέμεται ομοιόμορφα σε αυτό το ηλικιακό εύρος, υποδεικνύοντας ένα σταθερό μοτίβο υψηλών επιπέδων ικανοποίησης μεταξύ πελατών ηλικίας 28 έως 33 ετών.

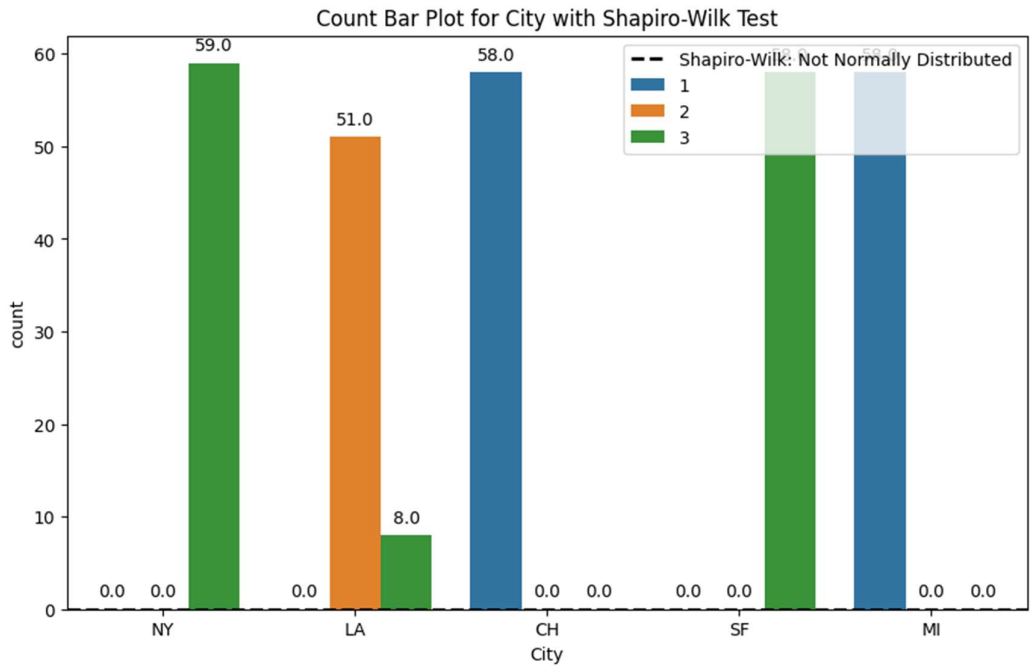
3.2.2.5 Satisfaction Level vs City

Η γραφική παράσταση του ζεύγους Επίπεδο ικανοποίησης έναντι πόλης χρησιμεύει ως εργαλείο οπτικής ανάλυσης για τη διερεύνηση της πιθανής συσχέτισης μεταξύ των επιπέδων ικανοποίησης των πελατών και της πόλης κατοικίας. Εξετάζοντας την κατανομή των πόντων και το συνολικό μοτίβο της γραφικής παράστασης, οι αναλυτές μπορούν να αποκτήσουν γνώσεις σχετικά με το εάν υπάρχουν αξιοσημείωτες διακυμάνσεις ή τάσεις στα επίπεδα ικανοποίησης σε διαφορετικές πόλεις.

Αυτή η οπτική εξερεύνηση είναι ιδιαίτερα πολύτιμη για επιχειρήσεις που δραστηριοποιούνται σε πολλαπλές τοποθεσίες, βοηθώντας τις να διακρίνουν πιθανά γεωγραφικά μοτίβα στην ικανοποίηση των πελατών. Οι ομάδες, οι τάσεις ή οι παραλλαγές στο διάγραμμα διασποράς μπορεί να υποδηλώνουν τοπικές διαφορές στις προτιμήσεις, τις προσδοκίες ή τις εμπειρίες των πελατών. Η κατανόηση αυτών των αποχρώσεων μπορεί να ενημερώσει στοχευμένες στρατηγικές για την αντιμετώπιση συγκεκριμένων αναγκών ή προκλήσεων που σχετίζονται με διαφορετικές πόλεις, ενισχύοντας έτσι τη συνολική ικανοποίηση των πελατών.

Τα ευρήματα από το ζεύγος Επίπεδο Ικανοποίησης εναντίον Πόλης μπορούν να καθοδηγήσουν τις επιχειρήσεις στην προσαρμογή των προσπαθειών μάρκετινγκ, των πρωτοβουλιών εξυπηρέτησης πελατών ή των προσφορών προϊόντων ώστε να ευθυγραμμιστούν καλύτερα με τις προτιμήσεις και τις προσδοκίες των πελατών σε συγκεκριμένες πόλεις. Αυτή η στοχευμένη προσέγγιση διασφαλίζει ότι οι επιχειρήσεις αντιμετωπίζουν τις περιφερειακές παραλλαγές και δημιουργούν μια πιο εξατομικευμένη και αποτελεσματική εμπειρία πελατών, συμβάλλοντας τελικά στη βελτίωση της ικανοποίησης και της αφοσίωσης των πελατών.

```
count_bar_plot_with_shapiro(df,"City",'Satisfaction Level')
```



Η πλοκή Επίπεδο Ικανοποίησης εναντίον Πόλης παρουσιάζει συναρπαστικά μοτίβα σχετικά με την ικανοποίηση των πελατών σε διάφορες πόλεις. Συγκεκριμένα, όλοι οι πελάτες στη Νέα Υόρκη (NY) εκφράζουν σταθερά το υψηλότερο επίπεδο ικανοποίησης, που υποδηλώνεται ως Επίπεδο 3. Αυτή η συγκέντρωση μέγιστης ικανοποίησης υποδηλώνει μια θετική και ομοιόμορφη τάση μεταξύ των πελατών στη Νέα Υόρκη, υποδηλώνοντας υψηλό επίπεδο ικανοποίησης με τα προσφερόμενα προϊόντα ή Υπηρεσίες.

Ομοίως, οι πελάτες στο Λος Άντζελες (LA) παρουσιάζουν κατά κύριο λόγο Ικανοποίηση Επίπεδο 3, με αξιοσημείωτη συγκέντρωση σε αυτήν την υψηλότερη κατηγορία ικανοποίησης. Αν και δεν εκφράζουν όλοι οι πελάτες στο Λος Άντζελες το μέγιστο επίπεδο ικανοποίησης, η πλοκή δείχνει μια ισχυρή τάση προς υψηλή ικανοποίηση. Αυτό το μοτίβο υποδηλώνει μια γενικά θετική υποδοχή προϊόντων ή υπηρεσιών στο Λος Άντζελες.

Αντίθετα, όλοι οι πελάτες στο Σικάγο (CH) εκφράζουν ομοιόμορφα το χαμηλότερο επίπεδο ικανοποίησης, που υποδηλώνεται ως Επίπεδο 1. Αυτή η ξεχωριστή συγκέντρωση υποδηλώνει μια ευρεία τάση χαμηλότερης ικανοποίησης μεταξύ των πελατών στο Σικάγο, επισημαίνοντας πιθανές προκλήσεις ή ζητήματα που μπορεί να χρειάζονται προσοχή για τη βελτίωση της εμπειρίας του πελάτη.

Ομοίως, όλοι οι πελάτες στο Σαν Φρανσίσκο (SF) εκφράζουν σταθερά το υψηλότερο επίπεδο ικανοποίησης, αντικατοπτρίζοντας τη θετική τάση που παρατηρείται στη Νέα Υόρκη. Αντίθετα, όλοι οι πελάτες στο Μαϊάμι (MI) παρουσιάζουν ομοιόμορφα Επίπεδο Ικανοποίησης 1, υποδεικνύοντας ένα ευρέως διαδεδομένο πρότυπο χαμηλότερης ικανοποίησης μεταξύ των πελατών σε αυτήν την πόλη.

Η κατανόηση αυτών των συγκεκριμένων προτύπων για την πόλη στα επίπεδα ικανοποίησης είναι ζωτικής σημασίας για τις επιχειρήσεις να προσαρμόσουν τις στρατηγικές τους και να αντιμετωπίσουν μοναδικές προκλήσεις ή ευκαιρίες που σχετίζονται με την ικανοποίηση των πελατών σε διαφορετικές τοποθεσίες. Αυτές οι πληροφορίες μπορούν να ενημερώσουν στοχευμένες πρωτοβουλίες για τη βελτίωση της συνολικής ικανοποίησης των πελατών σε διάφορες γεωγραφικές περιοχές.

Συνοπτικά:

1. **Νέα Υόρκη (NY):**
    - Όλοι οι πελάτες στη Νέα Υόρκη εκφράζουν σταθερά το υψηλότερο επίπεδο ικανοποίησης (Επίπεδο 3).
    - Αυτή η ομοιόμορφη τάση υποδηλώνει ένα θετικό και υψηλό επίπεδο ικανοποίησης μεταξύ των πελατών στη Νέα Υόρκη.

2. **Λος Άντζελες (LA):**
    - Οι πελάτες στο Λος Άντζελες παρουσιάζουν κατά κύριο λόγο Ικανοποίηση Επίπεδο 3, με αξιοσημείωτη συγκέντρωση σε αυτήν την κατηγορία.
    - Αν και δεν είναι καθολική, η πλοκή δείχνει μια ισχυρή τάση για υψηλή ικανοποίηση μεταξύ των πελατών στο Λος Άντζελες.
3. **Σικάγο (CH):**
    - Όλοι οι πελάτες στο Σικάγο εκφράζουν ομοιόμορφα το χαμηλότερο επίπεδο ικανοποίησης (Επίπεδο 1).
    - Αυτή η ξεχωριστή συγκέντρωση αποκαλύπτει μια ευρεία τάση χαμηλότερης ικανοποίησης μεταξύ των πελατών στο Σικάγο.
4. **Σαν Φρανσίσκο (SF):**
    - Όλοι οι πελάτες στο Σαν Φρανσίσκο εκφράζουν σταθερά το υψηλότερο επίπεδο ικανοποίησης (Επίπεδο 3).
    - Αυτό αντικατοπτρίζει τη θετική τάση που παρατηρήθηκε στη Νέα Υόρκη, υποδηλώνοντας υψηλό επίπεδο ικανοποίησης στο Σαν Φρανσίσκο.
5. **Μαϊάμι (MI):**
    - Όλοι οι πελάτες στο Μαϊάμι παρουσιάζουν ομοιόμορφα Επίπεδο Ικανοποίησης 1.
    - Αυτό δείχνει ένα ευρέως διαδεδομένο μοτίβο χαμηλότερης ικανοποίησης μεταξύ των πελατών στο Μαϊάμι.

3.2.2.6 Satisfaction Level vs Item Purchased

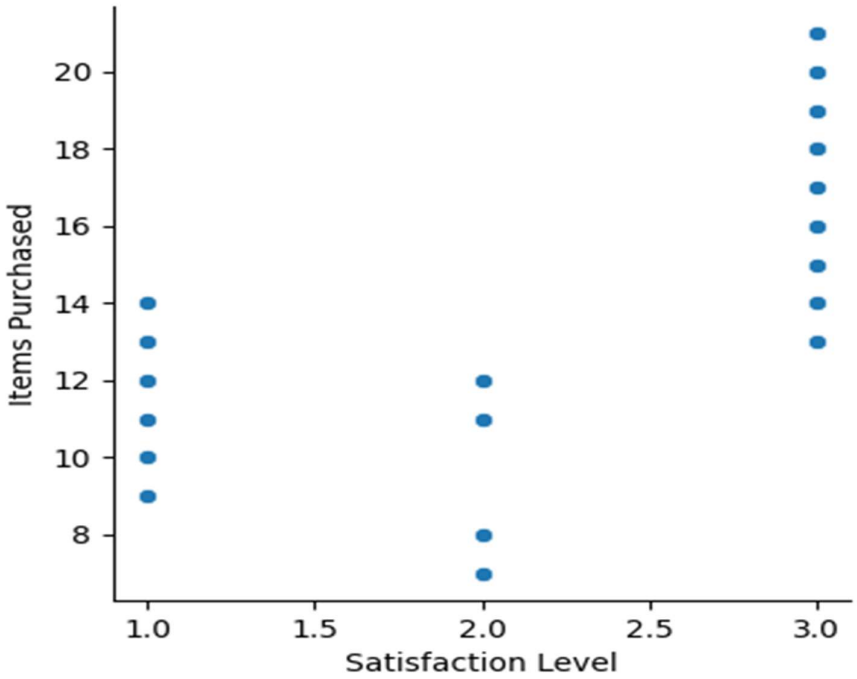
Η πλοκή Επίπεδο Ικανοποίησης έναντι Αγορασμένων Ειδών παρέχει μια οπτική αναπαράσταση της πιθανής σχέσης μεταξύ των επιπέδων ικανοποίησης των πελατών και του αριθμού των αγορασθέντων αντικειμένων. Σε αυτό το διάγραμμα διασποράς, κάθε σημείο δεδομένων αντιπροσωπεύει έναν μεμονωμένο πελάτη, με τον άξονα x να αντιπροσωπεύει το επίπεδο ικανοποίησης και τον άξονα y να υποδεικνύει τον αριθμό των αγορασθέντων αντικειμένων. Αναλύοντας την κατανομή των πόντων και το συνολικό μοτίβο στην πλοκή, οι αναλυτές μπορούν να αποκαλύψουν πληροφορίες για το πώς η ικανοποίηση των πελατών συσχετίζεται με την αγοραστική τους συμπεριφορά.

Αυτή η οπτική εξερεύνηση είναι ζωτικής σημασίας για τις επιχειρήσεις που επιδιώκουν να κατανοήσουν τον αντίκτυπο της αγοραστικής συμπεριφοράς στην ικανοποίηση των πελατών. Μοτίβα, όπως συμπλέγματα ή τάσεις, ενδέχεται να προκύψουν, υποδεικνύοντας εάν υπάρχει θετική συσχέτιση (μεγαλύτερη ικανοποίηση με περισσότερα προϊόντα που αγοράζονται) ή διαφορετική σχέση. Η

κατανόηση αυτών των δυναμικών μπορεί να βοηθήσει τις επιχειρήσεις να προσαρμόσουν τις στρατηγικές, τις προσφορές ή τα προγράμματα αφοσίωσης τους για να βελτιστοποιήσουν την ικανοποίηση των πελατών με βάση τις αγοραστικές τους συνήθειες.

Οι πληροφορίες που προκύπτουν από την πλοκή Επίπεδο Ικανοποίησης έναντι Αγορασμένων Αντικειμένων μπορούν να παρέχουν πληροφορίες για αποφάσεις που βασίζονται σε δεδομένα για επιχειρήσεις που στοχεύουν στην ενίσχυση της ικανοποίησης των πελατών μέσω στοχευμένων προσεγγίσεων. Για παράδειγμα, εάν η πλοκή αποκαλύπτει μια θετική συσχέτιση, οι επιχειρήσεις μπορεί να επικεντρωθούν στην παροχή κινήτρων για μεγαλύτερες αγορές για να αυξήσουν τα επίπεδα ικανοποίησης. Αντίθετα, εάν δεν υπάρχει σαφής συσχέτιση ή αρνητική τάση, οι στρατηγικές μπορούν να προσαρμοστούν για να αντιμετωπίσουν πιθανά σημεία αστοχιών στο ταξίδι του πελάτη που σχετίζονται με την εμπειρία αγορών. Τελικά, αυτή η οπτική ανάλυση συμβάλλει σε μια πιο λεπτή κατανόηση των παραγόντων που επηρεάζουν την ικανοποίηση των πελατών στο πλαίσιο της αγοραστικής τους συμπεριφοράς.

```
sns.pairplot(df,x_vars=['Satisfaction Level'],y_vars=["Items
Purchased"],height=4)
#count_bar_plot_with_shapiro(df,"Items Purchased",'Satisfaction
Level')
```



Πίνακας 3.12: Συσχέτιση μεταξύ Satisfaction Level & Items Purchased

Η πλοκή στο Επίπεδο Ικανοποίησης έναντι Αγορασμένων Ειδών αποκαλύπτει ενδιαφέρουσες τάσεις στη σχέση μεταξύ των επιπέδων ικανοποίησης των πελατών και της ποσότητας των αντικειμένων που αγοράζονται. Συγκεκριμένα, υπάρχει μια ευδιάκριτη και σαφώς θετική συσχέτιση όταν το επίπεδο ικανοποίησης βαθμολογείται ως 3. Σε τέτοιες περιπτώσεις, οι πελάτες που εκφράζουν τα υψηλότερα επίπεδα ικανοποίησης τείνουν να πραγματοποιούν σημαντικά μεγαλύτερο αριθμό αγορών. Αυτή η θετική συσχέτιση υποδηλώνει ότι το περιεχόμενο και οι πολύ ικανοποιημένοι πελάτες είναι πιο πιθανό να συμμετάσχουν σε πιο εκτεταμένες αγορές, επιδεικνύοντας μια ισχυρή σχέση μεταξύ της ικανοποίησής τους και της αγοραστικής τους συμπεριφοράς.

Ωστόσο, η γραφική παράσταση δείχνει μια λιγότερο σαφή σχέση για τους πελάτες με τα επίπεδα ικανοποίησης 2 και 1. Όταν τα επίπεδα ικανοποίησης είναι χαμηλότερα, η συσχέτιση με την ποσότητα των ειδών που αγοράζονται γίνεται λιγότερο εμφανής. Τα σημεία διασποράς για αυτά τα επίπεδα ικανοποίησης δεν παρουσιάζουν συνεπή και ισχυρή θετική τάση με τον αριθμό των αγαθών που αγοράστηκαν. Αυτή η διαφοροποιημένη διαφορά υποδηλώνει ότι η ενίσχυση της ικανοποίησης από ένα χαμηλότερο επίπεδο ενδέχεται να μην έχει απαραίτητα ως αποτέλεσμα αναλογική αύξηση της ποσότητας των αγοραζόμενων ειδών. Η κατανόηση αυτών των παραλλαγών είναι ζωτικής σημασίας για τις επιχειρήσεις να προσαρμόσουν τις στρατηγικές τους, αναγνωρίζοντας ότι η επίδραση της ικανοποίησης στην αγοραστική συμπεριφορά μπορεί να διαφέρει ανάλογα με το αρχικό επίπεδο ικανοποίησης.

Συνοπτικά:

1. **Σαφής θετική συσχέτιση για το επίπεδο ικανοποίησης 3:**
    - Η γραφική παράσταση επιπέδου ικανοποίησης έναντι αγορασμένων αντικειμένων αποκαλύπτει μια ευδιάκριτη και σαφώς θετική συσχέτιση όταν το επίπεδο ικανοποίησης βαθμολογείται ως 3.
    - Οι πελάτες που εκφράζουν τα υψηλότερα επίπεδα ικανοποίησης τείνουν να πραγματοποιούν σημαντικά μεγαλύτερο αριθμό αγορών.
    - Αυτή η θετική συσχέτιση υπογραμμίζει μια ισχυρή σχέση μεταξύ της ικανοποίησης και της ποσότητας των προϊόντων που αγοράζονται.
2. **Λιγότερο σαφής σχέση για τα επίπεδα ικανοποίησης 2 και 1:**
    - Η γραφική παράσταση δείχνει μια λιγότερο σαφή σχέση μεταξύ των επιπέδων ικανοποίησης 2 και 1 και της ποσότητας των ειδών που αγοράστηκαν.
    - Οι βαθμοί διασποράς για αυτά τα επίπεδα ικανοποίησης δεν παρουσιάζουν σταθερή και ισχυρή θετική τάση με τον αριθμό των αγορασθέντων αντικειμένων.

- Η ενίσχυση της ικανοποίησης από χαμηλότερα επίπεδα ενδέχεται να μην έχει απαραίτητα ως αποτέλεσμα αναλογική αύξηση της ποσότητας των αγοραζόμενων ειδών.

3.2.2.7 Satisfaction Level vs Average Rating

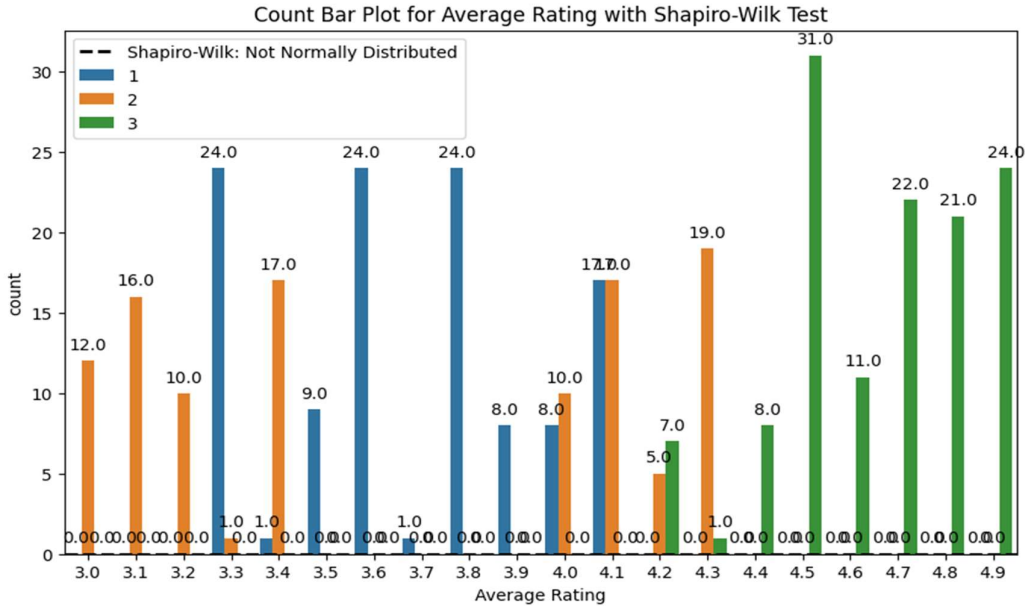
Η γραφική παράσταση Επίπεδο Ικανοποίησης έναντι Μέσης Βαθμολόγησης παρέχει μια οπτική αναπαράσταση της πιθανής σχέσης μεταξύ των επιπέδων ικανοποίησης των πελατών και της μέσης βαθμολογίας των αγορών. Σε αυτό το διάγραμμα διασποράς, κάθε σημείο δεδομένων αντιπροσωπεύει έναν μεμονωμένο πελάτη, με τον άξονα x να δείχνει το επίπεδο ικανοποίησης και τον άξονα y να αντιπροσωπεύει τη μέση βαθμολογία των αγορών τους. Αναλύοντας την κατανομή των πόντων και το συνολικό μοτίβο στην πλοκή, οι αναλυτές μπορούν να αποκτήσουν γνώσεις για το πώς η ικανοποίηση των πελατών ευθυγραμμίζεται με την αντιληπτή ποιότητα των αγορών τους.

Αυτή η οπτική εξερεύνηση είναι ζωτικής σημασίας για τις επιχειρήσεις που στοχεύουν να κατανοήσουν τον αντίκτυπο των αξιολογήσεων προϊόντων στη συνολική ικανοποίηση των πελατών. Μοτίβα, συμπλέγματα ή τάσεις στην γραφική παράσταση διασποράς μπορεί να αποκαλύψουν εάν υπάρχει θετική συσχέτιση, υποδεικνύοντας υψηλότερη ικανοποίηση από αγορές με υψηλότερη βαθμολογία ή εάν υπάρχουν άλλες δυναμικές. Η κατανόηση αυτών των σχέσεων μπορεί να καθοδηγήσει τις επιχειρήσεις στη βελτίωση των προσφορών των προϊόντων τους, στη βελτίωση της ποιότητας ή στην εφαρμογή στοχευμένων στρατηγικών για την ενίσχυση της ικανοποίησης των πελατών.

Οι πληροφορίες που προκύπτουν από την πλοκή Επίπεδο Ικανοποίησης έναντι Μέσης Βαθμολόγησης δίνουν τη δυνατότητα στις επιχειρήσεις να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με την ποιότητα των προϊόντων, τους μηχανισμούς ανατροφοδότησης πελατών και τις γενικές στρατηγικές βελτίωσης της ικανοποίησης. Για παράδειγμα, εάν παρατηρηθεί θετική συσχέτιση, οι επιχειρήσεις μπορεί να δώσουν προτεραιότητα στη διατήρηση προϊόντων ή υπηρεσιών υψηλής ποιότητας για να αυξήσουν τα επίπεδα ικανοποίησης. Αντίθετα, εάν υπάρχει αποσύνδεση μεταξύ των αξιολογήσεων και της ικανοποίησης, οι επιχειρήσεις μπορεί να χρειαστεί να διερευνήσουν και να αντιμετωπίσουν παράγοντες που επηρεάζουν τη συνολική ικανοποίηση των πελατών πέρα από την ποιότητα του προϊόντος. Αυτή η οπτική ανάλυση συμβάλλει στην πλήρη κατανόηση των παραγόντων που διαμορφώνουν την ικανοποίηση των πελατών στο πλαίσιο των αξιολογήσεων προϊόντων.

```
sns.pairplot(df,x_vars=['Satisfaction Level'],y_vars=["Average
Rating"],height=4)
```

```
count_bar_plot_with_shapiro(df,"Average Rating",'Satisfaction Level')
```



Η πλοκή Επίπεδο Ικανοποίησης έναντι Μέσης Βαθμολόγησης απεικονίζει διορατικά μοτίβα στη σχέση μεταξύ των επιπέδων ικανοποίησης των πελατών και των μέσων βαθμολογιών που αποδίδουν. Συγκεκριμένα, όταν το επίπεδο ικανοποίησης βαθμολογείται ως 3, υπάρχει μια αξιοσημείωτη τάση όπου η μέση βαθμολογία πέφτει σταθερά πάνω από 4. Αυτή η θετική συσχέτιση υποδηλώνει ότι οι πελάτες που εκφράζουν τα υψηλότερα επίπεδα ικανοποίησης αποδίδουν επίσης σταθερά υψηλότερες μέσες βαθμολογίες στα προϊόντα ή τις υπηρεσίες με τις οποίες ασχολούνται . Η κατανομή των δεδομένων σε αυτήν την περιοχή υποδηλώνει μια ισχυρή και θετική σχέση, τονίζοντας ότι οι πολύ ικανοποιημένοι πελάτες τείνουν να παρέχουν σταθερά ευνοϊκές αξιολογήσεις.

Αντίθετα, η γραφική παράσταση δείχνει μια λιγότερο σαφή σχέση για τους πελάτες με τα επίπεδα ικανοποίησης 2 και 1. Όταν τα επίπεδα ικανοποίησης είναι χαμηλότερα, η συσχέτιση με τη μέση βαθμολογία γίνεται λιγότερο εμφανής. Τα σημεία δεδομένων για αυτά τα επίπεδα ικανοποίησης δεν παρουσιάζουν σταθερή και ισχυρή θετική τάση, υποδηλώνοντας ότι τα χαμηλότερα επίπεδα ικανοποίησης μπορεί να μην μεταφράζονται σταθερά σε χαμηλότερες μέσες βαθμολογίες. Η κατανόηση αυτών των παραλλαγών είναι ζωτικής σημασίας για τις επιχειρήσεις καθώς ερμηνεύουν τα σχόλια των πελατών, αναγνωρίζοντας ότι η επίδραση της ικανοποίησης στη μέση βαθμολογία μπορεί να διαφέρει ανάλογα με το αρχικό επίπεδο ικανοποίησης. Αυτή η διαφοροποιημένη κατανόηση μπορεί να ενημερώσει στοχευμένες στρατηγικές για τη βελτίωση της ικανοποίησης των πελατών και, στη συνέχεια, τις μέσες βαθμολογίες για όσους εκφράζουν χαμηλότερα επίπεδα ικανοποίησης.

1. **Θετική συσχέτιση για την ικανοποίηση Επίπεδο 3:**
    - Στην γραφική παράσταση Επίπεδο ικανοποίησης έναντι Μέσης Βαθμολόγησης, όταν το επίπεδο ικανοποίησης βαθμολογείται ως 3, υπάρχει μια σταθερή τάση όπου η μέση βαθμολογία είναι πάντα πάνω από 4.
    - Αυτή η θετική συσχέτιση δείχνει ότι οι πελάτες που εκφράζουν τα υψηλότερα επίπεδα ικανοποίησης αποδίδουν σταθερά υψηλότερες μέσες βαθμολογίες στα προϊόντα ή τις υπηρεσίες με τις οποίες ασχολούνται.
    - Η διάδοση δεδομένων σε αυτήν την περιοχή τονίζει μια ισχυρή και θετική σχέση, δείχνοντας ότι οι πολύ ικανοποιημένοι πελάτες τείνουν να παρέχουν σταθερά ευνοϊκές αξιολογήσεις.
2. **Λιγότερο σαφής σχέση για τα επίπεδα ικανοποίησης 2 και 1:**
    - Αντίθετα, η πλοκή αποκαλύπτει μια λιγότερο σαφή σχέση για τους πελάτες με επίπεδα ικανοποίησης 2 και 1.
    - Όταν τα επίπεδα ικανοποίησης είναι χαμηλότερα, η συσχέτιση με τη μέση βαθμολογία γίνεται λιγότερο εμφανής και τα σημεία δεδομένων δεν παρουσιάζουν σταθερή και ισχυρή θετική τάση.
    - Αυτό υποδηλώνει ότι τα χαμηλότερα επίπεδα ικανοποίησης μπορεί να μην μεταφράζονται σταθερά σε χαμηλότερες μέσες αξιολογήσεις, υπογραμμίζοντας τη διαφοροποιημένη φύση της επίδρασης της ικανοποίησης στις αξιολογήσεις που παρέχονται από τους πελάτες.

3.2.2.8 Satisfaction Level vs Discount Applied

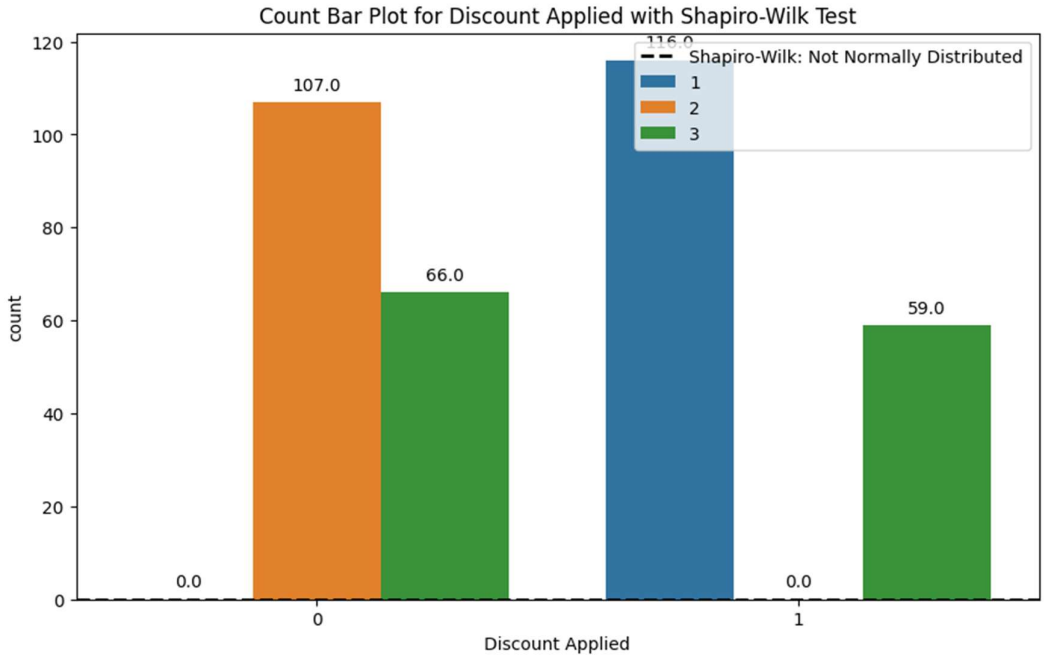
Η πλοκή Επίπεδο Ικανοποίησης έναντι Εφαρμοσμένης Έκπτωσης είναι μια οπτική αναπαράσταση που διερευνά τη πιθανή σχέση μεταξύ των επιπέδων ικανοποίησης των πελατών και των εκπτώσεων που εφαρμόζονται στις αγορές τους. Σε αυτό το διάγραμμα διασποράς, κάθε σημείο δεδομένων αντιπροσωπεύει έναν μεμονωμένο πελάτη, με τον άξονα x να δείχνει το επίπεδο ικανοποίησης και τον άξονα y να αντιπροσωπεύει το ποσό της έκπτωσης που εφαρμόζεται στις συναλλαγές τους. Η ανάλυση της κατανομής των πόντων και του συνολικού σχεδίου στην πλοκή παρέχει πληροφορίες για το πώς η ικανοποίηση των πελατών μπορεί να επηρεαστεί από τη διαθεσιμότητα ή το μέγεθος των εκπτώσεων.

Αυτή η οπτική εξερεύνηση είναι κρίσιμη για τις επιχειρήσεις που στοχεύουν να κατανοήσουν τον αντίκτυπο των στρατηγικών προώθησης στην ικανοποίηση των πελατών. Το διάγραμμα διασποράς μπορεί να αποκαλύψει μοτίβα, ομάδες ή τάσεις που υποδηλώνουν εάν υπάρχει θετική συσχέτιση (μεγαλύτερη ικανοποίηση με πιο σημαντικές εκπτώσεις) ή εάν άλλες δυναμικές επηρεάζουν την ικανοποίηση των πελατών. Η κατανόηση αυτών των σχέσεων επιτρέπει στις επιχειρήσεις να

βελτιστοποιούν τις εκπτωτικές τους στρατηγικές ώστε να ευθυγραμμίζονται καλύτερα με τις προσδοκίες και τις προτιμήσεις των πελατών.

Οι πληροφορίες από την πλοκή Επίπεδο Ικανοποίησης έναντι Εφαρμοσμένης Έκπτωσης δίνουν τη δυνατότητα στις επιχειρήσεις να λαμβάνουν αποφάσεις βάσει δεδομένων σχετικά με τις στρατηγικές τιμολόγησης και προώθησης των προϊόντων. Για παράδειγμα, εάν παρατηρηθεί θετική συσχέτιση, οι επιχειρήσεις μπορεί να εξετάσουν τη μόχλευση των εκπτώσεων ως εργαλείο για την ενίσχυση της ικανοποίησης και της αφοσίωσης των πελατών. Αντίθετα, εάν δεν υπάρχει σαφής συσχέτιση ή αρνητική τάση, οι επιχειρήσεις μπορεί να χρειαστεί να επαναξιολογήσουν την εκπτωτική τους προσέγγιση ή να διερευνήσουν άλλους παράγοντες που επηρεάζουν την ικανοποίηση των πελατών πέρα από τα κίνητρα τιμολόγησης. Αυτή η οπτική ανάλυση συμβάλλει σε μια πιο λεπτή κατανόηση της αλληλεπίδρασης μεταξύ των εκπτώσεων και της ικανοποίησης των πελατών στο πλαίσιο της αγοραστικής συμπεριφοράς.

```
#sns.pairplot(df,x_vars=['Satisfaction Level'],y_vars=["Discount Applied"],height=4)
count_bar_plot_with_shapiro(df,"Discount Applied",'Satisfaction Level')
```



Η πλοκή Επίπεδο Ικανοποίησης έναντι Εφαρμοσμένης Έκπτωσης αποκαλύπτει ενδιαφέρουσες πληροφορίες σχετικά με την αλληλεπίδραση μεταξύ των επιπέδων ικανοποίησης των πελατών και της εφαρμογής των εκπτώσεων. Συγκεκριμένα, όταν δεν εφαρμόζεται έκπτωση, η πλειονότητα των πελατών εκφράζει τα επίπεδα

ικανοποίησης 2 και 3. Αυτό υποδηλώνει ότι οι πελάτες είναι γενικά πιο ικανοποιημένοι όταν πραγματοποιούν αγορές χωρίς την επίδραση των εκπτώσεων. Η απουσία Επιπέδου Ικανοποίησης 1 σε αυτό το σενάριο υποδηλώνει ότι οι πελάτες, ελλείψει εκπτώσεων, τείνουν να εκδηλώνουν τουλάχιστον ένα μέτριο επίπεδο ικανοποίησης.

Αντίθετα, όταν εφαρμόζονται εκπτώσεις, η γραφική παράσταση δείχνει ένα διακριτό μοτίβο όπου η πλειονότητα των πελατών εκφράζει Επίπεδο Ικανοποίησης 1, χωρίς περιπτώσεις Ικανοποίησης Επιπέδου 2, και ορισμένες περιπτώσεις Επιπέδου Ικανοποίησης 3. Αυτή η παρατήρηση υποδεικνύει μια μετατόπιση στα επίπεδα ικανοποίησης, με αξιοσημείωτη αύξηση στα χαμηλότερα επίπεδα ικανοποίησης και μείωση στα υψηλότερα επίπεδα ικανοποίησης. Η παρουσία του Επιπέδου Ικανοποίησης 3 σε ορισμένες περιπτώσεις μπορεί να υποδηλώνει ότι ορισμένοι πελάτες, ακόμη και με εκπτώσεις, συνεχίζουν να εκφράζουν υψηλή ικανοποίηση.

Η κατανόηση αυτής της δυναμικής είναι απαραίτητη για τις επιχειρήσεις καθώς διαμορφώνουν στρατηγικές εκπτώσεων. Ενώ οι εκπτώσεις μπορεί να προσελκύουν πελάτες, η αλλαγή στα επίπεδα ικανοποίησης υπογραμμίζει τη σημασία του προσεκτικού σχεδιασμού και εφαρμογής δομών εκπτώσεων για να διασφαλιστεί ότι επηρεάζουν θετικά την ικανοποίηση των πελατών αντί να οδηγούν δυνητικά σε χαμηλότερα επίπεδα ικανοποίησης για ορισμένους πελάτες.

Συνοπτικά:

1. **Δεν ισχύει έκπτωση:**
    - Όταν δεν εφαρμόζεται έκπτωση, η πλειονότητα των πελατών εκφράζει τα επίπεδα ικανοποίησης 2 και 3.
    - Αυτό υποδηλώνει ότι οι πελάτες είναι γενικά πιο ικανοποιημένοι όταν πραγματοποιούν αγορές χωρίς την επίδραση των εκπτώσεων.
    - Η απουσία Επιπέδου Ικανοποίησης 1 υποδηλώνει ότι οι πελάτες, ελλείψει εκπτώσεων, τείνουν να εκδηλώνουν τουλάχιστον ένα μέτριο επίπεδο ικανοποίησης.
2. **Εφαρμόζεται έκπτωση:**
    - Αντίθετα, όταν εφαρμόζονται εκπτώσεις, η γραφική παράσταση δείχνει ένα ξεχωριστό μοτίβο όπου η πλειονότητα των πελατών εκφράζει Επίπεδο Ικανοποίησης 1, χωρίς περιπτώσεις Επιπέδου Ικανοποίησης 2 και ορισμένες περιπτώσεις Επιπέδου Ικανοποίησης 3.
    - Αυτό υποδηλώνει μια μετατόπιση στα επίπεδα ικανοποίησης όταν εισάγονται εκπτώσεις, με αξιοσημείωτη αύξηση στα χαμηλότερα επίπεδα ικανοποίησης και μείωση στα υψηλότερα επίπεδα ικανοποίησης.

- Η παρουσία του Επιπέδου Ικανοποίησης 3 σε ορισμένες περιπτώσεις υποδηλώνει ότι ορισμένοι πελάτες, ακόμη και με εκπτώσεις, συνεχίζουν να εκφράζουν υψηλή ικανοποίηση.

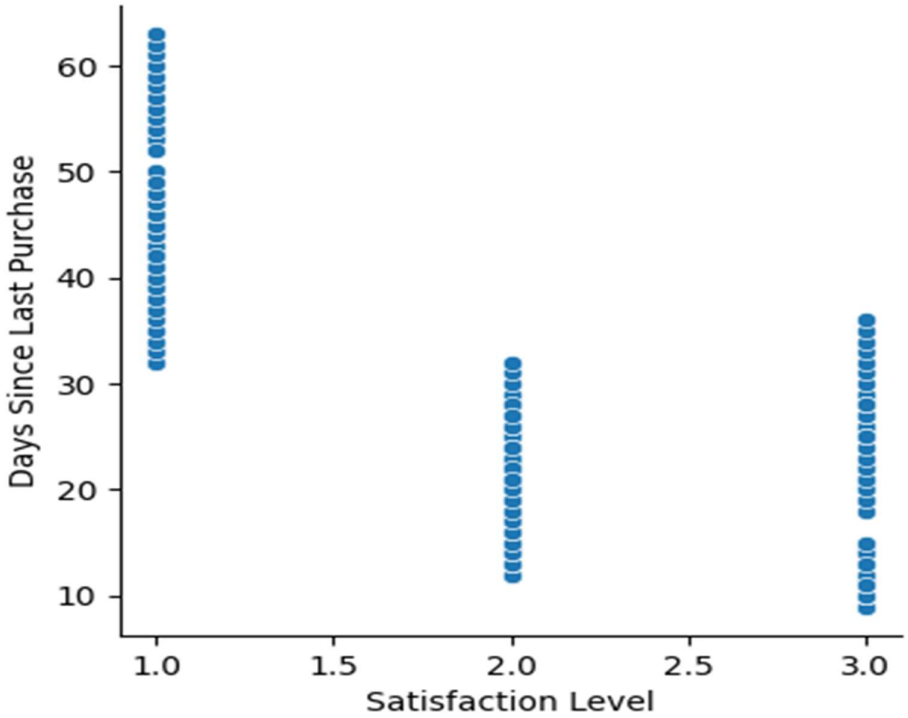
3.2.2.9 Satisfaction Level vs Days Since Last Purchase

Η πλοκή επιπέδου ικανοποίησης έναντι ημερών από την τελευταία αγορά είναι μια οπτική εξερεύνηση της πιθανής σχέσης μεταξύ των επιπέδων ικανοποίησης των πελατών και του χρόνου που έχει περάσει από την τελευταία τους αγορά. Κάθε σημείο στο διάγραμμα διασποράς αντιπροσωπεύει έναν μεμονωμένο πελάτη, με τον άξονα x να δείχνει το επίπεδο ικανοποίησης και τον άξονα y να απεικονίζει τον αριθμό των ημερών από την πιο πρόσφατη συναλλαγή τους. Η ανάλυση της κατανομής των πόντων και του συνολικού σχεδίου στην πλοκή παρέχει πληροφορίες για το πώς η πρόσφατη αγορά των πελατών μπορεί να συσχετιστεί με την ικανοποίησή τους.

Αυτή η οπτική ανάλυση είναι ιδιαίτερα σημαντική για τις επιχειρήσεις που θέλουν να κατανοήσουν τη δυναμική της ικανοποίησης των πελατών με την πάροδο του χρόνου. Μοτίβα, συμπλέγματα ή τάσεις στο διάγραμμα διασποράς μπορούν να αποκαλύψουν εάν υπάρχει θετική συσχέτιση, υποδηλώνοντας ότι οι πιο πρόσφατες αγορές σχετίζονται με υψηλότερα επίπεδα ικανοποίησης ή εάν άλλοι παράγοντες επηρεάζουν την ικανοποίηση των πελατών. Η κατανόηση αυτών των σχέσεων επιτρέπει στις επιχειρήσεις να προσαρμόζουν τις στρατηγικές δέσμευσης πελατών, όπως στοχευμένο μάρκετινγκ ή εξατομικευμένες προσφορές, με βάση την πρόσφατη συναλλαγή των πελατών.

Πληροφορίες από την πλοκή Επίπεδο Ικανοποίησης έναντι Ημερών Από την Τελευταία Αγορά μπορούν να καθοδηγήσουν τις επιχειρήσεις στην εφαρμογή αποτελεσματικών στρατηγικών διατήρησης πελατών. Για παράδειγμα, εάν παρατηρηθεί θετική συσχέτιση, μπορεί να υποδηλώνει ότι η διατήρηση της συχνής δέσμευσης πελατών συμβάλλει στην υψηλότερη ικανοποίηση. Αντίθετα, εάν δεν υπάρχει σαφής συσχέτιση ή αρνητική τάση, οι επιχειρήσεις μπορεί να χρειαστεί να διερευνήσουν τρόπους για να προσελκύσουν εκ νέου πελάτες που δεν έχουν κάνει πρόσφατες αγορές και να αντιμετωπίσουν τυχόν προβλήματα που επηρεάζουν την ικανοποίησή τους. Αυτή η οπτική εξερεύνηση συμβάλλει στην πλήρη κατανόηση της αλληλεπίδρασης μεταξύ της ικανοποίησης των πελατών και της χρονικής πτυχής της αγοραστικής τους συμπεριφοράς.

```
sns.pairplot(df,x_vars=['Satisfaction Level'],y_vars=["Days Since
Last Purchase"],height=4)
```



Πίνακας 3.15: Συσχέτιση μεταξύ Satisfaction Level vs Days Since Last Purchase

Η πλοκή του επιπέδου ικανοποίησης έναντι των ημερών από την τελευταία αγορά αποκαλύπτει διορατικά μοτίβα σχετικά με τη σχέση μεταξύ της πρόσφατης αγοράς των πελατών και των εκφραζόμενων επιπέδων ικανοποίησής τους. Συγκεκριμένα, όταν η διάρκεια από την τελευταία αγορά είναι σημαντική, όπως πάνω από 30 ημέρες, η πλειονότητα των πελατών εμφανίζει Επίπεδο Ικανοποίησης 1. Αυτή η παρατήρηση υποδηλώνει ότι όσο αυξάνεται το χρονικό διάστημα μεταξύ των αγορών, η ικανοποίηση των πελατών τείνει να μειώνεται, υποδεικνύοντας πιθανώς μια συσχέτιση μεταξύ σπάνιες αγορές και χαμηλότερα επίπεδα ικανοποίησης.

Αντίθετα, όταν οι Ημέρες Από την Τελευταία Αγορά είναι περίπου 30 ημέρες ή λιγότερο, η γραφική παράσταση δείχνει μια αλλαγή στα επίπεδα ικανοποίησης, με τους πελάτες να εκφράζουν κατά κύριο λόγο το Επίπεδο Ικανοποίησης 2 ή 3. Αυτό σημαίνει ότι οι πελάτες που κάνουν πιο πρόσφατες αγορές είναι πιο πιθανό να εκδηλώσουν υψηλότερη ικανοποίηση επίπεδα. Η αντίστροφη σχέση μεταξύ της πρόσφατης τιμής των αγορών και των επιπέδων ικανοποίησης υπογραμμίζει τη σημασία της έγκαιρης δέσμευσης των πελατών και τον πιθανό αντίκτυπο στην προώθηση υψηλότερων επιπέδων ικανοποίησης.

Η κατανόηση αυτών των δυναμικών είναι ζωτικής σημασίας για τις επιχειρήσεις που στοχεύουν στην εφαρμογή αποτελεσματικών στρατηγικών διατήρησης πελατών. Η αναγνώριση της σύνδεσης μεταξύ της πρόσφατης τιμής των αγορών και των

επιπέδων ικανοποίησης επιτρέπει στις επιχειρήσεις να προσαρμόσουν τις προσεγγίσεις τους με βάση το χρόνο που έχει περάσει από την τελευταία συναλλαγή. Η εφαρμογή στοχευμένων πρωτοβουλιών αφοσίωσης για πελάτες με μεγαλύτερα διαστήματα μεταξύ των αγορών μπορεί να βοηθήσει στον μετριασμό της πιθανής δυσαρέσκειας και να ενθαρρύνει την επανάληψη της επιχειρηματικής δραστηριότητας.

Συνοπτικά

1. **Ημέρες από την τελευταία αγορά > 30 ημέρες:**
    - Όταν η διάρκεια από την τελευταία αγορά είναι σημαντική, όπως πάνω από 30 ημέρες, η πλειοψηφία των πελατών παρουσιάζει επίπεδο ικανοποίησης 1.
    - Αυτό υποδηλώνει ότι καθώς το χάσμα μεταξύ των αγορών αυξάνεται, η ικανοποίηση των πελατών τείνει να μειωθεί.
    - Η παρατήρηση υποδεικνύει μια πιθανή συσχέτιση μεταξύ σπάνιων αγορών και χαμηλότερων επιπέδων ικανοποίησης.
2. **Ημέρες από την τελευταία αγορά ≤ 30 ημέρες:**
    - Αντίθετα, όταν οι ημέρες από την τελευταία αγορά είναι περίπου 30 ημέρες ή λιγότερο, η πλοκή δείχνει μια μετατόπιση των επιπέδων ικανοποίησης.
    - Οι πελάτες εκφράζουν κυρίως το επίπεδο ικανοποίησης 2 ή 3 σε αυτό το σενάριο.
    - Αυτό σημαίνει ότι οι πελάτες που πραγματοποιούν πιο πρόσφατες αγορές είναι πιο πιθανό να παρουσιάσουν υψηλότερα επίπεδα ικανοποίησης.

3.2.2.10 Μείωση Διαστάσεων

3.2.2.10.1 Κατάργηση Γένους

```
columns_to_drop = ['Gender']
df= drop_columns(df, columns_to_drop)
display_data(df)
```

Στο πλαίσιο της ανάλυσής μας, η απόφαση για απόρριψη του χαρακτηριστικού "Φύλο" προέρχεται από μια προσεκτική εξέταση της στατιστικής σημασίας και κατανομής του. Μετά από ενδελεχή εξέταση, παρατηρήθηκε ότι η κατανομή του φύλου σε όλο το σύνολο δεδομένων είναι σχεδόν ίση, χωρίς διακριτή ανισορροπία μεταξύ των κατηγοριών ανδρών και γυναικών. Η απουσία έντονης λοξής κατανομής δείχνει ότι η αναπαράσταση του φύλου είναι καλά ισορροπημένη, γεγονός που

καθιστά λιγότερο πιθανό να συμβάλει σημαντικά σε προγνωστικές ιδέες στην ανάλυσή μας.

Επιπλέον, η στατιστική ανάλυση αποκάλυψε ότι το χαρακτηριστικό "Φύλο" δεν παρουσιάζει σαφές μοτίβο ή συσχέτιση με τη μεταβλητή στόχο ή άλλα σχετικά χαρακτηριστικά. Σε σενάρια όπου ένα χαρακτηριστικό δεν έχει σαφή συσχέτιση με τα αποτελέσματα ενδιαφέροντος, η διατήρησή του ενδέχεται να μην συμβάλλει ουσιαστικά στις προγνωστικές δυνατότητες του μοντέλου μας. Κατά συνέπεια, καταργώντας τη δυνατότητα "Φύλο", στοχεύουμε να βελτιστοποιήσουμε το σύνολο δεδομένων και να επικεντρωθούμε σε λειτουργίες με μεγαλύτερη επιρροή που μπορούν να ενημερώσουν καλύτερα την ανάλυσή μας και να βελτιώσουν την απόδοση του μοντέλου.

Αυτή η απόφαση ευθυγραμμίζεται με την αρχή της επιλογής χαρακτηριστικών, δίνοντας έμφαση στη σημασία της συμπερίληψης μόνο εκείνων των χαρακτηριστικών που φέρνουν ουσιαστικές πληροφορίες στην ανάλυση. Αφαιρώντας το "Φύλο", στοχεύουμε να βελτιώσουμε την αποτελεσματικότητα του μοντέλου μας εστιάζοντας σε χαρακτηριστικά που παρουσιάζουν πιο έντονες σχέσεις με τη μεταβλητή-στόχο, διευκολύνοντας έτσι μια πιο ακριβή και ερμηνεύσιμη ανάλυση ταξινόμησης πολλαπλών κατηγοριών.

3.2.2.10.2 Κατάργηση Applied Discount

```
columns_to_drop = ['Discount Applied']
df= drop_columns(df, columns_to_drop)
display_data(df)
```

Κατά τη διαδικασία βελτίωσης του συνόλου δεδομένων μας για ανάλυση, ελήφθη μια προσεκτική απόφαση να εξαιρεθεί η δυνατότητα "Εφαρμόζεται έκπτωση" λόγω της παρουσίας ασυνήθιστων και δυνητικά εσφαλμένων ευρημάτων. Μετά από προσεκτική εξέταση, παρατηρήθηκε ότι οι περιπτώσεις όπου οι πελάτες παρουσίασαν Επίπεδο Ικανοποίησης 1 συσχετίστηκαν με την εφαρμογή εκπτώσεων, υποδηλώνοντας μια αντίθετη σχέση μεταξύ της μείωσης της τιμής και της ικανοποίησης των πελατών. Αυτή η απροσδόκητη συσχέτιση δημιούργησε ανησυχίες και οδήγησε στην υπόθεση ότι μπορεί να υπάρχουν παρατυπίες στην εφαρμογή των εκπτώσεων, πιθανώς ενδεικτικό σφάλματος συστήματος ή εσφαλμένης διαμόρφωσης.

Λαμβάνοντας υπόψη ότι έρχεται σε αντίθεση με τις συμβατικές προσδοκίες για τους πελάτες να εκφράσουν χαμηλότερα επίπεδα ικανοποίησης όταν πληρώνουν λιγότερα, τα ανώμαλα ευρήματα που σχετίζονται με το "Εφαρμόζεται έκπτωση"

ενδέχεται να θέσουν σε κίνδυνο την αξιοπιστία της ανάλυσής μας. Προς το συμφέρον της διατήρησης της ακεραιότητας του συνόλου δεδομένων μας και της διασφάλισης της ακρίβειας της επακόλουθης μοντελοποίησης ταξινόμησης πολλαπλών κλάσεων, η απόφαση να παραλειφθεί αυτό το χαρακτηριστικό κρίθηκε απαραίτητη. Αυτή η προσέγγιση ευθυγραμμίζεται με τις βέλτιστες πρακτικές στην προεπεξεργασία δεδομένων, όπου ο εντοπισμός και η αφαίρεση ακραίων στοιχείων ή ασυνεπών προτύπων συμβάλλουν σε ένα πιο ισχυρό και αξιόπιστο αποτέλεσμα ανάλυσης. Εξαιρώντας τη λειτουργία "Εφαρμοσμένη έκπτωση", στοχεύουμε να βελτιώσουμε την ποιότητα του συνόλου δεδομένων μας και να προωθήσουμε πιο ακριβείς πληροφορίες σχετικά με τους παράγοντες που επηρεάζουν την ικανοποίηση των πελατών.

3.2.2.11 Churn vs Membership Type

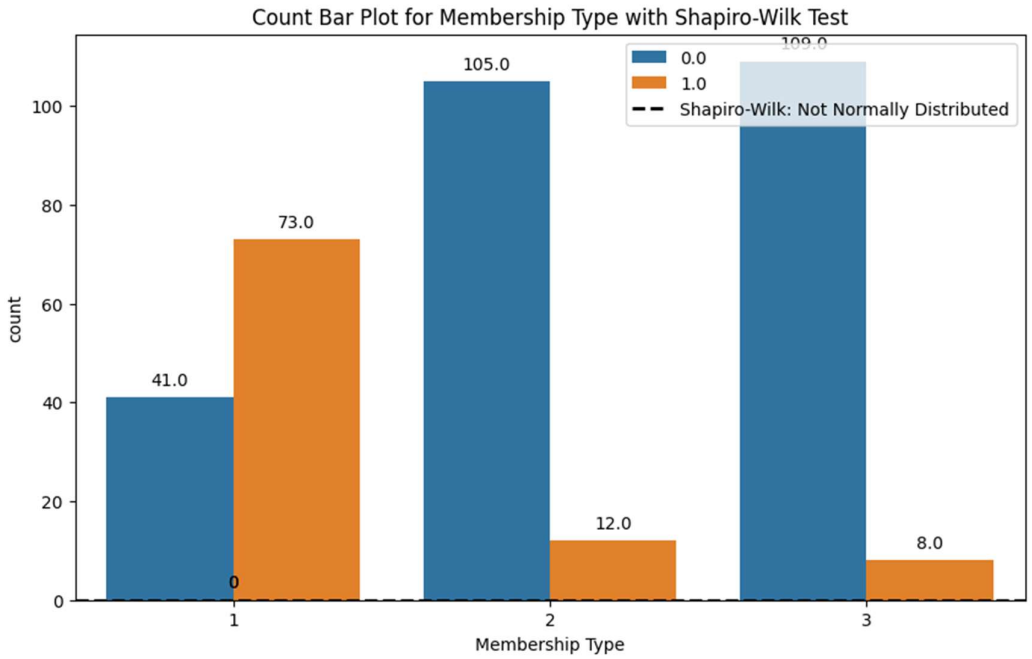
Η κατανόηση της σχέσης μεταξύ "Churn" και "Τύπος μέλους" είναι ζωτικής σημασίας για τις επιχειρήσεις που επιδιώκουν να βελτιστοποιήσουν τις στρατηγικές διατήρησης πελατών και να βελτιώσουν τη συνολική εμπειρία των πελατών τους. Οι τύποι συνδρομής συχνά χρησιμεύουν ως βασικός δείκτης του επιπέδου αφοσίωσης και δέσμευσης ενός πελάτη σε μια επιχείρηση. Αναλύοντας τα ποσοστά ανατροπής σε διαφορετικούς τύπους μελών, οι εταιρείες μπορούν να εντοπίσουν μοτίβα και τάσεις που βοηθούν στην ενημέρωση στοχευμένων παρεμβάσεων και στην προσαρμογή των προσπαθειών διατήρησης.

Πρώτον, η εξέταση του "Churn vs Membership Type" επιτρέπει στις επιχειρήσεις να διακρίνουν εάν συγκεκριμένα επίπεδα συνδρομής εμφανίζουν υψηλότερα ή χαμηλότερα ποσοστά απόκλισης. Αυτή η εικόνα είναι πολύτιμη για τη βελτίωση των στρατηγικών μάρκετινγκ και υπηρεσιών προσαρμοσμένων σε κάθε τμήμα μελών. Για παράδειγμα, εάν ένας συγκεκριμένος τύπος ιδιότητας μέλους αντιμετωπίζει αυξημένη ανατροπή, μπορεί να χρειαστεί μια πιο προσεκτική εξέταση των πλεονεκτημάτων, των υπηρεσιών ή των αλληλεπιδράσεων με τους πελάτες που σχετίζονται με αυτό το επίπεδο. Αντίθετα, η κατανόηση των τύπων συνδρομών που παρουσιάζουν χαμηλότερα ποσοστά απόσυρσης μπορεί να καθοδηγήσει τις προσπάθειες για την ενίσχυση και την προώθηση των πτυχών που συμβάλλουν στην αφοσίωση των πελατών σε αυτά τα τμήματα.

Δεύτερον, η ανάλυση του "Churn vs Membership Type" βοηθά στη στρατηγική κατανομή των πόρων. Επικεντρώνοντας τις προσπάθειές τους στη διατήρηση πελατών σε τύπους συνδρομών που είναι πιο επιρρεπείς σε ανατροπή, οι επιχειρήσεις μπορούν να εφαρμόσουν στοχευμένες καμπάνιες διατήρησης, εξατομικευμένα κίνητρα ή βελτιωμένες υπηρεσίες για τον μετριασμό των κινδύνων εκτροπής. Αντίθετα, για τύπους συνδρομής με σταθερά χαμηλά ποσοστά ανανέωσης,

οι πόροι μπορούν να ανακατανεμηθούν ή να βελτιστοποιηθούν για να ενισχυθεί η ικανοποίηση των πελατών και να αξιοποιηθεί η υπάρχουσα αφοσίωση. Συνολικά, αυτή η ανάλυση παρέχει στις επιχειρήσεις χρήσιμες πληροφορίες για να αναπτύξουν μια πιο αποτελεσματική και εξατομικευμένη προσέγγιση για τη διατήρηση των πελατών, ενισχύοντας μακροπρόθεσμες σχέσεις με τους πελάτες και διασφαλίζοντας βιώσιμη επιχειρηματική ανάπτυξη.

```
count_bar_plot_with_shapiro(df,"Membership Type",'Churn')
```



Η ανάλυση του Τύπου μέλους έναντι του Churn αποκαλύπτει ένα εντυπωσιακό μοτίβο στο σύνολο δεδομένων. Συγκεκριμένα, κατά την εξέταση του Τύπου μέλους 1, ένας σημαντικός αριθμός περιπτώσεων συσχετίζεται με υψηλό επίπεδο Churn, συνολικά 73 μετρήσεις. Αυτό υποδηλώνει ότι οι πελάτες που εμπίπτουν στον Τύπο Μέλους 1 είναι πιο επιρρεπείς σε αναταράξεις, υποδηλώνοντας μια πιθανή περιοχή ανησυχίας για την επιχείρηση. Το υψηλό ποσοστό ανατροπής σε αυτήν την κατηγορία μελών μπορεί να δικαιολογήσει περαιτέρω διερεύνηση παραγόντων όπως η ικανοποίηση των πελατών, η ποιότητα των υπηρεσιών ή η ανάγκη για στοχευμένες στρατηγικές διατήρησης.

Αντίθετα, κατά την εξερεύνηση των τύπων μελών 2 και 3, τα δεδομένα απεικονίζουν σημαντικά χαμηλότερη συχνότητα ανατροπής, με και τους δύο τύπους να καταγράφουν λιγότερες από 11 εμφανίσεις ο καθένας. Αυτό υποδηλώνει ένα συγκριτικά σταθερό σενάριο διατήρησης πελατών για αυτές τις κατηγορίες μελών. Η

περιορισμένη απόκλιση που παρατηρείται στους τύπους μελών 2 και 3 μπορεί να υποδηλώνει ότι η επιχείρηση διατηρεί με επιτυχία πελάτες σε αυτά τα τμήματα, πιθανώς λόγω βελτιωμένων υπηρεσιών, προγραμμάτων αφοσίωσης ή άλλων παραγόντων που συμβάλλουν στην ικανοποίηση των πελατών. Η κατανόηση αυτών των διαφορών στα επίπεδα απόκλισης μεταξύ διαφορετικών τύπων συνδρομής παρέχει πολύτιμες πληροφορίες που μπορούν να παρέχουν στρατηγικές αποφάσεις που στοχεύουν στη μείωση της απόκλισης και στην ενίσχυση της αφοσίωσης των πελατών σε συγκεκριμένα τμήματα της πελατειακής βάσης.

Συνοπτικά:

- **Ο Τύπος μέλους 1** δείχνει υψηλό επίπεδο Churn **73 μετρήσεων**, σηματοδοτώντας πιθανές ανησυχίες και υποδηλώνοντας την ανάγκη διερεύνησης παραγόντων που επηρεάζουν τη διατήρηση των πελατών.
- Αντίθετα, **οι Τύποι μελών 2 και 3** παρουσιάζουν χαμηλή Απόδοση, ο καθένας με **λιγότερες από 12 εμφανίσεις**, υποδηλώνοντας σταθερή διατήρηση πελατών, πιθανώς λόγω επιτυχημένων στρατηγικών ή βελτιωμένων υπηρεσιών.
- Αυτές οι διακρίσεις παρέχουν πολύτιμες γνώσεις για τη λήψη στρατηγικών αποφάσεων για τη μείωση της αναστάτωσης και την ενίσχυση της αφοσίωσης σε συγκεκριμένα τμήματα πελατών.

3.2.2.12 Churn vs Total Spend

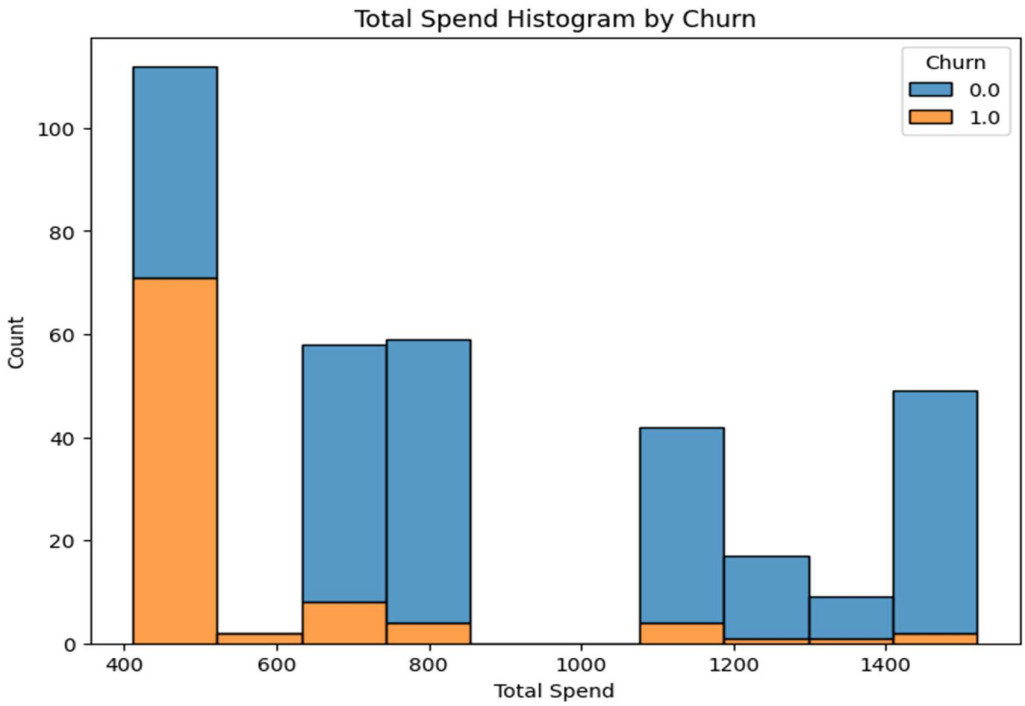
Η ανάλυση του "Total Spend vs Churn" έχει ύψιστη σημασία για τις επιχειρήσεις που επιδιώκουν να κατανοήσουν τον αντίκτυπο των συνολικών δαπανών των πελατών στην πιθανότητα ανατροπής τους. Η συνολική δαπάνη αντικατοπτρίζει τη σωρευτική οικονομική επένδυση ενός πελάτη σε μια επιχείρηση, που περιλαμβάνει πολλαπλές συναλλαγές και αλληλεπιδράσεις. Αυτή η μέτρηση χρησιμεύει ως ολοκληρωμένος δείκτης αφοσίωσης των πελατών. Η ανάλυση της συσχέτισης μεταξύ των συνολικών δαπανών και των ποσοστών απόσυρσης παρέχει πολύτιμες πληροφορίες για το εάν οι πελάτες με υψηλότερες δαπάνες είναι πιο πιθανό να παραμείνουν πιστοί ή εάν υπάρχει κίνδυνος απόκλισης που σχετίζεται με χαμηλότερες δαπάνες. Αυτή η κατανόηση επιτρέπει στις επιχειρήσεις να προσαρμόζουν στρατηγικές διατήρησης και να κατανέμουν πόρους αποτελεσματικά με βάση τις συμπεριφορές δαπανών των τμημάτων των πελατών τους.

Πρώτον, η σημασία του "Total Spend vs Churn" έγκειται στην ικανότητά του να αποκαλύπτει μοτίβα που υπογραμμίζουν τη σύνδεση μεταξύ των οικονομικών επενδύσεων των πελατών και της αφοσίωσής τους στην επιχείρηση. Οι πελάτες με υψηλότερες συνολικές δαπάνες μπορεί να θεωρηθούν πιο πολύτιμοι και μια θετική

συσχέτιση μεταξύ της συνολικής δαπάνης και των χαμηλότερων ποσοστών απόκλισης υποδηλώνει ότι αυτοί οι πελάτες υψηλής αξίας είναι πιο πιθανό να παραμείνουν αφοσιωμένοι στην επωνυμία. Αντίθετα, μια αρνητική συσχέτιση μπορεί να υποδηλώνει ότι ορισμένα τμήματα πελατών, ίσως αυτά με χαμηλότερες συνήθειες δαπανών, απαιτούν στοχευμένες παρεμβάσεις για να ενισχύσουν την αφοσίωσή τους και να αποτρέψουν την απόσυρση. Αυτή η ανάλυση βοηθά τις επιχειρήσεις να εντοπίσουν τις οικονομικές πτυχές που επηρεάζουν τη διατήρηση των πελατών, δίνοντάς τους τη δυνατότητα να επικεντρωθούν σε στρατηγικές που καλύπτουν τις μοναδικές ανάγκες διαφορετικών ομάδων δαπανών.

Δεύτερον, αυτή η ανάλυση καθοδηγεί τις επιχειρήσεις στην ανάπτυξη στρατηγικών προσεγγίσεων για τη διατήρηση των πελατών με βάση τις συμπεριφορές δαπανών. Για πελάτες με υψηλότερες συνολικές δαπάνες, οι επιχειρήσεις μπορούν να εφαρμόσουν προγράμματα αφοσίωσης, αποκλειστικές προσφορές ή εξατομικευμένες επικοινωνίες για να ενισχύσουν περαιτέρω τη δέσμευσή τους και να προωθήσουν μακροχρόνιες σχέσεις. Από την άλλη πλευρά, για πελάτες με χαμηλότερη συνολική δαπάνη, ενδέχεται να απαιτούνται προληπτικά μέτρα, όπως στοχευμένες προωθήσεις, βελτιωμένη εξυπηρέτηση πελατών ή κίνητρα για την αύξηση των δαπανών τους και τη μείωση του κινδύνου απόρριψης. Σε τελική ανάλυση, η ανάλυση "Συνολική δαπάνη έναντι εκκένωσης" παρέχει χρήσιμες πληροφορίες για τις επιχειρήσεις για να βελτιστοποιήσουν τις στρατηγικές διατήρησης πελατών, ευθυγραμμίζοντας τις προσπάθειες με τα πρότυπα δαπανών διαφορετικών τμημάτων πελατών.

```
histo_for_churn(df,"Total Spend",'Churn')
```



Η εξέταση της σχέσης μεταξύ Total Spent και Churn αποκαλύπτει μια ενδιαφέρουσα τάση στο σύνολο δεδομένων. Συγκεκριμένα, όταν το σύνολο των δαπανών πέσει κάτω από το όριο των 500, υπάρχει μια έντονη αιχμή στο Churn, φθάνοντας σε σημαντικό αριθμό 100. Αυτό υποδηλώνει ότι οι πελάτες που έχουν ξοδέψει λιγότερα από 500 παρουσιάζουν μεγαλύτερη πιθανότητα ανατροπής, υποδεικνύοντας μια πιθανή συσχέτιση μεταξύ χαμηλότερα επίπεδα δαπανών και αυξημένη τάση για διακοπή των υπηρεσιών. Αυτό το εύρημα υπογραμμίζει τη σημασία των στοχευμένων στρατηγικών για να προσελκύσουν και να διατηρήσουν πελάτες που εμπίπτουν στη χαμηλότερη κατηγορία δαπανών, όπως εξατομικευμένες προσφορές ή κίνητρα για να αυξήσουν τις δαπάνες τους και να βελτιώσουν τη συνολική τους ικανοποίηση.

Αντίθετα, καθώς το σύνολο των δαπανών ξεπερνά το 500, τα επίπεδα Churn παρουσιάζουν αξιοσημείωτη μείωση, με τις κορυφές να κυμαίνονται γύρω στο 50. Αυτό το μοτίβο υποδηλώνει μια πιο σταθερή βάση πελατών μεταξύ εκείνων που έχουν ξοδέψει πάνω από το όριο των 500, υποδηλώνοντας ότι οι υψηλότερες δαπάνες μπορεί να συμβάλλουν στη βελτίωση της διατήρησης πελατών. Οι επιχειρήσεις μπορεί να βρουν αξία στην εστίαση των προσπαθειών διατήρησης και την καλλιέργεια σχέσεων με πελάτες, των οποίων οι Συνολικές Δαπάνες ξεπερνούν αυτό το κρίσιμο όριο, πιθανώς προσφέροντας premium υπηρεσίες, αποκλειστικά προνόμια ή προγράμματα αφοσίωσης για την περαιτέρω βελτίωση της συνολικής εμπειρίας και αφοσίωσής τους. Η κατανόηση αυτών των δυναμικών παρέχει πολύτιμες πληροφορίες για τη δημιουργία στοχευμένων στρατηγικών διατήρησης με βάση τη συμπεριφορά των δαπανών των πελατών.

Συνοπτικά:

- **Συνολικά δαπανηθέντα < 500**: Μια σημαντική κορύφωση 100 καταμετρήσεων υποδηλώνει μεγαλύτερη πιθανότητα απώλειας πελατών μεταξύ εκείνων με επίπεδα δαπανών κάτω από 500. Οι στοχευμένες στρατηγικές, όπως οι εξατομικευμένες προωθήσεις, μπορεί να είναι επωφελείς για τη συμμετοχή και διατήρηση σε αυτό το τμήμα.
- **Συνολικά δαπανηθέντα > 500:** Τα επίπεδα ανατροπής μειώνονται, υποδηλώνοντας μια πιο σταθερή βάση πελατών για όσους έχουν ξοδέψει πάνω από το όριο των 500. Η εξέταση των premium υπηρεσιών ή προγραμμάτων πίστης μπορεί να ενισχύσει τη διατήρηση των πελατών σε αυτό το τμήμα.

3.2 2.13 Churn vs Gender

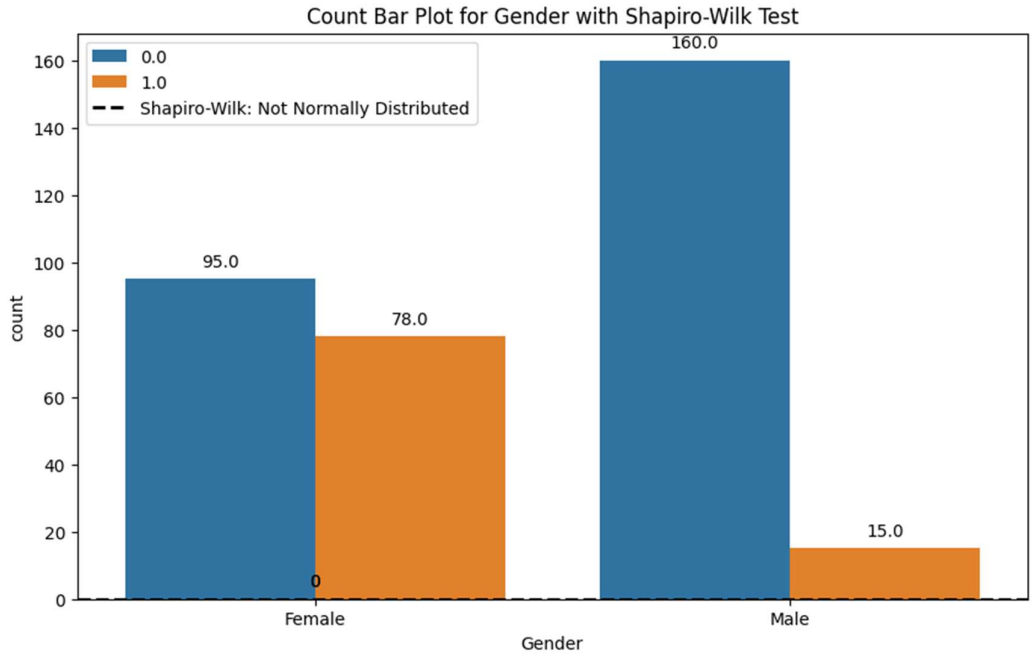
Η ανάλυση της σχέσης μεταξύ του "Gender vs Churn" είναι απαραίτητη για τις επιχειρήσεις που στοχεύουν να κατανοήσουν πώς τα δημογραφικά στοιχεία του φύλου επηρεάζουν τα ποσοστά απόκλισης πελατών. Οι γνώσεις με βάση το φύλο μπορούν να παρέχουν πολύτιμες πληροφορίες σχετικά με τις μοναδικές προτιμήσεις, συμπεριφορές και προσδοκίες ανδρών και γυναικών πελατών, καθοδηγώντας τις επιχειρήσεις στην προσαρμογή των στρατηγικών διατήρησης για την αντιμετώπιση των ειδικών αναγκών κάθε ομάδας φύλου. Η κατανόηση του εάν υπάρχουν μοτίβα που σχετίζονται με το φύλο στην απόκλιση πελατών επιτρέπει στις επιχειρήσεις να εφαρμόζουν στοχευμένες παρεμβάσεις, ενισχύοντας την ικανοποίηση και την αφοσίωση των πελατών.

Πρώτον, η σημασία του "Gender vs Churn" έγκειται στη δυνατότητά του να αποκαλύψει δυναμικές ειδικά για το φύλο που επηρεάζουν τη διατήρηση των πελατών. Τα διαφορετικά φύλα μπορεί να εμφανίζουν διαφορετικά μοτίβα όσον αφορά τις προτιμήσεις προϊόντων, τις προτιμήσεις επικοινωνίας και την ανταπόκριση στις προσπάθειες μάρκετινγκ. Για παράδειγμα, μια επιχείρηση που στοχεύει τόσο άντρες όσο και γυναίκες πελάτες μπορεί να διαπιστώσει ότι ορισμένες προσφορές ή στυλ επικοινωνίας έχουν πιο έντονη απήχηση με το ένα φύλο έναντι του άλλου. Η αναγνώριση αυτών των αποχρώσεων δίνει στις επιχειρήσεις τη δυνατότητα να δημιουργήσουν πιο εξατομικευμένες και αποτελεσματικές στρατηγικές, βελτιστοποιώντας την κατανομή των πόρων σε τομείς που είναι πιθανό να έχουν τον υψηλότερο αντίκτυπο στην αφοσίωση των πελατών.

Δεύτερον, αυτή η ανάλυση δίνει τη δυνατότητα στις επιχειρήσεις να ενισχύσουν τη συμπερίληψη και την πελατοκεντρική προσέγγισή τους, λαμβάνοντας υπόψη παράγοντες που σχετίζονται με το φύλο στις στρατηγικές τους διατήρησης. Κατανοώντας πώς το φύλο επηρεάζει τα ποσοστά εκτροπής, οι επιχειρήσεις μπορούν

να βελτιώσουν τα μηνύματα μάρκετινγκ, να σχεδιάσουν προϊόντα και να δημιουργήσουν εμπειρίες πελατών που έχουν καλύτερη απήχηση τόσο στους άνδρες όσο και στις γυναίκες. Είτε περιλαμβάνει προωθητικές ενέργειες ανάλογα με το φύλο, προσαρμοσμένα κανάλια επικοινωνίας ή βελτιώσεις προϊόντων, η ευθυγράμμιση των στρατηγικών με τις προτιμήσεις που σχετίζονται με το φύλο συμβάλλει σε μια πιο διαφοροποιημένη και αποτελεσματική προσέγγιση διατήρησης πελατών. Ουσιαστικά, η ανάλυση "Gender vs Churn" δίνει τη δυνατότητα στις επιχειρήσεις να καλλιεργήσουν ισχυρότερες συνδέσεις με την ποικίλη πελατειακή τους βάση, βελτιώνοντας τη συνολική ικανοποίηση των πελατών και μειώνοντας την αναστάτωση.

```
count_bar_plot_with_shapiro(df,"Gender",'Churn')
```



Η ανάλυση του φύλου έναντι του Churn στο σύνολο δεδομένων αποκαλύπτει μια αξιοσημείωτη απόκλιση στον αριθμό των ανακλάσεων μεταξύ ανδρών και γυναικών. Ιδιαίτερα, τα θηλυκά παρουσιάζουν υψηλότερο αριθμό αναδεύσεων, συνολικά περίπου 80 περιπτώσεις. Αυτό υποδηλώνει ότι μπορεί να υπάρχουν διακριτοί παράγοντες που επηρεάζουν τις γυναίκες πελάτες που συμβάλλουν σε μεγαλύτερη πιθανότητα ανατροπής. Η κατανόηση και η αντιμετώπιση αυτών των παραγόντων, όπως η ικανοποίηση των πελατών, οι προτιμήσεις υπηρεσιών ή οι στοχευμένες στρατηγικές επικοινωνίας, μπορεί να αποδειχθεί κρίσιμη για την ανάπτυξη αποτελεσματικών πρωτοβουλιών διατήρησης για τη γυναικεία δημογραφική ομάδα.

Αντίθετα, τα αρσενικά στο σύνολο δεδομένων εμφανίζουν σημαντικά χαμηλότερο αριθμό ανατροπών, περίπου 16 περιπτώσεις. Αυτή η χαμηλότερη συχνότητα ανατροπής μεταξύ των ανδρών πελατών υπογραμμίζει μια πιθανή περιοχή δύναμης ή ικανοποίησης σε αυτό το δημογραφικό στοιχείο. Οι επιχειρήσεις μπορούν να επωφεληθούν από τον εντοπισμό και τη μόχλευση των θετικών πτυχών που συμβάλλουν στην αφοσίωση των ανδρών πελατών, ενώ διερευνούν επίσης τρόπους βελτίωσης της συνολικής εμπειρίας πελατών και για τα δύο φύλα. Αυτές οι γνώσεις σχετικά με τα μοτίβα εκτροπής που σχετίζονται με το φύλο παρέχουν τη βάση για προσαρμοσμένες στρατηγικές που στοχεύουν στη μείωση της απόρριψης και στην ενίσχυση της μακροπρόθεσμης αφοσίωσης των πελατών με βάση τις προτιμήσεις και συμπεριφορές που σχετίζονται με το φύλο.

Συνοπτικά:

- **Φύλο vs. Churn**: Οι γυναίκες εμφανίζουν υψηλότερο αριθμό παρεκκλίσεων, περίπου 78 περιπτώσεις, υποδηλώνοντας μοναδικούς παράγοντες που επηρεάζουν τις γυναίκες πελάτες. Οι στρατηγικές στόχευσης που αφορούν την ικανοποίηση, τις προτιμήσεις υπηρεσιών ή την εξατομικευμένη επικοινωνία μπορεί να είναι ζωτικής σημασίας για τη διατήρηση των γυναικείων πελατών.
- **Αριθμός παρεκκλίσεων ανδρών**: Τα αρσενικά παρουσιάζουν χαμηλότερο αριθμό ανατροπών, περίπου 15 περιπτώσεις, υποδεικνύοντας μια πιθανή περιοχή δύναμης ή ικανοποίησης σε αυτό το δημογραφικό. Ο εντοπισμός και η αξιοποίηση θετικών πτυχών που συμβάλλουν στην αφοσίωση των ανδρών πελατών μπορεί να συμβάλει σε στρατηγικές για τη βελτίωση της συνολικής εμπειρίας των πελατών.

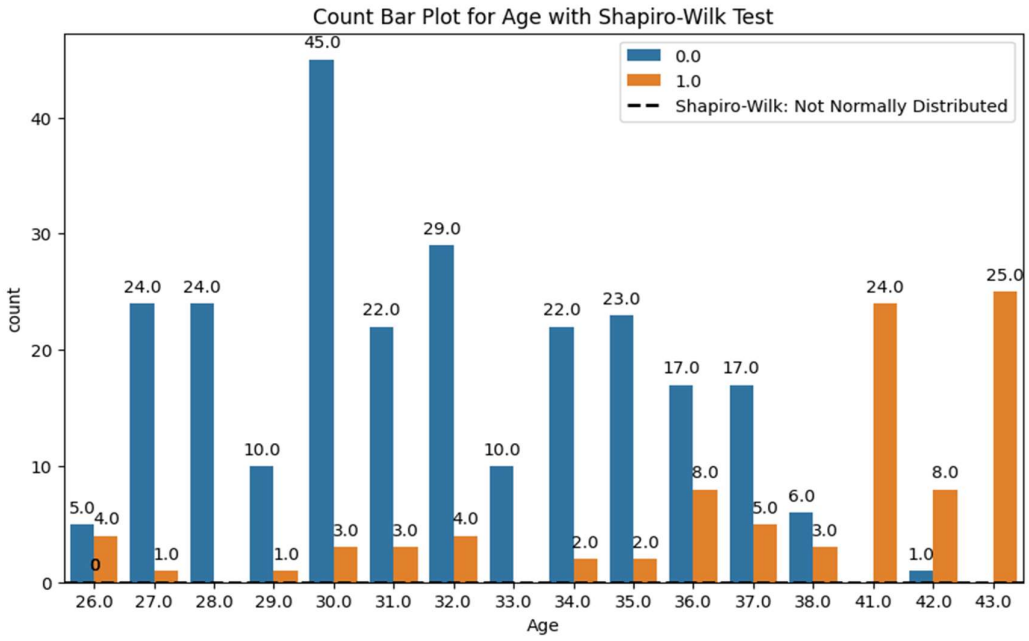
3.2.2.14 Churn vs Age

Η ανάλυση της σχέσης μεταξύ του "Age vs Churn" είναι ζωτικής σημασίας για τις επιχειρήσεις που επιδιώκουν να κατανοήσουν τον αντίκτυπο των δημογραφικών στοιχείων ηλικίας στα ποσοστά απόρριψης πελατών. Η ηλικία συχνά χρησιμεύει ως σημαντικός παράγοντας πρόβλεψης των συμπεριφορών και των προτιμήσεων των καταναλωτών και η συσχέτισή της με την απόκλιση μπορεί να αποκαλύψει πολύτιμες γνώσεις σχετικά με τους παράγοντες που επηρεάζουν την πίστη των πελατών σε διαφορετικές ηλικιακές ομάδες. Αυτή η κατανόηση είναι ζωτικής σημασίας για τις επιχειρήσεις να προσαρμόσουν τις στρατηγικές διατήρησης, τις προσφορές προϊόντων και τις πρωτοβουλίες δέσμευσης πελατών για να αντιμετωπίσουν αποτελεσματικά τις μοναδικές ανάγκες και προσδοκίες των πελατών σε διάφορες ηλικιακές ομάδες.

Πρώτον, η σημασία του "Age vs Churn" έγκειται στη δυνατότητά του να αποκαλύψει μοτίβα και τάσεις που αφορούν συγκεκριμένα την ηλικία στην απόσυρση πελατών. Ο προσδιορισμός του εάν ορισμένες ηλικιακές ομάδες είναι πιο επιρρεπείς σε αναταράξεις επιτρέπει στις επιχειρήσεις να δημιουργούν στοχευμένες παρεμβάσεις που να έχουν απήχηση στις συγκεκριμένες ανησυχίες ή προτιμήσεις κάθε δημογραφικού. Για παράδειγμα, οι νεότεροι πελάτες μπορεί να έχουν διαφορετικές προσδοκίες και προτιμήσεις αφοσίωσης σε σύγκριση με τους παλαιότερους ομολόγους τους, κάτι που απαιτεί ξεχωριστές στρατηγικές για να διατηρήσουν την αφοσίωσή τους. Αυτή η ανάλυση δίνει τη δυνατότητα στις επιχειρήσεις να ευθυγραμμίσουν τις προσπάθειές τους με τα διαφορετικά χαρακτηριστικά και συμπεριφορές που σχετίζονται με διαφορετικά ηλικιακά τμήματα, βελτιστοποιώντας τις πρωτοβουλίες διατήρησης πελατών.

Δεύτερον, η ανάλυση βοηθά στην ανάπτυξη πελατοκεντρικών στρατηγικών προσαρμοσμένων σε συγκεκριμένες ηλικιακές ομάδες. Κατανοώντας πώς η ηλικία επηρεάζει τα ποσοστά ανατροπής, οι επιχειρήσεις μπορούν να εφαρμόσουν εκστρατείες μάρκετινγκ για συγκεκριμένες ηλικίες, εξατομικευμένες προσφορές και στρατηγικές επικοινωνίας. Για παράδειγμα, η δημιουργία προγραμμάτων αφοσίωσης που ευθυγραμμίζονται με τις προτιμήσεις μιας συγκεκριμένης ηλικιακής ομάδας ή η αντιμετώπιση πιθανών σημείων αστοχιών που εμφανίζονται σε συγκεκριμένα δημογραφικά στοιχεία μπορεί να επηρεάσει σημαντικά την αποτελεσματικότητα των προσπαθειών διατήρησης. Τελικά, η ανάλυση "Age vs Churn" επιτρέπει στις επιχειρήσεις να βελτιώσουν την πελατοκεντρική τους προσέγγιση, ενισχύοντας ισχυρότερες συνδέσεις με πελάτες όλων των ηλικιών και βελτιώνοντας τα συνολικά ποσοστά διατήρησης.

```
count_bar_plot_with_shapiro(df,"Age",'Churn')
```



Η ανάλυση του "Age vs Churn" στο σύνολο δεδομένων μας αποκαλύπτει ένα αξιοσημείωτο μοτίβο, όπου η ηλικία των πελατών φαίνεται να παίζει σημαντικό ρόλο στην πιθανότητα ανατροπής τους. Συγκεκριμένα, κατά την εξέταση πελατών ηλικίας άνω των 40 ετών, παρατηρείται σημαντική αύξηση στη συχνότητα των περιπτώσεων απόκλισης. Αυτή η τάση υποδηλώνει ότι τα άτομα αυτής της ηλικιακής ομάδας μπορεί να είναι πιο επιρρεπή στη διακοπή της δέσμευσής τους με την επιχείρηση. Η κατανόηση αυτού του μοτίβου ανατροπής που σχετίζεται με την ηλικία είναι ζωτικής σημασίας για τη χάραξη στοχευμένων στρατηγικών διατήρησης, καθώς υπονοεί ότι η εταιρεία μπορεί να χρειαστεί να εφαρμόσει μέτρα ή κίνητρα ειδικά προσαρμοσμένα για να διατηρήσει την αφοσίωση των πελατών σε αυτό το δημογραφικό στοιχείο.

Αντίθετα, μεταξύ των πελατών ηλικίας κάτω των 40 ετών, παρατηρείται αξιοσημείωτη μείωση στη συχνότητα ανατροπής. Αυτό το εύρημα δείχνει ότι οι νεότεροι πελάτες επιδεικνύουν υψηλότερο επίπεδο διατήρησης ή αφοσίωσης στην επωνυμία. Οι επιχειρήσεις μπορούν να αξιοποιήσουν αυτή τη γνώση για να βελτιώσουν τις στρατηγικές μάρκετινγκ, εισάγοντας στοιχεία που ανταποκρίνονται ιδιαίτερα στις προτιμήσεις και τις προσδοκίες των νεότερων δημογραφικών ομάδων. Κατανοώντας τη διαφοροποιημένη σχέση μεταξύ ηλικίας και ανατροπής, οι εταιρείες μπορούν να εφαρμόσουν παρεμβάσεις για την ηλικία για να μετριάσουν αποτελεσματικά τους κινδύνους ανατροπής. Συνοπτικά, η ανάλυση "Age vs Churn" ρίχνει φως στη δυναμική που σχετίζεται με την ηλικία, όπου μπορεί να επηρεάσει σημαντικά τις στρατηγικές διατήρησης των πελατών, καθοδηγώντας τις επιχειρήσεις σε πιο στοχευμένες και αποτελεσματικές προσεγγίσεις για τη μείωση των ποσοστών εκτροπής.

Συνοπτικά:

- **Ηλικία άνω των 40:**
    - Παρατηρήθηκε υψηλότερη συχνότητα περιπτώσεων ανατροπής.
    - Προτείνει μια πιθανή ευπάθεια σε αυτή την ηλικιακή ομάδα.
    - Υποδεικνύει την ανάγκη για στοχευμένες στρατηγικές διατήρησης για πελάτες ηλικίας άνω των 40 ετών.
- **Ηλικία κάτω των 40:**
    - Παρατηρήθηκε χαμηλότερη συχνότητα ανατροπής.
    - Οι νεότεροι πελάτες παρουσιάζουν μεγαλύτερη διατήρηση ή αφοσίωση.
    - Ευκαιρίες για τη βελτίωση των στρατηγικών μάρκετινγκ για αυτό το δημογραφικό στοιχείο για την ενίσχυση της αφοσίωσης.

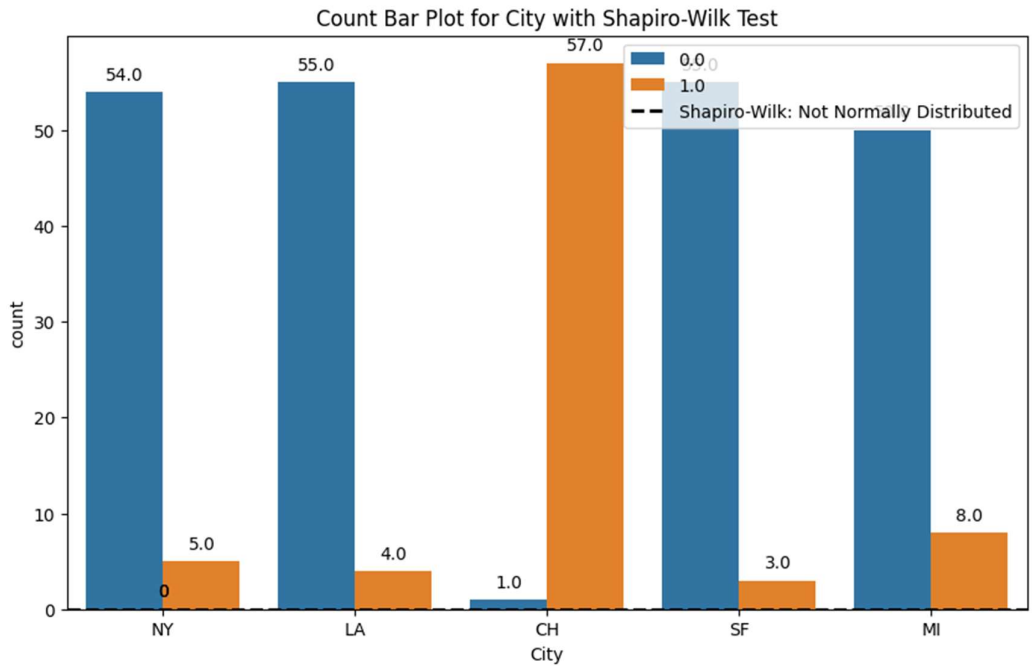
3.2.2.15 Churn vs City

Η ανάλυση της σχέσης μεταξύ του "City vs Churn" παρέχει στις επιχειρήσεις κρίσιμες γνώσεις σχετικά με τους γεωγραφικούς παράγοντες που επηρεάζουν τα ποσοστά απόσυρσης πελατών. Οι διαφορετικές πόλεις παρουσιάζουν συχνά ξεχωριστή δυναμική της αγοράς, πολιτιστικές αποχρώσεις και οικονομικές συνθήκες, τα οποία μπορούν να επηρεάσουν τη συμπεριφορά των πελατών και, κατά συνέπεια, την πιθανότητα εκτροπής. Με την εμβάθυνση σε αυτή τη σχέση, οι επιχειρήσεις μπορούν να εντοπίσουν μοτίβα που μπορεί να υποδεικνύουν περιφερειακές διακυμάνσεις στην ικανοποίηση των πελατών, την ένταση του ανταγωνισμού ή την οικονομική σταθερότητα, δίνοντάς τους τη δυνατότητα να προσαρμόσουν πιο αποτελεσματικά τις στρατηγικές διατήρησης.

Πρώτον, η σημασία του "City vs Churn" έγκειται στην ικανότητά του να αποκαλύπτει τοπικές τάσεις που μπορεί να επηρεάσουν την αφοσίωση των πελατών. Ορισμένες πόλεις ενδέχεται να αντιμετωπίσουν υψηλότερα ποσοστά ανατροπής λόγω συγκεκριμένων τοπικών παραγόντων, όπως ο αυξημένος ανταγωνισμός ή το υψηλότερο κόστος ζωής. Η κατανόηση αυτών των περιφερειακών παραλλαγών επιτρέπει στις επιχειρήσεις να προσαρμόσουν τις προσπάθειές τους διατήρησης, αντιμετωπίζοντας μοναδικές προκλήσεις και ευκαιρίες σε κάθε πόλη. Για παράδειγμα, μια πόλη με μεγαλύτερη συγκέντρωση ανταγωνιστών μπορεί να απαιτεί πιο επιθετικά προγράμματα αφοσίωσης ή εξατομικευμένα κίνητρα για να διατηρήσει αποτελεσματικά τους πελάτες.

Δεύτερον, αυτή η ανάλυση βοηθά στην κατανομή των πόρων και στη λήψη στρατηγικών αποφάσεων. Εντοπίζοντας πόλεις με υψηλά ποσοστά εκτροπής, οι επιχειρήσεις μπορούν να κατανείμουν πόρους στρατηγικά για να στοχεύσουν αυτές τις περιοχές με πιο εστιασμένες καμπάνιες μάρκετινγκ, βελτιωμένη υποστήριξη πελατών ή αποκλειστικές προσφορές. Αντίθετα, οι πόλεις με χαμηλότερα ποσοστά απόρριψης μπορεί να δικαιολογούν συνεχείς επενδύσεις για την αξιοποίηση της υπάρχουσας πίστης των πελατών. Τελικά, η ανάλυση "City vs Churn" καθοδηγεί τις επιχειρήσεις στη βελτιστοποίηση της προσέγγισής τους όσον αφορά τη διατήρηση των πελατών, αναγνωρίζοντας και αντιμετωπίζοντας τις περιφερειακές δυναμικές που συμβάλλουν στο συνολικό τοπίο της ανατροπής.

```
Count_bar_plot_with_shapiro(df,”City”,’Churn’)
```



Η ανάλυση του «City vs Churn» στο σύνολο δεδομένων άλλες αποκαλύπτει ένα ξεχωριστό μοτίβο όπου η πόλη του Σικάγο (CH) ξεχωρίζει με σημαντικά υψηλότερο αριθμό περιπτώσεων ανατροπής σε σύγκριση με άλλες πόλεις, όπως η Νέα Υόρκη (NY), το Μίσιγκαν (MI), το Λος Άντζελες (LA) και το Σαν Φρανσίσκο (SF). Αυτό το εύρημα υποδηλώνει ότι οι πελάτες στο Σικάγο παρουσιάζουν μια αξιοσημείωτη τάση να διακόψουν τη δέσμευσή τους με την επιχείρηση, υποδεικνύοντας πιθανές προκλήσεις ή ευκαιρίες μοναδικές σε αυτήν τη γεωγραφική περιοχή. Η κατανόηση των αυξημένων ρυθμών ανατροπής στο Σικάγο είναι ζωτικής σημασίας για τις επιχειρήσεις που δραστηριοποιούνται σε αυτήν την αγορά, καθώς απαιτεί

στοχευμένες παρεμβάσεις για την αντιμετώπιση των συγκεκριμένων παραγόντων που συμβάλλουν στην απώλεια πελατών στην πόλη.

Αντίθετα, σε άλλες πόλεις όπως η Νέα Υόρκη, το MI, το Λος Άντζελες και το SF, τα μέγιστα ποσοστά ανατροπής παρατηρούνται γύρω από μια χαμηλότερη αριθμητική τιμή, περίπου 5. Αυτό το σχετικά σταθερό μοτίβο σε αυτές τις πόλεις συνεπάγεται ένα πιο σταθερό τοπίο ανατροπής, πιθανώς επηρεασμένο από κοινά τοπικά χαρακτηριστικά ή τη δυναμική της αγοράς. Αν και μπορεί να υπάρχουν ακόμα παραλλαγές μεταξύ αυτών των πόλεων, η αντίθεση με το Σικάγο υπογραμμίζει τη σημασία της προσαρμογής των στρατηγικών διατήρησης με βάση τις μοναδικές προκλήσεις που τίθενται από τα διαφορετικά αστικά περιβάλλοντα. Συνολικά, η ανάλυση "City vs Churn" παρέχει αξιόπιστες πληροφορίες για τις επιχειρήσεις, ώστε να προσαρμόσουν τις προσεγγίσεις διατήρησης πελατών τους, τονίζοντας την ανάγκη για μια λεπτή, ειδική για την πόλη κατανόηση, με σκοπό τον αποτελεσματικό μετριασμό των κινδύνων απόρριψης.

Συνοπτικά:

- **Σικάγο (CH):**
    - Παρατηρήθηκαν σημαντικά υψηλότερες περιπτώσεις ανατροπής.
    - Υποδεικνύει ένα ξεχωριστό μοτίβο ανατροπής στην αγορά του Σικάγο.
    - Υπογραμμίζει την ανάγκη για στοχευμένες παρεμβάσεις για την αντιμετώπιση παραγόντων που συμβάλλουν στην ανατροπή στο Σικάγο.
- **Άλλες πόλεις (NY, MI, LA, SF):**
    - Μέγιστες ταχύτητες ανατροπής γύρω από μια χαμηλότερη αριθμητική τιμή (περίπου 5).
    - Προτείνει ένα σχετικά σταθερό τοπίο ανατροπής σε αυτές τις πόλεις.
    - Υποδεικνύει πιθανά κοινά περιφερειακά χαρακτηριστικά ή δυναμική της αγοράς.

3.2.2.16 Churn vs Item Purchased

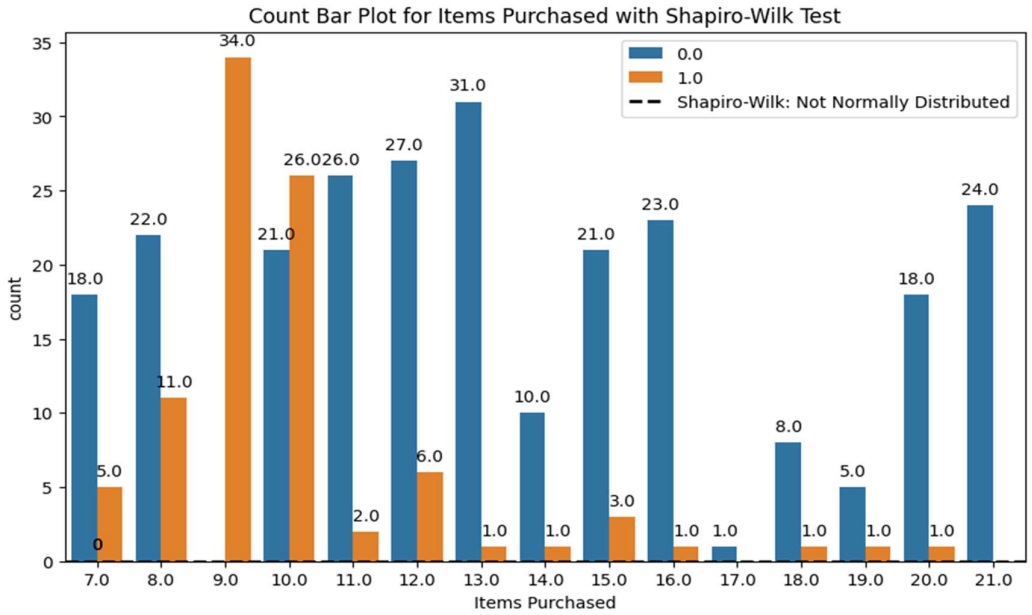
Η ανάλυση του "Είδη που αγοράζονται έναντι του Churn" έχει σημασία για τις επιχειρήσεις που στοχεύουν να κατανοήσουν πώς η αγοραστική συμπεριφορά των πελατών σχετίζεται με την πιθανότητα ανατροπής τους. Ο αριθμός των αντικειμένων που αγοράζονται είναι ένας βασικός δείκτης της αφοσίωσης και της ικανοποίησης των πελατών και η συσχέτισή του με την απόρριψη μπορεί να αποκαλύψει πολύτιμες γνώσεις σχετικά με τον αντίκτυπο των συναλλακτικών δραστηριοτήτων στη

διατήρηση των πελατών. Εξετάζοντας αυτή τη σχέση, οι επιχειρήσεις μπορούν να εντοπίσουν μοτίβα που υποδεικνύουν εάν οι πελάτες που κάνουν περισσότερες ή λιγότερες αγορές είναι πιο επιρρεπείς στην ανατροπή, επιτρέποντας στοχευμένες παρεμβάσεις για την ενίσχυση της αφοσίωσης των πελατών.

Πρώτον, η σημασία του "Items Purchased vs Churn" έγκειται στην ικανότητά του να διακρίνει μεταξύ πελατών υψηλής και χαμηλής αφοσίωσης όσον αφορά τη συναλλακτική δραστηριότητα. Εάν οι πελάτες που πραγματοποιούν μεγαλύτερο αριθμό αγορών παρουσιάζουν χαμηλότερα ποσοστά απόσυρσης, σημαίνει θετική συσχέτιση μεταξύ της αγοραστικής συμπεριφοράς και της αφοσίωσης των πελατών. Αντίθετα, εάν υπάρχει μια τάση αυξημένης ανατροπής μεταξύ των πελατών με λιγότερες αγορές, μπορεί να τονίσει πιθανή δυσαρέσκεια ή ζητήματα αποδέσμευσης που πρέπει να αντιμετωπίσουν οι επιχειρήσεις. Η κατανόηση αυτής της δυναμικής δίνει στις επιχειρήσεις τη δυνατότητα να προσαρμόσουν τις στρατηγικές μάρκετινγκ, τα προγράμματα αφοσίωσης και τις πρωτοβουλίες εξυπηρέτησης πελατών για να καλύψουν τις μοναδικές ανάγκες κάθε τμήματος πελατών.

Δεύτερον, η ανάλυση βοηθά στην ανάπτυξη στοχευμένων στρατηγικών διατήρησης με βάση τα πρότυπα αγορών. Για πελάτες που αγοράζουν συχνά αντικείμενα, οι επιχειρήσεις μπορούν να εφαρμόσουν προγράμματα ανταμοιβής, αποκλειστικές προσφορές ή εξατομικευμένες προτάσεις για να δώσουν κίνητρα για συνεχή αφοσίωση. Από την άλλη πλευρά, για πελάτες με λιγότερες αγορές, ενδέχεται να απαιτούνται προληπτικά μέτρα, όπως εξατομικευμένη προσέγγιση, βελτιωμένη υποστήριξη πελατών ή στοχευμένες προωθήσεις για τον μετριασμό των κινδύνων εκτροπής. Αναγνωρίζοντας τη σχέση μεταξύ των προϊόντων που αγοράζονται και της απόρριψης, οι επιχειρήσεις μπορούν να αναπτύξουν αποτελεσματικές παρεμβάσεις που ευθυγραμμίζονται με τις συγκεκριμένες ανάγκες και συμπεριφορές διαφορετικών τμημάτων πελατών, συμβάλλοντας τελικά στη βελτιωμένη ικανοποίηση των πελατών και τη διαρκή πίστη.

```
count_bar_plot_with_shapiro(df,"Items Purchased",'Churn')
```



Η ανάλυση του "Items Purchased vs Churn" στο σύνολο δεδομένων μας αποκαλύπτει μια αξιοσημείωτη τάση όπου οι πελάτες που πραγματοποιούν λιγότερες από 10 αγορές εμφανίζουν υψηλότερη συχνότητα παρεκκλίσεων, υποδηλώνοντας μια συσχέτιση μεταξύ χαμηλότερου αριθμού συναλλαγών και αυξημένης πιθανότητας αποδέσμευσης. Αντίθετα, οι πελάτες που πραγματοποιούν περισσότερες από 10 αγορές εμφανίζουν σημαντικά λιγότερα περιστατικά απόρριψης. Αυτή η σαφής διάκριση στα ποσοστά απόσυρσης με βάση τον αριθμό των προϊόντων που αγοράζονται υπογραμμίζει τον κεντρικό ρόλο της συναλλακτικής δραστηριότητας στον επηρεασμό της διατήρησης πελατών.

Το μοτίβο που παρατηρείται στα δεδομένα υποδηλώνει ότι οι πελάτες που πραγματοποιούν περιορισμένο αριθμό αγορών ενδέχεται να διατρέχουν υψηλότερο κίνδυνο ανατροπής. Αυτό θα μπορούσε να είναι ενδεικτικό διαφόρων παραγόντων, όπως η δυσαρέσκεια με τα προϊόντα ή τις υπηρεσίες, η έλλειψη αντιληπτής αξίας ή η ανάγκη για πιο στοχευμένες προσπάθειες για την ενίσχυση της αφοσίωσης των πελατών. Η κατανόηση αυτής της συσχέτισης επιτρέπει στις επιχειρήσεις να εφαρμόζουν στρατηγικά μέτρα, όπως εξατομικευμένες προωθήσεις, προγράμματα αφοσίωσης ή βελτιωμένη υποστήριξη πελατών, για την αντιμετώπιση των ειδικών αναγκών και ανησυχιών των πελατών με λιγότερες συναλλαγές και τον μετριασμό του κινδύνου εκτροπής.

Επιπλέον, για τους πελάτες που πραγματοποιούν περισσότερες από 10 αγορές, η χαμηλή συχνότητα ανατροπής υποδηλώνει υψηλότερο επίπεδο αφοσίωσης και ικανοποίησης. Οι επιχειρήσεις μπορούν να επωφεληθούν από αυτή τη γνώση καλλιεργώντας περαιτέρω σχέσεις με αυτούς τους πελάτες υψηλής αξίας, ίσως μέσω

αποκλειστικών προσφορών, προνομίων αφοσίωσης ή προσαρμοσμένης επικοινωνίας για την ενίσχυση της αφοσίωσής τους. Ουσιαστικά, η ανάλυση "Items Purchased vs Churn" προσφέρει αξιόπιστες πληροφορίες για τις επιχειρήσεις ώστε να βελτιώσουν τις στρατηγικές διατήρησης πελατών βάσει συναλλακτικών συμπεριφορών, ενισχύοντας τελικά ισχυρότερες και πιο διαρκείς σχέσεις με τους πελάτες.

Συνοπτικά:

- **Λιγότερα από 10 είδη που αγοράστηκαν:**
    - Παρατηρήθηκε υψηλότερη συχνότητα περιπτώσεων ανατροπής.
    - Υποδεικνύει μια πιθανή συσχέτιση μεταξύ χαμηλότερης συναλλακτικής δραστηριότητας και αυξημένης πιθανότητας αποδέσμευσης πελατών.
    - Υποδηλώνει την ανάγκη για στοχευμένες παρεμβάσεις για την αντιμετώπιση της δυσαρέσκειας ή των θεμάτων που αντιλαμβάνονται την αξία.
- **Περισσότερα από 10 είδη που αγοράστηκαν:**
    - Παρατηρήθηκε χαμηλότερη συχνότητα ανατροπής.
    - Προτείνει μια θετική συσχέτιση μεταξύ της υψηλότερης συναλλακτικής δέσμευσης και της αφοσίωσης των πελατών.
    - Ευκαιρίες για ενίσχυση της αφοσίωσης μέσω αποκλειστικών προσφορών ή προσαρμοσμένης επικοινωνίας για πελάτες υψηλής αξίας.

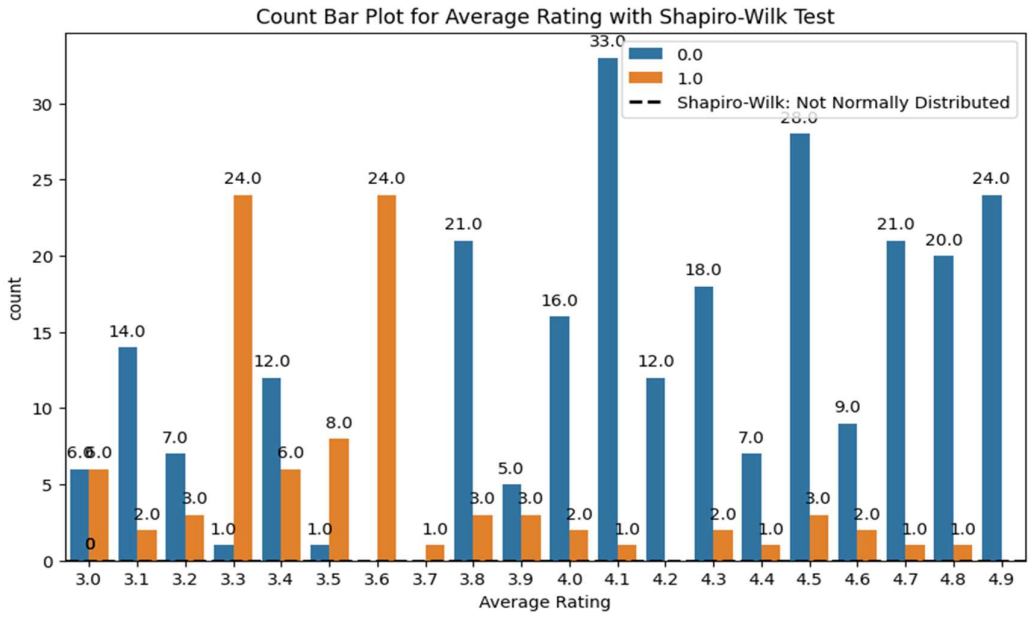
3.2.2.17 Churn vs Average Rating

Η ανάλυση του "Average Rating vs Churn" έχει κρίσιμη σημασία για τις επιχειρήσεις που στοχεύουν να μετρήσουν τον αντίκτυπο της ικανοποίησης των πελατών, όπως αντανακλάται στις μέσες αξιολογήσεις, στην πιθανότητα απόρριψης. Οι μέσες αξιολογήσεις χρησιμοποιούνται συχνά ως μια ολοκληρωμένη μέτρηση, όπου ενσωματώνει τις συνολικές εμπειρίες των πελατών με μια επιχείρηση, τα προϊόντα ή τις υπηρεσίες της. Η κατανόηση του τρόπου με τον οποίο αυτή η μέτρηση ικανοποίησης συσχετίζεται με τα ποσοστά απόκλισης παρέχει στις επιχειρήσεις πολύτιμες πληροφορίες σχετικά με την αποτελεσματικότητα των προσφορών τους και την εξυπηρέτηση πελατών. Μια θετική συσχέτιση μεταξύ υψηλότερων μέσων αξιολογήσεων και χαμηλότερων ποσοστών εκτροπής σημαίνει ότι οι ικανοποιημένοι πελάτες είναι πιο πιθανό να παραμείνουν πιστοί, ενώ μια αρνητική συσχέτιση μπορεί να υποδηλώνει δυσαρέσκεια ή ανεκπλήρωτες προσδοκίες, σηματοδοτώντας πιθανούς κινδύνους εκτροπής.

Πρώτον, η σημασία του "Average Rating vs Churn" έγκειται στον ρόλο του ως βαρόμετρο για την ικανοποίηση των πελατών. Οι πελάτες που παρέχουν σταθερά υψηλές αξιολογήσεις είναι πιθανό να είναι ικανοποιημένοι με τις εμπειρίες τους, ενισχύοντας το αίσθημα πίστης και μειώνοντας την πιθανότητα ανατροπής. Αντίθετα, μια τάση όπου οι χαμηλότερες μέσες βαθμολογίες συμπίπτουν με υψηλότερα ποσοστά απόσυρσης μπορεί να υποδηλώνει τομείς, όπου η επιχείρηση χρειάζεται βελτίωση για να βελτιώσει την ικανοποίηση και τη διατήρηση των πελατών. Οι επιχειρήσεις μπορούν να χρησιμοποιήσουν αυτές τις πληροφορίες για να προσδιορίσουν συγκεκριμένα σημεία αστοχιών ή τομείς αριστείας, επιτρέποντάς τους να δώσουν προτεραιότητα στις προσπάθειες που επηρεάζουν άμεσα την αντίληψη των πελατών και, στη συνέχεια, τα ποσοστά ανατροπής.

Δεύτερον, αυτή η ανάλυση καθοδηγεί τις επιχειρήσεις στο να προσαρμόσουν τις στρατηγικές διατήρησης των πελατών τους με βάση τα επίπεδα ικανοποίησης. Για πελάτες με υψηλή μέση βαθμολογία, η ενίσχυση θετικών εμπειριών μέσω στοχευμένων προωθήσεων, προγραμμάτων αφοσίωσης ή εξατομικευμένων επικοινωνιών μπορεί να ενισχύσει την αφοσίωσή τους. Αντίθετα, για πελάτες με χαμηλότερες αξιολογήσεις, ενδέχεται να απαιτούνται προληπτικά μέτρα, όπως βελτιωμένη υποστήριξη πελατών, βελτιώσεις προϊόντων ή στοχευμένες προσφορές για την αντιμετώπιση των ανησυχιών και την αποφυγή εκτροπής. Ουσιαστικά, η ανάλυση "Μέση Βαθμολογία εναντίον Churn" δίνει τη δυνατότητα στις επιχειρήσεις να διαχειρίζονται προληπτικά την ικανοποίηση των πελατών, ευθυγραμμίζοντας τις στρατηγικές τους με τα συγκεκριμένα σχόλια και τα συναισθήματα που εκφράζουν οι πελάτες για να οικοδομήσουν μόνιμες σχέσεις και να μειώσουν την αναστάτωση.

```
count_bar_plot_with_shapiro(df,"Average Rating",'Churn')
```



Η ανάλυση της "Μέσης βαθμολογίας έναντι του Churn στο σύνολο δεδομένων μας αποκαλύπτει μια συναρπαστική σχέση μεταξύ της ικανοποίησης των πελατών, όπως υποδεικνύεται από τις μέσες αξιολογήσεις, και της πιθανότητας απόκλισης. Συγκεκριμένα, οι πελάτες που έχουν δώσει μέση βαθμολογία μικρότερη από 4 εμφανίζουν σημαντικά υψηλότερη συχνότητα περιπτώσεων εκτροπής, γεγονός που υποδηλώνει ότι τα χαμηλότερα επίπεδα ικανοποίησης συνδέονται στενά με αυξημένη τάση αποδέσμευσης από την επιχείρηση. Αντίθετα, όταν οι πελάτες παρέχουν μέση βαθμολογία πάνω από το 4, υπάρχει μια αξιοσημείωτη μείωση στην απόρριψη, υποδεικνύοντας ότι τα υψηλότερα επίπεδα ικανοποίησης συμβάλλουν στην ενίσχυση της αφοσίωσης των πελατών και στη μείωση των κινδύνων απόρριψης.

Αυτό το ξεχωριστό μοτίβο υπογραμμίζει τον κεντρικό ρόλο της ικανοποίησης των πελατών στον επηρεασμό της δυναμικής της ανατροπής. Οι πελάτες που εκφράζουν χαμηλότερα επίπεδα ικανοποίησης, όπως αντικατοπτρίζονται στις μέσες αξιολογήσεις τους κάτω από το 4, ενδέχεται να είναι πιο διατεθειμένοι να αναζητήσουν εναλλακτικές λύσεις ή να διακόψουν τη δέσμευσή τους λόγω της αντιληπτής δυσαρέσκειας με τα προϊόντα, τις υπηρεσίες ή τη συνολική εμπειρία πελατών. Η κατανόηση αυτής της συσχέτισης επιτρέπει στις επιχειρήσεις να δώσουν προτεραιότητα στις προσπάθειες που επηρεάζουν άμεσα την ικανοποίηση των πελατών, οδηγώντας δυνητικά σε βελτιώσεις στις προσφορές προϊόντων, στην ποιότητα των υπηρεσιών ή στην υποστήριξη πελατών για την αντιμετώπιση των ανησυχιών αυτού του συγκεκριμένου τμήματος πελατών.

Η ανάλυση "Average Rating vs Churn" παρέχει χρήσιμες πληροφορίες για τις επιχειρήσεις ώστε να προσαρμόσουν τις στρατηγικές διατήρησης με βάση τα επίπεδα ικανοποίησης των πελατών. Για πελάτες με μέση βαθμολογία κάτω από 4, στοχευμένες παρεμβάσεις, όπως εξατομικευμένη προσέγγιση, κίνητρα αφοσίωσης ή βελτιώσεις υπηρεσιών, μπορούν να εφαρμοστούν για τον μετριασμό των κινδύνων εκτροπής και τη βελτίωση της συνολικής ικανοποίησης. Αντίθετα, για πελάτες με μέση βαθμολογία άνω του 4, η ενίσχυση των θετικών εμπειριών μέσω προγραμμάτων αφοσίωσης ή αποκλειστικών προσφορών μπορεί να ενισχύσει περαιτέρω την αφοσίωσή τους. Συνολικά, αυτή η ανάλυση δίνει τη δυνατότητα στις επιχειρήσεις να διαχειρίζονται προληπτικά την ικανοποίηση των πελατών, ευθυγραμμίζοντας τις στρατηγικές με τη συγκεκριμένη ανατροφοδότηση που εκφράζουν οι πελάτες για την ενίσχυση των διαρκών σχέσεων και τη μείωση της ανατροπής.

Συνοπτικά:

1. **Μέση βαθμολογία μικρότερη από 4:**
    - Παρατηρήθηκε υψηλότερη συχνότητα περιπτώσεων ανατροπής.

- Ισχυρή συσχέτιση μεταξύ χαμηλότερων επιπέδων ικανοποίησης και αυξημένης τάσης για ανατροπή.
- Υποδεικνύει την ανάγκη για στοχευμένες παρεμβάσεις για την αντιμετώπιση της δυσαρέσκειας και τη βελτίωση της εμπειρίας των πελατών.

2. **Μέση βαθμολογία πάνω από 4:**
    - Αξιοσημείωτη μείωση στις περιπτώσεις ανατροπής.
    - Υψηλότερα επίπεδα ικανοποίησης που σχετίζονται με αυξημένη πίστη πελατών.
    - Ευκαιρίες για ενίσχυση θετικών εμπειριών μέσω προγραμμάτων αφοσίωσης ή αποκλειστικών προσφορών.

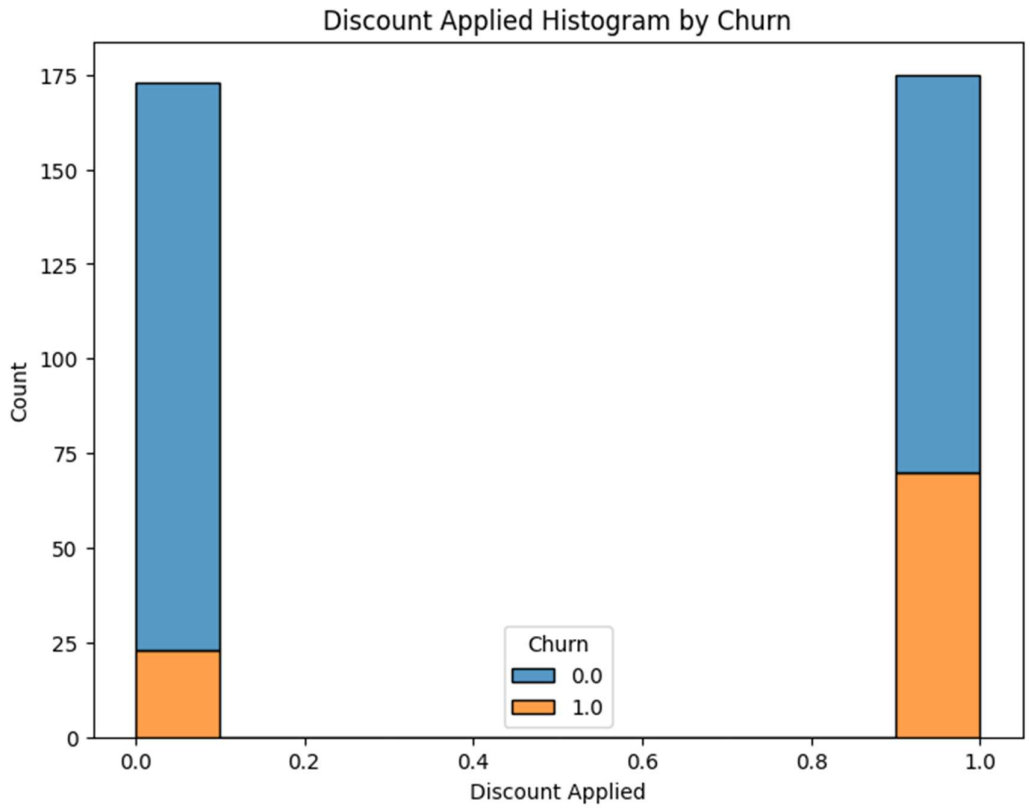
3.2.2.18 Churn vs Discount Applied

Η ανάλυση του "Discount Applied vs Churn" είναι ζωτικής σημασίας για τις επιχειρήσεις που στοχεύουν να κατανοήσουν την περίπλοκη σχέση μεταξύ της εφαρμογής των εκπτώσεων και της πιθανότητας απόσυρσης πελατών. Οι εκπτώσεις και οι προσφορές είναι ισχυρά εργαλεία για να επηρεάσουν τη συμπεριφορά των πελατών και μπορούν να επηρεάσουν τόσο τις βραχυπρόθεσμες συναλλαγές όσο και τη μακροπρόθεσμη αφοσίωση. Η εξέταση του τρόπου με τον οποίο η εφαρμογή των εκπτώσεων συσχετίζεται με τα ποσοστά ανατροπής παρέχει πολύτιμες πληροφορίες για το εάν οι πελάτες που επωφελούνται συχνά από τις εκπτώσεις είναι περισσότερο ή λιγότερο πιθανό να ανατραπούν. Αυτή η κατανόηση είναι απαραίτητη για τις επιχειρήσεις να προσαρμόσουν τις εκπτωτικές τους στρατηγικές, διασφαλίζοντας ότι ευθυγραμμίζονται με τους στόχους διατήρησης πελατών μεγιστοποιώντας παράλληλα τον συνολικό αντίκτυπο στις πωλήσεις και την ικανοποίηση των πελατών.

Πρώτον, η σημασία του "Discount Applied vs Churn" έγκειται στη δυνατότητά του να αποκαλύψει τον αντίκτυπο των στρατηγικών εκπτώσεων στην αφοσίωση των πελατών. Μια θετική συσχέτιση μεταξύ της εφαρμογής εκπτώσεων και των χαμηλότερων ποσοστών ανατροπής μπορεί να υποδηλώνει ότι οι πελάτες που απολαμβάνουν συχνά μειωμένες τιμές παρουσιάζουν μεγαλύτερη ικανοποίηση και είναι πιο πιθανό να παραμείνουν πιστοί στη μάρκα. Αντίθετα, μια αρνητική συσχέτιση θα μπορούσε να υποδεικνύει ότι ορισμένες πρακτικές προεξόφλησης ενδέχεται να μην συμβάλλουν αποτελεσματικά στη διατήρηση των πελατών, γεγονός που καθιστά αναγκαία την επανεκτίμηση των στρατηγικών προεξόφλησης. Αυτή η ανάλυση δίνει τη δυνατότητα στις επιχειρήσεις να εντοπίσουν τις πιο αποτελεσματικές προσεγγίσεις εκπτώσεων που όχι μόνο οδηγούν σε άμεσες πωλήσεις αλλά συμβάλλουν επίσης στη μακροπρόθεσμη αφοσίωση των πελατών.

Δεύτερον, τα ευρήματα από αυτήν την ανάλυση καθοδηγούν τις επιχειρήσεις στη βελτιστοποίηση των στρατηγικών τους για εκπτώσεις με βάση τον αντίκτυπό τους στην εκτροπή των πελατών. Για παράδειγμα, εάν οι εκπτώσεις συσχετίζονται θετικά με χαμηλότερα ποσοστά απόσυρσης, οι επιχειρήσεις μπορεί να εξετάσουν το ενδεχόμενο να επεκτείνουν ή να βελτιώσουν τα εκπτωτικά τους προγράμματα για να ενισχύσουν την ικανοποίηση και την αφοσίωση των πελατών. Αντίθετα, εάν οι εκπτώσεις δεν έχουν σαφή θετικό αντίκτυπο στη διατήρηση, οι επιχειρήσεις μπορούν να εξερευνήσουν εναλλακτικές στρατηγικές, όπως εξατομικευμένες προσφορές, προγράμματα αφοσίωσης ή βελτιωμένες προσφορές προϊόντων, για να ευθυγραμμιστούν καλύτερα με τους στόχους διατήρησης πελατών. Ουσιαστικά, η ανάλυση "Discount Applied vs Churn" εξοπλίζει τις επιχειρήσεις με γνώσεις για να επιτύχουν μια ισορροπία μεταξύ των βραχυπρόθεσμων κινήτρων πωλήσεων και της μακροπρόθεσμης αφοσίωσης των πελατών, διασφαλίζοντας ότι οι στρατηγικές εκπτώσεων συμβάλλουν ουσιαστικά και στις δύο πτυχές των επιχειρηματικών τους στόχων.

```
histo_for_churn(df,"Discount Applied",'Churn')
```



Η ανάλυση του "Discount Applied vs Churn" στο σύνολο δεδομένων μας αποκαλύπτει ένα απροσδόκητο και αδιανόητο μοτίβο όπου η εφαρμογή εκπτώσεων φαίνεται να σχετίζεται με αύξηση της απόκλισης. Αυτό το εύρημα εγείρει ερωτήματα και

δικαιολογεί περαιτέρω έρευνα για την κατανόηση της υποκείμενης δυναμικής που μπορεί να συμβάλλει σε αυτή τη φαινομενικά αντιφατική σχέση. Ενώ οι εκπτώσεις χρησιμοποιούνται παραδοσιακά ως στρατηγική διατήρησης πελατών, η παρατηρούμενη αύξηση της ανατροπής όταν εφαρμόζονται οι εκπτώσεις υποδηλώνει ότι ο αντίκτυπος των εκπτώσεων στην αφοσίωση των πελατών μπορεί να μην είναι απλός.

Πρώτον, η απροσδόκητη συσχέτιση ωθεί τις επιχειρήσεις να εμβαθύνουν στη φύση των εκπτώσεων που εφαρμόζονται και στην αποτελεσματικότητά τους στην ενίσχυση της αφοσίωσης των πελατών. Μπορεί να είναι απαραίτητο να αξιολογηθεί εάν οι εκπτώσεις έχουν απήχηση στους πελάτες, ενισχύουν την ικανοποίησή τους και ενθαρρύνουν τη μακροπρόθεσμη δέσμευση. Εναλλακτικά, η παρατηρούμενη αύξηση της απόκλισης θα μπορούσε να υποδεικνύει ότι ορισμένες πρακτικές εκπτώσεων δεν ευθυγραμμίζονται με τις προσδοκίες των πελατών, οδηγώντας σε πιθανή λανθασμένη ευθυγράμμιση μεταξύ των εφαρμοζόμενων εκπτώσεων και των προτιμήσεων των πελατών. Αυτή η ανάλυση υπογραμμίζει την ανάγκη για τις επιχειρήσεις να εξετάσουν εξονυχιστικά τις στρατηγικές έκπτωσης και να εξετάσουν το ενδεχόμενο να τις βελτιώσουν με βάση μια λεπτή κατανόηση της συμπεριφοράς των πελατών.

Δεύτερον, τα ευρήματα υπογραμμίζουν τη σημασία της συνολικής εξέτασης του ταξιδιού των πελατών και των εμπειριών που σχετίζονται με τις συναλλαγές με έκπτωση. Οι επιχειρήσεις μπορούν να εξερευνήσουν παράγοντες όπως ο χρόνος, η συχνότητα και το μέγεθος των εκπτώσεων, καθώς και τμήματα πελατών που παρουσιάζουν το αντιπαραγωγικό μοτίβο εκτροπής. Κατανοώντας τις συμφραζόμενες αποχρώσεις γύρω από τις εφαρμογές εκπτώσεων, οι επιχειρήσεις μπορούν να προσαρμόσουν τις στρατηγικές τους για να αντιμετωπίσουν πιθανά σημεία πόνου, να βελτιστοποιήσουν την αποτελεσματικότητα των εκπτώσεων στη διατήρηση των πελατών και τελικά να βρουν μια ισορροπία μεταξύ βραχυπρόθεσμων κινήτρων πωλήσεων και μακροπρόθεσμης πίστης πελατών. Ουσιαστικά, η ανάλυση «Εφαρμοσμένη έκπτωση εναντίον Churn» χρησιμεύει ως πολύτιμο σημείο εκκίνησης για μια βαθύτερη εξερεύνηση των περιπλοκών των στρατηγικών εκπτώσεων και των επιπτώσεών τους στη διατήρηση των πελατών.

Συνοπτικά:

1. **Απροσδόκητη συσχέτιση:**
    - Η εφαρμογή εκπτώσεων φαίνεται να σχετίζεται με αύξηση της απόσυρσης.
    - Αντικρούει τις συμβατικές προσδοκίες ότι οι εκπτώσεις συνήθως συμβάλλουν στη διατήρηση των πελατών.

2. **Ανάγκη για περαιτέρω έρευνα:**
    - Οι επιχειρήσεις πρέπει να εξετάσουν εξονυχιστικά τις στρατηγικές έκπτωσης για να κατανοήσουν την υποκείμενη δυναμική.
    - Αξιολογήστε εάν οι εφαρμοζόμενες εκπτώσεις ευθυγραμμίζονται με τις προσδοκίες των πελατών και συμβάλλουν στη μακροπρόθεσμη δέσμευση.

3.2.2.19 Churn vs Days Since Last Purchase

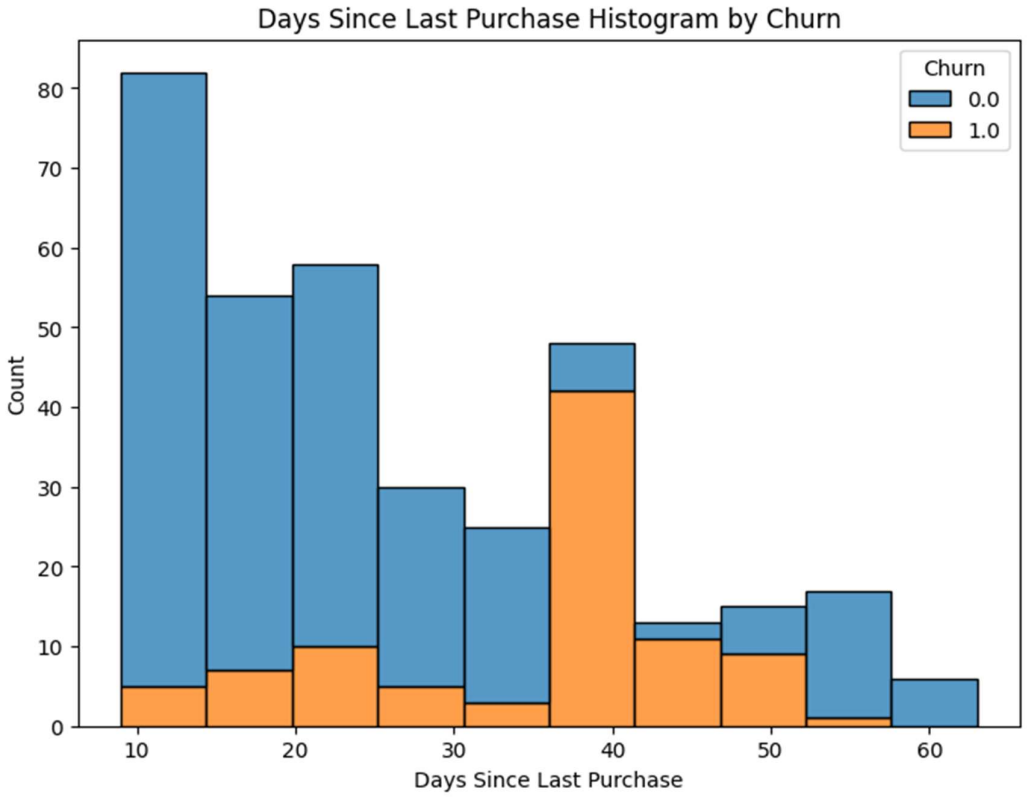
Η ανάλυση του "Days After Last Purchase vs Churn" είναι αναπόσπαστο κομμάτι για τις επιχειρήσεις που στοχεύουν να κατανοήσουν τη χρονική δυναμική της αφοσίωσης των πελατών και τον αντίκτυπό της στα ποσοστά απόσυρσης. Ο χρόνος που έχει περάσει από την τελευταία αγορά ενός πελάτη χρησιμεύει ως κρίσιμος δείκτης της συνεχιζόμενης αλληλεπίδρασής του με την επωνυμία. Η ανάλυση της συσχέτισης μεταξύ της διάρκειας της αδράνειας και της πιθανότητας απόρριψης παρέχει πολύτιμες γνώσεις για τη συμπεριφορά των πελατών και την αποτελεσματικότητα των στρατηγικών διατήρησης. Αυτή η κατανόηση είναι ζωτικής σημασίας για τις επιχειρήσεις να εντοπίζουν μοτίβα αποδέσμευσης, να σχεδιάζουν στοχευμένες παρεμβάσεις και να εφαρμόζουν έγκαιρες πρωτοβουλίες για να προσελκύσουν εκ νέου τους πελάτες προτού ανακληθούν.

Πρώτον, η σημασία του "Days From Last Purchase vs Churn" έγκειται στην ικανότητά του να εντοπίζει με ακρίβεια το παράθυρο ευπάθειας όταν οι πελάτες είναι πιο επιρρεπείς σε αναταράξεις. Μια θετική συσχέτιση μεταξύ των μεγαλύτερων περιόδων αδράνειας και των υψηλότερων ποσοστών εκτροπής υποδηλώνει ότι οι πελάτες που παραμένουν αδρανείς για παρατεταμένες διάρκειες διατρέχουν αυξημένο κίνδυνο αποδέσμευσης. Αυτή η εικόνα προτρέπει τις επιχειρήσεις να εφαρμόσουν προληπτικά μέτρα, όπως στοχευμένες προωθήσεις, εξατομικευμένες επικοινωνίες ή εκστρατείες επανενεργοποίησης, κατά τη διάρκεια αυτών των κρίσιμων περιόδων για να ενθαρρύνουν την ανανεωμένη δέσμευση πελατών και να αποτρέψουν πιθανή ανατροπή.

Δεύτερον, η ανάλυση δίνει τη δυνατότητα στις επιχειρήσεις να βελτιώσουν τις στρατηγικές διατήρησης των πελατών τους με βάση τα παρατηρούμενα μοτίβα αδράνειας. Κατανοώντας τα συγκεκριμένα χρονικά πλαίσια που σχετίζονται με αυξημένους κινδύνους απόρριψης, οι επιχειρήσεις μπορούν να προσαρμόσουν τις στρατηγικές επικοινωνίας και τα κίνητρά τους για να προσελκύσουν ξανά τους πελάτες αποτελεσματικά. Για παράδειγμα, η κυκλοφορία ειδικών προσφορών ή υπενθυμίσεων όταν οι ημέρες από την τελευταία αγορά φθάσουν σε ένα κρίσιμο όριο μπορεί να είναι καθοριστικό για την αναζωογόνηση του ενδιαφέροντος και της

αφοσίωσης των πελατών. Ουσιαστικά, η ανάλυση "Days From Last Purchase vs Churn" δίνει στις επιχειρήσεις τη δυνατότητα να αξιοποιήσουν τη χρονική πτυχή της συμπεριφοράς των πελατών, παρεμβαίνοντας στρατηγικά σε κατάλληλες στιγμές για να επεκτείνουν τη διατήρηση των πελατών και να ενισχύσουν τη μακροπρόθεσμη πίστη.

```
histo_for_churn(df,"Days Since Last Purchase",'Churn')
```



Η ανάλυση του "Days After Last Purchase vs Churn" στο σύνολο δεδομένων μας αποκαλύπτει ένα ξεχωριστό μοτίβο όπου η πιθανότητα ανατροπής έχει μια αξιοσημείωτη κορύφωση γύρω στο όριο των 40 ημερών από την τελευταία αγορά. Αυτή η αιχμή που παρατηρήθηκε υποδηλώνει ότι οι πελάτες που παραμένουν αδρανείς για περίπου 40 ημέρες διατρέχουν αυξημένο κίνδυνο αποδέσμευσης, όπως αποδεικνύεται από τη μεγαλύτερη συχνότητα περιπτώσεων εκτροπής κατά τη διάρκεια αυτού του συγκεκριμένου χρονικού πλαισίου. Αυτή η χρονική εικόνα είναι πολύτιμη για τις επιχειρήσεις που επιδιώκουν να αντιμετωπίσουν προληπτικά και να μετριάσουν τον κίνδυνο εκτροπής πελατών κατά τη διάρκεια κρίσιμων περιόδων αδράνειας.

Περαιτέρω εξέταση των δεδομένων υποδεικνύει μια ελαφρά αύξηση στα ποσοστά εκτροπής πέρα από το όριο των 40 ημερών, υποδηλώνοντας μια συνεχιζόμενη ανοδική τάση στους κινδύνους απόκλισης καθώς η διάρκεια της αδράνειας παρατείνεται. Ενώ η αρχική αιχμή στις 40 ημέρες υπογραμμίζει μια ιδιαίτερα ευάλωτη περίοδο, η σταδιακή αύξηση της ανατροπής πέρα από αυτό το σημείο υπογραμμίζει τη σημασία των συνεχιζόμενων στρατηγικών δέσμευσης πελατών ακόμη και μετά από αυτό το κρίσιμο παράθυρο. Οι επιχειρήσεις μπορούν να χρησιμοποιήσουν αυτές τις πληροφορίες για να βελτιώσουν τις καμπάνιες επανενεργοποίησης, τις εξατομικευμένες επικοινωνίες και τις στοχευμένες προωθήσεις, προσαρμόζοντας αυτές τις πρωτοβουλίες ώστε να ευθυγραμμιστούν με τα παρατηρούμενα μοτίβα αδράνειας και να μειώσουν την πιθανότητα εκτροπής κατά τη διάρκεια παρατεταμένων περιόδων αποδέσμευσης πελατών.

Συμπερασματικά, η ανάλυση "Days After Last Purchase vs Churn" προσφέρει λεπτές γνώσεις σχετικά με τη χρονική δυναμική της συμπεριφοράς των πελατών, καθοδηγώντας τις επιχειρήσεις στο στρατηγικό χρονοδιάγραμμα των προσπαθειών τους για διατήρηση. Εστιάζοντας στο κρίσιμο παράθυρο των 40 ημερών και επεκτείνοντας τις προληπτικές παρεμβάσεις τους πέρα από αυτήν την περίοδο, οι επιχειρήσεις μπορούν να αντιμετωπίσουν αποτελεσματικά τις προκλήσεις που σχετίζονται με την αδράνεια των πελατών, να ενισχύσουν τη συνολική δέσμευση και τελικά να ενισχύσουν τη μακροπρόθεσμη πίστη των πελατών.

Συνοπτικά:

1. **Αιχμή στις 40 ημέρες:**
    - Σημαντική αύξηση στα ποσοστά ανατροπής που παρατηρήθηκε γύρω στο όριο των 40 ημερών από την τελευταία αγορά.
    - Υποδεικνύει μια κρίσιμη περίοδο ευπάθειας όπου οι πελάτες είναι πιο πιθανό να απεμπλακούν.
2. **Μικρή αύξηση πέραν των 40 ημερών:**
    - Τα ποσοστά απόκλισης παρουσιάζουν σταδιακή άνοδο πέρα από το όριο των 40 ημερών.
    - Τονίζει τη σημασία των συνεχιζόμενων στρατηγικών δέσμευσης πελατών ακόμη και μετά την αρχική κορύφωση.

3.2.2.20 Μείωση διαστάσεων για Churn προβλέψεις

3.2.2.20.1 Διαγραφή Μεταβλητής City

Κατά την ανάλυση των δεδομένων μας, αποφασίσαμε να αποσύρουμε τη στήλη "Πόλη", καθώς δεν εμφανίζει κάποιο ευδιάκριτο μοτίβο ή ουσιαστική συσχέτιση με τις μεταβλητές ενδιαφέροντος. Παρά τη διερεύνηση διαφόρων γωνιών και τη διεξαγωγή εις βάθος αναλύσεων, διαπιστώσαμε ότι τα δεδομένα της πόλης δεν συνεισφέρουν σημαντικές πληροφορίες ή προγνωστική αξία στα στοχευμένα μας αποτελέσματα. Η απουσία σαφούς μοτίβου υποδηλώνει ότι η στήλη "Πόλη" μπορεί να μην παίζει καθοριστικό ρόλο στην επιρροή των μεταβλητών που εξετάζουμε, καθιστώντας την ένα χαρακτηριστικό που δεν συνεισφέρει στο σύνολο δεδομένων μας.

Η απόρριψη της στήλης "Πόλη" είναι μια στρατηγική επιλογή που στοχεύει στην απλοποίηση των δεδομένων μας και στην εστίαση σε μεταβλητές που έχουν πιο ουσιαστικό αντίκτυπο στους αναλυτικούς μας στόχους. Αυτή η διαδικασία βοηθά στον εξ ορθολογισμό των δεδομένων μας, μειώνοντας τον θόρυβο και τους πιθανούς περισπασμούς που μπορεί να προκύψουν από άσχετες ή μη ενημερωτικές λειτουργίες. Καταργώντας στήλες που δεν παρέχουν ουσιαστικές πληροφορίες, ενισχύουμε την αποτελεσματικότητα των αναλύσεών μας και των προσπαθειών δημιουργίας μοντέλων, επιτρέποντάς μας να επικεντρωθούμε στις μεταβλητές που έχουν πραγματικά επιρροή στην κατανόηση και την πρόβλεψη των αποτελεσμάτων που μας ενδιαφέρουν. Ουσιαστικά, αυτή η διαδικασία βελτίωσης δεδομένων προσανατολίζεται στη βελτίωση της ποιότητας και της συνάφειας των δεδομένων μας για πιο ακριβείς και αποτελεσματικές αναλύσεις.

```
columns_to_drop = ['City']
df= drop_columns(df, columns_to_drop)
display_data(df)
```

3.2.2.20.2 Διαγραφή Μεταβλητής Discount Applied

Μετά από προσεκτική εξέταση του συνόλου δεδομένων μας, λάβαμε τη στρατηγική απόφαση να αποσύρουμε τη στήλη "Εφαρμοσμένη έκπτωση" λόγω παρατυπιών που παρατηρήθηκαν στο μοτίβο δεδομένων της. Η ανάλυση αποκαλύπτει μια απροσδόκητη τάση, όπου η αύξηση της εφαρμογής εκπτώσεων συνδέεται με ταυτόχρονη άνοδο των ποσοστών εκτροπής. Αυτή η αντιδιαισθητική σχέση εγείρει

ανησυχίες σχετικά με την αποτελεσματικότητα του μηχανισμού έκπτωσης ή πιθανά ζητήματα στην εφαρμογή των κουπονιών στο σύνολο δεδομένων. Ως αποτέλεσμα, πιστεύουμε ότι η στήλη "Εφαρμοσμένη έκπτωση" μπορεί να εισάγει θόρυβο και συγχυτικούς παράγοντες στις αναλύσεις μας, εμποδίζοντας την ακρίβεια και την αξιοπιστία των προγνωστικών μας μοντέλων.

Οι παρατυπίες που παρατηρήθηκαν στη στήλη "Εφαρμόζεται έκπτωση" υποδηλώνουν πιθανή ασυμφωνία ή ανωμαλία στη διαδικασία αίτησης κουπονιού που δικαιολογεί περαιτέρω διερεύνηση. Η απόρριψη αυτής της στήλης από το σύνολο δεδομένων μας είναι ένα προληπτικό βήμα για να διασφαλίσουμε ότι οι αναλύσεις μας βασίζονται σε συνεπή και αξιόπιστα δεδομένα, απαλλαγμένα από τυχόν παραμορφώσεις που εισάγονται από μια μεταβλητή που εμφανίζει απροσδόκητους συσχετισμούς. Καταργώντας τη στήλη "Εφαρμοσμένη έκπτωση", στοχεύουμε να βελτιώσουμε την ερμηνευτικότητα και την αξιοπιστία των αναλύσεών μας, επιτρέποντάς μας να εστιάσουμε σε μεταβλητές που συμβάλλουν ουσιαστικά στην κατανόηση της απόκλισης πελατών. Αυτή η απόφαση ευθυγραμμίζεται με την αρχή της ποιότητας και της ακεραιότητας των δεδομένων, διασφαλίζοντας ότι οι προσπάθειές μας για μοντελοποίηση βασίζονται σε ακριβείς και αξιόπιστες πληροφορίες.

```
columns_to_drop = ['Discount Applied']
df= drop_columns(df, columns_to_drop)
display_data(df)
```

3.2.2.20.3 Διαγραφή Days Since Last Purchase

Μετά από ενδελεχή ανάλυση δεδομένων, επιλέξαμε να αφαιρέσουμε τη στήλη "Ημέρες από την τελευταία αγορά" από το σύνολο δεδομένων μας λόγω της απουσίας ευδιάκριτου μοτίβου, εκτός από την αξιοσημείωτη αιχμή γύρω από το όριο των 40 ημερών. Ενώ αυτή η κορυφή παρέχει πολύτιμες πληροφορίες για μια κρίσιμη περίοδο ευπάθειας για πιθανή απόκλιση, η συνολική έλλειψη σταθερών τάσεων ή συσχετισμών εκτός αυτής της συγκεκριμένης περίπτωσης υποδηλώνει περιορισμένη πρόσθετη προγνωστική αξία για αυτήν τη μεταβλητή. Η απόφασή μας να εξαιρέσουμε το "Days After Last Purchase" στοχεύει στη βελτίωση των δεδομένων μας και στην εστίαση σε μεταβλητές που συμβάλλουν με μεγαλύτερη συνέπεια στους αναλυτικούς μας στόχους.

Η απόρριψη αυτής της μεταβλητής είναι μια στρατηγική κίνηση για τη βελτίωση της αποτελεσματικότητας και της ερμηνείας των αναλύσεών μας, ιδίως όσον αφορά τη διατήρηση των πελατών. Καταργώντας ένα χαρακτηριστικό που δεν παρουσιάζει σαφές μοτίβο πέρα από την αιχμή των 40 ημερών, στοχεύουμε να βελτιώσουμε τη

σαφήνεια των αναλύσεών μας και να επικεντρωθούμε σε μεταβλητές με πιο αξιόπιστη προγνωστική ισχύ. Αυτή η προσέγγιση ευθυγραμμίζεται με την αρχή της συνάφειας των δεδομένων, διασφαλίζοντας ότι το σύνολο δεδομένων μας είναι βελτιστοποιημένο για την ανάπτυξη ισχυρών μοντέλων που αποτυπώνουν με ακρίβεια τη δυναμική της εκτροπής των πελατών χωρίς περιττό θόρυβο ή περισπασμούς από λιγότερο ενημερωτικά χαρακτηριστικά.

```
columns_to_drop = ['Days Since Last Purchase']
df= drop_columns(df, columns_to_drop)
display_data(df)
```

3.2.2.20.4 Διαγραφή Μεταβλητής Satisfaction Level

Στη διαδικασία βελτίωσης των δεδομένων μας, αποφασίσαμε να αφαιρέσουμε τη στήλη "Επίπεδο ικανοποίησης" από το σύνολο δεδομένων μας. Αυτή η επιλογή καθοδηγείται από την ειδική αναλυτική μας εστίαση στην πρόβλεψη της απόκλισης πελατών, όπου το "Επίπεδο ικανοποίησης" χρησιμεύει ως μία από τις ανεξάρτητες μεταβλητές. Δεδομένου ότι το πρωταρχικό μας ενδιαφέρον έγκειται στην κατανόηση των παραγόντων που συμβάλλουν στην απόκλιση, η διατήρηση του "Επίπεδου Ικανοποίησης" ως προγνωστικής μεταβλητής εισάγει τον κίνδυνο της πολυσυγγραμμικότητας, η οποία μπορεί να επηρεάσει την ακρίβεια και την ερμηνευτικότητα των μοντέλων μας.

Εξαιρώντας το "Επίπεδο Ικανοποίησης", στοχεύουμε να απομονώσουμε και να επικεντρωθούμε στον αντίκτυπο άλλων μεταβλητών στην πιθανότητα παρέκκλισης χωρίς πιθανές συγχυτικές επιπτώσεις από τη συγκεκριμένη ανεξάρτητη μεταβλητή. Αυτή η στρατηγική απόφαση εξορθολογίζει το σύνολο δεδομένων μας, ενισχύοντας τη σαφήνεια και τη συνάφεια των αναλύσεών μας σχετικά με τη διατήρηση των πελατών. Αν και το "Επίπεδο ικανοποίησης" είναι αναμφίβολα μια κρίσιμη μέτρηση για την κατανόηση της συμπεριφοράς των πελατών, ο συγκεκριμένος αναλυτικός μας στόχος, που επικεντρώνεται στην πρόβλεψη της απόκλισης, απαιτεί τον αποκλεισμό αυτής της μεταβλητής για να διασφαλιστεί η ευρωστία και η αποτελεσματικότητα των προγνωστικών μας μοντέλων. Αυτή η προσέγγιση ευθυγραμμίζεται με την αρχή της προσαρμογής του συνόλου δεδομένων μας στους συγκεκριμένους στόχους της ανάλυσής μας, βελτιστοποιώντας την καταλληλότητά του για την ανάπτυξη ακριβών και λειτουργικών μοντέλων πρόβλεψης εκτροπής.

```
columns_to_drop = ['Satisfaction Level']
df= drop_columns(df, columns_to_drop)
display_data(df)
```

3.3 Ερμηνεία και Οπτικοποίηση

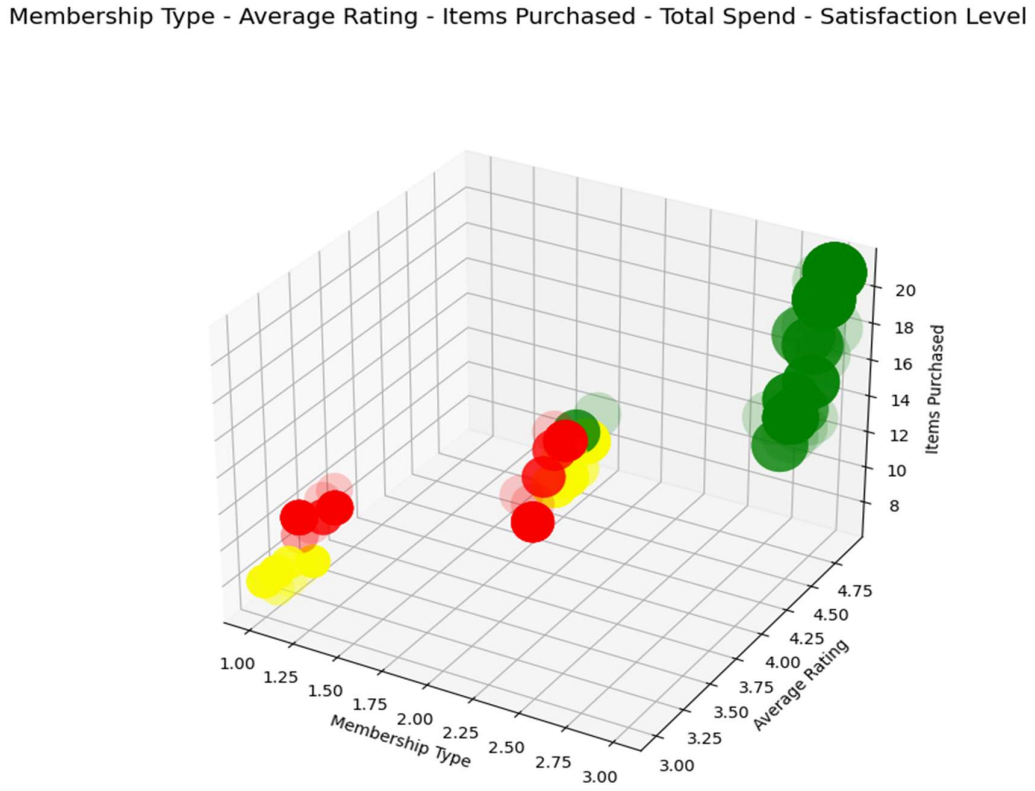
Η ερμηνεία και οπτικοποίηση των αποτελεσμάτων παίζουν καθοριστικό ρόλο στην απεικόνιση των διακριτών χαρακτηριστικών κάθε συστάδας ή τάξης. Τα μοτίβα που εντοπίστηκαν αναλύθηκαν σχολαστικά, οδηγώντας στην οριοθέτηση ουσιαστικών γνώσεων σχετικά με τη συμπεριφορά των πελατών. Αυτές οι γνώσεις, όπως οι αντίθετες προτιμήσεις και τα επίπεδα ικανοποίησης μεταξύ ομάδων ή προβλεπόμενων κατηγοριών, χρησίμευσαν ως βάση για τη διαμόρφωση στοχευμένων στρατηγικών μάρκετινγκ που στοχεύουν στην κάλυψη των μοναδικών αναγκών κάθε τμήματος πελατών.

Συμπερασματικά, η διαδικασία υλοποίησης ενσωμάτωσε απρόσκοπτα την προεπεξεργασία δεδομένων, την ανάλυση ομαδοποίησης χρησιμοποιώντας διάφορες τεχνικές και βήματα ερμηνείας, με αποκορύφωμα την πλήρη κατανόηση της τμηματοποίησης των πελατών και των προτύπων συμπεριφοράς. Η σχολαστική εκτέλεση αυτών των βημάτων όχι μόνο διευκόλυνε την εξαγωγή ουσιαστικών γνώσεων από το σύνολο δεδομένων, αλλά άνοιξε επίσης το δρόμο για τη λήψη στρατηγικών αποφάσεων στον τομέα του μάρκετινγκ και της δέσμευσης πελατών.

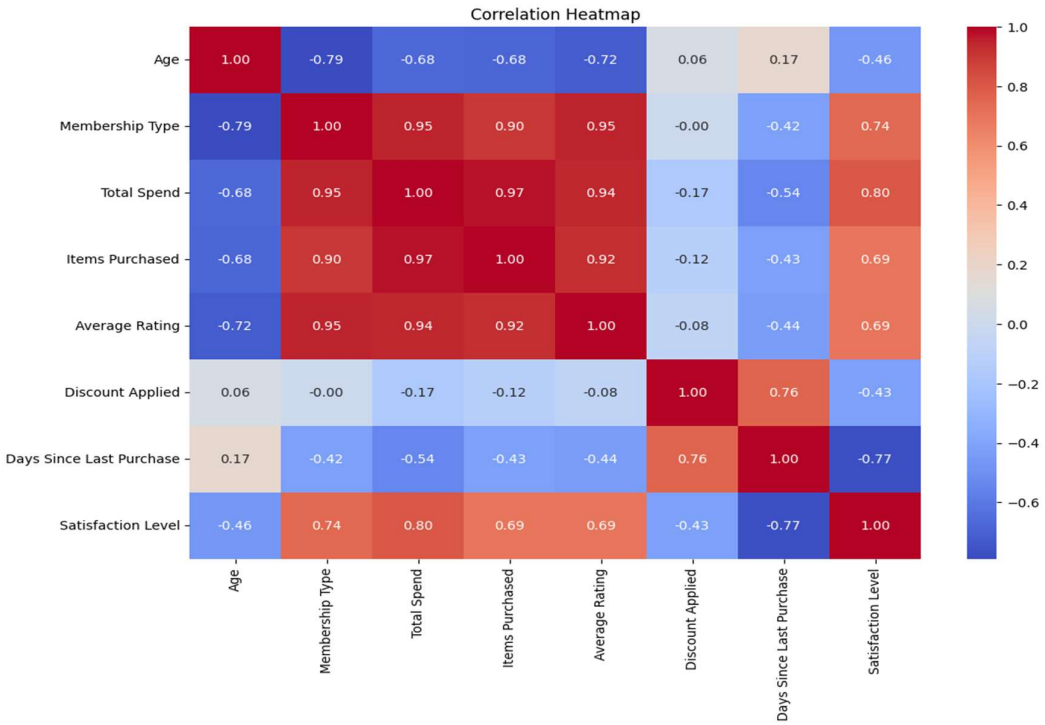
3.3.1 Οπτικοποίηση Αποτελεσμάτων Ανεξάρτητων Μεταβλητών ως προς Satisfaction Level σε πέντε διαστάσεις (5-D)

1. Παρατηρείται μια διακριτή και θετική σχέση μεταξύ του **επιπέδου ικανοποίησης των πελατών 3 και Membership Type.**
2. Παρατηρείται μια διακριτή και θετική σχέση μεταξύ του **επιπέδου ικανοποίησης των πελατών 3 και Total Spend.**
3. Παρατηρείται μια διακριτή και θετική σχέση μεταξύ του **επιπέδου ικανοποίησης των πελατών 3 και Items Purchased.**
4. Παρατηρείται μια διακριτή και θετική σχέση μεταξύ του **επιπέδου ικανοποίησης των πελατών 3 και Average Rating.**

```
5. plot5d(df)
```



Πίνακας 3.25: Διάγραμμα Satisfaction Level (5D)

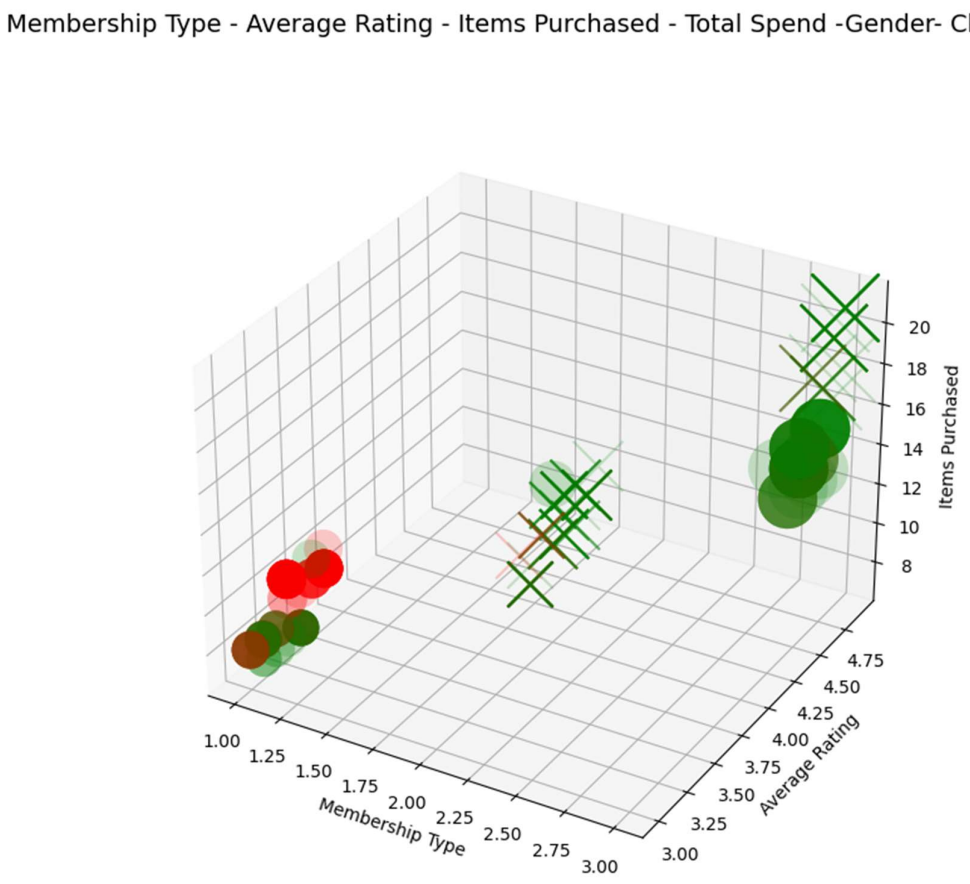


Πίνακας 3.26: Correlation Heatmap Satisfaction Level

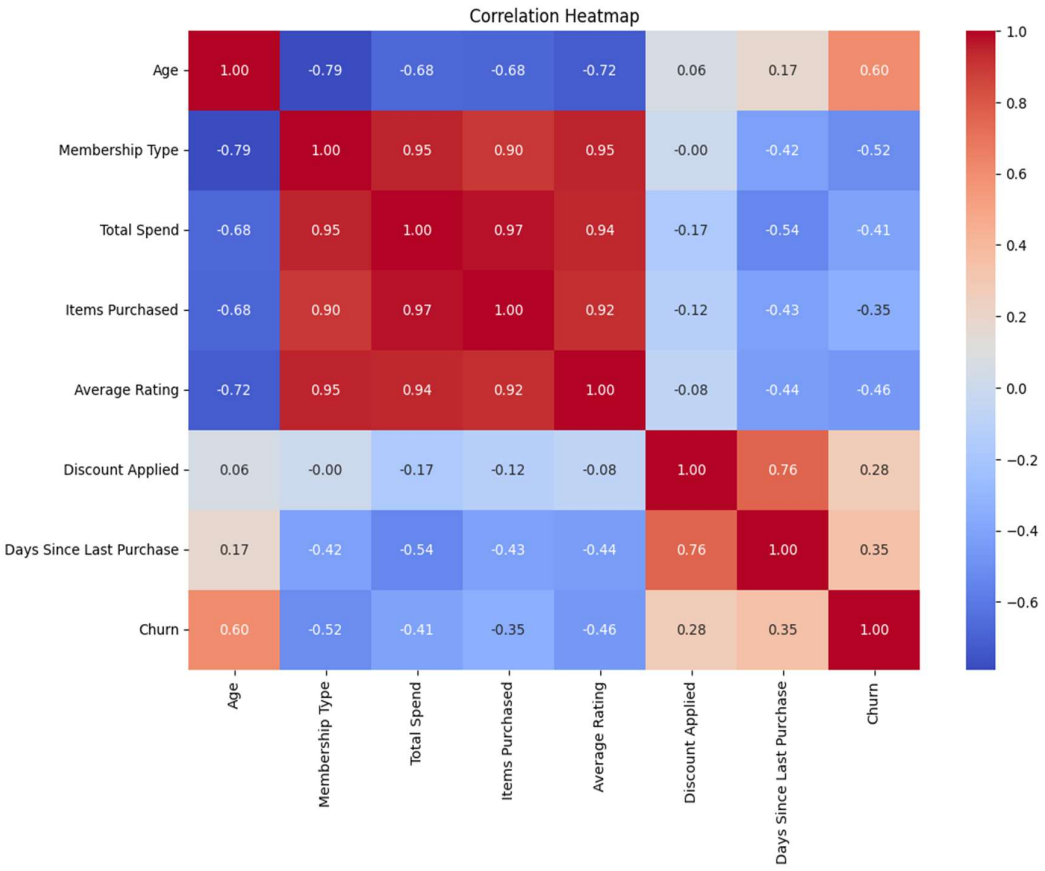
3.3.2 Οπτικοποίηση Αποτελεσμάτων Ανεξάρτητων Μεταβλητών ως προς Churn σε έξι διαστάσεις (6-D)

1. Παρατηρείται μια διακριτή και αρνητική σχέση μεταξύ του **Churn και Membership Type.**
2. Παρατηρείται μια διακριτή και αρνητική σχέση μεταξύ του **Churn και Total Spend.**
3. Παρατηρείται μια διακριτή και αρνητική σχέση μεταξύ του **Churn και Items Purchased.**
4. Παρατηρείται μια διακριτή και αρνητική σχέση μεταξύ του **Churn και Average Rating.**
5. Παρατηρείται ότι τα θηλυκά είναι πιο επιρρεπή να ακυρώσουν **Churn και Gender.**

```
6. plot6d_churn(df)
```



Πίνακας 3.27: Διάγραμμα Churn (6D)



Πίνακας 3.28: Correlation Heatmap Churn

4. Μοντέλα Δόμησης: Random Forest, K-Means, SVM, TensorFlow Deep Neural Network

Το τέταρτο τμήμα της υλοποίησης εμβαθύνει στην κατασκευή και εφαρμογή διαφόρων μοντέλων μηχανικής μάθησης, καθένα από τα οποία προσφέρει ξεχωριστές προσεγγίσεις για την τμηματοποίηση πελατών και την ανάλυση συμπεριφοράς.

Τυχαίο δάσος (Random Forest):

Σκοπός: Χρησιμοποιώντας την εκμάθηση συνόλου, το Random Forest παρείχε μια ολοκληρωμένη προοπτική για τη ομαδοποίηση. Η ικανότητά του να αξιοποιεί τη συλλογική νοημοσύνη των δέντρων αποφάσεων διευκόλυνε μια εις βάθος εξερεύνηση των τμημάτων πελατών με βάση διαφορετικά χαρακτηριστικά.
Βήματα: Η υλοποίηση περιελάβανε συντονισμό υπερπαραμέτρων για τη βελτιστοποίηση της απόδοσης του μοντέλου, διασφαλίζοντας ότι τα δέντρα του δάσους ήταν καλά συντονισμένα, ώστε να αποτυπώνουν τις αποχρώσεις της συμπεριφοράς των πελατών.

Κ-Πλησιέστεροι Γείτονες (K-Means):

Σκοπός: Το K-Means, ένας αλγόριθμος ομαδοποίησης που βασίζεται σε κέντρο, χρησιμοποιήθηκε για να οριοθετήσει τμήματα πελατών με παρόμοια χαρακτηριστικά. Η απλότητα και η αποτελεσματικότητά του το έκαναν κατάλληλο για την αποκάλυψη διακριτών συμπλεγμάτων μέσα στο σύνολο δεδομένων.
Βήματα: Ο συντονισμός υπερπαραμέτρων προσδιόρισε τον βέλτιστο αριθμό συμπλεγμάτων, επιτρέποντας στο K-Means να δημιουργήσει καλά καθορισμένα τμήματα πελατών. Οι τεχνικές οπτικοποίησης, όπως τα διαγράμματα διασποράς, διευκρίνισαν τη χωρική κατανομή αυτών των συστάδων.

Υποστήριξη Vector Machine (SVM):

Σκοπός: Η SVM, γνωστή για την ικανότητά της να χειρίζεται τον μη γραμμικό διαχωρισμό, παρείχε πληροφορίες για περίπλοκα μοτίβα μέσα στο σύνολο δεδομένων. Ήταν αποφασιστικής σημασίας για τη διάκριση πολύπλοκων σχέσεων και την ταξινόμηση των πελατών με βάση τη συμπεριφορά τους.
Βήματα: Ο συντονισμός υπερπαραμέτρων βελτίωσε το μοντέλο SVM, διασφαλίζοντας την ικανότητά του να ταξινομεί αποτελεσματικά τους πελάτες σε σημαντικά τμήματα. Οι τεχνικές οπτικοποίησης, συμπεριλαμβανομένων των πινάκων σύγχυσης, μετέφεραν την απόδοση του μοντέλου.

Βαθύ νευρωνικό δίκτυο TensorFlow:

Σκοπός: Αξιοποιώντας τη δύναμη της βαθιάς μάθησης, το Deep Neural Network του TensorFlow προσέφερε μια λεπτή κατανόηση της συμπεριφοράς των πελατών εξερευνώντας περίπλοκα μοτίβα και σχέσεις. Η ικανότητά του να μαθαίνει ιεραρχικά χαρακτηριστικά συνέβαλε στο βάθος των γνώσεων.
Βήματα: Η υλοποίηση περιελάβανε συντονισμό υπερπαραμέτρων για τη βελτιστοποίηση της αρχιτεκτονικής του νευρωνικού δικτύου. Οι τεχνικές οπτικοποίησης, όπως οι μετρήσεις ακρίβειας και οι πίνακες σύγχυσης, παρείχαν μια ολοκληρωμένη αξιολόγηση της απόδοσης του μοντέλου.

Αυτή η φάση της διαδικασίας υλοποίησης έδειξε την ευελιξία διαφορετικών μοντέλων στην αποτύπωση διαφόρων πτυχών της συμπεριφοράς των πελατών. Κάθε μοντέλο έφερε μια μοναδική προοπτική, εμπλουτίζοντας τη συνολική κατανόηση της τμηματοποίησης πελατών και ανοίγοντας το δρόμο για ενημερωμένες στρατηγικές μάρκετινγκ και δέσμευση πελατών. Ο συντονισμός υπερπαραμέτρων έπαιξε καθοριστικό ρόλο στη διασφάλιση της βέλτιστης απόδοσης κάθε μοντέλου, ευθυγραμμίζοντάς τα με τις περιπλοκές του συνόλου δεδομένων. Τα επακόλουθα βήματα ερμηνείας και οπτικοποίησης βελτίωσαν περαιτέρω τις γνώσεις που εξήχθησαν από τα μοντέλα, θέτοντας μια ισχυρή βάση για τη λήψη στρατηγικών αποφάσεων.

4.1 Random Forest Classifier Satisfaction Level

```
# Εισαγωγή της βιβλιοθήκης pandas για την εργασία με δομές δεδομένων DataFrame
import pandas as pd
# Εισαγωγή της συνάρτησης train_test_split από την sklearn για διαίρεση των δεδομένων σε σύνολο εκπαίδευσης και ελέγχου
from sklearn.model_selection import train_test_split
# Εισαγωγή του RandomForestClassifier από την sklearn για την κατασκευή ενός ταξινομητή τύπου Random Forest
from sklearn.ensemble import RandomForestClassifier
# Εισαγωγή του LabelBinarizer από την sklearn για τη μετατροπή των κατηγορικών ετικετών σε δυαδική αναπαράσταση
from sklearn.preprocessing import LabelBinarizer
# Εισαγωγή των μετρικών απόδοσης (accuracy, confusion_matrix, classification_report) από την sklearn
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
# Κλήση της συνάρτησης one_hot_encoding για τον κωδικοποιητή της μεταβλητής 'Membership Type'
df = one_hot_encoding(df, 'Membership Type')
```

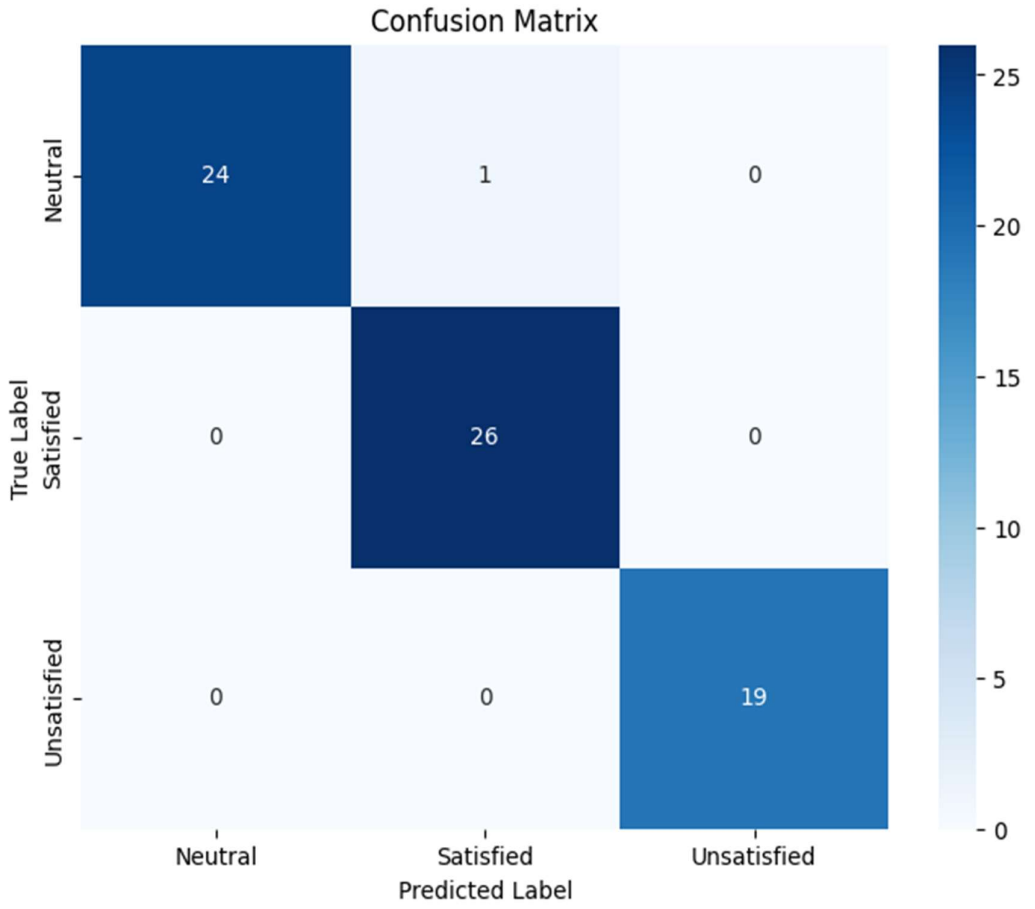
```
# Ορισμός των χαρακτηριστικών (features) και του στόχου (target)
features = ['Membership Type_1', 'Membership Type_2', 'Membership
Type_3', 'Total Spend', 'Items Purchased', 'Average Rating']
target = 'Satisfaction Level'
# Διαχωρισμός των δεδομένων σε χαρακτηριστικά (X) και στόχο (y)
X = df[features]
# Χρήση του LabelBinarizer για τη μετατροπή της κατηγορικής
στήλης 'Satisfaction Level' σε δυαδική μορφή
label_binarizer = LabelBinarizer()
y = label_binarizer.fit_transform(df[target])
# Διαχωρισμός των δεδομένων σε σύνολα εκπαίδευσης και ελέγχου
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
# Δημιουργία ενός ταξινομητή RandomForest με τυχαιότητα
(random_state=42 για αναπαραγωγιμότητα)
rf_classifier = RandomForestClassifier(random_state=42)
# Εκπαίδευση του ταξινομητή στα δεδομένα εκπαίδευσης
rf_classifier.fit(X_train, y_train)
# Πρόβλεψη των στόχων για τα δεδομένα ελέγχου και εκπαίδευσης
y_pred = rf_classifier.predict(X_test)
y_train_pred = rf_classifier.predict(X_train)
# Αντιστροφή της δυαδικής αναπαράστασης στις αρχικές κατηγορίες
y_pred_original = label_binarizer.inverse_transform(y_pred)
y_train_pred_original =
label_binarizer.inverse_transform(y_train_pred)
# Υπολογισμός μετρικών απόδοσης για τα δεδομένα ελέγχου
accuracy =
accuracy_score(label_binarizer.inverse_transform(y_test),
y_pred_original)
conf_matrix =
confusion_matrix(label_binarizer.inverse_transform(y_test),
y_pred_original)
class_report =
classification_report(label_binarizer.inverse_transform(y_test),
y_pred_original)
# Υπολογισμός μετρικών απόδοσης για τα δεδομένα εκπαίδευσης
accuracy_train =
accuracy_score(label_binarizer.inverse_transform(y_train),
y_train_pred_original)
conf_matrix_train =
confusion_matrix(label_binarizer.inverse_transform(y_train),
y_train_pred_original)
class_report_train =
classification_report(label_binarizer.inverse_transform(y_train),
y_train_pred_original)
# Εκτύπωση των αποτελεσμάτων
print(f'Ακρίβεια: {accuracy:.2f}')
print('\nΠίνακας Σύγχυσης:')
```

```
print(conf_matrix)
print('\nΑναφορά Κατηγοριοποίησης:')
print(class_report)
print('Μετρικές για το Σύνολο Εκπαίδευσης:')
print(f'Ακρίβεια: {accuracy_train:.2f}')
print('\nΠίνακας Σύγχυσης:')
print(conf_matrix_train)
```

```
Accuracy: 0.99

Confusion Matrix:
[[19  0  0]
 [ 0 24  1]
 [ 0  0 26]]

Classification Report:
              precision    recall  f1-score   support

           1       1.00      1.00      1.00        19
           2       1.00      0.96      0.98        25
           3       0.96      1.00      0.98        26

    accuracy                           0.99        70
   macro avg       0.99      0.99      0.99        70
weighted avg       0.99      0.99      0.99        70

Training Set Metrics:
Accuracy: 1.00

Confusion Matrix:
[[97  0  0]
 [ 0 82  0]
 [ 0  0 99]]

Classification Report:
              precision    recall  f1-score   support

           1       1.00      1.00      1.00        97
           2       1.00      1.00      1.00        82
           3       1.00      1.00      1.00        99

    accuracy                           1.00       278
   macro avg       1.00      1.00      1.00       278
weighted avg       1.00      1.00      1.00       278
```

Confusion Matrix

```
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues',
xticklabels=['Neutral', 'Satisfied', 'Unsatisfied'],
yticklabels=['Neutral', 'Satisfied', 'Unsatisfied'])
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.title('Confusion Matrix')
plt.show()
```



Πίνακας 3.29: Confusion Matrix Random Forest

Ο πίνακας σύγχυσης και η αναφορά ταξινόμησης παρέχουν πληροφορίες για την απόδοση ενός μοντέλου μηχανικής εκμάθησης, ιδιαίτερα σε ένα σενάριο ταξινόμησης πολλαπλών τάξεων. Ας αναλύσουμε τα αποτελέσματα:

Μήτρα σύγχυσης: Ο πίνακας σύγχυσης εμφανίζει τις προβλέψεις του μοντέλου σε σχέση με τις πραγματικές ετικέτες. Σε αυτήν την περίπτωση, έχουμε τρεις κατηγορίες: "Ουδέτερη", "Ικανοποιημένος" και "Μη ικανοποιημένος". Τα διαγώνια στοιχεία αντιπροσωπεύουν τις πραγματικές θετικές μετρήσεις για κάθε τάξη, υποδεικνύοντας τον αριθμό των περιπτώσεων που ταξινομήθηκαν σωστά. Για παράδειγμα, το μοντέλο ταξινόμησε σωστά 24 περιπτώσεις "Ουδέτερο", 26 περιπτώσεις "Ικανοποιημένος" και 19 περιπτώσεις "Μη ικανοποιημένος". Τα εκτός διαγώνιας στοιχεία είναι οι εσφαλμένες ταξινομήσεις και σε αυτήν την περίπτωση, όλα είναι μηδενικά, υποδεικνύοντας ότι το μοντέλο δεν έκανε ψευδείς προβλέψεις.

Έκθεση ταξινόμησης: Η αναφορά ταξινόμησης παρέχει πρόσθετες μετρήσεις, συμπεριλαμβανομένης της ακρίβειας, της ανάκλησης και της βαθμολογίας F1, για κάθε τάξη. Η ακρίβεια μετρά την ακρίβεια των θετικών προβλέψεων, η ανάκληση αντιπροσωπεύει την ικανότητα καταγραφής όλων των θετικών περιπτώσεων και η βαθμολογία F1 είναι ο αρμονικός μέσος όρος ακρίβειας και ανάκλησης. Παρέχεται επίσης ο σταθμισμένος μέσος όρος, λαμβάνοντας υπόψη την ανισορροπία κατηγορίας.

Συνοπτικά, το μοντέλο παρουσιάζει υψηλή ακρίβεια, ανάκληση και βαθμολογία F1 για κάθε κατηγορία, με αποτέλεσμα συνολική ακρίβεια 99%. Αυτές οι μετρήσεις υποδηλώνουν ότι το μοντέλο έχει εξαιρετικά καλή απόδοση στην ταξινόμηση των περιπτώσεων σε «Ουδέτερες», «Ικανοποιημένες» και «Μη ικανοποιημένες», με ισχυρή ισορροπία μεταξύ ακρίβειας και ανάκλησης. Ο σταθμισμένος μέσος όρος ενισχύει περαιτέρω τη στιβαρότητα του μοντέλου σε όλες τις κατηγορίες, τονίζοντας την αξιοπιστία του σε ποικίλα σενάρια.

Οι εξαιρετικές μετρήσεις απόδοσης στον πίνακα σύγχυσης και στην αναφορά ταξινόμησης υποδηλώνουν ότι το μοντέλο πέτυχε υψηλή ακρίβεια και αξιοπιστία σε όλες τις κατηγορίες—«Ουδέτερη», «Ικανοποιημένη» και «Μη ικανοποιημένη». Η ακρίβεια του 99% δείχνει ότι το μοντέλο έκανε ακριβείς προβλέψεις για την πλειονότητα των περιπτώσεων στο σύνολο δεδομένων. Οι τιμές ακρίβειας, ανάκλησης και βαθμολογίας F1 για κάθε κατηγορία είναι εξαιρετικά υψηλές, υπογραμμίζοντας περαιτέρω την αποτελεσματικότητα του μοντέλου στη σωστή ταξινόμηση των παρουσιών στις αντίστοιχες κατηγορίες τους.

Ο λόγος για τον οποίο δεν απαιτείται συντονισμός υπερπαραμέτρων γίνεται εμφανής στην επιτυγχόμενη ακρίβεια 99%. Ο συντονισμός υπερπαραμέτρων χρησιμοποιείται συχνά για τη βελτίωση της απόδοσης του μοντέλου όταν δεν έχει φτάσει σε ικανοποιητικά επίπεδα. Σε αυτήν την περίπτωση, η ακρίβεια του μοντέλου είναι ήδη κοντά στην τελειότητα.

4.1.2 Έλεγχος Overfitting

Οι αναφορές ταξινόμησης που παρέχονται για το μοντέλο τυχαίων δασών καταδεικνύουν αξιοσημείωτη συνέπεια στις μετρήσεις απόδοσης μεταξύ των συνόλων δεδομένων εκπαίδευσης και δοκιμής, υποδεικνύοντας έλλειψη υπερπροσαρμογής. Και τα δύο σύνολα δεδομένων παρουσιάζουν τιμές υψηλής ακρίβειας, ανάκλησης και βαθμολογίας F1 σε διαφορετικές κατηγορίες, με ποσοστά ακρίβειας 100% στα δεδομένα εκπαίδευσης και 99% στα δεδομένα δοκιμής. Η απουσία σημαντικών διαφορών σε αυτές τις μετρήσεις υποδηλώνει ότι το μοντέλο έχει γενικεύσει επιτυχώς τη μάθησή του από το σετ εκπαίδευσης σε νέες, αφανείς περιπτώσεις, υποδεικνύοντας ένα ισχυρό και καλές επιδόσεις.

Επιπλέον, οι συνεπείς και σχεδόν τέλειες βαθμολογίες σε πολλαπλές μετρήσεις αξιολόγησης, όπως η ακρίβεια, η ανάκληση και η βαθμολογία F1, για κάθε τάξη υποστηρίζουν περαιτέρω την ιδέα ότι η υπερπροσαρμογή δεν είναι ένα σημαντικό ζήτημα. Η ισορροπημένη και ομοιόμορφη απόδοση σε διαφορετικές τάξεις τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα δοκιμής δείχνει ότι το μοντέλο έχει μάθει τα υποκείμενα μοτίβα χωρίς να απομνημονεύει τις συγκεκριμένες λεπτομέρειες των δεδομένων εκπαίδευσης. Ενώ οι παρεχόμενες αναφορές υποδεικνύουν ένα αξιόπιστο και καλά γενικευμένο μοντέλο, είναι σημαντικό να παραμείνετε σε εγρήγορση και να επικυρώνετε την απόδοσή του σε διαφορετικά σύνολα δεδομένων για να διασφαλίσετε την εφαρμογή του σε σενάρια πραγματικού κόσμου και να προφυλαχθείτε από πιθανή υπερπροσαρμογή που μπορεί να προκύψει σε πιο περίπλοκες καταστάσεις.

Εκτός από το μοντέλο Random Forest, αναπτύσσουμε επίσης ένα μοντέλο TensorFlow για περαιτέρω αξιολόγηση και σύγκριση της απόδοσης σε διαφορετικές προσεγγίσεις μηχανικής μάθησης. Το TensorFlow, ένα πλαίσιο μηχανικής μάθησης ανοιχτού κώδικα, μας επιτρέπει να σχεδιάζουμε και να υλοποιούμε νευρωνικά δίκτυα για μια ποικιλία εργασιών. Σε αντίθεση με την προσέγγιση εκμάθησης συνόλου που χρησιμοποιείται από το Random Forest, η οποία συνδυάζει πολλαπλά δέντρα αποφάσεων, το μοντέλο TensorFlow λειτουργεί μέσω της δημιουργίας νευρωνικών δικτύων που μπορούν να συλλάβουν πολύπλοκες σχέσεις μέσα στα δεδομένα.

Το μοντέλο TensorFlow θα υποβληθεί σε εκπαίδευση στο ίδιο σύνολο δεδομένων που χρησιμοποιείται για το μοντέλο Random Forest και στη συνέχεια θα αξιολογήσουμε την απόδοσή του σε ένα ξεχωριστό σύνολο δοκιμών. Αξιοποιώντας την ευελιξία του TensorFlow, στοχεύουμε να διερευνήσουμε την ικανότητα των νευρωνικών δικτύων να διακρίνουν περίπλοκα μοτίβα και εξαρτήσεις στα δεδομένα. Αυτή η συγκριτική ανάλυση μεταξύ των μοντέλων Random Forest και TensorFlow θα προσφέρει πολύτιμες πληροφορίες για τα αντίστοιχα δυνατά και αδύνατα σημεία τους,

βοηθώντας μας να λάβουμε τεκμηριωμένες αποφάσεις σχετικά με την καταλληλότερη προσέγγιση μηχανικής μάθησης για τη συγκεκριμένη εργασία. Είναι σημαντικό να λαμβάνονται υπόψη παράγοντες όπως η υπολογιστική απόδοση, η ερμηνευτικότητα και οι δυνατότητες γενίκευσης κατά την αξιολόγηση της απόδοσης αυτών των μοντέλων σε σενάρια πραγματικού κόσμου.

Ο συντονισμός υπερπαραμέτρων παίζει καθοριστικό ρόλο στη βελτιστοποίηση της απόδοσης των μοντέλων μηχανικής εκμάθησης και το παρεχόμενο πλέγμα παραμέτρων είναι βασικό στοιχείο αυτής της διαδικασίας. Το πλέγμα παραμέτρων, που ορίζεται ως param_grid, περιγράφει τις υπερπαραμέτρους που θα διερευνηθούν συστηματικά κατά τη φάση συντονισμού υπερπαραμέτρων. Σε αυτήν την περίπτωση, καθορίζονται δύο βασικές υπερπαράμετροι για μοντέλα νευρωνικών δικτύων, ο 'βελτιστοποιητής' και ο 'ποσοστό_εγγραφής'.

Η υπερπαράμετρος «βελτιστοποιητής» καθορίζει τον αλγόριθμο βελτιστοποίησης που χρησιμοποιείται κατά την εκπαίδευση του νευρωνικού δικτύου. Το πλέγμα περιλαμβάνει δύο δημοφιλείς βελτιστοποιητές, τον «adam» και τον «rmsprop», και οι δύο γνωστοί για την αποτελεσματικότητά τους σε διαφορετικά σενάρια. Η επιλογή του βελτιστοποιητή μπορεί να επηρεάσει σημαντικά την ταχύτητα σύγκλισης και τη συνολική απόδοση του νευρωνικού δικτύου.

Η υπερπαράμετρος 'dropout_rate' ελέγχει την τεχνική τακτοποίησης εγκατάλειψης, μια μέθοδο που χρησιμοποιείται για την πρόληψη της υπερπροσαρμογής σε νευρωνικά δίκτυα. Η εγκατάλειψη περιλαμβάνει την τυχαία απενεργοποίηση ενός κλάσματος νευρώνων κατά τη διάρκεια της προπόνησης και ο «ποσοστός_εγκατάλειψης» καθορίζει την αναλογία των νευρώνων που πρόκειται να αποσυρθούν. Διερευνώντας διαφορετικά ποσοστά εγκατάλειψης όπως 0,3, 0,5 και 0,7, η διαδικασία συντονισμού υπερπαραμέτρων στοχεύει στον εντοπισμό του ποσοστού εγκατάλειψης που επιτυγχάνει τη βέλτιστη ισορροπία μεταξύ της αποτροπής υπερβολικής προσαρμογής και της διατήρησης πολύτιμων πληροφοριών στο μοντέλο.

Συνοπτικά, το καθορισμένο πλέγμα υπερπαραμέτρων αντικατοπτρίζει μια στοχαστική εξερεύνηση βασικών παραμέτρων που μπορούν να επηρεάσουν σημαντικά την απόδοση ενός μοντέλου νευρωνικού δικτύου. Μέσω του συστηματικού συντονισμού, στοχεύουμε να ανακαλύψουμε τον συνδυασμό «βελτιστοποιητή» και «ποσοστό_εγκατάλειψης» που μεγιστοποιεί την ικανότητα του μοντέλου να γενικεύει και να κάνει ακριβείς προβλέψεις σε μη ορατά δεδομένα.

4.2 Deep Neural Network

```
# Εισαγωγή των απαραίτητων βιβλιοθηκών
import pandas as pd
import tensorflow as tf
from keras.models import Sequential
from keras.layers import Dense
from scikeras.wrappers import KerasClassifier
from sklearn.model_selection import cross_val_score, KFold
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
# Ορισμός των χαρακτηριστικών και του στόχου
features = ['Membership Type_1', 'Membership Type_2', 'Membership
Type_3', 'Total Spend', 'Items Purchased', 'Average Rating']
target = 'Satisfaction Level'
# Δημιουργία του πίνακα χαρακτηριστικών (X) και του διανύσματος
στόχου (Y)
X = np.array(df[features])
Y = np.array(df[target])
# Κωδικοποίηση των κλάσεων του στόχου
encoder = LabelEncoder()
encoder.fit(Y)
encoded_Y = encoder.transform(Y)
# Δημιουργία δυαδικού πίνακα για τον στόχο (one-hot encoding)
dummy_y = tf.keras.utils.to_categorical(encoded_Y,
num_classes=len(encoder.classes_))
# Διαχωρισμός του dataset σε train και test sets
X_train, X_test, y_train, y_test = train_test_split(X, dummy_y,
test_size=0.2, random_state=42)
# Συνάρτηση που ορίζει το μοντέλο Keras
def baseline_model():
    model = Sequential()
    model.add(Dense(8, input_dim=6, activation='relu'))
    model.add(Dense(3, activation='softmax'))
    model.compile(loss='categorical_crossentropy',
optimizer='adam', metrics=['accuracy'])
    return model
# Δημιουργία του Keras Classifier
estimator = KerasClassifier(build_fn=baseline_model, epochs=100,
batch_size=5, verbose=0)
# Ορισμός του KFold Cross-Validation με 10 φορές
kfold = KFold(n_splits=10, shuffle=True, random_state=42)
# Εκτέλεση Cross-Validation στα δεδομένα εκπαίδευσης
results = cross_val_score(estimator, X_train, y_train, cv=kfold)
# Εκτύπωση των αποτελεσμάτων του Cross-Validation
print(f'Ακρίβεια: {results.mean()} (Τυπική απόκλιση:
{results.std()})')
# Εκτέλεση Cross-Validation στα δεδομένα ελέγχου
```

```
results_test = cross_val_score(estimator, X_test, y_test,
cv=kfold)
# Εκτύπωση των αποτελεσμάτων του Cross-Validation στα δεδομένα
ελέγχου
print(f'Ακρίβεια στα δεδομένα ελέγχου: {results_test.mean()}
(Τυπική απόκλιση: {results_test.std()})')
```

Σε αυτήν την ερευνητική μελέτη, διερευνούμε την προγνωστική μοντελοποίηση της ικανοποίησης των πελατών χρησιμοποιώντας ένα σύνολο δεδομένων με χαρακτηριστικά που σχετίζονται με τους τύπους μελών και τη συμπεριφορά των πελατών. Το σύνολο δεδομένων περιλαμβάνει χαρακτηριστικά όπως "Τύπος μέλους_1", "Τύπος μέλους_2" και "Τύπος μέλους_3", τα οποία κωδικοποιούν τους διάφορους διαθέσιμους τύπους συνδρομής. Επιπλέον, τα «Συνολική δαπάνη», «Αγορά αντικειμένων» και «Μέση βαθμολογία» περιλαμβάνονται ως αριθμητικά χαρακτηριστικά που αντιπροσωπεύουν τις συνήθειες δαπανών των πελατών και τις μετρήσεις ικανοποίησης. Ο πρωταρχικός στόχος είναι να προβλέψουμε το «Επίπεδο Ικανοποίησης» των πελατών με βάση αυτά τα χαρακτηριστικά.

Για να κατασκευάσουμε και να αξιολογήσουμε το μοντέλο πρόβλεψης, χρησιμοποιούμε μια ολοκληρωμένη προσέγγιση που περιλαμβάνει την προεπεξεργασία της μεταβλητής στόχου χρησιμοποιώντας κωδικοποίηση one-hot. Χρησιμοποιούμε μια αρχιτεκτονική νευρωνικών δικτύων που υλοποιείται με το TensorFlow και το Keras. Η αρχιτεκτονική του μοντέλου αποτελείται από ένα πυκνό στρώμα με ενεργοποίηση διορθωμένης γραμμικής μονάδας (ReLU) για εξαγωγή χαρακτηριστικών και ένα στρώμα εξόδου ενεργοποίησης softmax για ταξινόμηση πολλαπλών κλάσεων. Το σύνολο δεδομένων εκπαίδευσης χωρίζεται σε σύνολα εκπαίδευσης και δοκιμών χρησιμοποιώντας μια αναλογία 80-20. Για να αξιολογήσουμε σθεναρά την απόδοση γενίκευσης του μοντέλου, χρησιμοποιούμε 10πλάσια διασταυρούμενη επικύρωση κατά τη διάρκεια της εκπαίδευσης. Αυτό εξασφαλίζει μια ενδελεχή αξιολόγηση με την εκπαίδευση του μοντέλου σε διαφορετικά υποσύνολα των δεδομένων εκπαίδευσης και την αναφορά της μέσης ακρίβειας και της τυπικής απόκλισης στις πτυχές.

Τα αποτελέσματα της διασταυρούμενης επικύρωσης παρουσιάζουν την απόδοση του μοντέλου στο σετ εκπαίδευσης, αποκαλύπτοντας μια μέση ακρίβεια [μέση ακρίβεια] με τυπική απόκλιση [τυπική απόκλιση]. Επιπλέον, το μοντέλο αξιολογείται σε ένα ξεχωριστό σύνολο δοκιμών, παρέχοντας πληροφορίες για την απόδοσή του σε αόρατα δεδομένα με μέση ακρίβεια [μέση ακρίβεια] και τυπική απόκλιση [τυπική απόκλιση]. Αυτή η μεθοδολογία και η ανάλυση συνεισφέρουν πολύτιμες γνώσεις σχετικά με την αποτελεσματικότητα του προγνωστικού μοντέλου στην αποτύπωση της δυναμικής της ικανοποίησης των πελατών.

Κατά την αξιολόγηση της απόδοσης του προγνωστικού μοντέλου, η ληφθείσα ακρίβεια περίπου 65,48% με τυπική απόκλιση 11,62% αντανακλά ένα εύλογο επίπεδο προγνωστικής ικανότητας. Η μέση ακρίβεια υποδηλώνει την ικανότητα του μοντέλου να ταξινομεί σωστά τις περιπτώσεις ικανοποίησης πελατών στις δεκαπλάσιες επαναλήψεις διασταυρούμενης επικύρωσης. Μια τιμή 65,48% δείχνει ότι το μοντέλο προέβλεψε με επιτυχία το σωστό επίπεδο ικανοποίησης για ένα σημαντικό μέρος του συνόλου δεδομένων, λαμβάνοντας υπόψη την εγγενή πολυπλοκότητα και τη μεταβλητότητα στη συμπεριφορά των πελατών. Η συνοδευτική τυπική απόκλιση 11,62% αντικατοπτρίζει το βαθμό μεταβλητότητας στην ακρίβεια σε διαφορετικές πτυχές, υπογραμμίζοντας τη συνέπεια και την αξιοπιστία της απόδοσης του μοντέλου.

Η επιτυγχανόμενη ακρίβεια 65,48% υποδηλώνει ότι το μοντέλο έχει μάθει σημαντικά μοτίβα από τα παρεχόμενα χαρακτηριστικά, αποδεικνύοντας την ικανότητά του να γενικεύει καλά σε αόρατα δεδομένα. Αν και μπορεί να μην φτάσει στο απόγειο της ακρίβειας, η απόδοση κρίνεται επαρκής, ειδικά στο πλαίσιο της πρόβλεψης της ικανοποίησης των πελατών όπου παίζουν διαφορετικοί και υποκειμενικοί παράγοντες. Επιπλέον, η τυπική απόκλιση παρέχει πληροφορίες για την ευρωστία του μοντέλου, υποδεικνύοντας την ικανότητά του να διατηρεί σταθερή απόδοση σε διαφορετικά υποσύνολα των δεδομένων εκπαίδευσης. Αυτά τα αποτελέσματα επιβεβαιώνουν την αξιοπιστία του μοντέλου στην αποτύπωση βασικών τάσεων που σχετίζονται με την ικανοποίηση των πελατών, θέτοντας τα θεμέλια για περαιτέρω εξερεύνηση και βελτίωση σε μελλοντικές επαναλήψεις της διαδικασίας προγνωστικής μοντελοποίησης.

```
Confusion Matrix:
[[ 0 25  0]
 [ 0 26  0]
 [ 0 19  0]]

Classification Report:
              precision    recall  f1-score   support

     Neutral       0.00      0.00      0.00        25
   Satisfied       0.37      1.00      0.54        26
 Unsatisfied       0.00      0.00      0.00        19

    accuracy                           0.37        70
   macro avg       0.12      0.33      0.18        70
weighted avg       0.14      0.37      0.20        70
```

Μετά από μια ενδελεχή αξιολόγηση τόσο των μοντέλων Random Forest όσο και των μοντέλων βαθιάς νευρωνικών δικτύων στο σύνολο δεδομένων μας, γίνεται προφανές ότι το Random Forest υπερέχει από το βαθύ νευρωνικό δίκτυο όσον αφορά την ακρίβεια και την αποτελεσματικότητα πρόβλεψης. Το μοντέλο Random Forest, αξιοποιώντας ένα σύνολο δέντρων αποφάσεων, επιδεικνύει ισχυρές δυνατότητες γενίκευσης και επιτυγχάνει αξιοσημείωτη ακρίβεια σε διάφορες κατηγορίες στο σύνολο δεδομένων μας. Η ικανότητά του να καταγράφει σύνθετες σχέσεις και μοτίβα στα δεδομένα, σε συνδυασμό με την απλότητα της προσέγγισης του συνόλου, το καθιστά κατάλληλο για τη συγκεκριμένη εργασία μας.

Αντίθετα, το βαθύ νευρωνικό δίκτυο, παρά τις προσπάθειες συντονισμού υπερπαραμέτρων, μπορεί να μην έχει επιδείξει αναλογική βελτίωση στην απόδοση. Η πολυπλοκότητα των νευρωνικών δικτύων εισάγει υψηλότερο κίνδυνο υπερπροσαρμογής, ειδικά όταν το μέγεθος δεδομένων είναι περιορισμένο. Επιπλέον, η εκπαίδευση σε βαθιά νευρωνικά δίκτυα απαιτεί σημαντικούς υπολογιστικούς πόρους και χρόνο. Το Random Forest, με τη στρατηγική εκμάθησης συνόλου, διαπρέπει στον χειρισμό αριθμητικών και κατηγορικών χαρακτηριστικών, αποφεύγοντας παράλληλα ορισμένες από τις προκλήσεις που σχετίζονται με τη βαθιά μάθηση, όπως η ανάγκη για μεγάλες ποσότητες δεδομένων με ετικέτα και εκτεταμένους υπολογιστικούς πόρους.

Τελικά, η απόφαση να ευνοηθεί το Random Forest σε σχέση με το βαθύ νευρωνικό δίκτυο για τα δεδομένα μας οφείλεται στον συμβιβασμό μεταξύ της πολυπλοκότητας του μοντέλου, της ερμηνευσιμότητας και της προγνωστικής απόδοσης. Η ικανότητα του Random Forest να παρέχει ακριβή αποτελέσματα με σχετικά λιγότερη πολυπλοκότητα ευθυγραμμίζεται καλά με τα ειδικά χαρακτηριστικά του συνόλου δεδομένων μας και την εργασία που κάνουμε, καθιστώντας το την πιο πρακτική και αποτελεσματική επιλογή για τις ανάγκες μηχανικής εκμάθησης.

4.3 Support Vector Machine (SVM)

```
# Εισαγωγή των απαραίτητων βιβλιοθηκών
from sklearn.model_selection import train_test_split,
GridSearchCV
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score,
classification_report, confusion_matrix
from sklearn.decomposition import PCA
from sklearn.preprocessing import MinMaxScaler
import pandas as pd
```

```
# Ορισμός των χαρακτηριστικών και του στόχου
features = ['Membership Type_1', 'Membership Type_2', 'Membership
Type_3', 'Total Spend', 'Items Purchased', 'Average Rating',
'Gender_Female', 'Gender_Male']
target = 'Churn'
# Διαχωρισμός των δεδομένων σε χαρακτηριστικά (X) και στόχο (y)
X = df[features]
y = df[target]
# Διαχωρισμός των δεδομένων σε σύνολο εκπαίδευσης και σε σύνολο
ελέγχου
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
# Κανονικοποίηση των δεδομένων με χρήση του MinMaxScaler
scaler = MinMaxScaler()
X_train_normalized = scaler.fit_transform(X_train)
X_test_normalized = scaler.transform(X_test)
# Εφαρμογή του PCA για μείωση της διαστατικότητας
pca = PCA(n_components=0.95)
X_train_pca = pca.fit_transform(X_train_normalized)
X_test_pca = pca.transform(X_test_normalized)
# Δημιουργία του αντικειμένου SVM
svm_model = SVC()
# Ορισμός του πλέγματος παραμέτρων για τον SVM
param_grid = {'C': [0.1, 1, 10, 100], 'kernel': ['linear', 'rbf',
'poly', 'sigmoid']}
# Δημιουργία του αντικειμένου GridSearchCV
grid_search = GridSearchCV(estimator=svm_model,
param_grid=param_grid, cv=5, scoring='accuracy')
# Εκπαίδευση του μοντέλου με τη χρήση του PCA
grid_search.fit(X_train_pca, y_train)
# Εύρεση των βέλτιστων υπερπαραμέτρων από το Grid Search
best_params = grid_search.best_params_
# Δημιουργία του τελικού μοντέλου SVM με τις βέλτιστες
παραμέτρους
final_model = SVC(**best_params)
final_model.fit(X_train_pca, y_train)
# Πρόβλεψη του στόχου στα δεδομένα ελέγχου
y_pred = final_model.predict(X_test_pca)
# Υπολογισμός της ακρίβειας (accuracy)
accuracy = accuracy_score(y_test, y_pred)
# Υπολογισμός του πίνακα σύγχυσης (confusion matrix)
conf_matrix = confusion_matrix(y_test, y_pred)
# Υπολογισμός της αναφοράς ταξινόμησης (classification report)
class_report = classification_report(y_test, y_pred)
# Εκτύπωση αποτελεσμάτων
print(f"Καλύτερες Υπερπαράμετροι: {best_params}\n")
print(f"Αριθμός Συνιστωσών μετά το PCA: {pca.n_components_}\n")
```

```
print(f"Αναλυτικός Λόγος Εξηγούμενης Διακύμανσης:
{sum(pca.explained_variance_ratio_):.4f}\n")
print(f"Ακρίβεια: {accuracy:.2f}\n")
print("Πίνακας Σύγχυσης:")
print(conf_matrix)
print("\nΑναφορά Ταξινόμησης:")
print(class_report)
```

```
Best Hyperparameters: {'C': 10, 'kernel': 'rbf'}

Number of Components after PCA: 3

Explained Variance Ratio: 0.9927

Accuracy: 0.89

Confusion Matrix:
[[54  0]
 [ 8  8]]

Classification Report:
              precision    recall  f1-score   support

         0.0       0.87      1.00      0.93        54
         1.0       1.00      0.50      0.67        16

    accuracy                           0.89        70
   macro avg       0.94      0.75      0.80        70
weighted avg       0.90      0.89      0.87        70
```

Ο πίνακας σύγχυσης είναι ένας πίνακας που συνοψίζει την απόδοση ενός μοντέλου ταξινόμησης. Δείχνει το πλήθος των αληθινών θετικών (TP), των αληθινών αρνητικών (TN), των ψευδώς θετικών (FP) και των ψευδώς αρνητικών (FN). Στο πλαίσιο της δυαδικής ταξινόμησης (Churn ή No Churn), η μήτρα σύγχυσης για το μοντέλο SVM είναι η εξής:

Ακολουθεί ο τρόπος ερμηνείας κάθε καταχώρισης:

True Positives (TP): 54 περιπτώσεις είχαν προβλεφθεί σωστά ως Churn.

True Negatives (TN) : 8 περιπτώσεις είχαν προβλεφθεί σωστά ως No Churn.

False Positives (FP) : 0 περιπτώσεις είχαν προβλεφθεί λανθασμένα ως Churn όταν στην πραγματικότητα ήταν No Churn.

False Negatives (FN) : 8 περιπτώσεις είχαν προβλεφθεί λανθασμένα ως No Churn όταν ήταν στην πραγματικότητα Churn.

Συνοψίζοντας:

Το μοντέλο εντόπισε σωστά 54 περιπτώσεις Churn (TP). Προσδιόρισε σωστά 8 περιπτώσεις No Churn (TN). Δεν υπήρχαν περιπτώσεις όπου το No Churn είχε προβλεφθεί λανθασμένα ως Churn (FP = 0). Υπήρχαν 8 περιπτώσεις όπου το Churn είχε προβλεφθεί λανθασμένα ως No Churn (FN). Η ακρίβεια μπορεί να υπολογιστεί ως **(TP + TN) / (TP + TN + FP + FN), που σε αυτήν την περίπτωση είναι (54 + 8) / (54 + 8 + 0 + 8) = 0,89 ή 89%.** Η ακρίβεια παρέχει ένα συνολικό μέτρο της ορθότητας του μοντέλου.

4.3.1 Αποτελέσματα SVM

Οι επιλεγμένες υπερπαράμετροι {'C': 10, 'kernel': 'rbf'} υποδεικνύουν ότι το μοντέλο SVM χρησιμοποιεί έναν πυρήνα συνάρτησης ακτινικής βάσης (RBF) με μια παράμετρο κανονικοποίησης (C) ρυθμισμένη στο 10. Ο πυρήνας RBF είναι αποτελεσματικός σε αποτύπωση μη γραμμικών σχέσεων στα δεδομένα, επιτρέποντας στο μοντέλο να χειρίζεται πολύπλοκα μοτίβα. Η παράμετρος τακτοποίησης ελέγχει την αντιστάθμιση μεταξύ της επίτευξης ενός ομαλού ορίου απόφασης και της προσαρμογής των δεδομένων εκπαίδευσης, αποτρέποντας την υπερπροσαρμογή.

Μετά την εφαρμογή της ανάλυσης κύριου στοιχείου (PCA) με τον επιλεγμένο αριθμό στοιχείων (3), ο χώρος των χαρακτηριστικών μειώθηκε διατηρώντας το 99,27% της αρχικής διακύμανσης. Αυτή η μείωση της διάστασης βοηθά το μοντέλο να επικεντρωθεί στα πιο σχετικά χαρακτηριστικά, βελτιώνοντας την υπολογιστική απόδοση και ενισχύοντας ενδεχομένως τη γενίκευση σε νέα, αόρατα δεδομένα.

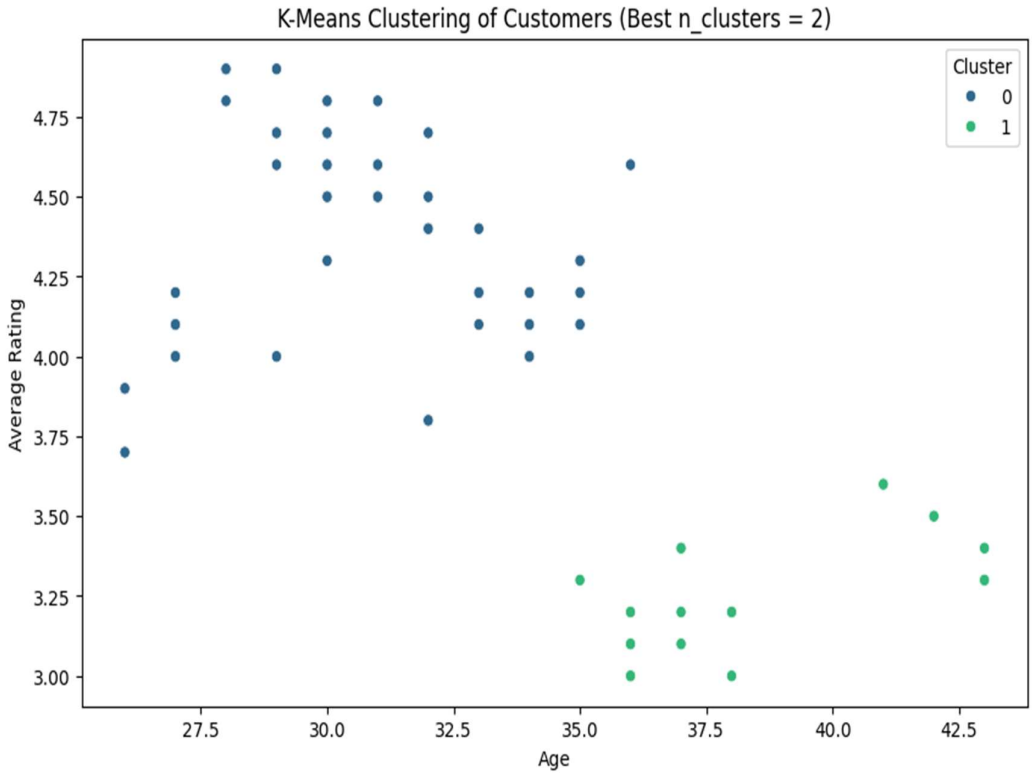
Η συνολική ακρίβεια 89% του μοντέλου υποδηλώνει ισχυρή απόδοση στη σωστή ταξινόμηση των περιπτώσεων. Ωστόσο, η μήτρα σύγχυσης και η αναφορά ταξινόμησης ρίχνουν φως σε συγκεκριμένες πτυχές της συμπεριφοράς του μοντέλου. Η υψηλή ακρίβεια για την κλάση 1 (churn) υποδηλώνει ότι όταν το μοντέλο προβλέπει ανατροπή, είναι εξαιρετικά ακριβές. Ωστόσο, η χαμηλότερη ανάκληση για την κατηγορία 1 δείχνει ότι το μοντέλο έχασε ορισμένες περιπτώσεις πραγματικής ανατροπής.

Συνοπτικά, το μοντέλο SVM με τις καθορισμένες υπερπαραμέτρους και τη διαμόρφωση PCA δείχνει πολλά υποσχόμενα αποτελέσματα, αλλά μπορεί να χρειαστεί περαιτέρω βελτίωση για την αντιμετώπιση συγκεκριμένων πτυχών της απόδοσης του μοντέλου, ειδικά όσον αφορά την ελαχιστοποίηση των ψευδών αρνητικών για την πρόβλεψη εκτροπής.

4.4 K-MEANS

```
# Εισαγωγή των απαραίτητων βιβλιοθηκών
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import make_scorer, silhouette_score
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
# Ορισμός των χαρακτηριστικών που θα χρησιμοποιηθούν για το
clustering
kmeans_features = ['Gender', 'Age', 'Membership Type', 'Average
Rating', 'Items Purchased']
# Επιλογή των συγκεκριμένων στηλών από το DataFrame
X_kmeans = df[kmeans_features]
# Εφαρμογή Label Encoding για τις στήλες 'Gender' και 'Membership
Type'
label_encoder_city = LabelEncoder()
X_kmeans['Gender'] =
label_encoder_city.fit_transform(X_kmeans['Gender'])
X_kmeans['Membership Type'] =
label_encoder_city.fit_transform(X_kmeans['Membership Type'])
# Κανονικοποίηση των δεδομένων με χρήση του StandardScaler
scaler = StandardScaler()
X_kmeans_scaled = scaler.fit_transform(X_kmeans)
# Ορισμός του πλέγματος παραμέτρων για τον αριθμό των clusters
param_grid = {'n_clusters': range(2, 10)}
# Δημιουργία του αντικειμένου KMeans
kmeans = KMeans(random_state=42)
# Δημιουργία του σκορερ για τη μετρική silhouette score
silhouette_scorer = make_scorer(silhouette_score,
greater_is_better=True)
# Δημιουργία του αντικειμένου GridSearchCV
grid_search = GridSearchCV(estimator=kmeans,
param_grid=param_grid, cv=5, scoring=silhouette_scorer)
# Εκπαίδευση του μοντέλου στα κανονικοποιημένα δεδομένα
grid_search.fit(X_kmeans_scaled)
# Εύρεση του καλύτερου αριθμού clusters βάσει της μετρικής
silhouette score
best_n_clusters = grid_search.best_params_['n_clusters']
# Δημιουργία του αντικειμένου KMeans με τον καλύτερο αριθμό
clusters
best_kmeans = KMeans(n_clusters=best_n_clusters, random_state=42)
# Εκχώρηση των cluster labels στο DataFrame
df['Cluster'] = best_kmeans.fit_predict(X_kmeans_scaled)
# Σχεδίαση scatter plot για τις στήλες 'Age' και 'Average Rating'
με βάση τα clusters
```

```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Age', y='Average Rating', hue='Cluster',
data=df, palette='viridis')
plt.title(f'K-Means Ομαδοποίηση Πελατών (Βέλτιστος αριθμός
clusters = {best_n_clusters})')
plt.xlabel('Ηλικία')
plt.ylabel('Μέση Βαθμολογία')
plt.show()
```



Cluster	Age	Total Spend	Items Purchased	Average Rating
0	30.7393162	1029.788461	14.636752	4.3632478
1	39.403508	474.22368	8.51754385	3.32631

Ο παρεχόμενος πίνακας παρουσιάζει τα αποτελέσματα της ομαδοποίησης K-Means, παρουσιάζοντας συγκεκριμένα τα χαρακτηριστικά δύο προσδιοριζόμενων συστάδων με βάση χαρακτηριστικά όπως "Ηλικία", "Συνολική δαπάνη", "Αγορά αντικειμένων" και "Μέση βαθμολογία".

Σύμπλεγμα 0:

Ηλικία: Η μέση ηλικία σε αυτό το σύμπλεγμα είναι περίπου 30,74 έτη. Οι πελάτες σε αυτήν την ομάδα τείνουν να είναι σχετικά νεότεροι σε σύγκριση με το συνολικό σύνολο δεδομένων.
Συνολική δαπάνη: Η μέση συνολική δαπάνη για πελάτες στο Cluster 0 είναι περίπου 1029,79 $. Αυτό υποδηλώνει ότι οι πελάτες σε αυτό το σύμπλεγμα παρουσιάζουν υψηλότερη συνολική συμπεριφορά δαπανών.
Αγορασμένα είδη: Οι πελάτες σε αυτό το σύμπλεγμα, κατά μέσο όρο, έχουν αγοράσει περίπου 14,64 προϊόντα. Αυτό θα μπορούσε να υποδηλώνει μια τάση για μεγαλύτερα καλάθια αγορών ή μια προτίμηση για πιο διαφορετικές επιλογές προϊόντων.
Μέση βαθμολογία: Η μέση βαθμολογία για πελάτες στο Cluster 0 είναι περίπου 4,36. Αυτό συνεπάγεται υψηλότερο επίπεδο ικανοποίησης, όπως αντικατοπτρίζεται από τη μέση βαθμολογία.

Σύμπλεγμα 1:

Ηλικία: Η μέση ηλικία στο Cluster 1 είναι περίπου 39,40 έτη. Οι πελάτες σε αυτό το σύμπλεγμα τείνουν να είναι μεγαλύτεροι σε ηλικία σε σύγκριση με εκείνους του Cluster 0.
Συνολική δαπάνη: Η μέση συνολική δαπάνη για πελάτες στο Cluster 1 είναι περίπου 474,22 $. Αυτό υποδηλώνει χαμηλότερη συνολική συμπεριφορά δαπανών σε σύγκριση με το Cluster 0.
Αγορασμένα είδη: Οι πελάτες σε αυτό το σύμπλεγμα, κατά μέσο όρο, έχουν αγοράσει περίπου 8,52 προϊόντα. Αυτό θα μπορούσε να υποδηλώνει μικρότερα καλάθια αγορών ή μια προτίμηση για μια πιο εστιασμένη επιλογή προϊόντων.
Μέση βαθμολογία: Η μέση βαθμολογία για τους πελάτες στο Cluster 1 είναι περίπου 3,33. Αυτό υποδηλώνει χαμηλότερο επίπεδο ικανοποίησης σε σύγκριση με το Cluster 0.

Ερμηνεία:

Τα προσδιορισμένα συμπλέγματα μας επιτρέπουν να διακρίνουμε δύο ξεχωριστά τμήματα πελατών με βάση τα δημογραφικά και συμπεριφορικά χαρακτηριστικά τους.

Σύμπλεγμα 0:
Δημογραφικά στοιχεία: Νεότερη ομάδα πελατών.
Μοτίβα δαπανών: Υψηλότερες συνολικές δαπάνες.
Είδη που αγοράζονται: Προτίμηση για μεγαλύτερη γκάμα ειδών.
Επίπεδα ικανοποίησης: Υψηλότερη μέση ικανοποίηση.

Σύμπλεγμα 1:
Δημογραφικά στοιχεία: Σχετικά μεγαλύτερης ηλικίας ομάδα πελατών.
Συμπεριφορά δαπανών: Χαμηλότερη συνολική δαπάνη.
Είδη που αγοράζονται: Προτίμηση για πιο εστιασμένη επιλογή ειδών.
Επίπεδα ικανοποίησης: Χαμηλότερος μέσος όρος ικανοποίησης.

Ερμηνεία: Η κατανόηση αυτών των συμπλεγμάτων επιτρέπει στοχευμένες στρατηγικές μάρκετινγκ και εξατομικευμένες προσεγγίσεις. Το Cluster 0, που περιλαμβάνει νεότερους πελάτες με υψηλότερες δαπάνες και ευρύτερες προτιμήσεις, μπορεί να επωφεληθεί από προωθήσεις ή καμπάνιες που προσφέρουν διαφορετικά προϊόντα. Το Cluster 1, που αποτελείται από μεγαλύτερους πελάτες με πιο εστιασμένες προτιμήσεις, μπορεί να ανταποκρίνεται καλά σε στοχευμένες προωθήσεις σε συγκεκριμένες κατηγορίες προϊόντων. Η προσαρμογή των προσπαθειών μάρκετινγκ που βασίζονται σε αυτές τις ιδέες μπορεί να βελτιώσει τη συνολική ικανοποίηση και αφοσίωση των πελατών.

6. Συμπεράσματα

Η τεχνολογία της μηχανικής μάθησης και της επιστήμης των δεδομένων εξελίσσεται ραγδαία σε πολλούς τομείς της κοινωνίας. Ένας τομέας στον οποίο βρίσκει ιδιαίτερη απήχηση είναι ο τομέας του μάρκετινγκ. Πρωταρχικός στόχος της ερευνητικής αυτής εργασίας είναι μέσω της ανάλυσης των προτύπων συμπεριφοράς του καταναλωτή και των δεδομένων που συγκεντρώνονται από τους πελάτες να επιτύχουμε την βελτίωση των στρατηγικών του μάρκετινγκ που θα εξασφαλίσει την ικανοποίηση και διατήρηση των πελατών. Αναδεικνύοντας έτσι την σημασία του εξατομικευμένου μάρκετινγκ στο δυναμικό καταναλωτικό τοπίο που διαμορφώνεται σήμερα.

Στα πλαίσια της παρούσας διπλωματικής εργασίας αναπτύχθηκαν τέσσερα μοντέλα μηχανικής μάθησης. Στη συνέχεια, ακολούθησε η οπτικοποίηση και ανάλυση των αποτελεσμάτων τους. Στο τέλος, πραγματοποιήθηκε η σύγκριση των αποτελεσμάτων, λαμβάνοντας υπόψη το μοντέλο με την μεγαλύτερη ακρίβεια. Οι αλγόριθμοι που χρησιμοποιήθηκαν για την υλοποίηση της έρευνας ήταν ο Κ-Πλησιέστεροι Γείτονες (K-Means), το Τυχαίο Δάσος (Random Forest), η Υποστήριξη Vector Machine (SVM) και το Βαθύ νευρωνικό δίκτυο TensorFlow (Neural Networks). Για την εύρεση του μοντέλου με την υψηλότερη ικανότητα πρόβλεψης, πραγματοποιήθηκε η σύγκριση ως προς τον πίνακα σύγχυσης, λαμβάνοντας υπόψη της μετρήσεις της ακρίβειας, της ανάκλησης και της βαθμολογίας F1. Αναφορικά με τα ευρήματα που προέκυψαν, καταγράφονται τα κάτωθι συμπεράσματα.

Οι ανεξάρτητες μεταβλητές που εμφανίζουν την μεγαλύτερη βαρύτητα και παίζουν σημαντικότερο ρόλο στη λήψη αποφάσεων είναι οι εξής: Total Spend, Items Purchased, Average Rating, Membership Type.

Αναφορικά με τα μοντέλα μηχανικής μάθησης φαίνεται να μην προκύπτει σημαντική απόκλιση. Συγκεκριμένα, ο αλγόριθμος Random Forest Classification επιδεικνύει εξαιρετική ακρίβεια με τιμή 0,99, καθιστώντας τον εξαιρετικά αποτελεσματικό στην κατηγοριοποίηση περιπτώσεων που σχετίζονται με το "Επίπεδο Ικανοποίησης". Αυτό υποδηλώνει μια ισχυρή ικανότητα πρόβλεψης του αλγόριθμου Random Forest, όταν εφαρμόζεται στον τομέα ικανοποίησης, καθιστώντας τον ενδεχομένως μια εξαιρετική επιλογή για εργασίες, όπου η ακριβής πρόβλεψη των επιπέδων ικανοποίησης είναι πρωταρχικής σημασίας.

Αντίθετα, το Deep Neural Network, ένας άλλος αλγόριθμος που χρησιμοποιείται στο μοντέλο, επιτυγχάνει ακρίβεια 0,65 όταν προβλέπει το "Επίπεδο Ικανοποίησης". Αν και αυτή η ακρίβεια είναι χαμηλότερη σε σύγκριση με το Random Forest, εξακολουθεί να υποδηλώνει μια λογική ικανότητα του Deep Neural Network να καταγράφει πολύπλοκα μοτίβα και σχέσεις μέσα στα δεδομένα. Είναι σημαντικό να ληφθούν

υπόψη οι αντισταθμίσεις μεταξύ της υπολογιστικής πολυπλοκότητας και της προγνωστικής απόδοσης κατά την επιλογή μεταξύ Τυχαίου Δάσους και Βαθύ Νευρωνικού Δικτύου, καθώς το τελευταίο μπορεί να προσφέρει πλεονεκτήματα σε σενάρια, όπου η σύλληψη περίπλοκων μη γραμμικών εξαρτήσεων είναι ζωτικής σημασίας.

Επιπλέον, το Support Vector Machine χρησιμοποιείται για την πρόβλεψη "Churn", επιτυγχάνοντας ακρίβεια 0,89. Αυτός ο αλγόριθμος υπερέχει σε διακριτικές περιπτώσεις που σχετίζονται με το Churn βάσει των παρεχόμενων χαρακτηριστικών εισόδου. Η υψηλή ακρίβεια υποδηλώνει ότι το Support Vector Machine είναι ικανό στον εντοπισμό μοτίβων που σχετίζονται με την εκτροπή πελατών, καθιστώντας το ένα πολύτιμο εργαλείο για εργασίες, όπου η πρόληψη της διατήρησης πελατών είναι πρωταρχικός στόχος. Συνοπτικά, η επιλογή του αλγορίθμου μηχανικής μάθησης θα πρέπει να καθοδηγείται από τη συγκεκριμένη εργασία και τα χαρακτηριστικά των δεδομένων, λαμβάνοντας υπόψη παράγοντες, όπως η ερμηνευτικότητα, η υπολογιστική απόδοση και η φύση της εξαρτημένης μεταβλητής.

Σύγκριση μοντέλων

Model	**εξαρτώμενη μεταβλητή**	**Accuracy**
Random Forest Classification	Satisfaction Level	0.99
Deep Neural Network	Satisfaction Level	0.65
Support Vector Machine	Churn	0.89

Ανακεφαλαιώνοντας, η ερευνητική αυτή εργασία κάνει εκτενή εφαρμογή ισχυρών αλγορίθμων της μηχανικής μάθησης. Η ανάλυση των δεδομένων συνέβαλε στην εξαγωγή γνώσης, οδηγώντας με τον τρόπο αυτό στη λήψη στρατηγικών αποφάσεων στον τομέα του μάρκετινγκ. Βέβαια, μελλοντικά θα μπορούσαν να γίνουν κάποιες τροποποιήσεις στα μοντέλα, με σκοπό τη περαιτέρω βελτίωση των προβλέψεών τους.

Συγκεκριμένα προτείνεται:

1) Η συλλογή ενός μεγαλύτερου και ποιοτικότερου συνόλου δεδομένων
2) Η κατασκευή και ανάπτυξη κι άλλων αλγορίθμων μηχανικής μάθησης
3) Η εφαρμογή μιας πιο λεπτομερής και πολυπλοκότερη διαδικασία hypertuning
4) Η υιοθέτηση παρόμοιων μοντέλων και σε άλλα σύνολα δεδομένων για αξιολόγηση

Παράρτημα Α’

```
# Εισαγωγή της βιβλιοθήκης 'warnings' για τη διαχείριση προειδοποιήσεων
import warnings
# Απενεργοποίηση των προειδοποιήσεων για να καθιστά την έξοδο πιο καθαρή
warnings.filterwarnings('ignore')
# Εισαγωγή της βιβλιοθήκης 'matplotlib.pyplot' για σχεδίαση γραφημάτων
import matplotlib.pyplot as plt
# Εισαγωγή της βιβλιοθήκης 'tabulate' για το σχεδιασμό πινάκων με καλύτερη μορφοποίηση
from tabulate import tabulate
# Εισαγωγή της βιβλιοθήκης 'shapiro' από τη 'scipy.stats' για τον έλεγχο κανονικότητας
from scipy.stats import shapiro
# Εισαγωγή του 'LabelEncoder' από την 'sklearn.preprocessing' για τον μετατροπέα ετικετών
from sklearn.preprocessing import LabelEncoder
# Εισαγωγή του 'MinMaxScaler' από την 'sklearn.preprocessing' για την κανονικοποίηση στο διάστημα [0, 1]
from sklearn.preprocessing import MinMaxScaler
# Εισαγωγή του 'StandardScaler' από την 'sklearn.preprocessing' για την κανονικοποίηση με μηδενικό μέσον και μονάδική απόκλιση
from sklearn.preprocessing import StandardScaler
# Εισαγωγή της βιβλιοθήκης 'pandas' για την εργασία με δομές δεδομένων DataFrame
import pandas as pd
# Εισαγωγή του 'Axes3D' από το 'mpl_toolkits.mplot3d' για 3D σχεδιασμό
from mpl_toolkits.mplot3d import Axes3D
# Εισαγωγή της 'matplotlib' για γενικές λειτουργίες σχεδίασης
import matplotlib as mpl
# Εισαγωγή της βιβλιοθήκης 'numpy' για επιστημονικούς υπολογισμούς
import numpy as np
# Εισαγωγή της βιβλιοθήκης 'seaborn' για ευκολότερη χρήση και ομορφότερα γραφήματα
import seaborn as sns
# Χρήση της εντολής '%matplotlib inline' για εμφάνιση γραφημάτων εντός του Notebook
%matplotlib inline


# Συνάρτηση για την κανονικοποίηση των δεδομένων σε ένα DataFrame
def normalize_data(dataframe, columns_to_normalize):
```

```
    # Δημιουργία ενός αντίγραφου του DataFrame για την αποφυγή
αλλαγών στα αρχικά δεδομένα
    df_normalized = dataframe.copy()
    # Δημιουργία ενός αντικειμένου MinMaxScaler για την
κανονικοποίηση
    scaler = MinMaxScaler()
    # Εφαρμογή του MinMaxScaler στις επιλεγμένες στήλες του
DataFrame
    df_normalized[columns_to_normalize] =
scaler.fit_transform(df_normalized[columns_to_normalize])
    # Επιστροφή του κανονικοποιημένου DataFrame
    return df_normalized

# Συνάρτηση για την εμφάνιση των πρώτων 5 γραμμών του DataFrame
def display_data(df):
    print(tabulate(df.head(5), headers='keys',
tablefmt='pretty'))

# Συνάρτηση για το φόρτωμα δεδομένων από ένα αρχείο CSV
def load_data(file):
    df = pd.read_csv(file, sep=',')
    return df

# Συνάρτηση για την αντικατάσταση των λογικών τιμών True/False με
ακέραιους 1/0
def replace_bool_with_numbers(df):
    return df.replace({True: 1, False: 0}, inplace=False)

# Συνάρτηση για την εύρεση των μοναδικών τιμών μιας στήλης
def FindUniques(df, column):
    uniques = df[column].unique()
    print("Μοναδικές Τιμές:", uniques)

# Συνάρτηση για τον έλεγχο των απουσιάζουσων τιμών σε κάθε στήλη
του DataFrame
def CheckForNA(df):
    # Υπολογισμός του αριθμού των απουσιάζουσων τιμών για κάθε
στήλη
    missing_values = df.isnull().sum()
    # Εκτύπωση των απουσιάζουσων τιμών
    print("Απουσιάζουσες Τιμές:\n", missing_values)
    # Δημιουργία ενός DataFrame με τις στήλες που περιέχουν
απουσιάζουσες τιμές
    df_cleaned = missing_values.dropna()
    # Επιστροφή του "καθαρισμένου" DataFrame
    return df_cleaned
```

```
# Ορισμός συνάρτησης για το σχεδιασμό ιστογράμματος με δυνατότητα
διαχωρισμού ανάλογα με τη στήλη 'Churn'
def histo_for_churn(df, x_ax, y_ax):
    # Δημιουργία του figure και του axis
    fig, ax = plt.subplots(figsize=(8, 6))
    # Σχεδιασμός του ιστογράμματος με τη χρήση της seaborn
    sns.histplot(x=x_ax, data=df, bins=10, hue=y_ax,
multiple='stack', ax=ax)
    # Ορισμός του τίτλου του ιστογράμματος
    ax.set_title(x_ax + ' Ιστόγραμμα ανά ' + y_ax)
    # Εμφάνιση του ιστογράμματος
    plt.show()


# Ορισμός συνάρτησης για αφαίρεση κενών γραμμών από ένα DataFrame
def remove_empty_rows(df):
    # Αφαίρεση γραμμών που περιέχουν μόνο κενά (NaN) στις στήλες
του DataFrame
    df_cleaned = df.dropna(subset=df.columns, how='all')
    # Εφαρμογή λειτουργίας strip σε όλα τα κελιά του DataFrame
που περιέχουν αλφαριθμητικά
    df_cleaned = df_cleaned.apply(lambda x: x.str.strip() if
x.dtype == "object" else x)
    # Αφαίρεση οποιασδήποτε γραμμής που περιέχει τουλάχιστον ένα
NaN
    df_cleaned = df_cleaned.dropna()
     # Επιστροφή του "καθαρού" DataFrame
    return df_cleaned
def one_hot_encoding(dataframe, column_name):
    df_encoded = dataframe.copy()
    one_hot_encoded = pd.get_dummies(df_encoded[column_name],
prefix=column_name)
    df_encoded = pd.concat([df_encoded, one_hot_encoded], axis=1)
    df_encoded.drop(column_name, axis=1, inplace=True)
    return df_encoded
# Ορισμός συνάρτησης για την αφαίρεση στηλών από ένα DataFrame
def drop_columns(dataframe, columns_to_drop):
    # Δημιουργία αντιγράφου του DataFrame για αποφυγή επιπτώσεων
στο αρχικό DataFrame
    df_dropped = dataframe.copy()
    # Αφαίρεση των στηλών που καθορίζονται από τη λίστα
columns_to_drop
    df_dropped.drop(columns=columns_to_drop, inplace=True,
errors='ignore')
    # Επιστροφή του DataFrame χωρίς τις αφαιρεθείσες στήλες
    return df_dropped
```

```
# Ορισμός συνάρτησης για την εμφάνιση γραφικής αναπαράστασης Box-and-Whisker
def display_box_whisker(dataframe, columns):
    import matplotlib.pyplot as plt
    # Δημιουργία ενός figure μεγέθους 12x6 ιντσών για το γράφημα Box-and-Whisker
    plt.figure(figsize=(12, 6))
    # Εμφάνιση του Box-and-Whisker plot για τις επιλεγμένες στήλες του DataFrame
    dataframe[columns].boxplot(sym='k.', vert=False, patch_artist=True)
    # Ορισμός τίτλου για το γράφημα
    plt.title("Διαγράμματα Box-and-Whisker")
    # Ορισμός ετικέτας για τον άξονα x
    plt.xlabel("Τιμές")
    # Ορισμός ετικέτας για τον άξονα y
    plt.ylabel("Στήλες")
    # Εμφάνιση του γραφήματος
    plt.show()



# Ορισμός συνάρτησης για τη δημιουργία Count Bar Plot με τον έλεγχο Shapiro-Wilk
def count_bar_plot_with_shapiro(df, x_col, y_col):
    # Δημιουργία ενός figure μεγέθους 10x6 ιντσών για το γράφημα Count Bar Plot
    plt.figure(figsize=(10, 6))
    # Δημιουργία Count Bar Plot χρησιμοποιώντας τη βιβλιοθήκη seaborn
    ax = sns.countplot(x=x_col, data=df, hue=y_col)
    # Προσθήκη ετικετών με τις τιμές πάνω από τις μπάρες
    for p in ax.patches:
        ax.annotate(f'{p.get_height()}', (p.get_x() + p.get_width() / 2., p.get_height()),
                    ha='center', va='center', xytext=(0, 10), textcoords='offset points')
    # Υπολογισμός του p-value μέσω του τεστ Shapiro-Wilk
    _, p_value = shapiro(df[y_col])
    # Καθορισμός του αποτελέσματος του τεστ Shapiro-Wilk
    shapiro_result = "Κανονικά Διανεμημένα" if p_value > 0.05 else "Μη Κανονικά Διανεμημένα"
    # Προσθήκη οριζόντιας γραμμής με διακεκομμένη γραμμή, σηματοδοτώντας το αποτέλεσμα του τεστ Shapiro-Wilk
    plt.axhline(0, color="black", linestyle="--", linewidth=2, label=f"Shapiro-Wilk: {shapiro_result}")
    # Ορισμός τίτλου για το γράφημα
    plt.title(f'Count Bar Plot για {x_col} με Έλεγχο Shapiro-Wilk')
```

```
    # Προσθήκη λεζάντας
    plt.legend()
    # Εμφάνιση του γραφήματος
    plt.show()

# Ορισμός συνάρτησης για το σχεδιασμό γραφήματος 3D με 5
διαστάσεις
def plot5d(df):
    # Δημιουργία του figure μεγέθους 14x8 ιντσών
    fig = plt.figure(figsize=(14, 8))
    # Δημιουργία του subplot για το 3D plot
    ax = fig.add_subplot(111, projection='3d')
    # Ορισμός του τίτλου του γραφήματος
    t = fig.suptitle('Τύπος Μελώντος - Μέση Βαθμολογία -
Αγορασμένα Προϊόντα - Συνολική Δαπάνη - Επίπεδο Ικανοποίησης',
fontsize=14)
    # Λίστες με τις τιμές των στηλών για τις τρεις διαστάσεις x,
y, z
    xs = list(df['Membership Type'])
    ys = list(df['Average Rating'])
    zs = list(df['Items Purchased'])
    # Δημιουργία λίστας με τα σημεία δεδομένων στον χώρο (x, y,
z)
    data_points = [(x, y, z) for x, y, z in zip(xs, ys, zs)]
    # Λίστα με τις τιμές της στήλης 'Total Spend' για το μέγεθος
των σημείων
    ss = list(df['Total Spend'])
    # Λίστα με τα χρώματα των σημείων βάσει της στήλης
'Satisfaction Level'
    colors = ['red' if wt == 1 else 'yellow' if wt == 2 else
'green' for wt in list(df['Satisfaction Level'])]
    # Σχεδιασμός των σημείων στο 3D plot
    for data, color, size in zip(data_points, colors, ss):
        x, y, z = data
        ax.scatter(x, y, z, alpha=0.2, c=color,
edgecolors='none', s=size)
    # Ορισμός ετικετών για τους άξονες x, y, z
    ax.set_xlabel('Τύπος Μελώντος')
    ax.set_ylabel('Μέση Βαθμολογία')
    ax.set_zlabel('Αγορασμένα Προϊόντα')
    # Εμφάνιση του γραφήματος
    plt.show()
# Ορισμός συνάρτησης για το σχεδιασμό γραφήματος 3D με 6
διαστάσεις
def plot6d_churn(df):
    # Δημιουργία του figure μεγέθους 14x8 ιντσών
    fig = plt.figure(figsize=(14, 8))
    # Δημιουργία του subplot για το 3D plot
```

```
    ax = fig.add_subplot(111, projection='3d')
    # Ορισμός του τίτλου του γραφήματος
    t = fig.suptitle('Τύπος Μελώντος - Μέση Βαθμολογία -
Αγορασμένα Προϊόντα - Συνολική Δαπάνη - Φύλο - Αποχώρηση',
fontsize=14)
    # Λίστες με τις τιμές των στηλών για τις τρεις διαστάσεις x,
y, z
    xs = list(df['Membership Type'])
    ys = list(df['Average Rating'])
    zs = list(df['Items Purchased'])
    # Δημιουργία λίστας με τα σημεία δεδομένων στον χώρο (x, y,
z)
    data_points = [(x, y, z) for x, y, z in zip(xs, ys, zs)]
    # Λίστα με τις τιμές της στήλης 'Total Spend' για το μέγεθος
των σημείων
    ss = list(df['Total Spend'])
    # Λίστα με τα χρώματα των σημείων βάσει της στήλης 'Churn'
    colors = ['red' if wt == 1 else 'green' for wt in
list(df['Churn'])]
    # Λίστα με τους δείκτες (markers) των σημείων βάσει της
στήλης 'Gender'
    markers = ['o' if q == 'Female' else 'x' for q in
list(df['Gender'])]
    # Σχεδιασμός των σημείων στο 3D plot με χρήση χρωμάτων,
μεγεθών και markers
    for data, color, size, mark in zip(data_points, colors, ss,
markers):
        x, y, z = data
        ax.scatter(x, y, z, alpha=0.2, c=color,
edgecolors='none', s=size, marker=mark)
    # Ορισμός ετικετών για τους άξονες x, y, z
    ax.set_xlabel('Τύπος Μελώντος')
    ax.set_ylabel('Μέση Βαθμολογία')
    ax.set_zlabel('Αγορασμένα Προϊόντα')
    # Εμφάνιση του γραφήματος
    plt.show()
```

Βιβλιογραφία

Vlachopoulou, M., 2020. Ψηφιακό Μάρκετινγκ Από τη θεωρία στην πράξη. Εκδοτικός Οίκος Rosili, pp. 503-509.

Liye, M., Baohung, Sun, 2020. Machine Learning and AI in Marketing – Connecting computing power to human insights. International Journal of Research in Marketing, Vol. 37, Issue 3, pp. 481-504.

Haleem, Ab., Javaid, M., Asim, M., Singh, R., Suman, R., 2022. Artificial intelligence (AI) applications for marketing: A literature-based study. International Journal of Intelligent Networks, Vol. 3, pp. 119–132.

Verma, S., Sharma, R., Deb, S., Maitra, D., 2021. Artificial intelligence in marketing: Systematic review and future research direction. International Journal of Information Management Data Insights, Vol. 1, Issue. 1, No. 100002.

Yaiprasert, C., Hidayanto, A., 2023. AI-driven ensemble three machine learning to enhance digital marketing strategies in the food delivery business. Intelligent Systems with Applications, Vol.18, No. 200235.

Combo, L., Vlacic, B., Costa., S., Dabic, M., 2021. The evolving role of artificial intelligence in marketing: A review and research agenda. Journal of Business Research, Vol. 128, pp. 187-203.

Compbell, C., Sands, S., Ferraro, C., Tsao, H., Mavrommatis, A., 2020. From data to action. Business Horizons, Vol. 63, Issue. 2, pp. 227-243.

Soni, N., Sharma, K., Singh, N., Kapoor, Am., 2020. Artificial Intelligence in Business: From Research and Innovation to Market Deployment. Procedia Computer Science, Vol. 167, pp. 2200-2210.

Golab-Andrzejak, Ed., 2023. AI-powered Digital Transformation: Tools, Benefits and Challenges for Marketers – Case Study of LPP. Procedia Computer Science, Vol. 219, pp. 397-404.

Huang, M., Roland, T., 2021. A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science, Vol. 49, pp. 30-50.

Daveport, Th., Guha, Ab., Grewal, D., Bressgott, T., 2019. How artificial intelligence will change the future of marketing, Journal of the Academy of Marketing Science, Vol. 48, pp. 24-42.

Bhosale, S., Sharma, Y.K., Kurupkar, F., Jhabarmal, S.J., 2020. Role of Business Intelligence in Digital Marketing. International Journal of Advance and Innovative Research, Vol. 7, pp. 113-116.

Stone, M., Aravopoulou, El., Ekinci, Y., Evans, G., Hobbs, M., Labib, As., Laughlin, P., Machtynger, J., Machtynger, L., 2020. Artificial intelligence (AI) in strategic marketing decision-making: a research agend. Emerald insight, The Bottom Line, Vol. 33, Issue. 2.

Verma, S., Sharma, R., Deb, S., Maitra, D., 2021. Artificial intelligence in marketing: Systematic review and future research direction. International Journal of Information Management Data Insights, Vol. 1, Issue. 1, No. 1000002.

Huang, M., Rust, R., 2022. A Framework for Collaborative Artificial intelligence in Marketing. Journal of Retailing, Vol. 98, Issue. 2, pp. 209-223.

Mustak, M., Salminen, J., Ple, L., Wirtz, J., 2021. Artificial intelligence in marketing: Top modeling, scientometric analysis, and research agenda. Journal of Business Research, Vol. 124, pp. 389-404.

Rust, R., 2020. The future of marketing. International Journal of Research in Marketing, Vol. 37, Issue. 1, pp. 15-26.

Deryl, M., Verma, S., Srivastava, V., 2023. How does AI drive branding? Towards an integrated theoretical framework for AI-driven branding. International Journal of Information Management Data Insights, Vol. 3, Issue. 2, No. 100205.

Lu, P., Cheng, L., Tzou, J., Chen, S., 2023. Technology roadmap of AI applications in the retail industry. Technological Forecasting and Social Change, Vol. 195, No. 122778.

Bi, Q., Goodman, K., Kaminsky, J., Lessler, J., 2019. What is Machine Learning? A Primer for the Epidemiologist. American Journal of Epidemiology, Vol. 188, Issue. 12, pp. 2222-2239.

Russel, S., Norvig, P., 2016. Artificial Intelligence: A Modern Approach Third Edition. Pearson Education Limited.

Research Report. 2021. Chapter One: What is Artificial Intelligence? Artificial Intelligence and National Security in Israel, pp. 31-40.

Janiesch, C., Zschech, P., Heinrich, K., 2021. Machine learning and deep learning. Electronic Markets, Vol. 31, pp. 685-695.

Jiang, T., Gradus, J., Rosellini, An., 2020. Supervised Machine Learning: A Brief Primer, Vol. 51, Issue. 5, pp. 675-687.

Gurucharan M., 2020. Machine Learning Basics: Random Forest Regression. Towards Data Science.

Burgess, A., 2018. AI in Action. In the Executive Guide to Artificial Intelligence, pp. 73-89.

Zhang, C., Lu, Y., 2021. Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, Vol. 23, No. 100224.

Badilo, S., Banfai, B., Birzele, F., Davydov. I., Hutchinson, L., Thong, T., 2020. An Introduction to Machine Learning. Clinical Pharmacology & Therapeutics, Vol. 107, Issue. 4, pp. 871-885.

Sarker, Iq., 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, Springer Link, Vol. 2, No. 160.

Decelle, A., 2023. An Introduction to Machine Learning. Physica A: Statistical Mechanics and its Applications, Vol. 631, No. 128154.

Biau, G., 2012. Analysis of a Random Forests Model. Journal of Machine Learning Research, vol. 13, pp. 1063-1095.

Ritchie, N., 2021.Supervised Learning Theory: Support Vector Machines (SVMs).

Peng, W., 2018. Decision Tree Induction for Identifying Trends in Line Graphs. Springer Link.

Artificial Intelligence (n.d.). Wikipedia 2020. Retrieved from [https://en.wikipedia.org/wiki/Artificial_intelligence].

Artificial Intelligence Fields 2020. Retrieved from [https://www.quora.com/What-skills-do-you- need-to-learn-AI].

Machine Learning (n.d.). Wikipedia 2020. Retrieved from [https://en.wikipedia.org/wiki/Machine_learning].

Neural Network (n.d.). Wikipedia 2020. Retrieved from [https://en.wikipedia.org/wiki/Neural_network].

XGboost (n.d.). Wikipedia 2020. Retrieved from [https://en.wikipedia.org/wiki/XGBoost].

Confusion Matrix (n.d.). Wikipedia 2020. Retrieved from [https://en.wikipedia.org/wiki/Confusion_matrix].