



ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΗΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ & ΠΛΗΡΟΦΟΡΙΚΗΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Εφαρμογή μηχανικής μάθησης σε διαδικτυακές διαφημίσεις και μάρκετινγκ

Υποψήφια: Μάλο Ερίσα (Π2016164)

Επιβλέπων: Ανδρέας Καναβός

Κέρκυρα, 2024

Εφαρμογή μηχανική μάθησης σε διαδικτυακές διαφημίσεις και μάρκετινγκ

Η πτυχιακή εργασία εγκρίθηκε από:

Ανδρέας Καναβός

Άννα Σωτηροπούλου

Στέργιος Παλαμάς

Περίληψη

Η τεχνητή νοημοσύνη και η μηχανική μάθηση αποτελούν πολύτιμα εργαλεία στον τομέα του μάρκετινγκ καθώς η δυνατότητά τους να επεξεργάζονται μεγάλα σύνολα δεδομένων γρήγορα και με ακρίβεια επιτρέπει στις εταιρείες να κατανοήσουν καλύτερα τον στόχο τους και να προσφέρουν εξατομικευμένες προτάσεις στους πελάτες τους βελτιώνοντας την εμπειρία τους και την αποτελεσματικότητα των καμπανιών τους. Η ανάπτυξη προηγμένων αλγορίθμων βασισμένων στη μηχανική μάθηση βοηθά τις επιχειρήσεις να προβλέπουν τις μελλοντικές τάσεις της αγοράς και να προσαρμόζονται γρήγορα σε αυτές. Στην παρούσα εργασία παρουσιάζεται μία βιβλιογραφική συλλογή που αποτελείται από μελέτες και έρευνες αναδεικνύοντας τη αναγκαιότητα της μηχανικής μάθησης στις στρατηγικές μάρκετινγκ των επιχειρήσεων προσφέροντας μια πληθώρα πρακτικών παραδειγμάτων και μεθοδολογιών. Η μικρή αυτή ατζέντα περιλαμβάνει έρευνες για την πρόβλεψη των καταναλωτικών συμπεριφορών, την τμηματοποίηση αγοράς, την απάτη κλικ, την χρήση chatbot, την ανάλυση συναισθήματος και τη δέσμευση πελατών και αποτελεί έναν οδηγό για την ανάπτυξη στρατηγικών μάρκετινγκ.

Λέξεις κλειδιά: Μηχανική μάθηση, μάρκετινγκ, πρόβλεψη συμπεριφοράς, τμηματοποίηση αγοράς, απάτη κλικ, chatbot, ανάλυση συναισθήματος, δέσμευση πελατών

Abstract

Artificial intelligence and machine learning are valuable tools in the field of marketing, as their ability to process large datasets quickly and accurately allows companies to understand their target audience and offer personalized recommendations to their campaigns, thereby improving their experience and the effectiveness of their campaigns. The development of advanced ML algorithms helps businesses predict future market trends and adapt quickly to them. This paper presents a bibliographic collection of studies and researches, highlighting the necessity of machine learning in marketing strategies of businesses providing practical examples and methodologies. This agenda includes research on predicting consumer behaviors, market segmentation, click fraud detection, chatbots usage, sentiment analysis, customer engagement and serves as a guide for developing marketing strategies.

Keywords: Machine learning, marketing, consumer behaviors, market segmentation, click fraud detection, chatbots, sentiment analysis, customer engagement

Κατάλογος περιεχομένων

Περίληψη	3
Κατάλογος εικόνων	7
Κατάλογος πινάκων	8

Κεφάλαιο 1 –Μηχανική Μάθηση

1.1 Εισαγωγή	9
1.2 Μηχανική μάθηση	9
1.3 Μέθοδοι μηχανικής μάθησης	10
1.4 Επιβλεπόμενη μάθηση	11
1.4.1 Εκμάθηση συνόλου	11
1.5 Μη επιβλεπόμενη μάθηση	11
1.6 Ενισχυτική μάθηση	11
1.7 Αλγόριθμοι μηχανικής μάθησης	11
1.7.1 Ιεραρχική Ομαδοποίηση	12
1.7.2 Ομαδοποίηση K-means	12
1.7.3 Δέντρα απόφασης	13
1.7.4 Νευρωνικά δίκτυα	14
1.7.5 Αφελείς Ταξινομητές Μπευζ	15
1.7.6 K-πλησιέστερος γείτονας	16
1.7.7 Μηχανές διανυσμάτων υποστήριξης	17
1.7.8 Λογιστική παλινδρόμηση	17
1.7.9 Extreme Gradient Boosting	18
1.7.10 Στοχαστικός αλγόριθμος ενίσχυσης	18
1.7.11 Τυχαίο Δάσος	19

Κεφάλαιο 2

2.1 Εξόρυξη δεδομένων	20
2.2 Τεχνικές διαχείρισης δεδομένων	21
2.2.1 Ταξινόμηση	21
2.2.2 Παλινδρόμηση	21
2.2.3 Συσταδοποίηση	21
2.3 Μετρικές αξιολόγησης	22
2.4 Weka	25
2.5 Εξόρυξη δεδομένων στο Μάρκετινγκ	25
2.6 Ηθική χρήση δεδομένων	26
2.8 Ηθική χρήση μηχανικής μάθησης	27
2.9 Python	29
2.10 Βιβλιοθήκες Python	29
2.11 Επεξεργασία Φυσικής γλώσσας	30

Κεφάλαιο 3: Στρατηγικές και προκλήσεις μάρκετινγκ

3.1. SEO	32
3.2 Email marketing	33
3.3 Πληρωμένη διαφήμιση	34
3.4 Διαχείριση πελατειακών σχέσεων	34
3.5 Ανάλυση δεδομένων	35
3.6 Προκλήσεις Μάρκετινγκ	35
3.7 Τύποι διαφημίσεων	36
3.7.1 Διαφήμιση μέσω προγραμματισμού	36
3.7.2 Διαδραστική διαφήμιση	37
3.7.3 Εγγενής διαφήμιση	38
3.7.4 AI powered διαφήμιση	39
3.7.5 Shoppable Content διαφήμιση	39
3.8. Διαφημιστική απάτη	40
3.9 Αποκλεισμός διαφημίσεων	40
3.10 Κανονισμοί απορρήτου	41
3.11 Ad View ability	42

Κεφάλαιο 4 -Εφαρμογές μηχανικής μάθησης στον τομέα του μάρκετινγκ

4.1 Πρόβλεψη συμπεριφοράς	43
4.2 Πρόβλεψη αποχώρησης πελατών	46
4.3 Τμηματοποίηση πελατών	48
4.4 Click through Rate	52
4.5 Click Fraud	56
4.6 Chabot	59
4.7 Ανάλυση συναισθήματος	63
4.8 Δέσμευση Πελάτη	67
4.9 Συγκεντρωτικός πίνακας αποτελεσμάτων.	69

Κεφάλαιο 5

Συμπεράσματα	70
Βιβλιογραφία	72

Κατάλογος εικόνων

1. Διάγραμμα ταξινόμησης
2. Διάγραμμα παλινδρόμησης
3. Διάγραμμα Συσταδοποίησης
4. Διαδικασία συσταδοποίησης
5. Πίνακας σύγχυσης (Confusion Matrix)\
6. Καμπύλη Roc
7. Καμπύλη AUC
8. Σύστημα Weka
9. Μηχανική μάθηση
10. Ιεραρχική ομαδοποίηση
11. K-means Ομαδοποίηση
12. Δέντρο απόφασης
13. Νευρωνικό Δίκτυο
14. Ταξινόμηση Naive Bayes
15. K-πλησιέστερος γείτονας
16. Μηχανές διανυσμάτων υποστήριξης
17. Σιγμοειδής συνάρτηση λογιστικής παλινδρόμησης
18. Αρχιτεκτονική XGBoost
19. Αρχιτεκτονική Τυχαίου Δάσους
20. Google Search Console
21. Διάγραμμα ροής email marketing
22. Σύγκριση απόδοσης πωλήσεων με Mobile CRM και χωρίς Mobile CRM
23. ROI μέτρηση απόδοσης
24. Διαδικασία προγραμματιστικής διαφήμισης
25. Διαφήμιση Paid Search
26. ROC αλγορίθμων
27. Διαδικασία προετοιμασίας δεδομένων
28. Αναλογία κλικ προς τον χρόνο που δαπανήθηκε.
29. Αναλογία κλικ προς την καθημερινή χρήση του διαδικτύου
30. Αναλογία κλικ προς την ηλικία του χρήστη

31. Αναλογία κλικ ως προς τα έσοδα των χρηστών
32. Αλληλεπίδραση χρήστη με ιστοσελίδα
33. Μεθοδολογία έρευνας
34. Διάγραμμα ροής chatbot
35. Διαδικασία ανάλυσης συναισθήματος
36. UML διάγραμμα κύριων τάξεων
37. Βασικές λειτουργικές ομάδες DSS
38. Σχέδιο ανάλυσης μεταξύ πηγών

Κατάλογος Πινάκων

1. Τύποι διαδραστικών διαφημίσεων\
2. Απάτες διαδικτύου
3. Αξιολόγηση απόδοσης αλγορίθμων για πρόβλεψη αποχώρησης πελατών.
4. Κατηγορίες επισκεπτών
5. Χαρακτηριστικά συνόλου δεδομένων
6. 10-fold Cross Validation
7. Αξιολόγηση απόδοσης μοντέλων για απάτη κλικ.
8. Ανάλυση καλύτερης απόδοσης μοντέλου σε δύο διαφορετικά σενάρια
9. Παραδείγματα ταξινόμησης σχολίων
10. Αξιολόγηση απόδοσης μοντέλων για δέσμευση πελατών
11. Αξιολόγηση αποτελεσμάτων ερευνών

Κεφάλαιο 1 – Μηχανική Μάθηση

1.1 Εισαγωγή

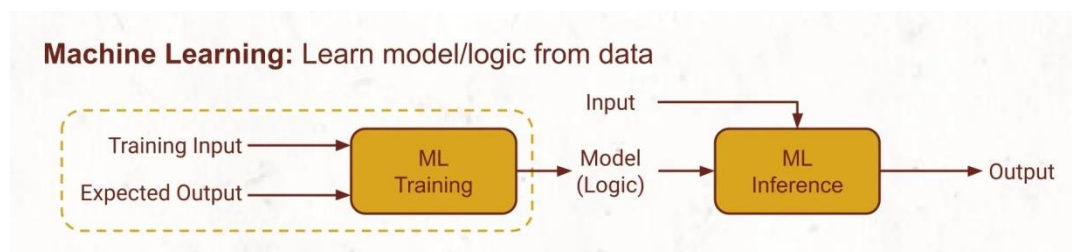
Τα τελευταία χρόνια, η τεχνητή νοημοσύνη έχει αλλάξει τον τρόπο που λειτουργεί μία κοινωνία του σήμερα, αντιμετωπίζοντας προκλήσεις που παλιά ήταν ανέφικτο να αντιμετωπιστούν αποτελεσματικά, αναπτύσσοντας τεχνολογίες που προσφέρουν λύσεις και οφέλη σε διάφορους τομείς. Η τεχνητή νοημοσύνη καθίσταται αναγκαία και απαραίτητη στη σύγχρονη κοινωνία διότι μπορεί να αντιμετωπίσει αποτελεσματικά προβλήματα που υπερβαίνουν την ανθρώπινη ικανότητα, καθώς διαθέτει την δυνατότητα αυτόματης επεξεργασίας δεδομένων, αναγνώρισης προτύπων αλλά και λήψης αποφάσεων που βασίζονται σε πολύπλοκες πληροφορίες. Ωστόσο, η τεχνητή νοημοσύνη χρειάζεται τη μηχανική μάθηση προκειμένου να μπορέσει να εξελιχθεί περαιτέρω. Η μηχανική μάθηση επιτρέπει στους υπολογιστές να εκπαιδεύονται βάσει των δεδομένων αναπτύσσοντας αλγορίθμους και μοντέλα που μπορούν να κάνουν προβλέψεις, να αναλύουν δεδομένα και να προσφέρουν προσαρμοστικές λύσεις. Χρησιμοποιείται σε διάφορους τομείς, όπως η υγεία, η ρομποτική, η χρηματοοικονομική ανάλυση το μάρκετινγκ και πολλοί άλλοι, βελτιώνοντας την ακρίβεια, την απόδοση και την αποτελεσματικότητα τους. Στον τομέα του μάρκετινγκ, η μηχανική μάθηση αποκτά έναν πολύ σημαντικό ρόλο στην αντιμετώπιση των προκλήσεων, καθώς ενισχύει την καινοτομία και την αποτελεσματικότητα των επιχειρήσεων παρέχοντας τους ένα ανταγωνιστικό πλεονέκτημα. Χρησιμοποιεί προηγμένες τεχνικές ανάλυσης και πρόβλεψης για την συμπεριφορά των καταναλωτών και δημιουργεί αποτελεσματικές στρατηγικές μάρκετινγκ διαχειρίζοντας μεγάλα σύνολα δεδομένων. Ουσιαστικά, χάρη σε αυτήν, οι επιχειρήσεις είναι σε θέση να αναλύουν μεγάλους όγκους δεδομένων, ανιχνεύοντας τις τάσεις στην αγορά, να προβλέπουν τις επιθυμίες και τις ανάγκες των πελατών βελτιώνοντας την εξυπηρέτησή τους, και να προσφέρουν ένα πιο εξατομικευμένο περιεχόμενο, βασισμένο στις ενέργειες του κάθε πελάτη, προσφέροντας τους εξατομικευμένες προσφορές. Η εφαρμογή της μηχανικής μάθησης στον τομέα του μάρκετινγκ καθίσταται ζωτικής σημασίας, καθώς με τη χρήση προηγμένων αλγορίθμων μηχανικής μάθησης βελτιώνεται η απόδοση των καμπανιών μιας επιχείρησης, μειώνεται αποτελεσματικά το κόστος και αυξάνεται η ανταπόκριση των πελατών. Πέρα από τις εξατομικευμένες προτάσεις και την πρόβλεψη συμπεριφοράς των καταναλωτών, με τη βοήθεια της μηχανικής μάθησης έχουν αναπτυχθεί εφαρμογές όμως προσωπικοί βοηθοί και εικονικοί βοηθοί, οι οποίες παρέχουν στους ανθρώπους εξατομικευμένες υπηρεσίες και πληροφορίες βελτιώνοντας την εμπειρία τους και προσφέροντας τους άμεση εξυπηρέτηση ανά πάσα ώρα και στιγμή.

1.2 Μηχανική μάθηση

Με τον όρο Μάθηση (Learning) εννοούμε την απόκτηση γνώσεων μέσα από διάφορες γνωστικές διαδικασίες. Οι επιστήμονες που ασχολούνται με την Τεχνητή Νοημοσύνη επιχείρησαν να φτιάξουν υπολογιστικά συστήματα τα οποία θα μπορούν να μάθουν, δημιουργώντας έτσι την Μηχανική Μάθηση (Machine Learning). Η Μηχανική Μάθηση βασίζεται στην δημιουργία μηχανών ικανών να μαθαίνουν, δηλαδή βάσει των δεδομένων που τους παρέχει ο άνθρωπος να βελτιώνουν την απόδοσή τους αποκτώντας γνώση και εμπειρία. Επιπλέον μία μηχανή είναι ικανή να δημιουργεί νέα μοντέλα και

πρότυπα μάθησης από ειδικά δεδομένα και παραδείγματα. Συνεπώς, με τον όρο Μηχανική Μάθηση ονομάζουμε την ικανότητα ενός υπολογιστικού συστήματος βάσει ενός συνόλου δεδομένων να δημιουργεί μοντέλα ή πρότυπα μάθησης.⁽¹⁸⁾

Ως ένας κλάδος της επιστήμης των υπολογιστών, η Μηχανική Μάθηση μελετά σε βάθος την κατασκευή των αλγορίθμων οι οποίοι μπορούν να εξάγουν πρότυπα και να δημιουργήσουν μοντέλα ικανά για να κάνουν χρήσιμες προβλέψεις, αναπτύσσοντας γνώση από σύνολα δεδομένων. Κάθε αλγόριθμος προκειμένου να μάθει και να μπορεί να κάνει προβλέψεις σε δεδομένα χρειάζεται εκπαίδευση. Κατασκευάζεται ένα σύνολο δεδομένων εκπαίδευσης (Training Set) με στόχο τη δημιουργία νέας γνώσης και στη συνέχεια ακολουθεί η φάση της πιστοποίησης της νέας γνώσης με βάση τα δεδομένα ελέγχου (Test Data). Έτσι ο αλγόριθμος μαθαίνει να εντοπίζει μοτίβα και για κάθε νέο δεδομένο να υλοποιεί έλεγχο του περιεχομένου του και να κάνει μια πρόβλεψη, δίνοντας τη νέα γνώση σε εφαρμογές που την έχουν ανάγκη για την λύση προβλημάτων.^[5]



Εικόνα 1. Μηχανική μάθηση

Ορισμός

Σύμφωνα με τον Mitchell (1997), « Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μία κλάση εργασιών T και ένα μέτρο απόδοσης P , αν η επίδοση του σε εργασίες της κλάσης T , όπως αποτιμάται από το μέτρο P , βελτιώνεται με την εμπειρία E ». ⁽¹⁸⁾

1.3 Μέθοδοι μηχανικής μάθησης

Οι μέθοδοι μηχανικής μάθησης εξαρτώνται από τον τρόπο με τον οποίο γίνεται η μάθηση στους αλγορίθμους. Για κάθε μέθοδο μάθησης προκειμένου να επιλυθεί ένα τυχαίο πρόβλημα χρησιμοποιείται ο αλγόριθμος που έχει εκπαιδευτεί κατάλληλα για πρόβλημα αυτό. Ο τρόπος με τον οποίο υπολογίζεται το μοντέλο πρόβλεψης αφορά την καταλληλότητα των αλγορίθμων που διακρίνεται στο είδος του προβλήματος που χρήζει επίλυσης, το σύνολο των δεδομένων και το πλήθος περιπτώσεων.

1.4 Επιβλεπόμενη Μάθηση (Supervised learning)

Στη μάθηση με επίβλεψη προτείνεται ένα είδος μάθησης το οποίο εξετάζει μια άγνωστη συνάρτηση χρησιμοποιώντας δεδομένα τα οποία καταγράφουν με ακρίβεια τα αποτελέσματα της συνάρτησης. Στους αλγόριθμους παρέχονται πληροφορίες εισόδου ενός συνόλου δεδομένων και γνωστά αποτελέσματα και σύμφωνα με τις σχέσεις τους προσδιορίζεται με ακρίβεια το μοντέλο που τα περιγράφει. Πιο συγκεκριμένα, ο αλγόριθμος μηχανικής μάθησης εκπαιδεύεται προκειμένου να δημιουργήσει μία συνάρτηση η οποία τοποθετεί τις εισόδους ενός συνόλου δεδομένων ανάλογα με τις γνωστές επιθυμητές εξόδους. Εν συνεχεία, εισάγεται ένα νέο σύνολο δεδομένων εισόδου, στο οποίο δεν είναι γνωστές οι τιμές εξόδων τους και ο αλγόριθμος μπορεί να κάνει προβλέψεις. Σκοπός της επιβλεπόμενης μάθησης δηλαδή είναι η πρόβλεψη των αποτελεσμάτων αναλόγως των μετρήσεων των δεδομένων εισόδου. Χρησιμοποιείται για προβλήματα ταξινόμησης και προβλήματα παλινδρόμησης.⁽¹⁹⁾

1.4.1 Εκμάθηση συνόλου

Η εκμάθηση συνόλου αποτελεί μια μέθοδο η οποία χρησιμοποιεί πολλούς αλγόριθμους εκμάθησης με στόχο την επίτευξη μιας καλύτερης προγνωστικής απόδοσης σε σύγκριση με τις αποδόσεις των μεμονωμένων αλγορίθμων.⁽²⁰⁾

1.5 Μη επιβλεπόμενη Μάθηση (Unsupervised Learning)

Στη μάθηση χωρίς επίβλεψη, οι αλγόριθμοι δέχονται πληροφορίες εισόδου ενός συνόλου δεδομένων χωρίς να είναι γνωστά τα αποτελέσματα εξόδου. Πιο συγκεκριμένα, το σύστημα επιχειρεί να βρει συσχετίσεις και ομάδες χρησιμοποιώντας μόνο τις ιδιότητες των δεδομένων και βρίσκει πρότυπα εισόδου που περιγράφουν ένα μέρος από τα δεδομένα χωρίς να γνωρίζει τις τιμές εξόδου. Στόχος της επιβλεπόμενης μάθησης είναι η αναγνώριση και η ομαδοποίηση των δεδομένων. Χρησιμοποιείται σε προβλήματα ανάλυσης συσχετισμών και συσταδοποίησης.⁽²¹⁾

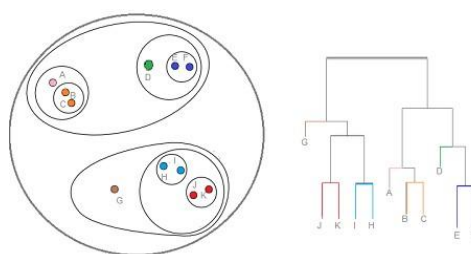
1.6 Ενισχυτική μάθηση

Στην ενισχυτική μάθηση (LR), το σύστημα μάθησης βασίζεται σε τεχνητή νοημοσύνη (πράκτορες) και μιμείται την ανθρώπινη νοημοσύνη δηλαδή τη διαδικασία μάθησης δοκιμής και λάθους που χρησιμοποιείται από τον άνθρωπο. Κατά την επεξεργασία δεδομένων, χρησιμοποιείται επίσης ένα παράδειγμα "τιμωρίας" ή "ανταμοιβής" μαθαίνοντας έτσι από την ανάδραση της κάθε δράσης και ανακαλύπτοντας μόνοι τους με ποιες ενέργειες θα επιτύχουν καλύτερα αποτελέσματα.⁽²²⁾

1.7 Αλγόριθμοι μηχανικής μάθησης

1.7.1 Ιεραρχική Ομαδοποίηση

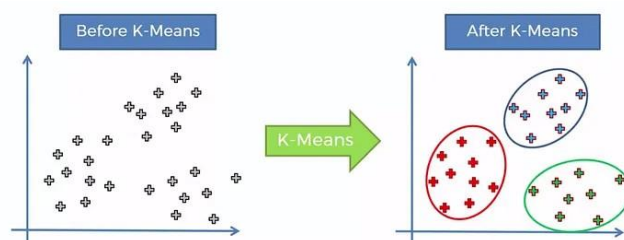
Η ιεραρχική ομαδοποίηση είναι μια μεθοδολογία για την ανάλυση συστάδων που χρησιμοποιείται για την οργάνωση σημείων δεδομένων σε μια ιεραρχία συστάδων, σχηματίζοντας έτσι ένα δενδρόγραμμα, μια δομή που μοιάζει με δέντρο όπου αναπαριστά οπτικά τις σχέσεις και τις ομαδοποιήσεις μεταξύ των δεδομένων. Η διαδικασία περιλαμβάνει γενικά μια αθροιστική προσέγγιση, όπου κάθε σημείο δεδομένων ξεκινά ως μεμονωμένο σύμπλεγμα και τα συμπλέγματα συγχωνεύονται διαδοχικά μέχρι να σχηματιστεί ένα ενιαίο γενικό σύμπλεγμα. Ο αλγόριθμος ξεκινά με τον υπολογισμό της ανομοιοτητας ή της απόστασης μεταξύ όλων των ζευγών συστάδων. Αυτό περιλαμβάνει την επιλογή μιας κατάλληλης μέτρησης απόστασης με βάση τα χαρακτηριστικά των δεδομένων. Έπειτα δημιουργείται ένας πίνακας εγγύτητας που χρησιμοποιείται για την αποθήκευση των ανομοιοτήτων. Στη συνέχεια ο αλγόριθμος προσδιορίζει τα πλησιέστερα συμπλέγματα με βάση το επιλεγμένο κριτήριο σύνδεσης. Τα κοινά κριτήρια σύνδεσης περιλαμβάνουν ενιαία σύνδεση, η οποία μετρά την απόσταση μεταξύ των πλησιέστερων μελών των συστάδων. Την πλήρη σύνδεση, η οποία λαμβάνει υπόψη την απόσταση μεταξύ των πιο απομακρυσμένων μελών, και μέση απόσταση μεταξύ όλων των μελών. Το ζεύγος των συστάδων με τη μικρότερη ανομοιοτητα συγχωνεύεται σε ένα νέο σύμπλεγμα και ο πίνακας εγγύτητας ενημερώνεται ανάλογα. Αυτή η διαδικασία αναγνώρισης, συγχώνευσης και ενημέρωσης του πίνακα εγγύτητας επαναλαμβάνεται μέχρι να παραμείνει μόνο ένα σύμπλεγμα, που περιλαμβάνει όλα τα σημεία δεδομένων. Το αποτέλεσμα είναι ένα δενδρόγραμμα, όπου το ύψος κάθε σύντηξης στο δέντρο αντιπροσωπεύει την ανομοιοτητα μεταξύ των συγχωνευμένων συστάδων. Ένα δενδρόγραμμα αναπαριστά οπτικά τις ιεραρχικές σχέσεις μεταξύ των δεδομένων. Είναι αρκετά ευέλικτο όσον αφορά τον αριθμό των συστάδων καθώς δεν χρειάζεται να καθοριστεί από την αρχή. Ωστόσο, ο τρόπος που υπολογίζεται η απόσταση μεταξύ των συστάδων κάνει την ιεραρχική ομαδοποίηση υπολογιστικά απαιτητική ειδικά για μεγάλα σύνολα δεδομένων. Δεν παύει όμως να αποτελεί ένα πολύτιμο εργαλείο για την ανάλυση δεδομένων και την ανακάλυψη προτύπων. ^[6]



Εικόνα 2. Ιεραρχική ομαδοποίηση

1.7.2 Ομαδοποίηση K-means

Η ομαδοποίηση K-means είναι ένας επαναληπτικός αλγόριθμος μηχανικής μάθησης χωρίς επίβλεψη που έχει σχεδιαστεί για να χωρίζει ένα σύνολο δεδομένων σε ένα προκαθορισμένο αριθμό συστάδων, με στόχο την ομαδοποίηση παρόμοιων σημείων δεδομένων. Ο αλγόριθμος ξεκινά επιλέγοντας τον επιθυμητό αριθμό συστάδων (K) και αρχικοποιώντας τυχαία τις K συστάδες κεντροειδών, χρησιμοποιώντας ως αρχικές εκτιμήσεις για τα κέντρα συμπλέγματος. Στη συνέχεια, κάθε σημείο δεδομένων εκχωρείται στο σύμπλεγμα του οποίου το κέντρο είναι πιο κοντά, τοπικά προσδιοριζόμενο χρησιμοποιώντας τη μέτρηση της ευκλείδειας απόστασης.⁽²³⁾ Ο στόχος είναι να ελαχιστοποιηθεί το άθροισμα των τετραγωνικών αποστάσεων από κάθε σημείο στο εκχωρημένο κέντρο συστάδας του. Μετά την ανάθεση, ο αλγόριθμος προχωρά στην ενημέρωση των κεντροειδών υπολογίζοντας τα ως μέσο όρο όλων των σημείων δεδομένων που ανήκουν στις αντίστοιχες συστάδες. Αυτή η διαδικασία εκχώρησης και ενημέρωσης κεντροειδούς επαναλαμβάνεται μέχρι τη σύγκλιση, που ορίζεται είτε από μία σταθερή διαμόρφωση κεντροειδούς είτε με τη επίτευξη ενός προκαθορισμένου αριθμού επαναλήψεων. Το τελικό συμπέρασμα είναι ένα σύνολο συστάδων με τα αντίστοιχα κεντροειδή τους, που προσφέρουν πληροφορίες για την εγγενή δομή των δεδομένων. Είναι σημαντικό να σημειωθεί ότι η ομαδοποίηση K-means μπορεί να συγκλίνει σε ένα τοπικό ελάχιστο και η απόδοσή του μπορεί να είναι ευαίσθητη στην αρχική τοποθέτηση του κέντρου. Συνεπώς, μια κοινή πρακτική του αλγορίθμου κατά την εκκίνηση εκτέλεσης είναι η χρήση διαφορετικών αρχικοποιήσεων.^[30] Επιπλέον, ο προσδιορισμός του βέλτιστου αριθμού συστάδων (K) συχνά περιλαμβάνει γνώση του τομέα ή τεχνικές όπως η μέθοδος Elbow ή η ανάλυση silhouette.⁽²⁴⁾



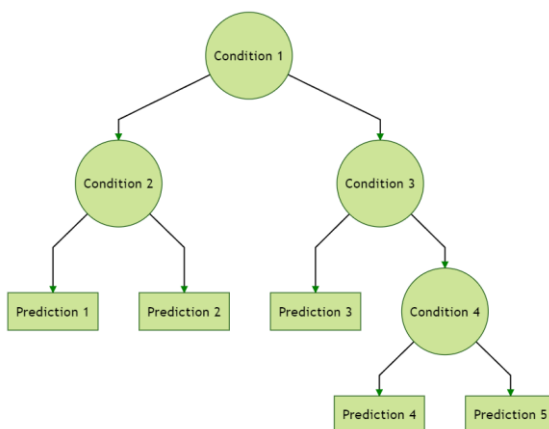
Εικόνα 3. K-means Ομαδοποίηση

1.7.3 Δέντρα Απόφασης (Decision Trees)

Ο αλγόριθμος δέντρων αποφάσεων (DT) θεωρείται από τους πιο γνωστούς αλγορίθμους ταξινόμησης επιβλεπόμενης μάθησης λόγω της σαφήνειας και της απλότητας. Χρησιμοποιείται για την πρόβλεψη και την εξαγωγή συμπερασμάτων βάσει ενός συνόλου παρατηρήσεων. Ένα δέντρο απόφασης αναπαρίσταται με μια δομή δέντρου και αποτελείται από τον κόμβο ρίζα, τους εσωτερικούς κόμβους και τους τελικούς κόμβους-φύλλα. Η ρίζα βρίσκεται στην κορυφή με ακμές προσανατολισμένες προς χαμηλότερου επιπέδου κόμβους και διαχωρίζει το σύνολο εκπαίδευσης. Για την μετάβαση από έναν κόμβο σε κάποιον άλλον διεξάγεται έλεγχος μιας συνθήκης και δημιουργούνται οι εσωτερικοί κόμβοι οι οποίοι με την σειρά τους χωρίζουν το κάθε υποσύνολο σε μικρότερα υποσύνολα μέχρι να εισαχθούν όλα τα χαρακτηριστικά. Τέλος, τα φύλλα αντιπροσωπεύουν την πρόβλεψη της τιμής του χαρακτηριστικού βάσει του συνόλου δεδομένων εκπαίδευσης. Χρησιμοποιούνται επαναληπτικοί διαμερισμοί των δεδομένων προκειμένου να βρεθούν τα χαρακτηριστικά που διαχωρίζουν καλύτερα τις κλάσεις.^[8] Η διαδικασία ανάπτυξης του δέντρου σταματά όταν δεν μπορεί να γίνει άλλος διαχωρισμός υποσυνόλου ή όταν εκπληρωθεί κάποιο κριτήριο.⁽²⁵⁾ Η διαδικασία "κλαδέματος" διαγράφει τα φύλλα που δεν χρειάζονται δημιουργώντας ένα δέντρο που προβλέπει με μεγαλύτερη ακρίβεια σε νέα δεδομένα.⁽²⁶⁾

Δέντρα ταξινόμησης καλούνται τα μοντέλα δέντρων στα οποία η μεταβλητή στόχος μπορεί να δεχτεί ένα διακριτό σύνολο τιμών.

Δέντρα παλινδρόμησης καλούνται τα δέντρα απόφασης όπου ο στόχος μπορεί να λάβει συνεχείς τιμές (συνήθως πραγματικούς αριθμούς).⁽²⁵⁾



Εικόνα 4. Δέντρο απόφασης

1.7.4 Νευρωνικά δίκτυα (Neural Networks)

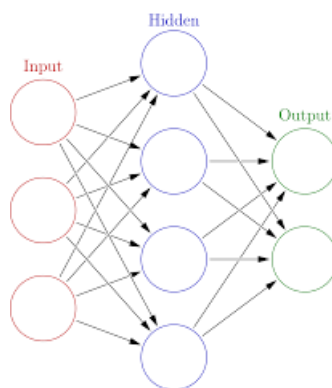
Τα τεχνητά νευρωνικά δίκτυα (ANN) αποτελούν ένα μοντέλο μηχανικής μάθησης με επίβλεψη και έχει σχεδιαστεί με τέτοιο τρόπο ώστε να μιμείται τον τρόπο με τον οποίο λειτουργεί το βιολογικό νευρικό σύστημα του ανθρώπου και πιο συγκεκριμένα τον ανθρώπινο εγκέφαλο, να μαθαίνει και από τα λάθη του να διορθώνεται. Τα ANN αποτελούνται από ένα σύνολο κόμβων οι οποίοι καλούνται νευρώνες και τις μεταξύ τους συνδέσεις που αναπαριστούν τις σχέσεις μεταξύ των νευρώνων. Οι συνδέσεις μεταφέρουν πληροφορίες από τους κόμβους εισόδου στους κόμβους εξόδου.⁽²⁵⁾ Πιο συγκεκριμένα ένα βασικό νευρωνικό δίκτυο αναπαρίσταται σε τρία επίπεδα:

Κόμβοι εισόδου: Αφορούν το πρώτο επίπεδο που δέχεται τα αρχικά δεδομένα εισόδου και στην συνέχεια αφού τα επεξεργαστούν και τα αναλύσουν, τα μεταφέρουν στο επόμενο επίπεδο.

Κρυμμένο επίπεδο: Οι κόμβοι βρίσκονται ανάμεσα από τους κόμβους εισόδου και εξόδου και ουσιαστικά σε αυτό το επίπεδο πραγματοποιείται η μάθηση και το μεγαλύτερο μέρος του υπολογισμού.

Κόμβοι εξόδου: Πρόκειται για το τελευταίο επίπεδο σε ένα νευρωνικό δίκτυο στο οποίο ο αριθμός των νευρώνων είναι μεταβαλλόμενο καθώς εξαρτάται από το ζητούμενο.

Τα νευρωνικά δίκτυα είναι ικανά για την ταχεία ταξινόμηση μεγάλου όγκου δεδομένων και ιδανικά να χρησιμοποιηθούν για την αναγνώριση ομιλίας και εικόνας και την ανάλυση συναισθήματος.⁽²⁶⁾



Εικόνα 5. Νευρωνικό Δίκτυο

1.7.5 Αφελείς ταξινομητές Μπεϋζ (Naive Bayesian Classifiers)

Ο πιθανοτικός αλγόριθμος ταξινόμησης Naive Bayes βασίζεται στο θεμελιωμένο θεώρημα του Bayes και λειτουργεί με μερικές βασικές προϋποθέσεις ανεξάρτητων χαρακτηριστικών. Οι προβλέψεις ενός μοντέλου Naive Bayes είναι υπό όρους ανεξάρτητες με οποιοδήποτε άλλο χαρακτηριστικό του μοντέλου. Παρόλο που είναι ένας απλός και γρήγορος αλγόριθμος, ο οποίος μπορεί να εφαρμοστεί σε δυαδικές ή και περισσότερες κλάσεις για ταξινόμηση, δεν είναι ικανός να καθορίσει τη συσχέτιση μεταξύ των χαρακτηριστικών.^[9]

Ορισμός θεωρήματος Bayes

Ορίζεται ως μαθηματική εξίσωση⁽²⁸⁾:

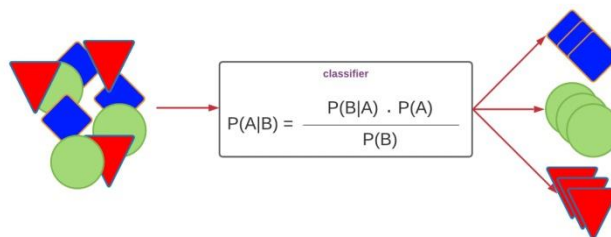
$$P(A|B) = P(B|A) \times \frac{P(A)}{P(B)}$$

όπου A και B είναι τα γεγονότα

P(A) και P(B) είναι οι πιθανότητες των δεδομένων A και B που είναι ανεξάρτητα μεταξύ τους.

P(A|B) είναι η πιθανότητα του A δεδομένου το B να είναι αληθές.

P(B|A) είναι η πιθανότητα του B δεδομένου το A να είναι αληθές.



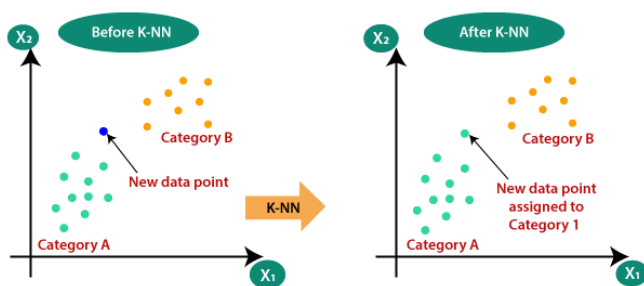
Εικόνα 6. Ταξινόμηση Naive Bayes

Bayesian Networks

Πρόκειται για ένα πιθανοτικό μοντέλο το οποίο παρουσιάζει μέσω ενός κατευθυνόμενου ακυκλικού γραφήματος (DAG) ένα σύνολο από μεταβλητές υπό όρους εξαρτήσεις. Είναι ιδανικά για την πρόβλεψη της πιθανότητας επαναγοράς ενός προϊόντος.⁽²⁹⁾

1.7.6 K-πλησιέστερος γείτονας (K-nearest Neighbor)

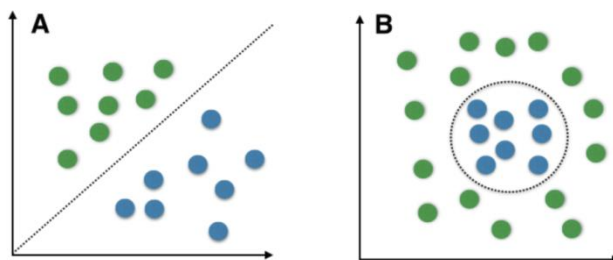
Ο αλγόριθμος k-nearest Neighbor (KNN) αποτελεί μια εποπτευόμενη μηχανική μάθηση που βασίζεται σε παραδείγματα. Η βασική ιδέα του αλγορίθμου περιλαμβάνει τη μέτρηση της ομοιότητας μεταξύ των δεδομένων, χρησιμοποιώντας συνήθως την "Ευκλείδεια απόσταση", κατηγοριοποιώντας τα σε ομάδες. Για κάθε νέα εκχώρηση, επιλέγεται η καταλληλότερη ομάδα ανάλογα με την ομοιότητα των δεδομένων της κάθε ομάδας. Πιο συγκεκριμένα, υπολογίζεται το μέτρο απόστασης και στη συνέχεια προσδιορίζεται ο k πλησιέστερος γείτονας ενός δεδομένου. Μπορεί να χρησιμοποιηθεί για προβλήματα ταξινόμησης αλλά και παλινδρόμησης. Για εργασίες ταξινόμησης, η ετικέτα κλάσης καθορίζεται με πλειοψηφία μεταξύ των k γειτόνων, ενώ για εργασίες παλινδρόμησης γίνεται πρόβλεψη της μέσης τιμής των k γειτόνων. Η επιλογή του k καθορίζεται από τον χρήστη και επηρεάζει την αποτελεσματικότητα του αλγορίθμου. Ένα μικρότερο k μπορεί να οδηγήσει σε υπερπροσαρμογή ενώ ένα μεγαλύτερο είναι πιθανό να ξεπεράσει τα όρια απόφασης. Ο αλγόριθμος KNN δεν εκπαιδεύεται άμεσα από το σύνολο εκπαίδευσης, αλλά αποθηκεύει τα δεδομένα εκπαίδευσης και εκτελείται μόνο κατά τη διαδικασία της πρόβλεψης. Θεωρείται εύκολος όσον αφορά την κατανόηση του και απλός στην εφαρμογή του και παρά την ευαισθησία που παρουσιάζει στην κλιμάκωση των χαρακτηριστικών, είναι υπολογιστικά ακριβής ειδικά σε μεγάλα σύνολα δεδομένων.⁽³⁰⁾



Εικόνα 7. K-πλησιέστερος γείτονας

1.7.7 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machine)

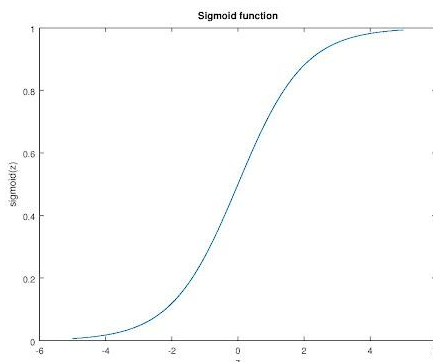
Οι μηχανές διανυσμάτων υποστήριξης (SVM) αναπτύχθηκαν την δεκαετία του 1990 από τον Vladimir N Vapnik και αποτελούν μοντέλα μηχανικής μάθησης με επίβλεψη. Ο αλγόριθμος SVM χρησιμοποιείται σε εργασίες ταξινόμησης και παλινδρόμησης όμως προτιμάται συνήθως σε προβλήματα ταξινόμησης. Στόχος του αλγορίθμου είναι να βρεθεί ένα βέλτιστο υπερεπίπεδο που διαχωρίζει ένα σύνολο δεδομένων σε διακριτές κλάσεις, το οποίο μεγαλώνει την απόσταση μεταξύ των πλησιέστερων σημείων δεδομένων των αντίθετων κλάσεων, βάζοντας στο μέλλον το νέο σημείο δεδομένων σε σωστή κατηγορία.⁽³¹⁾



Εικόνα 8. Μηχανές διανυσμάτων υποστήριξης

1.7.8 Λογιστική παλινδρόμηση (Logistic regression)

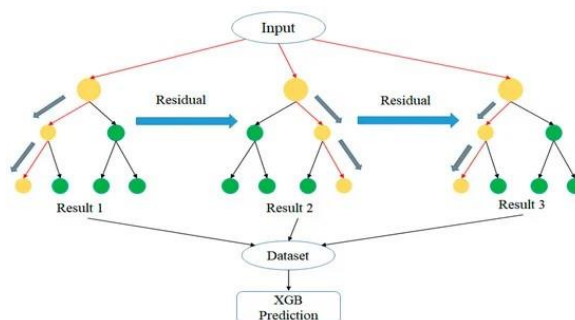
Η λογιστική παλινδρόμηση (LR) είναι ένας εποπτευόμενος αλγόριθμος ML, και παρόλο που πρόκειται για γραμμική παλινδρόμηση χρησιμοποιείται σε προβλήματα δυαδικής ταξινόμησης. Ουσιαστικά, κάνει χρήση μιας λογιστικής συνάρτησης για να μοντελοποιήσει μια δυαδική μεταβλητή εξόδου. Είναι κατάλληλος για περιπτώσεις στις οποίες οι ανεξάρτητες μεταβλητές είναι κατηγορικές ή συνεχείς και κατηγορικές, σε σύγκριση με την γραμμική παλινδρόμηση όπου η μεταβλητή εξόδου είναι αποκλειστικά ποσοτική.⁽³²⁾ Η λογιστική παλινδρόμηση χρησιμοποιεί μια λογιστική συνάρτηση γνωστή και ως σιγμοειδής συνάρτηση, η οποία σχηματίζεται με μια καμπύλη S με το εύρος της να περιορίζεται μεταξύ 0 και 1.⁽³³⁾



Εικόνα 9. Σιγμοειδής συνάρτηση λογιστικής παλινδρόμησης

1.7.9 XGBoost (Extreme Gradinet Boosting)

Το XGBoost είναι μια βιβλιοθήκη λογισμικού ανοιχτού κώδικα και έχει σχεδιαστεί για την αποτελεσματική εκπαίδευση μοντέλων μηχανικής μάθησης. Ο αλγόριθμος XGBoost αποτελεί μια μέθοδο εκμάθησης συνόλου στην οποία συνδυάζει τις προβλέψεις πολλών μεμονωμένων μοντέλων, συνήθως δέντρων αποφάσεων, για να κάνει μια πιο ισχυρή πρόβλεψη. Η ικανότητα του να χειρίζεται μεγάλα σύνολα δεδομένων και οι επιδόσεις του σε εργασίες ταξινόμησης και παλινδρόμησης τον καθιστούν έναν αποτελεσματικό αλγόριθμο και πιο ακριβή από έναν μεμονωμένο αλγόριθμο. Λόγω την αποτελεσματικής διαχείρισης των τιμών που λείπουν, μπορεί να χειρίζεται πραγματικά δεδομένα χωρίς να χρειάζεται σημαντική προεπεξεργασία. Επιπλέον περιλαμβάνει μια ενσωματωμένη υποστήριξη για παράλληλη επεξεργασία κανόνες τους υπολογισμούς πιο αποτελεσματικούς και την εκπαίδευση των μοντέλων σε πιο σύντομο χρονικό διάστημα. Το XGBoost κρίνεται ένας αλγόριθμος υψηλής απόδοσης πρόβλεψης και αρκετά ευέλικτος στη διαχείριση διαφορετικών συνόλων δεδομένων και χρησιμοποιείται σε διάφορους τομείς.⁽³⁴⁾



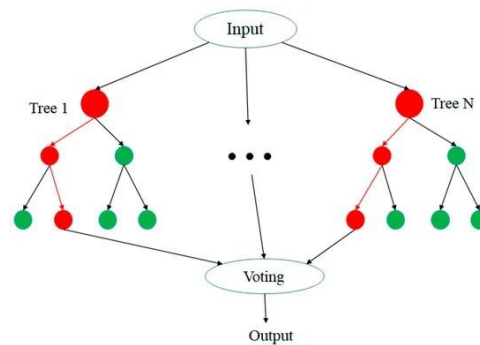
Εικόνα 10. Αρχιτεκτονική XGBoost

1.7.10 Στοχαστικός αλγόριθμος ενίσχυσης (Stochastic gradient Boost)

Ο στοχαστικός αλγόριθμος ενίσχυσης (SGB) αποτελεί μία μέθοδο ενισχυτικής μάθησης και χρησιμοποιείται για την κατασκευή μοντέλων πρόβλεψης. Η μέθοδος συνδυάζει την τεχνική ενίσχυσης με το στοχαστικό κύκλωμα, όπου σε κάθε επανάληψη του αλγορίθμου εκπαίδευσης χρησιμοποιείται μόνο ένα υποσύνολο της εκπαίδευσης. Αυτό κάνει τον στοχαστικό αλγόριθμο ταχύτερο στην εκπαίδευση και ιδιαίτερα αποτελεσματικό σε μεγάλα σύνολα δεδομένων κάνοντας ακριβείς προβλέψεις.^[10]

1.7.11 Τυχαίο Δάσος Απόφασης (Random Forest)

Το τυχαίο δάσος (RRF) βασίζεται στη μέθοδο εκμάθησης συνόλου και χρησιμοποιείται για ταξινόμηση και παλινδρόμηση κατασκευάζοντας ένα πλήθος δέντρων αποφάσεων κατά τη διάρκεια της εκπαίδευσης. Το κάθε δέντρο αποτελείται από ένα δείγμα εκκίνησης δεδομένων από ένα αρχικό σύνολο δεδομένων με τυχαία επιλογή δειγμάτων με αντικατάσταση. Αυτό βοηθάει στη μείωση της συσχέτισης μεταξύ των δέντρων απόφασης και συμβάλει στον μετριασμό της υπερβολικής προσαρμογής. Για εργασίες ταξινόμησης, η έξοδος του RF αφορά τη κλάση που επιλέγεται βάσει πλειοψηφίας από τα περισσότερα δέντρα ενώ για εργασίες παλινδρόμησης, οι προβλέψεις από κάθε δέντρο υπολογίζονται κατά μέσο όρο για να ληφθεί η τελική έξοδος. Παρόλο που θεωρείται μια χρονοβόρα και περίπλοκη διαδικασία η οποία απαιτεί περισσότερους πόρους, είναι ικανό στον χειρισμό συνόλων δεδομένων υψηλών διαστάσεων, στην καταγραφή σύνθετων σχέσεων και στην παροχή ισχυρών προβλέψεων. ⁽³⁵⁾



Εικόνα 11. Αρχιτεκτονική Τυχαίου Δάσους

Κεφάλαιο 2

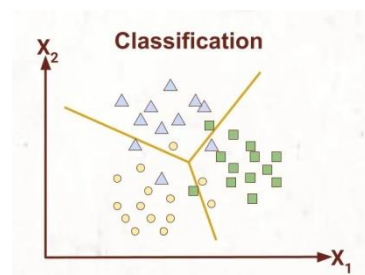
2.1 Εξόρυξη δεδομένων

Η εξόρυξη δεδομένων αποτελεί μια διαδικασία στην οποία γίνεται αναζήτηση και ανάλυση μεγάλων συνόλων δεδομένων ανακαλύπτοντας μοτίβα, συσχετισμούς και πολύτιμες πληροφορίες. Η διαδικασία ξεκινάμε τη συλλογή δεδομένων, όπου ποικίλες και σχετικές πληροφορίες προέρχονται από διάφορα κανάλια όπως βάσεις δεδομένων, αποθήκες δεδομένων και διαδικτυακά αποθετήρια. Μόλις συγκεντρωθούν, τα δεδομένα υποβάλλονται σε μία σχολαστική φάση καθαρισμού και προεπεξεργασίας, αντιμετωπίζοντας ζητήματα όπως σφάλματα, ασυνέπειες και τιμές που λείπουν για να διασφαλιστεί η ακρίβεια και η αξιοπιστία της επακόλουθης ανάλυσης. ^{[1](1)} Η Διερευνητική Ανάλυση Δεδομένων (EDA) χρησιμοποιεί οπτικές και στατιστικές τεχνικές για να εμβαθύνει στις περιπλοκές των δεδομένων, ανακαλύπτοντας τα χαρακτηριστικά των υποκειμένων και τις πιθανές σχέσεις μεταξύ των μεταβλητών. ^[2] Η επιλογή χαρακτηριστικών είναι ένα κρίσιμο βήμα, για τον εντοπισμό των πιο επιδραστικών μεταβλητών που συμβάλλουν σημαντικά στην ανάλυση και τη δημιουργία μοντέλων, ενώ παράλληλα μειώνουν τις διαστάσεις. Το πιο σημαντικό κομμάτι της εξόρυξης δεδομένων, βρίσκεται στη δημιουργία μοντέλων, όπου εφαρμόζεται μια ποικιλία αλγορίθμων και στατιστικών μοντέλων. Τα δέντρα αποφάσεων, οι τεχνικές ομαδοποίησης, η εξόρυξη κανόνων συσχέτισης, τα νευρωνικά δίκτυα και η ανάλυση παλινδρόμησης είναι μεταξύ των μεθόδων που χρησιμοποιούνται συνήθως. Αυτοί οι αλγόριθμοι εξετάζουν τα δεδομένα εντοπίζοντας μοτίβα και σχέσεις που μπορεί να μην είναι άμεσα εμφανή. Ακολουθεί η αξιολόγηση όπου η απόδοση των μοντέλων αξιολογείται χρησιμοποιώντας τις μετρήσεις accuracy, precision, recall και F1score. Η αυστηρή αξιολόγηση διασφαλίζει την αξιοπιστία και την αποτελεσματικότητα των γνώσεων που προέρχονται από τα δεδομένα. Μετά την επιτυχή επικύρωση του μοντέλου, η διαδικασία εξόρυξης δεδομένων εισέρχεται στη φάση ανακάλυψης γνώσης. Εδώ, εξάγονται σημαντικά μοτίβα και γνώσεις, που κυμαίνονται από τον εντοπισμό των τάσεων και συσχετισμών έως την αποκάλυψη ανωμαλιών και ακραίων στοιχείων. Η ερμηνεία αυτών των ευρημάτων είναι ζωτικής σημασίας, καθώς περιλαμβάνει τη μετάφραση των ανακαλυφθέντων μοτίβων σε γνώση που μπορεί να γίνει πράξη που μπορεί να ενημερώσει τις διαδικασίες λήψης αποφάσεων.^[4] Η εξόρυξη δεδομένων έχει βρει ευρείες εφαρμογές σε διάφορους κλάδους. Στα χρηματοοικονομικά, βοηθά στον εντοπισμό της απάτης και στην αξιολόγηση κινδύνου. Στην υγειονομική περίθαλψη, υποστηρίζει τη λήψη κλινικών αποφάσεων και την ανάλυση των αποτελεσμάτων των ασθενών. Επίσης στο μάρκετινγκ, καθώς επιτρέπει την τμηματοποίηση των πελατών και τη στοχευόμενη διαφήμιση. Οι τηλεπικοινωνίες και το ηλεκτρονικό εμπόριο επίσης επωφελοούνται από την εξόρυξη δεδομένων μέσω βελτιωμένης και βελτιστοποίησης δικτύου και εξατομικευμένων συστάσεων. ^[3] Ωστόσο, οι ηθικοί προβληματισμοί και οι ανησυχίες για το απόρρητο είναι πρωταρχικής σημασίας, ειδικά όταν πρόκειται για μεγάλα σύνολα δεδομένων που περιέχουν ευαίσθητες πληροφορίες. Η εξερεύνηση της ισορροπίας μεταξύ την εξαγωγής πολύτιμων πληροφοριών και της προστασίας μεταξύ του ατομικού απορρήτου είναι ζωτικής σημασίας για τις ευθύνες και ηθικές πρακτικές εξόρυξης δεδομένων. ⁽¹⁾

2.2 Τεχνικές διαχείρισης δεδομένων

2.2.1 Κατηγοριοποίηση / Ταξινόμηση (Classification)

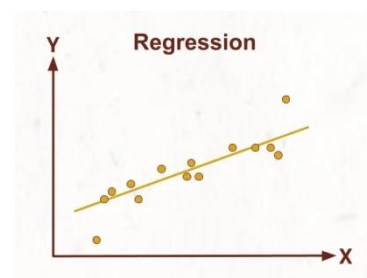
Η κατηγοριοποίηση ή αλλιώς ταξινόμηση αφορά μέθοδο μάθησης με επίβλεψη και είναι μία από τις πιο βασικές τεχνικές της μηχανικής μάθησης. Η διαδικασία της κατηγοριοποίησης αφορά ένα σύνολο παρατηρήσεων που γίνονται από τα χαρακτηριστικά (features) ή τις ιδιότητες (attributes) που περιλαμβάνονται σε ένα σύνολο κλάσεων που έχει καθοριστεί εκ των προτέρων. Στην συνέχεια χρησιμοποιείται ένα σύνολο εκπαίδευσης (training set) με γνωστές κλάσεις για να εκπαιδευτεί ο αλγόριθμος και να κατασκευαστεί ένα μοντέλο που θα ταξινομεί νέες εισόδους στις αντίστοιχες κλάσεις.⁽²⁾



Εικόνα 12. Διάγραμμα ταξινόμησης

2.2.2 Παλινδρόμηση (Regression)

Η παλινδρόμηση είναι μια μέθοδος παρόμοια με την κατηγοριοποίηση και χρησιμοποιείται στη μάθηση με επίβλεψη σε προβλήματα στα οποία η μεταβλητή εξόδου δέχεται συνεχόμενες τιμές. Ο αλγόριθμος εξετάζει την σχέση της εξαρτώμενης μεταβλητής ή αλλιώς της μεταβλητής στόχου με την ανεξάρτητη μεταβλητή, την μεταβλητή πρόβλεψης σε ένα σύνολο δεδομένων με σκοπό να κατασκευαστεί ένας μηχανισμός υπολογισμού αριθμητικών τιμών.⁽³⁾



Εικόνα 2. Διάγραμμα παλινδρόμησης

2.2.3 Συσταδοποίηση (Clustering)

Η συσταδοποίηση ή αλλιώς ομαδοποίηση αποτελεί μια προσέγγιση μη επιβλεπόμενης μάθησης κατά την οποία ένα σύνολο "αντικειμένων" διακρίνεται σε ένα σύνολο από λογικές ομάδες. Ο όρος ομοιότητα αφορά την καταχώρηση αντικειμένων του συνόλου δεδομένων σε ίδια ομάδα ενώ αντίστροφα, ο όρος ανομοιότητα αφορά αντικείμενα που ανήκουν σε διαφορετική ομάδα. Η Στόχος των αλγορίθμων συσταδοποίησης είναι να μεγιστοποιήσουν την ομοιότητα εντός των



Εικόνα 13. Διάγραμμα Συσταδοποίησης

ομάδων και την ανομοιότητα μεταξύ των ομάδων. Η διαφορά μεταξύ συσταδοποίησης και ταξινόμησης έγκειται στο γεγονός ότι η ταξινόμηση χρησιμοποιεί προκαθορισμένες κλάσεις στις οποίες εκχωρούνται αντικείμενα, ενώ η ομαδοποίηση ορίζει ομοιότητες μεταξύ αντικειμένων, τα οποία ομαδοποιεί σύμφωνα με κοινά χαρακτηριστικά, και τα διαφοροποιεί από άλλες συστάδες αντικειμένων. Η συσταδοποίηση μπορεί να εφαρμοστεί αποτελεσματικά για την τμηματοποίηση αγοράς παρέχοντας πιο εξατομικευμένο περιεχόμενο.⁽⁴⁾

Ορισμός

Δοθέντος ενός συνόλου διανυσμάτων $X = \{x_1, x_2, x_3, \dots, x_n\}$ ζητούνται m σύνολα-ομάδες $C_1, C_2, \dots, C_m$, $C_1, C_2, \dots, C_m$ με $m \ll n$ έτσι ώστε :

$$C_i \neq \emptyset = 1, 2, 3 \dots m$$

οι m ομάδες αποτελούν διαμέριση του συνόλου X .

Ο ορισμός αυτός αναφέρεται στην αυστηρή συσταδοποίηση διότι κάθε διάνυσμα ανήκει σε μία και μόνο ομάδα. Εναλλακτικά μπορεί να οριστεί η ασαφής συσταδοποίηση. Κάνοντας χρήση των ασαφών συνόλων μπορούμε να ορίσουμε m συναρτήσεις συμμετοχής $u_j: X \rightarrow [0, 1]$ για $j=1, 2, \dots, m$. Οι συναρτήσεις $u(\cdot)$ ποσοτικοποιούν την βεβαιότητα που έχουμε για το αν κάποιο διάνυσμα i ανήκει στην ομάδα j .



Εικόνα 14. Διαδικασία συσταδοποίησης

2.3 Μετρικές Αξιολόγησης Ταξινόμησης ή παλινδρόμησης

Accuracy (Ακρίβεια)

Η μετρική Accuracy χρησιμοποιείται για την αξιολόγηση μοντέλων ταξινόμησης και μετρά το ποσοστό των σωστών προβλέψεων. Η ακρίβεια είναι η αναλογία των σωστών προβλέψεων ως προς τον αριθμό όλων των περιπτώσεων και ορίζεται από τον τύπο:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- TP (True Positive) : Σωστές θετικές προβλέψεις
- TN (True Negative) : Σωστές αρνητικές προβλέψεις
- FP (False Positive) : Λανθασμένες θετικές προβλέψεις
- FN (False Negative) : Λανθασμένες αρνητικές προβλέψεις
- Σε περιπτώσεις που παρουσιάζεται ανισορροπία ή μεγάλη διαφορά μεταξύ των θετικών και των αρνητικών προβλέψεων είναι πιθανό να παρουσιαστεί μια αδυναμία στην απόδοση με αποτέλεσμα το συμπέρασμα να είναι αναξιόπιστο. Έτσι κρίνεται αναγκαία η χρήση ενός πίνακα σύγχυσης (Confusion Matrix) ο οποίος οπτικοποιεί την απόδοση του αλγορίθμου αντιμετωπίζοντας την

ανομοιομορφία για να επιτύχει ένα μεγαλύτερο ποσοστό απόδοσης και ένα ασφαλές συμπέρασμα.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Εικόνα 15. Πίνακας σύγχυσης (Confusion Matrix)

Precision

Η μετρική Precision εξετάζει πόσο ακριβής είναι μια πρόβλεψη και ορίζεται ως ο αριθμός των σωστών θετικών προβλέψεων προς το σύνολο των προβλεπόμενων θετικών.

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{Total\ Predictive\ positive}$$

Recall (Ανάκληση)

Στην συγκεκριμένη μέθοδο υπολογίζεται πόσες από τις πραγματικές θετικές περιπτώσεις θεωρήθηκαν θετικές από το μοντέλο πρόβλεψης και ορίζεται ως ο αριθμός των σωστών θετικών προβλέψεων προς το σύνολο των πραγματικών θετικών.

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{Total\ actual\ positive}$$

F1-score

Η βαθμολογία F1 συνδυάζει τις μετρήσεις Precision και Recall σε περιπτώσεις που παρουσιάζεται ανισορροπία μεταξύ των δεδομένων και υπάρχουν πολλές πραγματικά αρνητικές τιμές. Όταν η μετρική Precision είναι ίση με την μετρική Recall η βαθμολογία F1 παίρνει την μέγιστη τιμή.

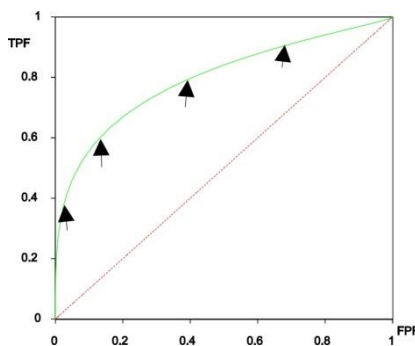
Ορίζεται από τον τύπο:

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Καμπύλη ROC

Μια καμπύλη ROC αναπαριστάται σε ένα γράφημα και δείχνει την απόδοση ενός μοντέλου ταξινόμησης ως προς το :

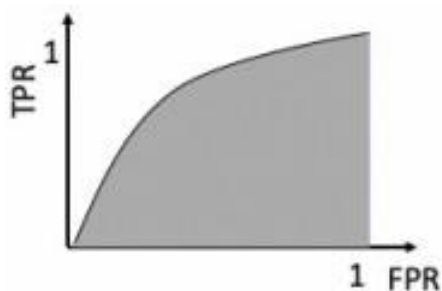
Αληθινό θετικό ποσοστό: $(TPR) = \frac{TP}{TP+FN}$, Ψευδώς θετικό ποσοστό: $(FPR) = \frac{FP}{FP+TN}$



Εικόνα 16. Καμπύλη ROC

AUC

Το AUC βρίσκεται κάτω από την καμπύλη ROC και μετρά ολόκληρη την περιοχή κάτω από την καμπύλη. Όσο μεγαλύτερη είναι η AUC τόσο καλύτερη απόδοση παρουσιάζει ένα μοντέλο μεταξύ θετικών και αρνητικών κλάσεων. ⁽⁵⁾

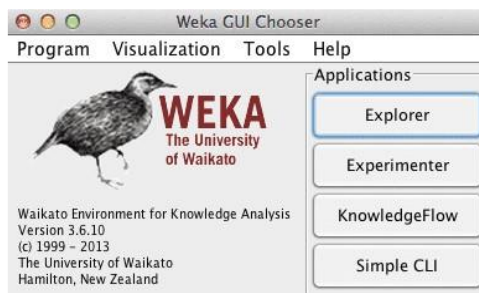


Εικόνα 17. Καμπύλη AUC

- Όταν $AUC = 1$, ο ταξινομητής διακρίνει τέλεια τα θετικά και αρνητικά σημεία κλάσης.
- Όταν $AUC = 0$, ο ταξινομητής προβλέπει τα αρνητικά ως θετικά και τα θετικά ως αρνητικά.
- Όταν $AUC = 0,5$, ο ταξινομητής δεν μπορεί να διακρίνει μεταξύ θετικών και αρνητικών κλάσεων. ⁽⁶⁾

2.4 WEKA

Το πρόγραμμα WEKA αναπτύχθηκε στο πανεπιστήμιο του Waikato της Νέας Ζηλανδίας και αποτελεί ένα δωρεάν λογισμικό μηχανικής μάθησης και ανάλυσης δεδομένων γραμμένο σε Java. Διαθέτει ένα σύνολο από αλγορίθμους μηχανικής μάθησης για την εξόρυξη της γνώσης από δεδομένα και τη μέθοδο πρόβλεψης.^[11] Πιο συγκεκριμένα, περιλαμβάνει μια συλλογή εργαλείων για προεπεξεργασία δεδομένων, ταξινόμηση, ομαδοποίηση, παλινδρόμηση, οπτικοποίηση και επιλογή χαρακτηριστικών. Στηρίζεται στην υπόθεση ότι τα δεδομένα είναι διαθέσιμα ως ένα ενιαίο αρχείο και περιγράφονται από ένα σταθερό αριθμό χαρακτηριστικών. Επίσης, υποστηρίζει αρχεία της μορφής .arff και παρέχει πρόσβαση σε βάσεις δεδομένων SQL χρησιμοποιώντας τη συνδεσιμότητα των βάσεων δεδομένων της Java διαθέτοντας την ικανότητα να επεξεργαστεί το αποτέλεσμα που επιστρέφεται. Θεωρείται εύκολο στη χρήση λόγω των γραφικώς διεπαφών χρήστη και της δυνατότητάς του να λειτουργεί σε διάφορες πλατφόρμες υπολογιστών, με την προϋπόθεση ότι υλοποιείται σε Java.⁽⁸⁾



Εικόνα 18. Σύστημα Weka

2.5 Εξόρυξη δεδομένων στο μάρκετινγκ

Στον τομέα του μάρκετινγκ, η εξόρυξη δεδομένων αξιοποιεί αλγορίθμους μηχανικής μάθησης για την εξαγωγή πρακτικών πληροφοριών από τεράστια σύνολα δεδομένων, συμβάλλοντας σε πιο αποτελεσματικές και στοχευόμενες στρατηγικές. Μια κοινή εφαρμογή είναι η τμηματοποίηση πελατών, όπου οι αλγόριθμοι μηχανικής μάθησης αναλύουν τη συμπεριφορά, τις προτιμήσεις και τα δημογραφικά στοιχεία των πελατών για να ομαδοποιήσουν άτομα με παρόμοια χαρακτηριστικά. Οι αλγόριθμοι ομαδοποίησης, όπως το K-means, βοηθούν στον εντοπισμό διακριτών τμημάτων πελατών, επιτρέποντας στους επαγγελματίες του μάρκετινγκ να προσαρμόσουν τις καμπάνιες τους σε συγκεκριμένο κοινό. Μία άλλη πολύτιμη χρήση της μηχανικής μάθησης στο μάρκετινγκ είναι η προγνωστική μοντελοποίηση. Αλγόριθμοι όπως η ανάλυση παλινδρόμησης και τα δέντρα αποφάσεων μπορούν να προβλέψουν τη συμπεριφορά των πελατών, όπως μοτίβα αγορών ή απόκριση σε καμπάνιες μάρκετινγκ. Αυτό επιτρέπει στους επαγγελματίες του μάρκετινγκ να βελτιστοποιούν τις στρατηγικές τους, να κατανέμουν αποτελεσματικά τους πόρους και να εξατομικεύουν τις αλληλεπιδράσεις με βάση τις προβλεπόμενες προτιμήσεις των πελατών. Συστήματα συστάσεων που

τροφοδοτούνται από μηχανική μάθηση κυριαρχούν και στο μάρκετινγκ. Το συνεργατικό φιλτράρισμα και οι αλγόριθμοι συστάσεων βάσει περιεχομένου αναλύουν τις προτιμήσεις και τις συμπεριφορές των πελατών για να προτείνουν προϊόντα ή υπηρεσίες που ευθυγραμμίζονται με τις μεμονωμένες προτιμήσεις. Αυτό βελτιώνει την εμπειρία του πελάτη, αυξάνει την αφοσίωση και συχνά οδηγεί σε υψηλότερα ποσοστά μετατροπών. Επιπλέον, οι αλγόριθμοι μηχανικής μάθησης διαδραματίζουν κρίσιμο ρόλο στην πρόβλεψη της εκτροπής. Αναλύοντας ιστορικά δεδομένα και αλληλεπιδράσεις με τους πελάτες, τα μοντέλα μπορούν να προβλέψουν την πιθανότητα ανατροπής ενός πελάτη, επιτρέποντας στους επαγγελματίες του μάρκετινγκ να εφαρμόσουν στοχευόμενες στρατηγικές διατήρησης. Αυτή η προληπτική προσέγγιση βοηθά στις προσπάθειες διατήρησης πελατών, μειώνοντας τελικά τα ποσοστά φθοράς των πελατών. Η ανάλυση συναισθήματος, μια εφαρμογή επεξεργασίας φυσικής γλώσσας (NLP), είναι άλλο ένα ισχυρό εργαλείο στο μάρκετινγκ. Τα μοντέλα μηχανικής μάθησης αναλύουν δεδομένα κειμένου από πηγές όπως μέσα κοινωνικής δικτύωσης, κριτικές πελατών και σχόλια για να μετρήσουν το συναίσθημα. Οι έμποροι μπορούν στη συνέχεια να προσαρμόσουν τις στρατηγικές τους με βάση την αντίληψη του κοινού για την επωνυμία ή τα προϊόντα τους, εντοπίζοντας τομείς προς βελτίωση ή αξιοποιώντας το θετικό συναίσθημα. Συνοπτικά, η εξόρυξη δεδομένων στο μάρκετινγκ χρησιμοποιώντας αλγορίθμους ML δίνει τη δυνατότητα στις επιχειρήσεις να κατανοούν καλύτερα τους πελάτες τους, να προβλέπουν συμπεριφορές, να εξατομικεύουν τις αλληλεπιδράσεις και να βελτιστοποιούν στρατηγικές μάρκετινγκ. Η εφαρμογή αυτών των τεχνικών έχει ως αποτέλεσμα πιο αποτελεσματικές και στοχευόμενες καμπάνιες συμβάλλοντας τελικά στην αύξηση της ικανοποίησης των πελατών και στη βελτιωμένη απόδοση επένδυσης (ROI) για πρωτοβουλίες μάρκετινγκ. ^[14]

2.6 Ηθική χρήση δεδομένων

Στο δυναμικό τοπίο του χειρισμού των δεδομένων πελατών, οι περιπλοκές του απορρήτου των δεδομένων και οι ηθικοί παράγοντες παρουσιάζουν προκλήσεις που απαιτούν διαφοροποιημένες προσεγγίσεις. Μία από τις κύριες προκλήσεις έγκειται στην εξασφάλιση ενημερωμένης συναίνεσης, όπου η πολυπλοκότητα των πρακτικών συλλογής δεδομένων μπορεί να διαφεύγει από την κατανόηση των χρηστών. Η ηθική ευθύνη απαιτεί διαφάνεια, υποχρεώνοντας τους οργανισμούς να διατυπώνουν, με σαφείς όρους, τους τύπους των δεδομένων που συλλέγονται και τους επιδιωκόμενους σκοπούς, διασφαλίζοντας ότι οι χρήστες μπορούν να λαμβάνουν τεκμηριωμένες αποφάσεις. Η διασφάλιση της ασφάλειας των δεδομένων των πελατών είναι μια διαρκής πρόκληση, με τη διαρκώς διαφαινόμενη απειλή παραβίασης δεδομένων και κυβερνοεπιθέσεων. Δεοντολογικά, οι οργανισμοί έχουν το καθήκον να εφαρμόζουν αυστηρά μέτρα ασφαλείας, προστατεύοντας τις πληροφορίες των πελατών από μη εξουσιοδοτημένη πρόσβαση. Η επίτευξη ισορροπίας μεταξύ της συλλογής βασικών δεδομένων και της ελαχιστοποίησης της εισβολής είναι μια διαρκής πρόκληση, η οποία απαιτεί μια λεπτή βαθμονόμηση των πρακτικών συλλογής δεδομένων για τον σεβασμό του απορρήτου των χρηστών. Η ακρίβεια των δεδομένων, μια άλλη βασική ανησυχία, απαιτεί συνεχή επαγρύπνηση για τη διατήρηση ενημερωμένων και αξιόπιστων πληροφοριών. Η ηθική υποχρέωση εδώ είναι ο μετριασμός των κινδύνων που σχετίζονται με παρωχημένα ή ανακριβή δεδομένα, τα οποία θα

μπορούσαν να οδηγήσουν σε εσφαλμένη λήψη αποφάσεων. Η συνεργασία με τρίτους παρόχους υπηρεσιών εισάγει πρόσθετα επίπεδα πολυπλοκότητας. Οι ηθικοί λόγοι απαιτούν προσεκτικό έλεγχο αυτών των συνεργατών για να διασφαλιστεί η ευθυγράμμιση με τα πρότυπα απορρήτου και να αποτραπεί η ακούσια κατάχρηση δεδομένων. Ο προσδιορισμός της ιδιοκτησίας των δεδομένων δεν αποτελεί απλώς νομικό κριτήριο, αλλά και ηθική επιταγή. Απαιτούνται σαφείς πολιτικές και συμφωνίες για την πλοήγηση στα συχνά θολά νερά της ιδιοκτησίας και των δικαιωμάτων χρήσης δεδομένων. Η ηθική πρόκληση της απόφασης για τη διάρκεια της διατήρησης δεδομένων περιλαμβάνει την επίτευξη ισορροπίας μεταξύ της χρησιμότητας και των πιθανών κινδύνων της παρατεταμένης αποθήκευσης, υπογραμμίζοντας την ανάγκη για προσεκτικές πρακτικές διαχείρισης δεδομένων. Η εισαγωγή αλγορίθμων στις διαδικασίες λήψης αποφάσεων φέρνει στο προσκήνιο ηθικούς προβληματισμούς που σχετίζονται με την προκατάληψη. Οι οργανισμοί είναι δεοντολογικά δεσμευμένοι να εντοπίζουν και να διορθώνουν προκαταλήψεις στους αλγόριθμους για να διασφαλίσουν τη δικαιοσύνη και να αποτρέψουν τα μεροληπτικά αποτελέσματα. Η πλοήγηση στο παγκόσμιο ρυθμιστικό τοπίο, γεμάτο με ποικίλους και εξελισσόμενους νόμους περί προστασίας δεδομένων, αποτελεί μια άλλη πρόκληση. Οι οργανισμοί πρέπει να παραμείνουν ενήμεροι με αυτούς τους κανονισμούς, συμμορφώνοντας τις νομικές απαιτήσεις, ενώ παράλληλα δίνουν προτεραιότητα στις πρακτικές δεοντολογικών δεδομένων. Η λογοδοσία και η διαφάνεια αποτελούν το θεμέλιο του δεοντολογικού χειρισμού δεδομένων. Η καλλιέργεια μιας κουλτούρας ευθύνης εντός των οργανισμών είναι απαραίτητη, διασφαλίζοντας ότι φέρουν την ευθύνη για τις πρακτικές δεδομένων τους. Οι ηθικοί λόγοι υπαγορεύουν το άνοιγμα στις ερωτήσεις των πελατών, ενισχύοντας τη διαφάνεια σχετικά με τη χρήση δεδομένων και τις διαδικασίες χειρισμού. Με την επιτυχή πλοήγηση σε αυτές τις πολύπλευρες προκλήσεις, οι οργανισμοί όχι μόνο προστατεύουν την εμπιστοσύνη των πελατών αλλά συμβάλλουν επίσης στη δημιουργία ενός υπεύθυνου και βιώσιμου οικοσυστήματος δεδομένων, όπου το απόρρητο και οι ηθικές αρχές τηρούνται ως πρωταρχικής σημασίας.^[15]

2.7 Ηθική χρήση μηχανικής μάθησης

Η διασφάλιση της ηθικής χρήσης της μηχανικής μάθησης στο μάρκετινγκ περιλαμβάνει μια πολύπλευρη προσέγγιση για την οικοδόμηση εμπιστοσύνης και την ιεράρχηση του απορρήτου των χρηστών. Η διαφάνεια είναι θεμελιώδης για αυτήν την προσπάθεια, απαιτώντας ολοκληρωμένες στρατηγικές επικοινωνίας που διευκρινίζουν την ενσωμάτωση της μηχανικής μάθησης στις πρακτικές μάρκετινγκ. Αυτό περιλαμβάνει τη λεπτομέρεια των συγκεκριμένων αλγορίθμων που χρησιμοποιούνται, τη λογική πίσω από τη χρήση τους και τις πιθανές επιπτώσεις για τις εμπειρίες των χρηστών. Η ενημερωμένη συναίνεση είναι μια αρχή ακρογωνιαίου λίθου, η οποία απαιτεί από τους οργανισμούς να λαμβάνουν ρητή άδεια από τους χρήστες πριν συλλέξουν και αξιοποιήσουν τα δεδομένα τους για σκοπούς μάρκετινγκ. Αυτό περιλαμβάνει όχι μόνο τη σαφή άρθρωση των τύπων δεδομένων που συλλέγονται, αλλά και την εξήγηση του τρόπου με τον οποίο θα χρησιμοποιηθούν αυτά τα δεδομένα, με τους χρήστες να έχουν την αυτονομία να εξαιρεθούν εάν το επιλέξουν.

Η στρατηγική τήρηση των κανονισμών προστασίας δεδομένων, όπως ο GDPR, είναι αδιαπραγμάτευτη. Αυτό συνεπάγεται όχι μόνο τον ασφαλή χειρισμό των δεδομένων πελατών, αλλά και την εφαρμογή μέτρων για την προστασία από μη εξουσιοδοτημένη πρόσβαση, διασφαλίζοντας τα υψηλότερα πρότυπα ασφάλειας δεδομένων. Μια ισχυρή στρατηγική μετριασμού της μεροληψίας είναι επιτακτική για την αντιμετώπιση πιθανών προκαταλήψεων που ενσωματώνονται στα μοντέλα μηχανικής μάθησης. Οι τακτικοί έλεγχοι, που περιλαμβάνουν ενδεδειγμένες αναθεωρήσεις πηγών δεδομένων, αλγοριθμικές διαδικασίες λήψης αποφάσεων και τη συνολική απόδοση της καμπάνιας, είναι ζωτικής σημασίας για τον εντοπισμό και τη διόρθωση τυχόν προκαταλήψεων ή ηθικών ανησυχιών που μπορεί να προκύψουν. Η καθιέρωση λογοδοσίας εντός του οργανισμού είναι πρωταρχικής σημασίας, με σαφώς καθορισμένους ρόλους και ευθύνες για τα άτομα που εμπλέκονται στην ανάπτυξη και την ανάπτυξη της μηχανικής μάθησης στο μάρκετινγκ. Θα πρέπει να εφαρμόζονται τακτικές πρακτικές παρακολούθησης και ελέγχου για να διασφαλίζεται η συνεχής συμμόρφωση με τις δεοντολογικές κατευθυντήριες γραμμές. Η ανθρώπινη εποπτεία στις διαδικασίες λήψης αποφάσεων είναι ένα συμπληρωματικό μέτρο, το οποίο διασφαλίζει ότι τα αυτοματοποιημένα συστήματα συμπληρώνονται από ανθρώπινη τεχνογνωσία, ειδικά σε κρίσιμα σενάρια. Η εκπαίδευση των πελατών σχετικά με τα οφέλη του εξατομικευμένου μάρκετινγκ, ενώ αντιμετωπίζει προληπτικά τις ανησυχίες σχετικά με το απόρρητο και την ασφάλεια, αποτελεί αναπόσπαστο μέρος μιας στρατηγικής ηθικού μάρκετινγκ. Η εφαρμογή ισχυρών μέτρων ασφαλείας, όπως η κρυπτογράφηση και η ασφαλής αποθήκευση, είναι απαραίτητη για την προστασία των δεδομένων πελατών από μη εξουσιοδοτημένη πρόσβαση. Επιπλέον, λαμβάνοντας υπόψη τον περιβαλλοντικό αντίκτυπο των διαδικασιών μηχανικής μάθησης και τη βελτιστοποίηση αλγορίθμων και υποδομών για ενεργειακή απόδοση συμβάλλουν στη συνολική βιωσιμότητα. Η προσπάθεια για δικαιοσύνη και συμπερίληψη είναι ένα προληπτικό μέτρο, που διασφαλίζει ότι το μάρκετινγκ οι καμπάνιες δεν επηρεάζουν δυσανάλογα συγκεκριμένες δημογραφικές ομάδες. Αυτό περιλαμβάνει προσεκτική εξέταση των πιθανών προκαταλήψεων τόσο στα δεδομένα όσο και στους αλγόριθμους να δημιουργήσετε καμπάνιες που να περιλαμβάνουν αποκλεισμούς και να σέβονται τη διαφορετικότητα.^[15]

Συνοπτικά, περιλαμβάνει ένα ολοκληρωμένο ηθικό πλαίσιο το οποίο αποτελείται από:

- Διαφάνεια
- Ενημερωμένη συναίνεση
- Απόρρητο δεδομένων
- Μετριασμός μεροληψίας
- Λογοδοσία
- Εκπαίδευση πελατών
- Τακτικοί έλεγχοι
- Μέτρα ασφαλείας
- Ανθρώπινη επίβλεψη
- Βιωσιμότητα

- Δικαιοσύνη

Και δημιουργεί μια ηθική και υπεύθυνη προσέγγιση της μηχανικής μάθησης στο μάρκετινγκ.

2.8 Python

Η Python αποτελεί μια από τις πιο δημοφιλείς γλώσσες προγραμματισμού στην επιστήμη των δεδομένων, η οποία χρησιμοποιεί έναν μεγάλο αριθμό χρήσιμων βιβλιοθηκών εκ των οποίων κάποιες είναι ιδανικές για αλγορίθμους μηχανικής μάθησης.⁽⁹⁾

2.9 Βιβλιοθήκες Python

NumPy

Η βιβλιοθήκη NumPy δημιουργήθηκε από τον Travis Oliphant το 2005 και θεωρείται το βασικότερο θεμέλιο για την κατασκευή των εργαλείων στην Python. Παρέχει πλήρη υποστήριξη για τη διαχείριση πολυδιάστατων πινάκων καθώς επίσης παρέχει χρήσιμες αριθμητικές πράξεις και συναρτήσεις υψηλού επιπέδου οι οποίες μπορούν να εφαρμοστούν αποτελεσματικά σε πίνακες ολόκληρων δεδομένων. Πολύ σημαντικό είναι ότι τα δεδομένα μεταφέρονται εύκολα σε εξωτερικές βιβλιοθήκες χαμηλού επιπέδου και επιστρέφουν δεδομένα Numpy Arrays.

Pandas

Η βιβλιοθήκη ανοιχτού κώδικα Pandas δημιουργήθηκε το 2008 από τον Wes McKinney και είναι βασισμένη στη βιβλιοθήκη NumPy. Χρησιμοποιείται για εργασίες με σύνολα δεδομένων και αποτελεί ένα απαραίτητο εργαλείο για την εισαγωγή την διαχείριση και την ανάλυση των δεδομένων σε εφαρμογές που έχουν αναπτυχθεί με Python. Περιέχει ευέλικτες δομές δεδομένων και εργαλεία που αναλύουν δεδομένα με ευκολία και ταχύτητα. Για την αποθήκευση των δεδομένων χρησιμοποιείται το Data Frame το οποίο έχει έτοιμες συναρτήσεις ιδανικές για ανάκτηση και αποθήκευση δεδομένων.

Scikit-learn

Η βιβλιοθήκη ελεύθερου λογισμού Scikit-learn χρησιμοποιείται για αλγορίθμους μηχανικής μάθησης και παρέχει εργαλεία μετρήσεων απόδοσης και μεθόδους αξιολόγησης αποδοτικότητας. Διαθέτει εργαλεία για εξόρυξη δεδομένων, εργαλεία προεπεξεργασίας των δεδομένων αλλά και ανάλυσης. Αποτελεί μια απαραίτητη και χρήσιμη βιβλιοθήκη καθώς περιλαμβάνει εργαλεία εκμάθησης μηχανών για την επίλυση προβλημάτων ταξινόμησης, ομαδοποίησης και παλινδρόμησης καθώς επίσης και για την επιλογή του βέλτιστου μοντέλου.

NLTK (Natural Language Toolkit)

Η βιβλιοθήκη ανοιχτού κώδικα NLTK δημιουργήθηκε το 2001 από τους Edward Loper, Steven Bird και Ewan Klein. Θεωρείται το πιο δημοφιλές εργαλείο επεξεργασίας φυσικής γλώσσας (NLP) το οποίο περιλαμβάνει πάνω από 50 λεξικά και είναι πολύ εύκολο στην εκμάθηση. Διαθέτει ένα πλήθος βιβλιοθηκών για την ανάλυση της φυσικής γλώσσας καθώς επίσης και εργαλεία επεξεργασίας κειμένου και γραφικής απεικόνισης των δεδομένων.⁽¹⁰⁾

2.10 Επεξεργασία φυσικής γλώσσας (NLP)

Η επεξεργασία φυσικής γλώσσας (National Language Processing ή NLP) αποτελεί μια διαδικασία η οποία μπορεί να επεξεργάζεται και να παράγει την φυσική γλώσσα μέσω Η/Υ. Δεν βοηθά απλά τις μηχανές να αναγνωρίζουν τη γλώσσα αλλά ακόμη και να αλληλεπιδρούν με αυτήν (π.χ Chatbot). Χάρη στη μηχανική μάθηση οι υπολογιστές είναι σε θέση να αναγνωρίζουν μοτίβα, να προσδιορίζουν συναισθηματικά στοιχεία και να καταλήγουν σε ένα συμπέρασμα με βάση το περιεχόμενο ενός κειμένου.

Στάδια επεξεργασίας

Tokenization: Αποτελεί μια διαδικασία που επεξεργάζεται και διαχωρίζει ένα τμήμα κειμένου όπως μια φράση ή μια παράγραφος σε ξεχωριστές λέξεις που ονομάζονται tokens. Για παράδειγμα, η φράση "I like this product" μετατρέπεται σε "(I, like, this, product)". Με αυτό τον τρόπο ο υπολογιστής κατανοεί και επεξεργάζεται την ανθρώπινη γλώσσα.⁽¹¹⁾

Stop-words: Είναι μια διεργασία κατά την οποία διαγράφονται κάποιες λέξεις όπως άρθρα, προθέσεις, αντωνυμίες κ.λπ. από ένα κείμενο καθώς δεν έχουν κάποια ιδιαίτερη σημασία και δεν προσθέτουν επιπλέον πληροφορίες στο κείμενο. Για παράδειγμα "the", "an", "so" κ.λπ. Με την αφαίρεση των λέξεων μειώνεται το μέγεθος του συνόλου δεδομένων, επομένως και ο χρόνος εκπαίδευσης.⁽¹²⁾

Lemmatization: Αφορά μια τεχνική προεπεξεργασίας κειμένου στην οποία μετατρέπεται μία λέξη στην αρχική της μορφή βρίσκοντας την ρίζα από την οποία προέρχεται. Για παράδειγμα οι λέξεις " learner" και "learning" θα αντικατασταθούν με τη ρίζα "learn".⁽¹³⁾

Stemming: Αφορά επίσης μια τεχνική στην οποία μετατρέπεται μια λέξη με την διαφορά ότι διατηρείται η ρίζα και αφαιρούνται κάποιες καταλήξεις. Για παράδειγμα οι λέξεις "applies" και "applying" θα μετατραπούν στις λέξεις "appli" και "apply" αντίστοιχα.⁽¹⁴⁾

Part of speech tagging (POS tagging): Η τεχνική αυτή κατηγοριοποιεί τις λέξεις αναλόγως τη μορφή τους και το μέρος του λόγου στο οποίο ανήκουν τοποθετώντας τους ετικέτες για την αναγνώριση τους. Οπότε ταξινομούνται ως ουσιαστικά, ρήματα, νούμερα κ.λπ. Παρόλα αυτά κάποιες λέξεις της αγγλικής γλώσσας μπορεί να ανήκουν σε παραπάνω από μία κατηγορίες όπως για παράδειγμα η λέξη "drink" ανήκει στην κατηγορία "ουσιαστικό" με την έννοια του ποτού και επίσης στην κατηγορία "ρήματα" με την έννοια του "πίνω".⁽¹⁵⁾

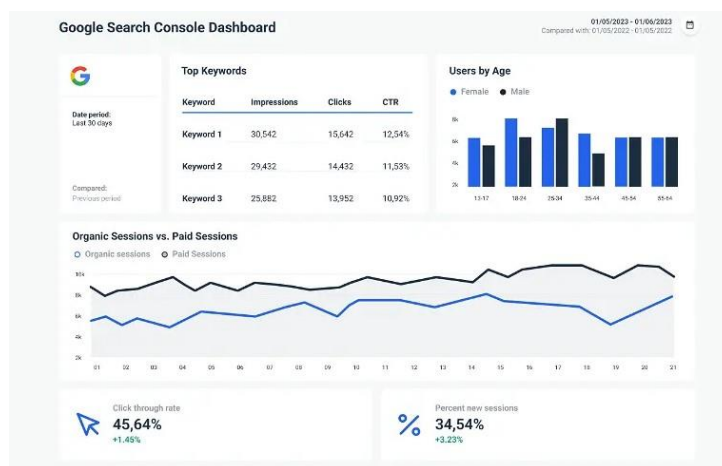
Named Entity Recognition (NER): Η μέθοδος αυτή εστιάζει στον εντοπισμό και την ταξινόμηση πληροφοριών γνωστών οντοτήτων. Οι ετικέτες που δίνονται στις οντότητες αυτές αφορούν ονόματα, ζώα, χρώματα τοποθεσίες και πολλά άλλα. Για παράδειγμα στη λέξη "Maria" θα δοθεί ετικέτα "Name" ενώ στην λέξη "duck" η ετικέτα "Animal". ⁽¹⁶⁾

Text cleaning: Είναι μια μέθοδος στην οποία πραγματοποιείται αφαίρεση των σημείων στίξης και των ειδικών συμβόλων και των emojis διότι δεν προσδίδουν τίποτα στην επεξεργασία της φυσικής γλώσσας. ⁽¹⁷⁾

Κεφάλαιο 3 – Στρατηγικές και προκλήσεις μάρκετινγκ

3.1 SEO

Βασικός στόχος μιας επιχείρησης αποτελεί η αυξημένη επισκεψιμότητα και αναγνωρισιμότητα του ιστότοπου και ο εύκολος εντοπισμός του από τους πελάτες. Αυτό επιτυγχάνεται με την βελτιστοποίηση των μηχανών αναζήτησης που περιλαμβάνεται από έναν συνδυασμό τεχνικών στρατηγικών. Για αρχή πρέπει να γίνει μια σωστή έρευνα λέξεων-κλειδιών, και να χρησιμοποιηθούν οι όροι που εμφανίζουν μεγάλο όγκο αναζητήσεων.^[11] Επιπλέον, το περιεχόμενο της ιστοσελίδας θα πρέπει να είναι ποιοτικό αλλά ταυτόχρονα ελκυστικό και μοναδικό, ιδανικό για τις ανάγκες του κοινού-στόχου και να ανανεώνεται συχνά. Πολύ σημαντικό παράγοντα αποτελεί η ταχύτητα της σελίδας, καθώς η ταχύτερη φόρτωση κάνει την εμπειρία του χρήστη καλύτερη και αποκτά υψηλότερη θέση ταξινόμησης στην αναζήτηση. Σύμφωνα με την Google ένας ισότοπος φιλικός προς τα κινητά βελτιώνει την εμπειρία του χρήστη και ανεβαίνει στα αποτελέσματα αναζήτησης.⁽³⁶⁾ Ιδιαίτερα χρήσιμοι για το SEO επίσης, είναι οι υπερσυνδέσεις από αξιόπιστους ιστότοπους που οδηγούν στην ιστοσελίδα μας, γνωστά και ως backlinks, ή η χρήση external link, δηλαδή υπερσυνδέσεις στην ιστοσελίδα μας οι οποίοι ανοίγουν όμως σε νέα καρτέλα.^[48] Πολύ σημαντικό ρόλο πλέον έχει και η ενεργή παρουσία της επωνυμίας στις πλατφόρμες μέσων κοινωνικής δικτύωσης βοηθώντας την αύξηση της προβολής της. Τέλος, το Google Analytics^[13] και ο Google Search Console⁽³⁷⁾ αποτελούν χρήσιμα εργαλεία για μια μακροπρόθεσμη επιτυχία SEO.

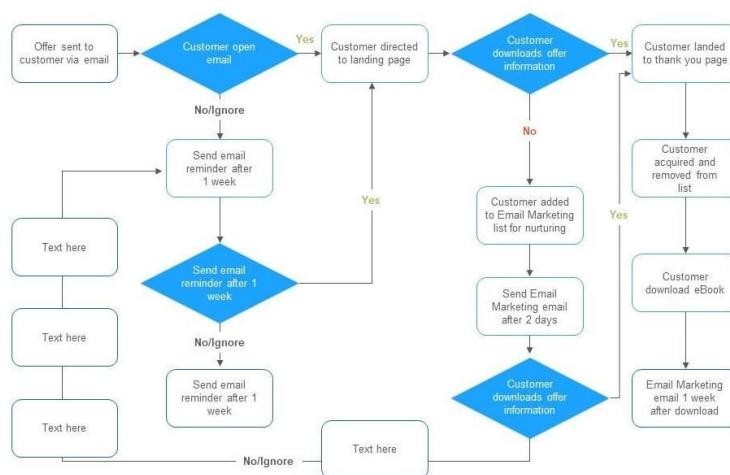


Εικόνα 19. Google Search Console

3.2 Email Marketing

Το email marketing αφορά μια ψηφιακή στρατηγική που επικεντρώνεται στην αποστολή στοχευόμενων μηνυμάτων στο κοινό-στόχο. Αποτελεί ένα χρήσιμο και οικονομικό εργαλείο, το οποίο προσεγγίζει εύστοχα μεγάλο αριθμό δυνητικών πελατών και χτίζει σχέσεις εμπιστοσύνης και αφοσίωσης. Ένας τρόπος προσέγγισης είναι η τμηματοποίηση κοινού, όπου βάσει των δεδομένων και την συμπεριφορά των πελατών χωρίζονται τα emails σε τμήματα παρέχοντας στον πελάτη ένα πιο εξατομικευμένο περιεχόμενο το οποίο ανταποκρίνεται στις ανάγκες και τα ενδιαφέροντα του. Επίσης η δημιουργία καμπανιών μέσω email με διάφορα ενημερωτικά δελτία και προωθητικές ενέργειες και προσφορές βοηθούν στην επίτευξη ορισμένων στόχων. Η αποστολή αυτοματοποιημένων μηνυμάτων τα οποία σχετίζονται με τις ενέργειες των χρηστών βοηθούν στην διατήρηση των πελατών και στην προσέλκυση νέων συνδρομητών. Βασική προϋπόθεση είναι η τήρηση των αρχών μάρκετινγκ που βασίζονται στην έγκριση των παραληπτών να λαμβάνουν οποιοδήποτε προωθητικό μήνυμα, χτίζοντας μια σχέση εμπιστοσύνης μεταξύ διαφημιστή και πελάτη και ακολουθώντας τους κανονισμούς προστασίας δεδομένων. Πολύ σημαντικό τα emails να είναι φιλικά και σε κινητές συσκευές καθώς ένα μεγάλο ποσοστό έχει πρόσβαση μέσω κινητής συσκευής.^[16]

Μία ακόμη αποτελεσματική στρατηγική είναι η αποστολή email με ένα κάλεσμα για δράση (CTA), το οποίο παροτρύνει τους καταναλωτές να δράσουν χρησιμοποιώντας επιτακτικά ρήματα όπως "καλέστε τώρα" ή χρονικούς περιορισμούς "Η προσφορά λήγει σύντομα". Για την βελτιστοποίηση του περιεχομένου των emails και του CTA, πραγματοποιούνται δοκιμές A/B. Οι δοκιμές A/B αποτελούν ένα τυχαιοποιημένο ελεγχόμενο πείραμα στο οποίο δύο εκδόσεις A και B ενός στοιχείου συγκρίνονται και καθορίζεται ποια έχει καλύτερη απόδοση, αποφέροντας καλύτερα αποτελέσματα όσον αφορά ένα CTA ή μια καμπάνια. Επίσης η χρήση ορισμένων εργαλείων ανάλυσης βοηθούν τους διαφημιστές να παρακολουθούν την απόδοση των απεσταλμένων email , τα ποσοστά αλληλεπίδρασης και την αύξηση των συνδρομητών. Τέλος η συγκέντρωση πληροφοριών μέσω σχολίων και ερευνών βοηθούν σημαντικά ένα μελλοντικό περιεχόμενο email βελτιώνοντας την εμπειρία του χρήστη.^[17]



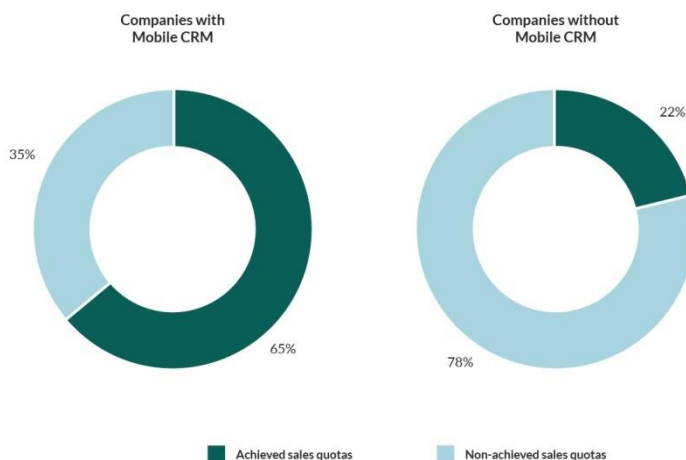
Εικόνα 20. Διάγραμμα ροής email marketing

3.3 Πληρωμένη διαφήμιση

Η πληρωμένη διαφήμιση είναι μια στρατηγική κατανομής προϋπολογισμού η οποία μπορεί να εφαρμοστεί σε διάφορα διαδικτυακά κανάλια, όπως τα μέσα κοινωνικής δικτύωσης, το Google Ads και τη διαφήμιση προβολής, στοχεύοντας στη προσέγγιση ενός ευρύτερου και στοχευόμενου κοινού, την αύξηση της επισκεψιμότητας, την ενίσχυση της αναγνωσιμότητας καθώς επίσης και τη δημιουργία δυνητικών πελατών ή πωλήσεων.^[18]

3.4 Διαχείριση πελατειακών σχέσεων (CRM)

Η διαχείριση πελατειακών σχέσεων είναι μια στρατηγική μέθοδος που εφαρμόζει εργαλεία και στρατηγικές για την επιτυχή διαχείριση και καλλιέργεια σχέσεων με τους πελάτες. Ουσιαστικά, το CRM συλλέγει δεδομένα πελατών για να παρέχει πλήρη κατανόηση των προτιμήσεων, των αλληλεπιδράσεων και των προηγούμενων αγορών τους. Αυτές οι πληροφορίες επιτρέπουν στις επιχειρήσεις να προσαρμόζουν τις αλληλεπιδράσεις, να τμηματοποιούν τους πελάτες για στοχευόμενη αφοσίωση και να βελτιστοποιούν την διαδικασία πωλήσεων μέσω αυτοματισμού. Το CRM είναι απαραίτητο για την παροχή εξαιρετικής εξυπηρέτησης πελατών, τη συλλογή σχολίων για την βελτίωση και τον εντοπισμό ευκαιριών για cross-selling και up selling.^[19] Με την ενσωμάτωση των εργαλείων ανάλυσης και αναφοράς, καθίσταται εύκολη η μέτρηση της αποτελεσματικότητας των αλληλεπιδράσεων, ενώ τα επεκτάσιμα και ασφαλή συστήματα CRM εξασφαλίζουν προσαρμοστικότητα και προστασία των δεδομένων. Το CRM συμβάλλει στην οικοδόμηση μακροχρόνιας αφοσίωσης των πελατών διευκολύνοντας την εξατομίκευση, την τμηματοποίηση των πελατών αλλά και την επικοινωνία σε όλα τα κανάλια.^[20]



Εικόνα 21. Σύγκριση απόδοσης πωλήσεων με Mobile CRM και χωρίς Mobile CRM

3.5 Ανάλυση δεδομένων

Η χρήση δεδομένων και αναλυτικών εργαλείων στην ανάλυση δεδομένων επιτρέπει στις επιχειρήσεις να αντλούν πολύτιμες γνώσεις σχετικά με τη συμπεριφορά των πελατών, να αξιολογούν την επιτυχία των εκστρατειών μάρκετινγκ και να λαμβάνουν καλά ενημερωμένες αποφάσεις. Οι επιχειρήσεις αξιοποιούν δεδομένα από διάφορες πηγές, που περιλαμβάνουν αλληλεπιδράσεις με πελάτες, συναλλαγές και διαδικτυακές δραστηριότητες για να εντοπίσουν μοτίβα, τάσεις και συσχετισμούς. Όσον αφορά τη συμπεριφορά των πελατών, η ανάλυση δεδομένων βοηθά στην κατανόηση των προτιμήσεων και στην πρόβλεψη μελλοντικών ενεργειών. Κατά την αξιολόγηση της αποτελεσματικότητας της καμπάνιας, τα αναλυτικά στοιχεία επικεντρώνονται σε βασικούς δείκτες απόδοσης (KPI), όπως ποσοστά μετατροπών, αναλογίες κλικ προς αριθμό εμφανίσεων και απόδοση επένδυσης (ROI). Αυτή η προσέγγιση με επίκεντρο τα δεδομένα δίνει τη δυνατότητα στις επιχειρήσεις να λαμβάνουν τεκμηριωμένες αποφάσεις, να βελτιστοποιούν τις στρατηγικές μάρκετινγκ και να κατανέμουν αποτελεσματικά τους πόρους, οδηγώντας τελικά σε ενισχυμένη αφοσίωση πελατών που είναι τόσο επιτυχημένη όσο και στοχευόμενη.⁽³⁸⁾

ROI Calculator	
Net Income Method	
Original Investment Value	\$1,000,000
Net Income	\$125,500
ROI	12.6%
Capital Gain Method	
Original Share Price	\$12.50
Current Share Price	\$15.20
ROI	21.6%
Total Return Method	
Original Share Price	\$12.50
Total Dividends Received	\$1.25
Current Share Price	\$15.20
ROI	31.6%
Annualized ROI	
Original Share Price	\$12.50
Current Share Price	\$15.20
Purchase Date	1/1/2017
Sale Date	8/24/2017
Annualized ROI	35.5%

Εικόνα 22. ROI μέτρηση απόδοσης

3.6 Προκλήσεις μάρκετινγκ

Στον τομέα του μάρκετινγκ, οι επιχειρήσεις αντιμετωπίζουν διάφορες προκλήσεις στην προσπάθειά τους να δημιουργήσουν μία μοναδική διαδικτυακή παρουσία και να πετύχουν τους στόχους που έχουν θέσει. Ο έντονος ανταγωνισμός που υπάρχει μεταξύ των βιομηχανιών απαιτεί καινοτόμες ιδέες και στρατηγικές για να ξεχωρίσουν και να προσελκύσουν περισσότερους πελάτες. Επίσης η συμπεριφορά των καταναλωτών παραμένει ένα διαρκές εμπόδιο για τους επαγγελματίες του μάρκετινγκ καθώς μεταβάλλεται συνεχώς αφού επηρεάζεται από τις τεχνολογικές εξελίξεις, τις τάσεις και τις κοινωνικές αλλαγές, με αποτέλεσμα οι επαγγελματίες να καλούνται να συμβαδίζουν με τις ταχέως εξελισσόμενες

συμπεριφορές των καταναλωτών. Από την άλλη η γρήγορη προσαρμογή σε νέες πλατφόρμες μέσω κοινωνικής δικτύωσης καθίσταται αναγκαία στο σύγχρονο μάρκετινγκ. Για την αποτελεσματική αξιοποίηση των πλατφορμών οι επαγγελματίες πρέπει να ενημερώνονται συνεχώς για τις τάσεις και τις γρήγορες αλλαγές στις προτιμήσεις των χρηστών προσαρμόζοντας τις στρατηγικές μάρκετινγκ με βάση τις τελευταίες τάσεις και βελτιστοποιώντας την απήχηση των καμπανιών.^[39] Επιπλέον η ταχύτητα με την οποία κυκλοφορούν ειδήσεις και απόψεις χρήζει προσεκτικής παρακολούθησης και άμεσης αντιμετώπισης οποιασδήποτε αρνητικής δημοσιότητας. Οι επαγγελματίες καλούνται να συμμετέχουν ενεργά σε διαδικτυακές συνομιλίες, σε κριτικές πελατών και μέσα κοινωνικής δικτύωσης για να εντοπίσουν έγκαιρα πιθανά προβλήματα, διατηρώντας και ενισχύοντας θετικά τη φήμη της επωνυμίας.^[21] Οι περιορισμένοι οικονομικοί πόροι μπορεί να έχουν μεγάλο αντίκτυπο στις διαφημιστικές εκστρατείες, την κατατομή των κεφαλαίων σε διάφορα κανάλια μάρκετινγκ καθώς μπορεί να επηρεάσουν την ικανότητα επένδυσης σε στρατηγικές μάρκετινγκ η και να εμποδίσουν το ανταγωνιστικό πλεονέκτημα μιας επιχείρησης. Με την σωστή διαχείριση των περιορισμένων προϋπολογισμών, βελτιστοποιώντας στρατηγικά τους διαθέσιμους πόρους και συνάμα αναζητώντας οικονομικά αποδοτικά κανάλια και δημιουργικές λύσεις μεγιστοποιείται και διασφαλίζεται η απόδοση της εταιρείας.^[22] Τέλος, η παγκοσμιοποίηση στο επιχειρηματικό πλαίσιο συνοδεύεται επίσης από ένα σύνολο προκλήσεων. Η κατανόηση των πολιτισμών και η προσαρμογή των καμπανιών σύμφωνα με τις τοπικές προτιμήσεις και γλώσσες είναι ζωτικής σημασίας για τη δημιουργία στρατηγικών μάρκετινγκ που έχουν απήχηση σε διαφορετικά κοινά. Η κατάλληλη διαχείριση των αυτών των προκλήσεων επιτρέπει στις επιχειρήσεις να εκμεταλλευτούν τις ευκαιρίες που παρουσιάζει η παγκοσμιοποίηση, προσεγγίζοντας ένα ευρύτερο κοινό, δείχνοντας ευαισθησία και προσαρμοστικότητα στην αγορά.^[60]

3.7 Τύποι διαφημίσεων

3.7.1 Διαφήμιση μέσω προγραμματισμού

Η διαφήμιση μέσω προγραμματισμού εκμεταλλεύεται θετικά την τεχνολογία προκειμένου η διαδικασία αγοράς των διαφημίσεων να γίνει πιο αποτελεσματική και ταχύτερη συμβάλλοντας στη βελτίωση της στόχευσης και την αποτελεσματικότητα της διαφημιστικής καμπάνιας. Πιο συγκεκριμένα αποτελεί μία αυτοματοποιημένη μέθοδο αγοράς και βελτιστοποίησης της διαδικτυακής διαφήμισης σε πραγματικό χρόνο. Στην διαφήμιση μέσω προγραμματισμού χρησιμοποιούνται αλγόριθμοι και λογισμικά για την αγορά διαφημίσεων καθιστώντας εφικτή την πρόσβαση σε πολλούς ιστότοπους εκδοτών ταυτόχρονα ενώ στην παραδοσιακή αγορά των διαφημίσεων απαιτούνται ανθρώπινες διαπραγματεύσεις με πολλούς αγοραστές και μη αυτόματες εντολές εισαγωγής.^[23]

Ανάλυση βασικών στοιχείων

Αυτοματοποιημένες αγορές: Αφορά την τεχνολογία υποβολής προσφορών σε πραγματικό χρόνο (RTB). Το RTB παρέχει τη δυνατότητα στις διαφημιστές να κάνουν προσφορές για εμφανίσεις εξετάζοντας την προβολή της πιο σχετικής διαφήμισης.

Στόχευση βάσει δεδομένων: Τα δεδομένα αποτελούν αναπόσπαστο κομμάτι αυτής της διαδικασίας καθώς

οι διαφημιστές στοχεύουν στοιχεία, συμπεριφορές και ενδιαφέροντα δημιουργώντας πιο στοχευόμενες διαφημίσεις.

Βελτιστοποίηση σε πραγματικό χρόνο: Δίνει τη δυνατότητα στους διαφημιζόμενους να τροποποιούν και να βελτιστοποιούν καμπάνιες σε πραγματικό χρόνο ανάλογα με τις μετρήσεις απόδοσης που έχουν πραγματοποιηθεί, όπως για παράδειγμα τα κλικ που έγιναν σε σχέση με τον αριθμό εμφανίσεων.

Πολλαπλές μορφές διαφήμισης: Χρησιμοποιείται ένας συνδυασμός διαφορετικών μορφών διαφήμισης όπως διαφημίσεις προβολής (banner), διαφημίσεις βίντεο, εγγενείς διαφημίσεις, διαφημίσεις συνδεδεμένες με τηλεόραση (CTV) και διαφημίσεις Digital Out-of-Home (DOOH).

Αποτελεσματικότητα και κόστος αποτελεσματικότητας: Η αυτοματοποιημένη διαδικασία βοηθά τους διαφημιστές να προσεγγίσουν το κοινό-στόχο τους με μεγαλύτερη ακρίβεια, βοηθώντας τους να επιλέξουν τις κατάλληλες εμφανίσεις και να μη ξοδέψουν πόρους σε άσχετες εμφανίσεις. Αυτό την καθιστά αποτελεσματική ως προς την διαδικασία αλλά και ως προς το κόστος.

Αυξημένη διαφάνεια: Οι πλατφόρμες διαφήμισης μέσω προγραμματισμού προσφέρουν στους διαφημιζόμενους διαφάνεια στην απόδοση της διαφημιστικής καμπάνιας καθώς προσφέρουν πολλές αναλυτικές πληροφορίες.



Εικόνα 23. Διαδικασία προγραμματιστικής διαφήμισης

Ο χρήστης επισκέπτεται έναν ιστότοπο, η πλατφόρμα του εκδότη αξιοποιεί τον διαφημιστικό χώρο και παράλληλα το λογισμικό Demand Side Platforms (DSP) χρησιμοποιεί τα δεδομένα δραστηριότητας προκειμένου να βρει τις διαφημίσεις που ταιριάζουν καλύτερα. Έπειτα το DSP υπολογίζει αυτόματα και τοποθετεί προσφορές για το διαφημιστικό σημείο βάσει δεδομένων. Στην συνέχεια η διαφήμιση που είχε την υψηλότερη προσφορά εμφανίζεται αμέσως στον επισκέπτη χρήστη. Τέλος, για κάθε εμφάνιση, συλλέγονται τα δεδομένα απόδοσης ώστε να βελτιστοποιηθούν οι μελλοντικές προσφορές και η στόχευση.⁽⁴¹⁾

3.7.2 Διαδραστική διαφήμιση

Οι διαδραστικές διαφημίσεις αποτελούν ψηφιακές διαφημίσεις πολύ διαφορετικές από τις παραδοσιακές στατικές διαφημίσεις που έχουμε συνηθίσει. Έχουν σχεδιαστεί για μεγιστοποιούν την προσήλωση των πελατών και να δημιουργούν μια πιο ελκυστική εμπειρία η οποία ενισχύει τους χρήστες να αλληλεπιδράσουν με την διαφήμιση. Ένας τρόπος για να μεγιστοποιηθεί η αφοσίωση του πελάτη είναι μια εξατομικευμένη προσέγγιση προσαρμόζοντας το περιεχόμενο ανάλογα με τα στοιχεία των χρηστών.

Επίσης, η ενσωμάτωση διαφόρων στοιχείων παιχνιδιού κάνουν την διαφήμιση πιο διασκεδαστική και συνάμα αυξάνουν τις πιθανότητες να αποτυπωθούν στο μυαλό του χρήστη. Εξίσου σημαντική είναι η δυναμική του περιεχομένου της διαφήμισης η οποία αλλάζει ανάλογα με τις ενέργειες των χρηστών, πέρα από αυτό όμως, οι διαφημίσεις που παροτρύνουν τον πελάτη σε δράση (CTA) , λειτουργούν ακόμη πιο αποτελεσματικά καθώς κάνουν τον πελάτη να προβεί σε μία ενέργεια (πχ. Click button) δημιουργώντας μια αλληλεπίδραση μεταξύ του χρήστη και της εταιρείας μετατρέποντας την αφοσίωση σε δράση. Με τις τωρινές τάσεις, πολλές διαδραστικές διαφημίσεις προβάλλονται στα μέσα κοινωνικής δικτύωσης από άτομα που ασκούν επιρροή σε χρήστες αυξάνοντας τις πιθανότητες απήχησης της διαφήμισης καθώς εμπιστεύονται τα άτομα αυτά. ⁽⁴²⁾

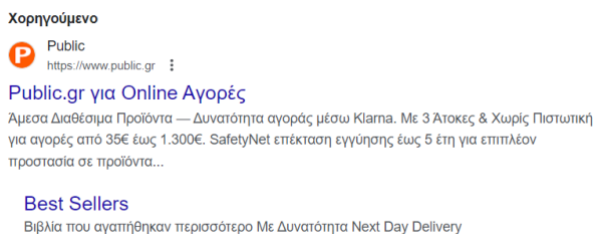
Τύποι διαδραστικών διαφημίσεων	Περιγραφή
Κουίζ και δημοσκοπήσεις	Ο χρήστης μπορεί να εκφράσει τις προτιμήσεις του για ένα προϊόν ή μια επωνυμία, συλλέγοντας σημαντικά δεδομένα.
Διαδραστικά βίντεο	Ο τύπος αυτής της διαφήμισης συνδυάζει την ισχύ ενός βίντεο με διαδραστικά στοιχεία επιτρέποντας στον χρήστη να αλληλεπιδράσει με το περιεχόμενο σε πραγματικό χρόνο.
Διαφήμιση με δυνατότητα αναπαραγωγής	Ο χρήστης έρχεται σε επαφή με ένα προϊόν μέσω μιας εμπειρίας που μοιάζει με παιχνίδι, διασκεδάζοντας τον ενώ ταυτόχρονα μεταφέρεται ένα μήνυμα μάρκετινγκ.
Διαφήμιση 360 μοιρών	Παρουσιάζονται γραφικά 360 μοιρών επιτρέποντας στον χρήστη να εξερευνήσει σύροντας με το δάχτυλο τους παρέχοντας του μια πιο διαδραστική και συναρπαστική εμπειρία.
Διαφήμιση επαυξημένης πραγματικότητας (AR)	Χρησιμοποιείται μια τεχνολογία επικάλυψης στοιχείων AR στο πραγματικό περιβάλλον του χρήστη δίνοντας του την ευκαιρία να οπτικοποιεί και να αλληλεπιδρά με εικονικά αντικείμενα.
Εικονική πραγματικότητα (VR)	Δημιουργείται μια ρεαλιστική εμπειρία προβολής του προϊόντος ή της υπηρεσίας.
Διαδραστική φωνητική διαφήμιση	Ο χρήστης αλληλεπιδρά με τη διαφήμιση ήχου μέσω βοηθών που ενεργοποιούνται με τη φωνή, όπως το Amazon Alexa, προσφέροντας μια εμπειρία hands-free.
Διαδραστικές διαφημίσεις banner	Πρόκειται για παραδοσιακές διαφημίσεις banner με πρόσθετα διαδραστικά χαρακτηριστικά και εφέ.

Πίνακας 1. Τύποι διαδραστικών διαφημίσεων ⁽⁴³⁾

3.7.3 Εγγενής διαφήμιση

Η εγγενής διαφήμιση είναι μία μορφή πληρωμένης ψηφιακής διαφήμισης η οποία ενσωματώνεται άψογα με το περιβάλλον της πλατφόρμας ή το μέσο που εμφανίζεται κάνοντας την να μη μοιάζει πραγματικά με διαφήμιση με αποτέλεσμα να κάνει την εμπειρία του χρήστη λιγότερο ενοχλητική. Κάποιες διαφημίσεις μπορεί να προβάλλονται στην ροή των ειδήσεων στα μέσα κοινωνικής δικτύωσης (Instagram, TikTok) γνωστές και ως διαφημίσεις “In Feed”. Άλλες, εμφανίζονται στην κορυφή των αποτελεσμάτων στις μηχανές αναζήτησης στις οποίες αναγράφεται από πάνω "χορηγούμενο", αυτές οι διαφημίσεις λέγονται “Paid Search”. Άλλη μια μορφή, αφορά διαφημίσεις με προτεινόμενο περιεχόμενο γνωστές και ως “Recommendation ads” και βρίσκονται συνήθως στα πλαϊνά των ιστοσελίδων και των μηχανών

αναζήτησης ή στο τέλος ενός άρθρου προτείνοντας ένα σχετικό περιεχόμενο στον χρήστη. Συνήθως αναγράφεται από πάνω “Προτείνεται για εσάς”.⁽⁴⁴⁾



Εικόνα 24. Διαφήμιση Paid Search

3.7.4 AI-powered διαφήμιση

Η βελτιστοποίηση διαφημίσεων με τεχνητή νοημοσύνη χρησιμοποιεί αλγόριθμους ΑΙ για την ανάλυση δεδομένων και τεχνικές μηχανικής μάθησης για την πραγματοποίηση προβλέψεων και την αυτόματη βελτιστοποίηση στοχεύοντας στη βελτίωση της απόδοσης και της αποτελεσματικότητας των διαφημιστικών καμπανιών. Κάποιοι αλγόριθμοι χρησιμοποιούνται για να αναλύουν μεγάλα σύνολα δεδομένων ώστε να εντοπίσουν μοτίβα ανάλογα με τη συμπεριφορά και τις προτιμήσεις των χρηστών προκειμένου να γίνει η στόχευση του κατάλληλου κοινού με εξατομικευμένο περιεχόμενο.⁽⁴⁵⁾ Επιπλέον, χρησιμοποιούνται αλγόριθμοι οι οποίοι αξιολογούν και υποβάλλουν προσφορές για τις εμφανίσεις διαφημίσεων σε πραγματικό χρόνο, εξασφαλίζοντας τη βέλτιστη τοποθέτηση για την υψηλότερη πιθανότητα αφοσίωσης των χρηστών. Η απόδοση της καμπάνιας παρακολουθείται και αυτή σε πραγματικό χρόνο, κάνοντας αυτόματες αλλαγές για τη μέγιστη απόδοση των αποτελεσμάτων. Οι αλγόριθμοι που χρησιμοποιούνται συμβάλλουν στην βελτιστοποίηση των προϋπολογισμών των διαφημίσεων και στην αποτελεσματικότερη τοποθέτηση τους. Επίσης εξετάζουν διαφορετικές παραλλαγές διαφημίσεων με σκοπό να βρουν τα πιο αποτελεσματικά στοιχεία και να βελτιστοποιήσουν την απόδοση. Με τη χρήση της τεχνητής νοημοσύνης πραγματοποιείται βελτιστοποίηση στο περιεχόμενο και τα γραφικά μιας διαφήμισης, με βάση τα δεδομένα των χρηστών, προκειμένου δημιουργηθούν διαφημίσεις που να ταιριάζουν με τις προτιμήσεις τους.⁽⁴⁶⁾

3.7.5 Shoppable Content διαφήμιση

Το περιεχόμενο με δυνατότητα αγοράς πρόκειται για έναν τύπο ψηφιακής εμπειρίας που δίνει τη δυνατότητα στον χρήστη να προβεί απευθείας στην αγορά του προϊόντος που εμφανίζεται στην διαφήμιση ή στα κοινωνικά μέσα. Κύριος στόχος είναι η διευκόλυνση του χρήστη στο να βρει το προϊόν χωρίς να μπει στην διαδικασία να το ψάξει, απλοποιώντας την διαδρομή από την ανακάλυψη στην αγορά. Η μείωση των βημάτων της αγοραστικής διαδικασίας κάνει την εμπειρία του χρήστη πιο ευχάριστη και τον οδηγεί σε παρορμητικές αγορές. Αρχικά ο χρήστης μπορεί να κάνει κλικ σε συγκεκριμένα στοιχεία που βρίσκονται σε μία εικόνα ή ένα βίντεο κατευθύνοντας τον απευθείας στον ιστότοπο για αγορά ή

ακόμη καλύτερα μπορεί να ολοκληρώνει αγορές απευθείας από το περιεχόμενο των μέσων κοινωνικής δικτύωσης χωρίς να φύγει από την πλατφόρμα που βρίσκεται. ⁽⁴⁷⁾

Επίσης αξιοποιώντας το influencer marketing οι influencer παρουσιάζουν διάφορα περιεχόμενα και με την χρήση συνδέσμων ο χρήστης αγοράζει απευθείας το προϊόν. Επιπλέον, ο χρήστης αποκτά πρόσβαση σε όλες τις λεπτομέρειες του προϊόντος χωρίς να οδηγηθεί στον ιστότοπο. Οι υπεύθυνοι μάρκετινγκ παρακολουθούν τις αλληλεπιδράσεις των χρηστών με το περιεχόμενο, και λαμβάνουν πληροφορίες για τις προτιμήσεις που συμβάλλουν στην βελτιστοποίηση των μελλοντικών καμπάνιες. ⁽⁴⁸⁾

3.8 Διαφημιστική απάτη

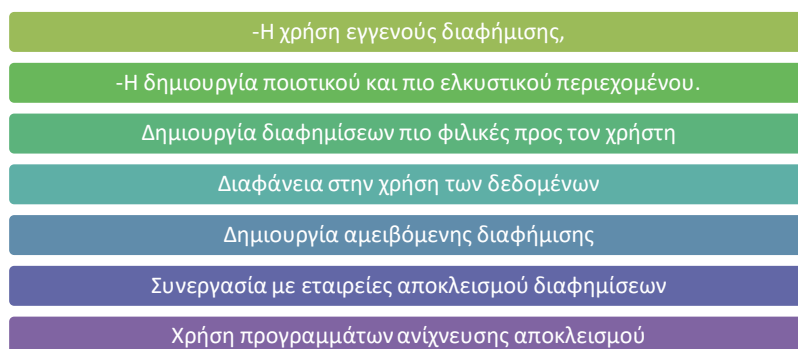
Η διαφημιστική απάτη αφορά παραπλανητικές και δόλιες δραστηριότητες που προσπαθούν να βλάψουν την ακεραιότητα και την αποτελεσματικότητα των διαφημιστικών καμπανιών. Οι απατεώνες αυξάνουν τα κέρδη τους με αθέμιτο τρόπο ενώ παράλληλα προκαλούν οικονομικές ζημιές στους διαφημιστές. ⁽⁴⁹⁾

Τύποι Απάτης στο διαδίκτυο	Περιγραφή
Απάτη κλικ	Αυτοματοποιημένα ρομπότ προσομοιώνουν κλικ σε ψηφιακές διαφημίσεις, αυξάνοντας τις αναλογίες κλικ με ψευδή αριθμό εμφανίσεων κάνοντας τους διαφημιστές να πληρώνουν για ψεύτικες αλληλεπιδράσεις με τους χρήστες
Απάτη με εντυπώσεις	Γίνεται παραπλάνηση των διαφημιζόμενων σχετικά με την πραγματική απήχηση των καμπανιών δημιουργώντας με bots ψευδείς προβολές διαφημίσεων.
Ψεύτικη επισκεψιμότητα	Χρησιμοποιούνται αυτοματοποιημένα ρομπότ δίνοντας την ψευδαίσθηση στον ιστότοπο ότι αυξάνεται η επισκεψιμότητα του.
Στοιβία διαφημίσεων	Τοποθετούνται πολλαπλές διαφημίσεις, η μία πάνω στην άλλη, σε μία μεμονωμένη τοποθέτηση διαφήμισης. Ο χρήστης βλέπει μόνο την κορυφαία διαφήμιση με αποτέλεσμα να καταγράφονται πολλαπλές εμφανίσεις οδηγώντας σε δόλια χρέωση.
Απάτη με cookie	Παραποιούνται τα δεδομένα και οι συμπεριφορές των χρηστών καταλήγοντας σε λάθος στόχευση και ανακριβή απόδοση.

Πίνακας 2. Απάτες διαδικτύου ⁽⁵⁰⁾

3.9 Αποκλεισμός διαφημίσεων

Ο αποκλεισμός των διαφημίσεων γίνεται με την χρήση λογισμικών με στόχο την απόκρυψη των διαφημίσεων. Ο χρήστης μπορεί να χρησιμοποιήσει λογισμικά αποκλεισμού ή επεκτάσεις προγράμματος περιήγησης για να αποτρέψει την εμφάνιση ψηφιακών διαφημίσεων από τον ιστότοπο ή μία εφαρμογή. Με την χρήση των προγραμμάτων αυτών γίνεται απαραίτητη η προσαρμογή των στρατηγικών από τους διαφημιστές για να προσεγγίσουν αποτελεσματικά το κοινό.⁽⁵¹⁾ Για την αντιμετώπιση του ad Block προτείνεται ⁽⁵²⁾:



3.10 Κανονισμοί απορρήτου

Οι κανονισμοί απορρήτου έχουν εφαρμοστεί για την προστασία των δεδομένων και την ασφάλεια των δικαιωμάτων των χρηστών καθώς επίσης και για την σωστή μεταχείριση των προσωπικών δεδομένων των χρηστών από τους οργανισμούς.⁽⁵³⁾

Γενικός Κανονισμός Προστασίας δεδομένων (GDPR)

Ο GDPR θεσπίστηκε στην Ευρωπαϊκή Ένωση το 2018 για την ενίσχυση της προστασίας των προσωπικών δεδομένων και των δικαιωμάτων των πολιτών που ανήκουν στην Ευρωπαϊκή Ένωση. Απαιτεί τη νόμιμη και διαφανή επεξεργασία δεδομένων καθώς επίσης σαφή και ρητή συγκατάθεση για την επεξεργασία τους.⁽⁵⁴⁾ Επιπλέον, επιβάλλει την ειδοποίηση για παραβιάσεις δεδομένων και καθιστά αναγκαίες τις εκτιμήσεις επιπτώσεων προστασίας δεδομένων σε περιπτώσεις επεξεργασίας υψηλού κινδύνου. Προκειμένου να δημιουργήσει ένα ισχυρό πλαίσιο προστασίας την ιδιωτικής ζωής των ατόμων επιβάλλει αυστηρά πρόστιμα σε όσους παραβιάζουν του κανονισμούς.⁽⁵⁵⁾

Νόμος απορρήτου της Καλιφόρνια (CCPA)

Στην πολιτεία της Καλιφόρνια των ΗΠΑ, το CCPA τέθηκε σε ισχύ το 2020, παρέχοντας στους κατοίκους διακριτά δικαιώματα και διασφαλίσεις που σχετίζονται με τα δεδομένα που συλλέγουν οι επιχειρήσεις.

⁽⁵⁶⁾ Μερικά κρίσιμα στοιχεία του CCPA ενσωματώνουν την εξουσία εξακρίβωσης των αποκτηθέντων

προσωπικών δεδομένων και το δικαίωμα να απαιτηθεί η αφαίρεση προσωπικών δεδομένων. Εγγυάται περαιτέρω το δικαίωμα άρνησης της εμπορίας δεδομένων και απαγορεύει την προκατάληψη έναντι των πελατών που ασκούν τα προνόμιά τους εμπιστευτικότητας. Οι επιχειρήσεις που πληρούν συγκεκριμένα πρότυπα εμπίπτουν στο πεδίο εφαρμογής του CCPA. Τα ετήσια ακαθάριστα έσοδα μιας επιχείρησης θα πρέπει να είναι υψηλότερα των 25 εκατομμυρίων δολαρίων, επίσης οι επιχειρήσεις πρέπει να χειρίζονται τα δεδομένα 50.000 ή και περισσότερων κατοίκων της Καλιφόρνια ετησίως και τέλος το 50% των ετήσιων εσόδων πρέπει να οφείλεται στην πώληση προσωπικών δεδομένων.⁽⁵⁷⁾

3.11 Ad Viewability

Η εμφανισιμότητα των διαφημίσεων αποτελεί σημαντική πρόκληση για τους διαφημιστές καθώς θα πρέπει να διασφαλίσουν ότι οι διαφημίσεις τους προβάλλονται από τα άτομα για τα οποία προορίζονται. Η εμφανισιμότητα της διαφήμισης αποτελεί μία μέτρηση που μετρά το ποσοστό μιας διαφήμισης που προβάλλεται από έναν χρήστη. Για την βελτίωση της εμφανισιμότητας των διαφημίσεων προτείνεται μία πολύπλευρη προσέγγιση που συνδυάζει την τοποθέτηση και τον σχεδιασμό της διαφήμισης την στόχευση του κοινού και την παρακολούθηση καμπάνιας. Η βελτιστοποίηση τοποθέτησης της διαφήμισης μπορεί να γίνει επιλέγοντας σελίδες ή εφαρμογές με υψηλή επισκεψιμότητα για την προβολή της ή τοποθετώντας τη σε σημείο που να είναι ορατή χωρίς κύλιση καθώς τείνουν να έχουν υψηλότερα ποσοστά εμφανισιμότητας. Επίσης, ο σχεδιασμός της διαφήμισης με καθηλωτικά γραφικά, σαφή ανταλλαγή μηνυμάτων και παρακίνηση του χρήστη για δράση βελτιώνουν τα ποσοστά εμφανισιμότητας. Επιπλέον, η στόχευση των κατάλληλων χρηστών , χρησιμοποιώντας δεδομένα για τις προτιμήσεις τους μπορούν να προσδιορίσουν τους χρήστες που είναι πιο πιθανό να αλληλεπιδράσουν με τις διαφημίσεις, βοηθώντας σημαντικά στην βελτίωση των ποσοστών εμφανισιμότητας. Τέλος, η ανάλυση των δεδομένων και των μετρήσεων βοηθούν τους διαφημιστές να προσαρμόζουν τις στρατηγικές τους και να βελτιστοποιήσουν τις καμπάνιες τους αυξάνοντας τα ποσοστά της εμφανισιμότητας.⁽⁵⁸⁾

Κεφάλαιο 4 -Εφαρμογές μηχανικής μάθησης

4.1 Πρόβλεψη συμπεριφοράς

Η έννοια της συμπεριφοράς του καταναλωτή^[24] αναφέρεται στην πορεία που ακολουθεί ένας καταναλωτής καθώς ερευνά, επιλέγει και αγοράζει ένα προϊόν ή μια υπηρεσία. Για μία επιχείρηση είναι κρίσιμο να κατανοεί τους λόγους που οδηγούν τους πελάτες στο να επιλέγουν ή να απορρίπτουν τα προϊόντα της. Αυτή η κατανόηση αποτελεί κλειδί για τον βέλτιστο σχεδιασμό των προϊόντων, με στόχο τη αύξηση των κερδών και την προαγωγή της εταιρικής ανάπτυξης. Η πρόβλεψη συμπεριφοράς του καταναλωτή απαιτείται για την επίτευξη αυξημένων κερδών. Πολλοί παράγοντες όπως η κουλτούρα, η εκπαίδευση και ο τρόπος ζωής επηρεάζουν αυτήν την συμπεριφορά. Η παραδοσιακή προσέγγιση περιλαμβάνει συνεντεύξεις με καταναλωτές και τμήματα εξυπηρέτησης πελατών, προκειμένου να κατανοηθούν οι ανάγκες τους. Ωστόσο, σε ορισμένες περιπτώσεις, οι καταναλωτές δυσκολεύονται να διατυπώσουν ακριβώς τις απαιτήσεις τους, και ενδέχεται να προτιμήσουν το προϊόν που τους ταιριάζει αν οι εταιρείες προσφέρουν επιλογές. Με την εξέλιξη της τεχνητής νοημοσύνης, η μηχανική μάθηση κυριαρχεί ως εργαλείο ανάλυσης αν και εξακολουθούν να υπάρχουν δυσκολίες στην καταμέτρηση εναλλακτικών μεθόδων, μερικές από τις οποίες μπορεί να έχουν μικρές διαφορές μεταξύ τους.

Η πρώτη μέθοδος πρόβλεψης της συμπεριφοράς του καταναλωτή δημιουργήθηκε το 1963 από τον Howard και εξελίχθηκε σε "Θεωρία συμπεριφοράς Αγοραστή" το 1969. Αυτό το μοντέλο παρέχει μεθόδους για την πρόβλεψη της συμπεριφοράς του καταναλωτή, ενισχύοντας την κατανόηση των καταναλωτικών πτυχών με αναλυτικές τεχνικές.

Ο αλγόριθμος του δέντρου απόφασης, πηγάζει από την ιδέα ενός μοναδικού δέντρου, η οποία αναπτύχθηκε παράλληλα με την εξέλιξη της ανάλυσης της συμπεριφοράς του καταναλωτή, από το 1987 και έπειτα. Αποτελεί ένα εύκολο αναλυτικό εργαλείο στη μηχανική μάθηση, εστιάζοντας σε παράγοντες που επηρεάζουν τα αποτελέσματα.

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) ανήκουν σε τεχνητά νευρωνικά δίκτυα με δυνατότητα επεξεργασίας ακολουθιών. Είναι πολύ δημοφιλή για την ανάλυση καταναλωτικής συμπεριφοράς, προσφέροντας ευελιξία και ευανάγνωστο αναλυτικό αλγόριθμο.

4.1.1 Βιβλιογραφική έρευνα

Στην πρώτη εργασία ο Cheng-Ju Liu^[25] προτείνει ένα γραμμικό μοντέλο παλινδρόμησης και το μοντέλο XGBoost που βασίζεται σε δέντρο αποφάσεων για ανάλυση συμπεριφοράς των πελατών σε πραγματικό χρόνο και πρόβλεψη των τάσεων αγοραστικής συμπεριφοράς των πελατών

Πηγή δεδομένων

Το προεπεξεργασμένο σύνολο δεδομένων που χρησιμοποιεί βασίζεται σε ένα μέγεθος παραθύρου 7 ημερών και διαμοιράζεται σε πολλά υποσύνολα ανά ημέρα. Στην συνέχεια εξάγονται ιστορικά δεδομένα από κάθε παράθυρο μέσω δείγματος εξαγωγής και μετατροπής χαρακτηριστικών των δεδομένων (user id,

item id). Κάθε παράθυρο συνιστάται από πολλαπλά υποσύνολα δειγμάτων, με τα δεδομένα να αποτελούνται από θετικά και αρνητικά δείγματα. Για την επιλογή των τελικών δεδομένων εκπαίδευσης εξάγεται ένα υποσύνολο των δειγμάτων από τα 10 παράθυρα, με περίπου 70.000 δείγματα ανά παράθυρο. Τέλος, ο αλγόριθμος επαληθεύεται ως προς την εγκυρότητα του και τα δεδομένα κατηγοριοποιούνται σε σύνολα δεδομένων πριν την επιλογή των χαρακτηριστικών $s(1 \leq i \leq 10)$ και σύνολα δεδομένων μετά την επιλογή χαρακτηριστικών $s'(1 \leq i \leq 10)$, με διαστάσεις 110 και 56 αντίστοιχα.

Μέθοδος αξιολόγησης

Για την αξιολόγηση της απόδοσης του μοντέλου ελέγχονται οι μετρικές Accuracy, AUC και F1

Εκπαίδευση μοντέλου

Κατά την εκπαίδευση του μοντέλου, χρησιμοποιείται σταδιακή παλινδρόμηση με βάση το κριτήριο AUC προκειμένου να αξιολογηθούν τα χαρακτηριστικά και να βελτιστοποιηθεί το μοντέλο. Η τιμή της ακρίβειας αυξάνεται όσο αυξάνεται ο αριθμός των δειγμάτων εκπαίδευσης. Ωστόσο, παρατηρείται πως η απόδοση του μοντέλου είναι καλύτερη με αναλογία θετικών προς αρνητικών δειγμάτων 1:3. Η τάση των τιμών δεν εμφανίζει έντονη διακύμανση, όμως όταν υπερσχύει το αρνητικό δείγμα, η τιμή AUC μειώνεται λόγω της αυξημένης πολυπλοκότητας της εκπαίδευσης σε μεγάλα δεδομένα. Ο αλγόριθμος λογιστικής παλινδρόμησης ρυθμίζεται με βάσει τη μέθοδο εκτίμησης μέγιστης πιθανότητας.

Στην συνέχεια ελέγχεται η τιμή AUC του μοντέλου σύντηξης, η οποία υπολογίζεται επαναληπτικά με τεχνητές τιμές. Το μοντέλο λογιστικής παλινδρόμησης φαίνεται να είναι γραμμικό και η πρόβλεψη ενός μόνο μοντέλου δεν είναι ιδιαίτερα αξιόπιστη. Το XGBoost όμως βάσει των τυπικών βαρών πολλών ομάδων δείχνει να επηρεάζει σημαντικά την εκπαίδευση των μοντέλων σύντηξης. Στο σετ εκπαίδευσης οι τιμή AUC του μοντέλου σύντηξης ξεπερνά το 0.76 ακολουθεί η τιμή AUC του XGBoost με 0.75 και τέλος η AUC λογιστικής παλινδρόμησης φτάνει 0.71.

Τα αποτελέσματα του μοναδικού μοντέλου στο σετ εκπαιδύονται με τον αλγόριθμο LR, και το τελικό μοντέλο σύντηξης κατασκευάζεται στο δοκιμαστικό μοντέλο. Η τιμή AUC του XGBoost και της λογιστικής παλινδρόμησης αυξάνεται στο 0.81, η AUC του XGBoost δεν δείχνει καμία σημαντική βελτίωση και παραμένει ίδια 0.75 ενώ η AUC λογιστικής παλινδρόμησης φτάνει στο 0.77.

Η σύγκριση μεταξύ δύο μεθόδων σύνθεσης δείχνει ότι η τεχνητή στάθμιση είναι πιο αποτελεσματική από τη γραμμική στάθμιση.

Ανάλυση αποτελεσμάτων μοντέλου πρόβλεψης ζήτησης πελατών

Χρησιμοποιώντας το ίδιο σύνολο δεδομένων αξιολογείται η ακρίβεια πρόβλεψης του μοντέλου δέντρου αποφάσεων, του τεχνητού νευρωνικού δικτύου και της μηχανής διανυσμάτων υποστήριξης ώστε συγκρίνοντας τα να επικυρωθεί η απόδοση του υβριδικού μοντέλου.

Τα αποτελέσματα δείχνουν ότι το υβριδικό μοντέλο υπερτερεί, προβλέποντας ακριβώς τις πωλήσεις και προσαρμόζοντας τα γραμμικά χαρακτηριστικά καλύτερα από τα άλλα μοντέλα. Με βάση τις προβλεπόμενες καμπύλες απόδοσης και με διαφορετικές παραμέτρους παρατηρείται ότι:

1. Ο αλγόριθμος ομαδοποίησης είναι αποτελεσματικός για υψηλής συχνότητας χρήστες, καθώς όσο αυξάνεται η τμηματοποίηση, βελτιώνεται η απόδοση του προγνωστικού μοντέλου.
2. Ο αλγόριθμος ομαδοποίησης είναι αποτελεσματικός και για χαμηλής συχνότητας χρήστες, καθώς όσο αυξάνεται η τμηματοποίηση, η απόδοση του μοντέλου επηρεάζεται από την σπανιότητα των δεδομένων.

Πρόβλεψη συμπεριφοράς αγοράς προϊόντος

Για την βελτίωση της ακρίβειας του μοντέλου, χρησιμοποιείται μια δεκαπλάσια μέθοδος διασταυρούμενης επικύρωσης στην οποία κάθε χρονικό παράθυρο χρησιμοποιείται ως δοκιμαστικό σετ, ενώ τα υπόλοιπα εννέα χρησιμοποιούνται για εκπαίδευση. Στην συνέχεια γίνεται η αξιολόγηση της απόδοσης με διάφορους αλγορίθμους μηχανικής μάθησης και χρησιμοποιείται η μέση τιμή F1 για σύγκριση. Μετά την επιλογή χαρακτηριστικών, η αξιολόγηση γίνεται σε διασταυρούμενο σύνολο δειγμάτων, και η μέση τιμή F1 χρησιμοποιείται για σύγκριση μεταξύ πλήρους και επιλεγμένου σετ δεδομένων. Παρατηρείται ότι με την επιλογή του αλγορίθμου SSP, η τιμή F1 του XGBoost και του τεχνητού νευρωνικού δικτύου έχει βελτιωθεί σημαντικά με την τιμή του F1 του XGBoost να ξεπερνάει το 0.90.

Προσθήκη θετικών και αρνητικών δειγμάτων

Τέλος, για να ελεγχθεί πώς η επιλογή του αλγορίθμου SSP επηρεάζει την απόδοση, συγκρίνονται οι τιμές F1 από τα μοντέλα πρόβλεψης και αναλύεται το πλάτος της διακύμανσης. Τα μοντέλα P/N & XGBoost και XGBoost δείχνουν έχουν σχετικά σταθερή απόδοση σε αντίθεση με τα υπόλοιπα τα οποία φαίνεται να έχουν χαμηλότερη ακρίβεια.

4.2 Πρόβλεψη αποχώρησης πελατών

Η πρόβλεψη αποχώρησης πελατών ⁽⁵⁹⁾ αποτελεί μια μεθοδολογία ανάλυσης των δεδομένων η οποία περιλαμβάνει τον εντοπισμό των πελατών που πιθανότατα έχουν σκοπό να διακόψουν μια συνδρομή ή να σταματήσουν να χρησιμοποιούν ένα προϊόν ή μια υπηρεσία εντός ενός συγκεκριμένου χρονικού διαστήματος.

Αυτή η τεχνική ανάλυσης των δεδομένων επιτρέπει στις επιχειρήσεις να αναλάβουν προληπτικά μέτρα για να διατηρήσουν τους πελάτες που ενδέχεται να αποχωρήσουν όπως να βελτιώσουν την εμπειρία τους, παρέχοντας τους πιο στοχευμένες καμπάνιες και δημιουργώντας ένα πιο συγκεκριμένο περιεχόμενο αυξάνοντας την διάρκεια παραμονής τους. Συνάμα, συμβάλλει στην αύξηση των κερδών μιας επιχείρησης καθώς η παροχή προϊόντων και υπηρεσιών κοστίζει λιγότερο σε ήδη υπάρχοντες πελάτες σε σύγκριση με την εύρεση και απόκτηση νέων πελατών. Επιπρόσθετα, η ικανότητα της επιχείρησης να μπορεί να κάνει προβλέψεις και να εντοπίσει τα προβλήματα που οδηγούν έναν πελάτη να αποχωρήσει, αντιμετωπίζοντάς τα έγκαιρα ενισχύεται την εμπιστοσύνη του πελάτη προς την επιχείρηση και προς τα προϊόντα ή τις υπηρεσίες που παρέχει.

Η μηχανική μάθηση συντελεί στην επίτευξη των παραπάνω οφελών, αναπτύσσοντας αποτελεσματικά μοντέλα πρόβλεψης αποχώρησης των πελατών, τα οποία αναγνωρίζουν πιο σύνθετα μοτίβα και διαχειρίζονται μεγαλύτερα σύνολα δεδομένων παρέχοντας μεγαλύτερη ακρίβεια στις προβλέψεις. Αποτελεί ένα απαραίτητο εργαλείο για την ανάπτυξη εξελιγμένων συστημάτων πρόβλεψης, στοχεύοντας στην αύξηση των εσόδων και την διατήρηση των πελατών.

Ο Brabadevi P.^[26] χρησιμοποιεί τεχνικές μηχανικής μάθησης αξιοποιώντας τις δυνατότητες πρόβλεψης ενός στοχαστικού αλγόριθμου ενίσχυσης για την πρόβλεψη απώλειας των πελατών.

Σύνολο δεδομένων

Το σύνολο δεδομένων λήφθηκε από το kaggle και αφορά έναν τηλεπικοινωνιακό οργανισμό ο οποίος περιλαμβάνει πληροφορίες 7044 πελατών. Αποτυπώνει τις ενέργειες των πελατών, ποιοι αποφάσισαν να αποχωρήσουν και ποιοι να συνεχίσουν τη συνδρομή τους.

Ανάλυση δεδομένων

Λαμβάνοντας υπόψη το σύνολο των πελατών που έχουν ήδη αποχωρήσει, αναλύονται τα χαρακτηριστικά και η συμπεριφορά τους προτού αποχωρήσουν και καταγράφονται κάποιες πληροφορίες για πρόβλεψη. Όπως ποιοι πελάτες αποχώρησαν, ποιες υπηρεσίες χρησιμοποίησε κάθε πελάτης και πληροφορίες συνδρομής.

Το 70% των δεδομένων απόκλισης πελατών χρησιμοποιείται για εκπαίδευση και το 30% για δοκιμές. Τα δεδομένα υποβάλλονται σε κατηγοριοποίηση και μετατρέπονται στην κατάλληλη μορφή. Κατά την επεξεργασία, χρησιμοποιούνται κωδικοποιητές και μετατροπές ετικετών για τη μετατροπή κατηγορικών ετικετών σε αριθμητικές ετικέτες και για την κανονικοποίηση των δεδομένων. Αυτή η διαδικασία είναι αναγκαία για την προετοιμασία του συνόλου δεδομένων για εκπαίδευση με διάφορα μοντέλα.

Μοντέλα πρόβλεψης

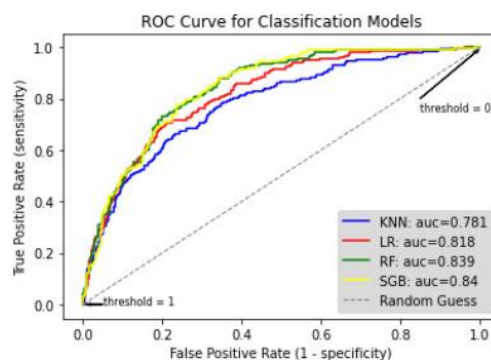
1. Stochastic gradient booster (SGB)
2. Random forest (RF)
3. K-nearest Neighbor (K-NN)
4. Logistic Regression (LR)

Αποτελέσματα

Μοντέλο πρόβλεψης	Accuracy	Accuracy δοκιμών	AUROC	Precision	Recall	F1score
KNN	0.78	0.77	0.78	0.81	0.92	0.86
LR	0.79	0.78	0.81	-	-	-
RF	0.803	0,787	0.892	0.82	0.91	0.86
SGB	0.806	0,791	0.839	0.82	0.91	0.87

Πίνακας 3. Αξιολόγηση απόδοσης αλγορίθμων για πρόβλεψη αποχώρησης πελατών.

Ο αλγόριθμος SGB κατέγραψε την υψηλότερη απόδοση με ακρίβεια 0.839, υποδεικνύοντας ένα υψηλό επίπεδο αξιοπιστίας στην πρόβλεψη. Οι αλγόριθμοι RF και LR ακολούθησαν με απόδοση ακρίβειας 0.829 και 0,826 αντίστοιχα, εμφανίζοντας μια ελαφρώς χαμηλότερη απόδοση όμως αρκετά ικανοποιητική. Ο αλγόριθμος με την χαμηλότερη απόδοση να φτάνει στα 0.781 είναι ο kNN, υποδεικνύοντας πως χρειάζεται περαιτέρω βελτίωση στην πρόβλεψη αποχώρησης πελατών.



Εικόνα 25. ROC αλγορίθμων

Εξετάζοντας την καμπύλη ROC και την καμπύλη AUC ο αλγόριθμος SGB εμφάνισε την καλύτερη απόδοση 0.84 ενώ ο αλγόριθμος KNN αν και εμφάνισε την χαμηλότερη απόδοση ακρίβειας η απόδοση της καμπύλης AUC ήταν αρκετά ικανοποιητική, ίση με 0,781. Η πρόβλεψη της αποχώρησης αποδεικνύεται πολύ σημαντική καθώς οι οργανισμοί μπορούν να προβλέψουν την συμπεριφορά των πελατών και να διατηρήσουν ακόμη περισσότερους πελάτες, μειώνοντας το κόστος.

4.3 Τμηματοποίηση αγοράς πελατών

Η τμηματοποίηση της αγοράς ⁽⁶⁰⁾ περιλαμβάνει τον χωρισμό των πελατών σε ομάδες βάσει κοινών χαρακτηριστικών. Πιο συγκεκριμένα, η τμηματοποίηση αντιπροσωπεύει τη διαδικασία εντοπισμού ομάδων πελατών που έχουν κοινές ανάγκες, επιθυμίες και συμπεριφορές. Αυτό επιτρέπει στις επιχειρήσεις να προσαρμόσουν προϊόντα, υπηρεσίες και στρατηγικές για να ανταποκριθούν στις διαφορετικές ανάγκες των διαφόρων ομάδων πελατών τους. Η τμηματοποίηση μπορεί να υλοποιηθεί με τη χρήση διαφόρων παραγόντων, όπως η ηλικία, το φύλο, το εισόδημα, η τοποθεσία, τα ενδιαφέροντα, και το ιστορικό προηγούμενων αγορών.

Η μηχανική μάθηση αναδεικνύεται ως ένα ισχυρό εργαλείο για την τμηματοποίηση πελατών, καθώς επιτρέπει τον εντοπισμό ομάδων πελατών με παρόμοια χαρακτηριστικά και συμπεριφορές. Η χρήση της μηχανικής μάθησης προσφέρει πολλά οφέλη στις επιχειρήσεις. Ένα από τα κυρίαρχα πλεονεκτήματα είναι η ικανότητα της να επεξεργάζεται γρήγορα και με ακρίβεια μεγάλες ποσότητες δεδομένων, οδηγώντας σε εξαιρετικά στοχευόμενη και ακριβή τμηματοποίηση. Οι αλγόριθμοι μηχανικής μάθησης που χρησιμοποιούνται είναι σε θέση να αναγνωρίζουν μοτίβα και να ανακαλύπτουν ιδέες στη συμπεριφορά των πελατών, πράγμα που θα ήταν δύσκολο να επιτευχθεί από τον άνθρωπο, επιτυγχάνοντας έτσι πιο ακριβείς προβλέψεις για τη μελλοντική συμπεριφορά των πελατών.

Επιπλέον, η μηχανική μάθηση εκδηλώνει την ικανότητα να προσαρμόζεται και να αντλεί πληροφορίες από νέα δεδομένα, καθιστώντας την αποτελεσματική για δυναμικές και εξελισσόμενες αγορές. Αυτή η προσέγγιση για τη τμηματοποίηση πελατών συμβάλλει στο να βοηθήσει τις εταιρείες να διατηρούν το προβάδισμα έναντι των ανταγωνιστών τους και να αντιδρούν άμεσα στις τάσεις της αγοράς. Επίσης, ενσωματώνει ποικίλες πηγές δεδομένων προσφέροντας μια πιο ολοκληρωμένη κατανόηση των αναγκών και των επιθυμιών των πελατών.



Εικόνα 26. Διαδικασία προετοιμασίας δεδομένων

Οι υπεύθυνοι του μάρκετινγκ δημιουργούν στρατηγικές για να προσελκύσουν ή και να διατηρήσουν πελάτες, εστιάζοντας σε εξατομικευμένες προσεγγίσεις. Το εξατομικευμένο μάρκετινγκ επικεντρώνεται στην δημιουργία σχέσεων μεταξύ επιχείρησης και πελάτη καθώς το κόστος για την απόκτηση ενός νέου

πελάτη είναι υψηλότερο από την διατήρηση του ενός (McIlroy & Barnett, 2000).^[27] Η τμηματοποίηση αγοράς κατηγοριοποιεί τους πελάτες με βάση την συμπεριφορά και κοινά χαρακτηριστικά. Χαρακτηριστικά τα οποία είναι απαραίτητα όπως η υπαρκτότητα, η προσβασιμότητα, η ενδεικτικότητα, η σταθερότητα, η ανταποκρισιμότητα και η εφαρμοσιμότητα, προκειμένου να αυξήσει την αποδοτικότητα και την αποτελεσματικότητα των προϋπολογισμών διασφαλίζοντας έτσι την σταθερότητα και την κερδοφορία των στρατηγικών μάρκετινγκ. Η τμηματοποίηση μπορεί να επιτευχθεί μέσω a priori τμηματοποίησης ή post-hoc τμηματοποίησης με την post-hoc να οδηγεί σε ανταγωνιστικό πλεονέκτημα. Τα τμήματα είναι αναγνωρίσιμα με βάση τα διαθέσιμα δεδομένα και τα χαρακτηριστικά των πελατών.

Η εργασία που προτείνει ο Rik van Leeuwen^[28] επικεντρώνεται στον αλγόριθμο ομαδοποίησης Cluster με ιεραρχική προσέγγιση καθώς χωρίζει τους πελάτες σε υψηλό επίπεδο χωρίς να γνωρίζει εκ των προτέρων τον αριθμό των ομάδων.

Δεδομένα

Τα δεδομένα που χρησιμοποιούνται αντλούνται από ένα Σύστημα Διαχείρισης Περιουσίας (PMS) στη βιομηχανία φιλοξενίας επικεντρώνοντας σε ένα ξενοδοχείο με ένα ιδιοκτησιακό σύστημα στο Άμστερνταμ από το 2015 έως το 2019. Το σύνολο των δεδομένων περιλαμβάνει κρατήσεις, προφίλ, και φυλλάδια, με ένα χρυσό προφίλ που δημιουργείται μέσω τεχνικών αντιστοίχισης και συγχώνευσης. Το χρυσό προφίλ συνδυάζει πληροφορίες επισκεπτών με ενδιαφέροντα και μετρικές. Ενσωματώνονται επιπλέον χαρακτηριστικά με βάση τα δεδομένα από τον πίνακα κρατήσεων όπως ο χρόνος προειδοποίησης της κράτησης και ο χρόνος παραμονής, ενώ τα χαρακτηριστικά των κρατήσεων ή των φυλλαδίων ταξινομούνται για πρόσθετη πληροφορία.

Ανάλυση δεδομένων

Η ανάλυση δεδομένων περιλαμβάνει περίπου 170,000 προφίλ, 230,000 κρατήσεις και 870,000 φύλλα που δημιουργήθηκαν μεταξύ συγκεκριμένων ημερομηνιών. Οι κρατήσεις κατηγοριοποιούνται ως διανομή, ακύρωση ή απουσία, με τα αντίστοιχα ποσοστά. νέα προφίλ, περίπου 34,000 ετησίως εγγράφονται και το 82,2% έχει μια κράτηση για διαμονή. Οι επαναλαμβανόμενοι επισκέπτες που αντιστοιχούν στο 14.3% ξοδεύουν κατά μέσο όρο 16,7% περισσότερο από τους επισκέπτες που έρχονται μία φορά. Το καθεστώς πιστότητας απονέμεται σε 1.17% των προφίλ που ξοδεύουν 4.6 φορές περισσότερο. Οι κρατήσεις μέσω άμεσων καναλιών ανέρχονται σε 44.49% με υψηλότερη δαπάνη 12.84%. Οι πιο συνήθεις λόγοι που ταξιδεύουν οι επισκέπτες είναι για αναψυχή ή για επαγγελματικούς λόγους. Επίσης παρουσιάζονται ποσοστά διατήρησης για προφίλ συνδεδεμένα με ομάδες, εταιρείες και πράκτορες. Η ανάλυση του χρόνου προειδοποίησης αποκαλύπτει μέσο χρόνο 31 ημερών. Συνολικά, αυτά τα σφαιρικά δεδομένα παρέχουν εισαγωγές για τη συμπεριφορά, τις προτιμήσεις και τα μοτίβα κρατήσεων των επισκεπτών.

Προετοιμασία δεδομένων

Τα δεδομένα μοντελοποιούνται και μετασχηματίζονται σε επίπεδο προφίλ με στόχο την ομαδοποίηση των επισκεπτών βάσει χαρακτηριστικών που καθορίζουν την συμπεριφορά τους. Τα αρχικά δεδομένα μετατρέπονται σε ένα σύνολο δεδομένων με 25 χαρακτηριστικά κατασκευασμένα με βάση τη λογική την επιχείρησης. Η μηχανική χαρακτηριστικών εισάγει νέα στοιχεία για χειροκίνητη αναθεώρηση και πρόσθετο πλαίσιο για το μοντέλο. Η επιχειρησιακή λογική καθορισμένη από την εταιρεία ξενοδοχείων, δημιουργεί χαρακτηριστικά δυαδικών τιμών που υποδεικνύουν την εκπλήρωση συγκεκριμένων συνθηκών σε επίπεδο προφίλ. Προκειμένου να αντιμετωπιστεί το "κακό της διαστασιμότητας" εφαρμόζονται μέθοδοι μείωσης διαστασιμότητας, εξετάζοντας ποσοτώσεις, εκτροπές και χαρακτηριστικά που σχετίζονται με τα έσοδα. Αυτοί οι κανόνες στοχεύουν στην βελτιστοποίηση της υπολογιστικής αποτελεσματικότητας και τη διατήρηση της σημαντικότητας των χαρακτηριστικών, ιδίως στον καθορισμό του καθεστώτος πιστότητας του επισκέπτη.

Η ανάλυση των προφίλ των επισκεπτών πραγματοποιείται με την χρήση της ιεραρχικής συσταδοποίησης, μιας μεθόδου μη εποπτευόμενης μηχανικής μάθησης. Αυτή η προσέγγιση ξεκινά με κάθε αντικείμενο δεδομένων ως ξεχωριστή συστάδα και συγχωνεύει επαναληπτικά τα πιο παρόμοια αντικείμενα βάσει μιας μετρικής απόστασης και μιας συνάρτησης σύνδεσης.

Η μετρική απόστασης του Gower χρησιμοποιείται για να διαχειριστεί δεδομένα που περιλαμβάνουν τόσο αριθμητικές όσο και κατηγορικές μεταβλητές. Η συνάρτηση σύνδεσης αποτελεί τη μέθοδο ελάχιστης διακύμανσης του Ward με στόχο τη δημιουργία ισορροπημένων συστάδων και την ανθεκτικότητα έναντι ακραίων τιμών.

Εκπαίδευση μοντέλου

Για την ιεραρχική συσταδοποίηση χρησιμοποιείται η συνάρτηση R `hclust` από το πακέτο `stats` με είσοδο έναν πίνακα απόστασης που υπολογίζεται εφαρμόζοντας τη μετρική απόσταση του Gower. Οι συστάδες συγχωνεύονται βάσει της μεθόδου ελάχιστης διακύμανσης του Ward. Στην συνέχεια εφαρμόζεται η μέθοδος του λαιμού μαζί με ένα κριτήριο που περιλαμβάνει το άθροισμα τετραγώνων για τον προσδιορισμό του βέλτιστου αριθμού συστάδων. Ο λαιμός σηματοδοτεί μια αλλαγή στην κλίση του συνολικού αθροίσματος τετραγώνων. Επιπλέον κριτήρια που περιλαμβάνουν τη σχετική δύναμη των αριθμών συστάδων χρησιμοποιούνται για την αύξηση της ακρίβειας στον προσδιορισμό του σωστού αριθμού συστάδων. Παρατηρείται πως υπάρχουν αρκετοί περιορισμοί στην ταξινόμηση ενός μεγάλου συνόλου προφίλ και για την αντιμετώπιση τους διενεργούνται πολλαπλές δοκιμές με υποσύνολα προφίλ και εφαρμόζεται ένας αλγόριθμος k-κοντινότερων γειτόνων για την κατάταξη του πλήρους συνόλου δεδομένων. Η σταθερότητα του αποτελέσματος αξιολογείται παρατηρώντας τη διασπορά του βέλτιστου αριθμού συστάδων σε διάφορες δοκιμές.

Αποτελέσματα

Εφαρμόζοντας τον αλγόριθμο μηχανικής μάθησης χωρίς επίβλεψη εντοπίζονται 8 κατάλληλες συστάδες από τις 15 εκτελέσεις που πραγματοποιήθηκαν. Εκτελώντας τον αλγόριθμο ιεραρχικής ομαδοποίησης ορίζονται οι ακόλουθες συστάδες μετά από αξιολόγηση των χαρακτηριστικών τους.

Τύποι επισκεπτών	Ανάλογο ποσοστό (%)	Περιγραφή χαρακτηριστικών
GIT	28.01	Επισκέπτες οι οποίοι αγοράζουν ένα ομαδικό πακέτο μέσω αντιπροσώπου. Οι επισκέπτες αυτοί συνήθως μένουν μόνο μία φορά συνεπώς.
GDS	13.70	Επισκέπτες οι οποίοι κάνουν την κράτησή τους μέσω αντιπροσώπου, συνήθως για επαγγελματικούς λόγους
Cancelled- Direct	3.89	Επισκέπτες που έκαναν κράτηση μέσω του ξενοδοχείου όμως την ακύρωσαν.
Transient- Direct	11.92	Επισκέπτες που κάνουν κρατήσεις μέσω του ξενοδοχείου για λόγους αναψυχής, οι κρατήσεις παρατηρούνται τα Σαββατοκύριακα.
MICE	16.51	Επισκέπτες που κάνουν κράτηση απευθείας με κωδικούς προσφοράς για συνέδρια και εκθέσεις
OTA	6.26	Επισκέπτες που κάνουν κράτηση μέσω έμμεσων καναλιών αλλά υπάρχει η δυνατότητα απόκτησης της άμεσης ηλεκτρονικής διεύθυνσης τους.
Cancelled- Indirect	9.66	Επισκέπτες που έκαναν κράτηση μέσω τρίτων καναλιών αλλά την ακύρωσαν.
Repeat	10.70	Επισκέπτες που έχουν κάνει κράτηση παραπάνω από μία φορά όμως το 50% από αυτούς έχει κάνει κράτηση απευθείας.

Πίνακας 4. Κατηγορίες επισκεπτών

Στην συνέχεια αξιολογούνται τα προφίλ των πελατών σε πολλές χρονικές σημάνσεις μεταξύ 2016-2020 χρησιμοποιώντας μηχανική μάθηση. Ένα μοντέλο εκπαιδευμένο σε δεδομένα του 2020 χρησιμοποιείται για την αξιολόγηση των προηγούμενων χρόνων. Ωστόσο οι αλλαγές στα υποκείμενα δεδομένα με την πάροδο του χρόνου σημαίνουν ότι ένα μοντέλο ενδέχεται να μην αξιολογεί με ακρίβεια τα προηγούμενα ή τα μελλοντικά δεδομένα. Η ανάλυση αυτή βοηθά στην κατανόηση των μεταβάσεων μεταξύ κερδοφόρων και λιγότερο επικερδών ομάδων πελατών. Αυτό καθοδηγεί τους επαγγελματίες του μάρκετινγκ σε ενέργειες για την βελτιστοποίηση της μετάβασης πελατών και του προϋπολογισμού μάρκετινγκ.

Στο τέλος , εξετάζονται τα νέα προφίλ επισκεπτών και η εκχώρηση τους στις συστάδες αποκαλύπτοντας αλλαγές και τάσεις με την πάροδο του χρόνου.

Οι επισκέπτες του ξενοδοχείου χωρίζονται σε επαγγελματίες και λάτρεις του ταξιδιού. Η στρατηγική μάρκετινγκ εξελίσσεται από γενικά δελτία σε εξατομικευμένα, στρατηγικά εστιασμένα σε συστάδες όπως Transient - Direct και Repeat. Η αποστολή ηλεκτρονικών δελτίων στοχεύει σε επισκέπτες με βάση τα ταξίδια και τις προτιμήσεις τους, συμμορφούμενη με το GDPR. Επιπλέον, σχεδιάζονται διαφημιστικές καμπάνιες σε πλατφόρμες όπως Google Ads, Instagram και Facebook, χρησιμοποιώντας δεδομένα για να κατανοήσουν τις συγκεκριμένες συνήθειες και προτιμήσεις των συστάδων προσαρμόζοντας την προώθηση για να προσελκύσουν νέους επισκέπτες.

4.4 Click through Rate

Με την ψηφιακή διαφήμιση να καταλαμβάνει τον μεγαλύτερο μέρος των διαφημιστικών εξόδων, οι επιχειρηματίες πρέπει να στοχεύουν στην συνεχή βελτιστοποίηση των καμπανιών τους. Ο δείκτης Click-Through Rate (CTR) ⁽⁶¹⁾ που μετρά τον αριθμό κλικ σε σχέση με τον αριθμό των εμφανίσεων, αναδεικνύει την αποτελεσματικότητα της ψηφιακής διαφήμισης. Για τους διαφημιστές που αποσκοπούν σε άμεση ανταπόκριση, ένα κλικ αντιπροσωπεύει συχνά το πρώτο βήμα προς την αγορά ενός προϊόντος ή την ένδειξη ενδιαφέροντος για την αγορά. Με την ανάλυση των στατιστικών CTR οι εταιρείες μπορούν να καθορίσουν εάν οι διαφημίσεις είναι αποτελεσματικές για τον στόχο τους και να προσαρμόσουν την στρατηγική τους. Ο δείκτης CTR υπολογίζεται ως το ποσοστό των κλικ προς τις εμφανίσεις, παρέχοντας ένα μέτρο αποδοτικότητας της διαφήμισης. Πολλοί παράγοντες επηρεάζουν το CTR, από τη συνάφεια με τα συμφραζόμενα(όπου προβάλλεται μια διαφήμιση, θέση σε μια σελίδα, μορφή διαφήμισης, κλπ.) έως χαρακτηριστικά χρήστη (ποια συσκευή χρησιμοποιεί ένα άτομο, την τοποθεσία, ή δημογραφικά στοιχεία). Τα συστήματα μηχανικής μάθησης (ML) έχουν εξελιχθεί σημαντικά στην ικανότητά τους να προβλέπουν αποτελεσματικά την πιθανότητα ενός κλικ. Η δημιουργία ενός μοντέλου πρόβλεψης, εκτιμά με ακρίβεια αν ένας χρήστης θα κάνει κλικ σε μια διαφήμιση, λαμβάνοντας υπόψη τα δημογραφικά στοιχεία του πελάτη. Συμβάλει στην βελτιστοποίηση των διαφημιστικών καμπανιών, εντοπίζοντας ποιες διαφημίσεις πιθανόν να έχουν υψηλότερο ποσοστό κλικ. Επίσης στη μεγιστοποίηση των εσόδων από διαφημίσεις, τοποθετώντας στρατηγικά υψηλά αποτελεσματικές διαφημίσεις. Και τέλος στη βελτίωση της απόδοσης των διαφημίσεων, εντοπίζοντας αναποτελεσματικές διαφημίσεις και λαμβάνοντας κατάλληλα μέτρα για την ενίσχυσή τους.

Χρησιμοποιείται ο αλγόριθμος Random Forest σε python κώδικα για να προβλέψει το ποσοστό κλικ σε σχέση με τον αριθμό των εμφανίσεων και ελέγχεται ως προς την ακρίβεια του. ⁽⁶²⁾

Γίνεται εισαγωγή των απαραίτητων βιβλιοθηκών. Η βιβλιοθήκη "plotly" χρησιμοποιείται για εύκολη οπτικοποίηση δεδομένων. Η ενότητα 'graph_objects' χρησιμοποιείται για τη δημιουργία και την απεικόνιση των γραφημάτων. Καλούμε την "RandomForestClassifier" για τη μοντελοποίηση της πρόβλεψης του CTR διαφημίσεων.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
import plotly.graph_objects as go
import plotly.express as px
import plotly.io as pio
pio.template.default = "plotly_white"
```

Χρησιμοποιείται ένα σύνολο δεδομένων που ανέβασε ο Gaurav Dutta στο Kaggle.

```
data = pd.read_csv('ad_10000records.csv')
data.head()
```

Ο κώδικας μετατρέπει το περιεχόμενο της στήλης "έκανε κλικ στη διαφήμιση" σε 1=Ναι και 0=Όχι.

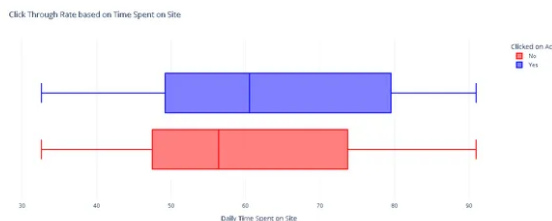
```
data["Clicked on Ad"] = data["Clicked on Ad"].map({0: "No", 1: "Yes"})
```

	Daily Time Spent on Site	Age	Area Income	Daily Internet Usage	Ad Topic Line	City	Gender	Country	Timestamp	Clicked on Ad
0	62.26	32.0	69481.85	172.83	Decentralized real-time circuit	Lisafort	Male	Svalbard & Jan Mayen Islands	2016-06-09 21:43:05	0
1	41.73	31.0	61840.26	207.17	Optional full-range projection	West Angelabury	Male	Singapore	2016-01-18 17:56:05	0
2	44.40	30.0	57877.15	172.83	Total 5thgeneration standardization	Reyesfurt	Female	Guadeloupe	2016-06-29 10:50:45	0
3	59.88	28.0	56180.93	207.17	Balanced empowering success	New Michael	Female	Zambia	2016-06-21 14:32:32	0
4	49.21	30.0	54324.73	201.58	Total 5thgeneration standardization	West Richard	Female	Qatar	2016-07-21 10:54:35	1

Πίνακας 5. Χαρακτηριστικά συνόλου δεδομένων

Ελέγχεται αν ο χρόνος που δαπανήθηκε επηρεάζει το CTR.

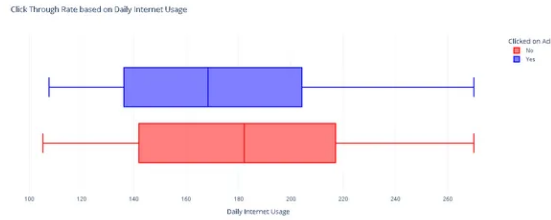
```
fig = px.box(data,
x = "Daily Time Spent on Site"
color = "Clicked on Ad",
title = "Click Through Rate based on Time Spent on Site",
color_discrete_map={'Yes': 'blue', 'No': 'red'})
fig.update_traces(quartilemethod="exclusive")
fig.show()
```



Εικόνα 27. Αναλογία κλικ προς τον χρόνο που δαπανήθηκε.

Βάσει αποτελεσμάτων όσο περισσότερο παραμένει ένας χρήστης στις σελίδες του διαδικτύου τόσα περισσότερα κλικ κάνουν σε διαφημίσεις. Έπειτα, πραγματοποιείται μια ανάλυση για να διαπιστωθεί αν η καθημερινή χρήση του διαδικτύου επηρεάζει το CTR.

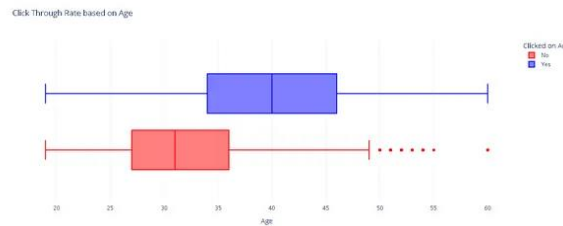
```
fig = px.box(data,
x = "Daily Internet Usage"
color = "Clicked on Ad",
title="Click Through Rate based on Daily Internet Usage",
color_discrete_map={'Yes': 'blue', 'No': 'red'})
fig.update_traces(quartilemethod="exclusive")
fig.show()
```



Εικόνα 28. Αναλογία κλικ προς την καθημερινή χρήση του διαδικτύου

Παρατηρείται πως το ποσοστό των καθημερινών χρηστών κάνουν πιο συχνά κλικ. Στην συνέχεια, εξετάζεται αν η ηλικία του χρήστη επηρεάζει την αναλογία των κλικ.

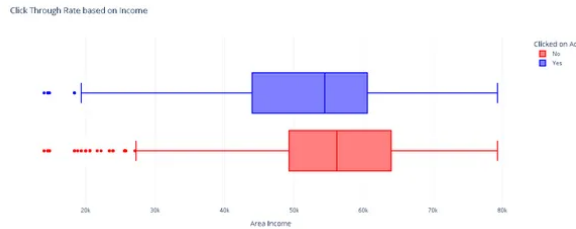
```
fig = px.box(data,
x="Age"
color="Clicked on Ad",
title="Click Through Rate based on Age",
color_discrete_map={'Yes': 'blue', 'No': 'red'})
fig.update_traces(quartilemethod="exclusive")
fig.show()
```



Εικόνα 29. Αναλογία κλικ προς την ηλικία του χρήστη

Οι χρήστες μεταξύ 35-45 ετών κάνουν περισσότερα κλικ στις διαφημίσεις. Τέλος ελέγχεται αν τα έσοδα των χρηστών επηρεάζουν το CTR.

```
fig = px.box(data,
x="Area income"
color="Clicked on Ad",
title="Click Through Rate based on Income",
color_discrete_map={'Yes': 'blue', 'No': 'red'})
fig.update_traces(quartilemethod="exclusive")
fig.show()
```



Εικόνα 30. Αναλογία κλικ ως προς τα έσοδα των χρηστών

Οι χρήστες με υψηλό εισόδημα είναι λιγότερο πιθανό να κάνουν κλικ. Για τον υπολογισμό του CTR πρώτα εξετάζεται η κατανομή των χρηστών και έπειτα προσδιορίζεται το CTR.

```
data["Clicked on Ad"].value_counts()
```

Βάσει αποτελεσμάτων οι 4917 απο τους 10000 έκαναν κλικ

```
click_through_rate = 4917 / 10000 * 100
print(click_through_rate)
```

Βαθμολογία CTR: 49,17

Στην συνέχεια δημιουργείται το μοντέλο που μπορεί να προβλέψει το ποσοστό κλικ σε σχέση με τον αριθμό των εμφανίσεων. Αρχικά, το σύνολο δεδομένων θα χωριστεί σε σύνολα εκπαίδευσης και δοκιμών. Προηγουμένως, οι τιμές της στήλης "Φύλο" θα πρέπει να μετατραπούν σε αριθμούς. Αντικαθιστώντας το "Ανδρας" με το "1" και το "0", κωδικοποιείται η μεταβλητή της κατηγορίας "φύλο" σε δυαδική μορφή για ταχύτερη ανάλυση. Επιπλέον, οι στήλες "Γραμμή θέματος διαφήμισης" και "Πόλη" από το σύνολο δεδομένων "x" θα πρέπει να αφαιρεθούν, καθώς δεν χρησιμοποιούν ως μεταβλητές εισόδου για το μοντέλο μηχανικής μάθησης.

```
data["Gender"] = data["Gender"].map({"Male": 1, "Female": 0})
x=data.iloc[:,0:7]
x=x.drop(['Ad Topic Line', 'City'], axis=1)
y=data.iloc[:,9]

xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2, random_state=33)
```

Αναπτύσσεται το μοντέλο ταξινόμησης για την εκπαίδευση των δεδομένα

```
model = RandomForestClassifier()
model.fit(x,y)
```

Έπειτα υπολογίζεται την ακρίβεια του μοντέλου

```
y_pred = model.predict(xtest)
print(accuracy_score(ytest,y_pred))
```

Η βαθμολογία ακρίβειας είναι 95,2%

Τέλος καταλήγουμε στο στάδιο δοκιμής μοντέλου κάνοντας προβλέψεις με βάση τα υπάρχοντα χαρακτηριστικά.

```
import warnings
warnings.filterwarnings("ignore")

print("Ads Click Through Rate Prediction : ")
a = float(input("Daily Time Spent on site: ") )
b = float(input("Age: ") )
c = float(input("Area Income: ") )
d = float(input("Daily Internet Usage: ") )
e = input("Gender ( Male = 1, Female = 0 ) : ")

features = no.array([[a,b,c,d,e]])
print("Will the user click on ad = " , model.predict(features) )
```

4.5 Απάτη κλικ (Fraud Click)

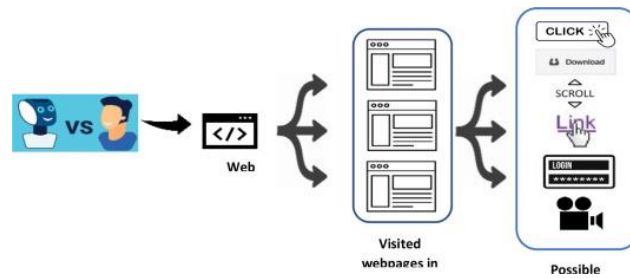
Η απάτη με κλικ ⁽⁶³⁾ αποτελεί μια παραβατική πρακτική στον τομέα της διαφήμισης pay-per-click (PPC) στην οποία ένα άτομο, ένα αυτοματοποιημένο σενάριο, ένα πρόγραμμα υπολογιστή ή ένα αυτόματο κλικ προσομοιώνει τη συμπεριφορά ενός κανονικού χρήστη, κάνοντας κλικ σε διαφημίσεις χωρίς πραγματικό ενδιαφέρον για το περιεχόμενο, με σκοπό να αυξήσουν τα έσοδα τους, ή να προκαλέσουν οικονομική ζημιά στους διαφημιστές. Βασικοί παραβάτες είναι οι ανταγωνιστές, οι επιδιώκουν να αποκομίσουν ανταγωνιστικά πλεονεκτήματα μέσω της εξάντλησης του προϋπολογισμού ανά κλικ. Επίσης οι διαχειριστές ιστοσελίδων, οι οποίοι αποκτούν παράνομα έσοδα κάνοντας αδικαιολόγητα κλικ σε διαφημίσεις για επιπλέον κέρδη και Τα κυκλώματα απάτης, τα οποία εκμεταλλεύονται ειδικά δίκτυα διαφημίσεων για γρήγορα χρηματικά κέρδη.

Μέθοδοι απάτης κλικ

- Crowd sourcing: Μέθοδος κατευθύνουν τους επισκέπτες στον ιστότοπο διαφήμισης με ψεύτικους συνδέσμους ή εικονίδια
- Click Farm: Μέθοδος στην οποία πείθουν ανθρώπους να κάνουν κλικ σε διαφημίσεις ανταλλάσσοντας χρήματα.
- Hit Inflation Attack: Μέθοδος που ανακατευθύνει χρήστες σε διαφημίσεις προτού φτάσουν στη σελίδα προορισμού.
- Botnet: Αποτελούν κακόβουλα λογισμικά που ελέγχουν υπολογιστές, καθοδηγώντας τους να κάνουν κλικ σε διαφημίσεις χωρίς τη γνώση των χρηστών.

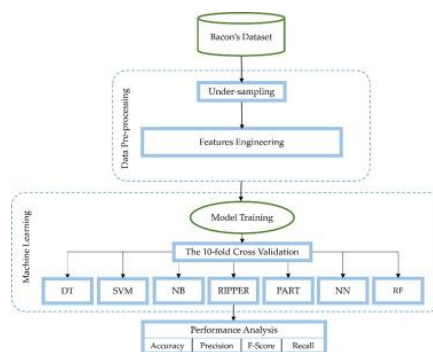
Η μηχανική μάθηση προσφέρει μια ισχυρή και προσαρμοστική λύση για την αντιμετώπιση της απάτης κλικ, καθώς κινητοποιώντας μεγάλα σύνολα δεδομένων και προηγμένους αλγόριθμους μπορεί να εντοπίσει μοτίβα και ανωμαλίες που υποδεικνύουν δόλια συμπεριφορά δίνοντας τη δυνατότητα στις επιχειρήσεις να εντοπίζουν και να αποτρέπουν την απάτη σε πραγματικό χρόνο. ⁽⁶⁴⁾

Οι πειραματισμοί της εργασίας που έκανε ο Malak Aljabri ^[29] πραγματοποιήθηκαν σε συνεργασία με την εταιρεία λογισμικού και υποστήριξης Beacon. Είναι διεθνώς γνωστή για τις λύσεις της στον τομέα του διαδικτυακού μάρκετινγκ και τη διαφημιστική απόδοση. Χρησιμοποιώντας ένα εργαλείο παρακολούθησης που αναπτύχθηκε από τη Beacon, αναλύονται τα δεδομένα συμπεριφοράς χρηστών σε ιστοσελίδες με τεχνικές μηχανικής μάθησης. Επικεντρώνεται στην πρόβλεψη εάν οι χρήστες είναι άνθρωποι ή bots βάσει της συμπεριφοράς τους πριν από την αλληλεπίδραση με διαφημίσεις. Κατά την περιήγηση των χρηστών σε ιστοσελίδες καταγράφηκαν οι ενέργειές τους και στη συνέχεια πραγματοποιήθηκε συσχέτιση με τις δραστηριότητές τους.



Εικόνα 31. Αλληλεπίδραση χρήστη με ιστοσελίδα

Η μεθοδολογία που επιλέχθηκε περιλαμβάνει την προεπεξεργασία των δεδομένων, την εκπαίδευση των μοντέλων ML και τη σύγκριση της απόδοσης των μοντέλων χρησιμοποιώντας σύνολο μετρήσεων.



Εικόνα 32. Μεθοδολογία έρευνας

Τα σύνολα δεδομένων περιέχουν λεπτομερείς πληροφορίες που σχετίζονται με τις ενέργειες των χρηστών κατά την πλοήγηση τους σε διάφορες ιστοσελίδες. Τα δεδομένα παρουσιάζονται σε μορφή JSON και επιστρέφονται στη συνέχεια σε μορφή που μπορεί να αναγνωριστεί από το λογισμικό WEKA, ένα αναγνωρισμένο εργαλείο ανοιχτού κώδικα για την ανάπτυξη ευφυών μοντέλων. Το WEKA περιλαμβάνει μια εκτενή συλλογή αλγορίθμων εξόρυξης δεδομένων και μηχανικής μάθησης, που χρησιμοποιούνται για την ανάλυση και την εξαγωγή συμπερασμάτων από τα δεδομένα. Έπειτα κατηγοριοποιούνται τρεις διαφορετικές κατηγορίες ταξιδιών, bot, Google Bot και άνθρωπος. Δίνεται περισσότερη έμφαση στη διάκριση μεταξύ των ανθρώπων και των bot, ενώ προηγείται η αφαίρεση οποιασδήποτε προσωπικής

πληροφορίας από τα δεδομένα προς ανάλυση. Επιπρόσθετα, λαμβάνοντας υπόψη τη σύσταση της Beacon, το Google bot θεωρείται "Καλό Bot", επομένως δεν συμπεριλήφθηκε στα δεδομένα προς ανάλυση.

Προεπεξεργασία δεδομένων

Το σύνολο δεδομένων των ανθρώπων είναι σημαντικά μεγαλύτερο από το σύνολο δεδομένων των bot με αποτέλεσμα τα δεδομένα να μην εμφανίζουν ισορροπία. Εφαρμόζεται η μέθοδος υποδειγματοληψίας, κατά την οποία επιλέχθηκαν τυχαία 86 δείγματα από το σύνολο δεδομένων των ανθρώπων και συνδυάστηκαν με το σύνολο των bot, με στόχο τη δημιουργία ενός ισορροπημένου συνόλου δεδομένων εκπαίδευσης.

Στη συνέχεια αναλύονται τα χαρακτηριστικά των δεδομένων και τη διαδικασία προεπεξεργασία τους που αφορούν τα χαρακτηριστικά παράγοντα χρήστη τον αριθμό των ιστοσελίδων που επισκέπτονται, τον χρόνο που δαπανάται σε μια διαφήμιση, οι ενέργειες που εκτελέστηκαν καθώς και ο αριθμός των ενεργειών και τέλος το χαρακτηριστικό που υποδεικνύει ότι η επίσκεψη σε μια ιστοσελίδα έχει ολοκληρωθεί σε συνδυασμό με τα παραπάνω χαρακτηριστικά.

Έπειτα ακολουθεί η διασταυρούμενη επικύρωση, δηλαδή μια διαδικασία αξιολόγησης της απόδοσης μαθηματικών μοντέλων στην επεξήγηση των σχέσεων μεταξύ χαρακτηριστικών σε ένα σύνολο δεδομένων. Η διαδικασία αυτή απαιτεί τη διαίρεση του συνόλου δεδομένων σε διαφορετικά τμήματα με ένα τμήμα να χρησιμοποιείται για την επικύρωση του μοντέλου και τα υπόλοιπα για την εκπαίδευση του. Σκοπός είναι να προσδιοριστεί πόσο καλά μπορεί το μοντέλο να γενικεύει σε νέα δεδομένα πέραν του συνόλου εκπαίδευσης. Η τεχνική αυτή συμβάλλει στη αποτίμηση της απόδοσης των μοντέλων μηχανικής μάθησης, αποφεύγοντας την υπερεκπαίδευση και προωθώντας την αξιοπιστία του μοντέλου.

Original Training Dataset										
Divide the Original Dataset into 10-Folds										
	Subset1	Subset2	Subset3	Subset4	Subset5	Subset6	Subset7	Subset8	Subset9	Subset10
Fold1	T	T	T	T	T	T	T	T	T	V
Fold2	T	T	T	T	T	T	T	T	V	T
Fold3	T	T	T	T	T	T	T	V	T	T
Fold4	T	T	T	T	T	T	V	T	T	T
Fold5	T	T	T	T	T	V	T	T	T	T
Fold6	T	T	T	T	V	T	T	T	T	T
Fold7	T	T	T	V	T	T	T	T	T	T
Fold8	T	T	V	T	T	T	T	T	T	T
Fold9	T	V	T	T	T	T	T	T	T	T
Fold10	V	T	T	T	T	T	T	T	T	T
T: Training subset										
V: Validation subset										

Πίνακας 6. 10-fold Cross Validation

Εξετάζονται οι αλγόριθμοι δέντρα αποφάσεων, μηχανής υποστήριξης (SVM), Naive Bayes (NB), Νευρωνικών Δικτύων (NN) και του Τυχαίου Δάσους (RF). Η χρήση των λειτουργιών Metalearner της WEKA χρησιμοποιώντας διασταυρούμενη επικύρωση, επιτρέπουν την βελτιστοποίηση των παραμέτρων για τη δημιουργία αποτελεσματικών μοντέλων.

Μέθοδοι αξιολόγησης

Το ποσοστό ακρίβειας ACC, η βαθμολογία ακρίβειας PRE, η βαθμολογία ανάκλησης Recall και η βαθμολογία F1 χρησιμοποιούνται για τη μέτρηση και την αξιολόγηση της απόδοσης των μοντέλων μηχανικής μάθησης.

Μοντέλο πρόβλεψης	Accuracy	Precision	Recall	F1score
DT	0.82	0.84	0.82	0.82
SVM	0.82	0.84	0.82	0.81
NN	0.83	0.86	0.83	0.83
RF	0.84	0.87	0.84	0.84

Πίνακας 7. Αξιολόγηση απόδοσης μοντέλων για απάτη κλικ.

Τα αποτελέσματα υποδεικνύουν ότι οι αλγόριθμοι εμφανίζουν μια ικανοποιητική απόδοση στην ταξινόμηση ενόψει της πρόκλησης που αντιμετωπίζεται στην αναγνώριση του συνόλου δεδομένου των ανθρώπων. Παρατηρείται ιδιαίτερα ότι ο αλγόριθμος RF υπερέβη τους υπόλοιπους σε όλες τις πτυχές αξιολόγησης με τις αναλογίες ανάκλησης των Bots και των ανθρώπων αποδεικνύουν ότι ο εντοπισμός των Bots είναι πιο ακριβής από τον εντοπισμό των ανθρώπων.

4.6 Chatbot

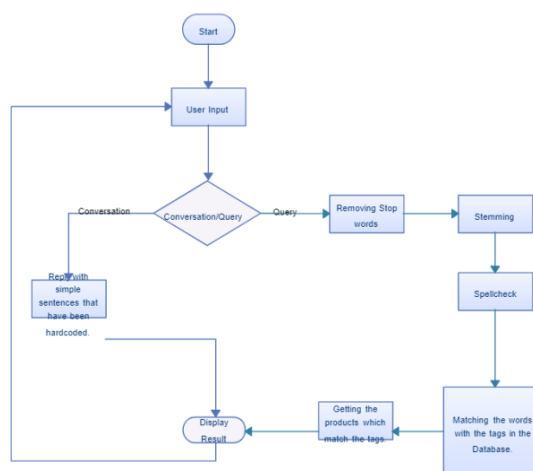
Οι αυτόματοι συνομιλητές μηχανικής εκμάθησης ⁽⁶⁵⁾ οι οποίοι είναι επίσης γνωστοί ως chatbots τεχνητής νοημοσύνης (AI), αντιπροσωπεύουν ένα εξαιρετικό εργαλείο στην εξέλιξη της τεχνολογίας επικοινωνίας. Επικοινωνούν με τους χρήστες απαντώντας ακόμη και τις πιο σύνθετες ερωτήσεις, δίνοντας όσο πιο φυσικές και αληθινές απαντήσεις γίνεται, συμβάλλοντας στη συνεχή βελτίωση της διαδικασίας.

Αξιοπαρατήρητο είναι το γεγονός ότι εξελίσσονται συνεχώς, αποκτώντας γνώσεις μέσω των διαδικτυακών αλληλεπιδράσεων και προσαρμόζοντας συνεχώς τις απαντήσεις τους για ακόμα καλύτερη αποτελεσματικότητα. Επιπλέον, αποτελεί μια εξειδικευμένη μορφή λογισμικού επικοινωνίας, το οποίο σχεδιάζεται για την αυτόματη επικοινωνία με τους χρήστες μέσω εφαρμογών. Αυτό το σύστημα εκμεταλλεύεται θετικά τις προηγμένες τεχνολογίες τεχνητής νοημοσύνης με σκοπό την αποτελεσματική ανταπόκριση στις ανάγκες των χρηστών.

Η μηχανική μάθηση εφαρμόζει πολύπλοκους αλγορίθμους και μοντέλα για την ανάκτηση και την επεξεργασία πληροφοριών. Με την πάροδο του χρόνου, αυτή η τεχνολογία βελτιώνει τις ικανότητες των chatbots ως προς την αποτελεσματικότητα τους και την φυσικότητα στις απαντήσεις τους. Η συνεχής ανάπτυξη παρέχει στους χρήστες πληροφορίες και υπηρεσίες με ακόμη μεγαλύτερη ακρίβεια. Μέσω της μηχανικής μάθησης, η NLP μετατρέπει την ανθρώπινη γλώσσα σε δεδομένα που μπορούν να κατανοηθούν από τους υπολογιστές, επιτρέποντας στα chatbots να ικανοποιήσουν τις ανάγκες των πελατών.

Η διαδικασία εκπαίδευσης των chatbots περιλαμβάνει τη χρήση ενός συνόλου συνομιλιών ως δεδομένων εκπαίδευσης και την ενσωμάτωση των βασικών πληροφοριών. Κατά την αλληλεπίδραση με τους χρήστες, η τεχνολογία NLP αναλύει τα μηνύματα προκειμένου να αναγνωρίσει τις προθέσεις τους και να ανταποκριθεί με ακρίβεια. Αυτή η διαδικασία συνεχίζεται, με τη σταδιακή βελτίωση του chatbot μέσω της επιστροφής πληροφοριών από τους χρήστες, προσφέροντας μια συνεχώς βελτιωμένη εμπειρία χρήσης

και μια σωστή αντιμετώπιση των αιτημάτων τους.



Εικόνα 33. Διάγραμμα ροής chatbot

Οι βασικοί στόχοι των chatbots περιλαμβάνουν την άμεση εξυπηρέτηση και υποστήριξη των πελατών ανά πάσα ώρα και στιγμή, την προσωποποίηση της εμπειρίας επικοινωνίας, τη συλλογή και την ανάλυση δεδομένων, καθώς επίσης τη διαφήμιση και προώθηση των προϊόντων. Με τη χρήση της μηχανικής μάθησης, τα chatbots είναι ικανά να προσαρμόσουν την επικοινωνία τους στις ιδιαίτερες ανάγκες του κάθε πελάτη. Αυτό δημιουργεί μια προσωποποιημένη εμπειρία ενισχύοντας τη σχέση μεταξύ του πελάτη και της επιχείρησης. Επίσης, η συλλογή και η ανάλυση των δεδομένων βοηθά τις επιχειρήσεις να αναπτύξουν μια στρατηγική για την βελτίωση της επικοινωνίας και των πωλήσεων. Επιπρόσθετα μέσω της διαφήμισης και της προώθησης προϊόντων, τα chatbots διευκολύνουν την προώθηση των προσφορών, συμβάλλοντας στην αύξηση της επισκεψιμότητας της ιστοσελίδας ή της εφαρμογής της επιχείρησης και φυσικά στις πωλήσεις. Τέλος, η δυνατότητα των chatbots να εξελίσσονται συνεχώς μέσω της μάθησης από την εμπειρία τους και τα δεδομένα που συγκεντρώνουν, δίνει τη δυνατότητα στις επιχειρήσεις να προσφέρουν πιο προσαρμοσμένες λύσεις στους πελάτες τους, ενισχύοντας την απόδοσή τους στον τομέα του μάρκετινγκ.

Οφέλη Chatbot

Η χρήση των chatbots στο μάρκετινγκ ⁽⁶⁶⁾ είναι ένα σημαντικό εργαλείο που προσφέρει πολλά οφέλη στις επιχειρήσεις. Καταρχάς, οι chatbots συμβάλλουν στην εξοικονόμηση χρόνου και ανθρώπινων πόρων, καθώς αναλαμβάνουν την αυτοματοποιημένη επικοινωνία με τους πελάτες. Με αυτόν τον τρόπο, οι επιχειρήσεις δεν χρειάζεται να διαθέτουν ανθρώπινο προσωπικό για την ανταπόκριση σε ερωτήσεις και ανάγκες των πελατών, εξοικονομώντας έτσι πόρους και μειώνοντας το κόστος λειτουργίας. Επίσης, η διαρκής διαθεσιμότητα των chatbots παρέχει στους πελάτες την αίσθηση ότι μπορούν να λάβουν υποστήριξη ανά πάσα στιγμή, ανεξάρτητα από την ώρα και την ημέρα. Αυτό συμβάλλει στην ενίσχυση

της εμπιστοσύνης των πελατών προς την επιχείρηση και στη βελτίωση της επικοινωνίας τους. Ένα ακόμη σημαντικό όφελος των chatbots είναι η δυνατότητα προσαρμογής στις ανάγκες και τις προτιμήσεις των πελατών μέσω της χρήσης μηχανικής μάθησης. Αυτό επιτρέπει στις επιχειρήσεις να παρέχουν μια εξατομικευμένη εμπειρία στους πελάτες τους, αυξάνοντας την πιθανότητα επιτυχίας των επικοινωνιακών τους προσπαθειών. Επιπλέον, οι chatbots συμβάλλουν στην αύξηση των πωλήσεων, καθώς διευκολύνουν τη διαδικασία αγοράς προϊόντων και υπηρεσιών μέσω της παροχής πληροφοριών και της προσφοράς προσφορών και εκπτώσεων. Ταυτόχρονα, η άμεση ανταπόκριση των chatbots σε ερωτήσεις και ανησυχίες των πελατών συμβάλλει στην ενίσχυση της εμπιστοσύνης και της επικοινωνίας με την επιχείρηση. Τέλος, η χρήση chatbots ως πηγή συλλογής δεδομένων και η ανάλυσή τους παρέχουν στρατηγική κατευθυντήρια γραμμή για το μάρκετινγκ και τις πωλήσεις. Επιπλέον, η μείωση των ανθρώπινων λαθών στην επικοινωνία με τους πελάτες αποτελεί ένα ακόμη όφελος της χρήσης chatbots. Μέσω αυτών των δυνατοτήτων, οι επιχειρήσεις επιτυγχάνουν βελτιωμένη αποτελεσματικότητα, αποδοτικότητα και επικοινωνία με τους πελάτες τους.

Η Eleni Adamopoulou ^[30] επιδιώκει να διερευνήσει τη δυνατότητα περαιτέρω εξατομίκευσης των αλληλεπιδράσεων ανθρώπου-υπολογιστή. Πιο συγκεκριμένα, εξετάζει τη δυνατότητα να αντιστοιχίσει τη προσωπικότητα του χρήστη και την προσωπικότητα των chatbots, χρησιμοποιώντας γλωσσικά μέσα και αναλύοντας τις επιδόσεις του.

Σύνολο δεδομένων

Χρησιμοποιήθηκε ένα δείγμα με περισσότερες από 57.000 αλληλεπιδράσεις συγκεντρώνοντας πληροφορίες για την προσωπικότητα του καταναλωτή, την προσωπικότητα των chatbot, την διάρκεια της συνομιλίας, την αγοραστική συμπεριφορά των καταναλωτών και την γλώσσα απόκρισης των chatbots.

Ανάλυση δεδομένων

Προσωπικότητα καταναλωτή

Για την πρόβλεψη της προσωπικότητας αναλύθηκαν μέσω τεχνητής νοημοσύνης οι αλληλεπιδράσεις με ένα chatbot σε θέματα όπως επαναφόρτιση, υπόλοιπο και λογαριασμό σε μια τηλεφωνική εταιρεία. Οι αλγόριθμοι μηχανικής μάθησης ανέλυσαν τα χαρακτηριστικά και τα ταξινομήσαν ανάλογα με το κυρίαρχο χαρακτηριστικό προσωπικότητας. Τα αποτελέσματα εστίασαν στην διαφοροποίηση μεταξύ εσωστρέφειας και εξωστρέφειας, με τους εσωστρεφείς-εξωστρεφείς να αντιστοιχούν στο 20,4% των ερωτηθέντων.

Προσωπικότητα Chatbot

Τα chatbots παρουσιάζουν διαφορετικές προσωπικότητες, προσανατολισμένες προς εσωστρέφεια και εξωστρέφεια, ανάλογα με τον τρόπο που αντιδρούν στις ερωτήσεις των χρηστών. Τα chatbot που παρουσιάζουν εσωστρέφεια χρησιμοποιούν μια γλώσσα που επικεντρώνεται στην αποτελεσματικότητα,

ενώ αυτά που παρουσιάζουν εξωστρέφεια χρησιμοποιούν μια πιο διεκδικητική γλώσσα που εστιάζει στη δημιουργία μιας εμπειρίας. Για να εξεταστεί η επίδραση της προσωπικότητας, οι πελάτες τυχαιοποιούνται με chatbots κατά τη διάρκεια της διαδικτυακής συνεδρίας χρησιμοποιώντας έναν τυχαίο αλγόριθμο επιλογής.

Διάρκεια αλληλεπίδρασης

Η διάρκεια της αλληλεπίδρασης με το Chatbot αντιπροσωπεύει το επίπεδο δέσμευσης του χρήστη και επηρεάζει τη συχνότητα των επισκέψεων και την απόδοση τους.

Αγοραστική συμπεριφορά

Η αγοραστική συμπεριφορά αξιολογείται με βάση τον αριθμό των υπηρεσιών επαναφόρτισης κινητών τηλεφώνων που προμηθεύτηκαν από πελάτες που επικοινωνήσαν με ένα chatbot.

Χειρισμός

Η συνολική συνέπεια της γλώσσας απόκρισης του chatbot ελέγχθηκε χρησιμοποιώντας ένα δείγμα από 72 πελάτες της Telco και 9 ειδικούς στη γλώσσα. Οι συμμετέχοντες ταξινομήσαν απαντήσεις του chatbot ως εσωστρεφείς ή εξωστρεφείς. Μετά την τμηματοποίηση τους και την τυχαία ομαδοποίηση τους, ένα δείγμα δεχόταν συμβατές απαντήσεις με την προσωπικότητα τους, ενώ το άλλο μη συναφείς απαντήσεις. Τα αποτελέσματα έδειξαν ότι όλοι οι συμμετέχοντες μπορούσαν να αναγνωρίσουν τη γλώσσα απόκρισης του chatbot.

Αποτελέσματα

Προκειμένου να εξεταστεί η σχέση μιας σύμφωνης προσωπικότητας και η αφοσίωση των chatbots για τη μετατροπή των πωλήσεων χρησιμοποιήθηκε η λογιστική παλινδρόμηση και η ανάλυση έδειξε ότι όταν η προσωπικότητα του χρήστη ταιριάζει με την προσωπικότητα του chatbot, η αφοσίωση και οι πωλήσεις αυξάνονται. Επίσης παρατηρήθηκε μεγαλύτερη επίδραση σε αυτή τη σχέση για τους εξωστρεφείς τύπους προσωπικότητας σε σύγκριση με τους εσωστρεφείς. Αντίθετα, όταν δεν υπήρχε αυτή η αντιστοίχιση προσωπικότητας, δεν παρατηρήθηκε σημαντική επίδραση.

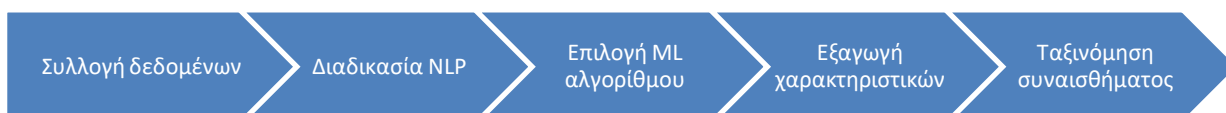
Η αφοσίωση στο Chatbot μετρήθηκε από την διάρκεια αλληλεπίδρασης των καταναλωτών, ενώ η αγοραστική συμπεριφορά καταγράφηκε από τις υπηρεσίες επαναφόρτισης που αγοράστηκαν. Η αντιστοίχιση προσωπικότητας καταναλωτή-chatbot οδηγεί σε υψηλότερη δέσμευση και πιθανότητα αγοράς επαναφόρτισης, με μεγαλύτερη επίδραση στους εξωστρεφείς. Παρατηρήθηκε ότι οι ενώ οι εσωστρεφείς δαπάνησαν περισσότερο χρόνο σε συνομιλίες φάνηκαν πιο διστακτικοί στις αποφάσεις αγοράς, πιθανώς λόγω ανάγκης για περισσότερες πληροφορίες ή χρόνια για ανάλυση προτάσεων.

Εξετάστηκαν οι επιπτώσεις της προσωπικότητας στη δέσμευση και την αγοραστική συμπεριφορά του chatbot βάσει της προσωπικότητας του καταναλωτή και του chatbot. Και στις δυο περιπτώσεις παρουσιάζονται καλύτερα αποτελέσματα για την εξωστρέφεια. Η αντιστοίχιση της προσωπικότητας του καταναλωτή με το chatbot οδηγεί σε υψηλότερα επίπεδα δέσμευσης και σε αυξημένες πωλήσεις.

Οι καταναλωτές ήταν πιο πιθανό να αγοράσουν αφού αλληλεπιδράσουν με ένα chatbot που ταιριάζει με τη προσωπικότητα τους. Τα αποτελέσματα δείχνουν θετικό αντίκτυπο στην αφοσίωση και την επαναφόρτιση των υπηρεσιών κινητής τηλεφωνίας.

4.7 Ανάλυση συναισθήματος

Η ανάλυση συναισθήματος ⁽⁶⁷⁾ αποτελεί μία μέθοδο την επεξεργασίας φυσικής γλώσσας NLP η οποία προσδιορίζει τον συναισθηματικό τόνο μέσα από την ανάλυση ενός ψηφιακού κειμένου και με την βοήθεια της εξόρυξης δεδομένων και της τεχνητής νοημοσύνης μπορεί να αναγνωρίσει εάν ο συναισθηματικός τόνο είναι θετικός, αρνητικός ή και ουδέτερος. Στον χώρο του μάρκετινγκ, η ανάλυση συναισθήματος είναι πολύ σημαντική καθώς επιτρέπει στις εταιρείες να αναλύουν μεγάλους όγκους δεδομένων, όπως σχόλια σε social media, κριτικές και emails βοηθώντας τους να λαμβάνουν ενημερωμένες αποφάσεις και να ενισχύουν την αφοσίωση των πελατών τους και τη φήμη του εκάστοτε brand.



Εικόνα 34. Διαδικασία ανάλυσης συναισθήματος

Για αυτήν την εργασία, ο Pietro Ducange ^[31] προτείνει ένα σύστημα υποστήριξης αποφάσεων για την ταξινόμηση των συναισθημάτων στον τομέα του ψηφιακού μάρκετινγκ και τα μέσα κοινωνικής δικτύωσης χρησιμοποιώντας τον αλγόριθμο SVM.

Ανάλυση απαιτήσεων

Για την ανάπτυξη του συστήματος υποστήριξης αποφάσεων (DSS) για ακρόαση μέσων κοινωνικής δικτύωσης ορίζονται με ακρίβεια οι απαιτήσεις που πρέπει να ικανοποιεί η πλατφόρμα λογισμικού. Με το σύστημα παρακολουθούνται και αναλύονται τα σχόλια των χρηστών που δημοσιεύονται σε δημόσιες σελίδες επιχειρήσεων στα μέσα κοινωνικής δικτύωσης. Με την ανάλυση των σχολίων εξάγεται το κάθε συναίσθημα και το σύστημα δημιουργεί αναφορές και στατιστικά στοιχεία βάσει συναισθημάτων, καθώς επίσης ειδοποιεί για αρνητικά σχόλια την επιχείρηση. Επιπλέον, προσφέρει βασικές λειτουργίες στους χρήστες, όπως ανέβασμα, κατέβασμα και ανάλυση δεδομένων σύμφωνα με την τάση των likes και των ακολούθων σε συγκεκριμένες σελίδες. Τέλος, μπορεί να διαχειρίζεται πολλούς λογαριασμούς σε κοινωνικά δίκτυα ανεξαρτήτως του κοινωνικού δικτύου.

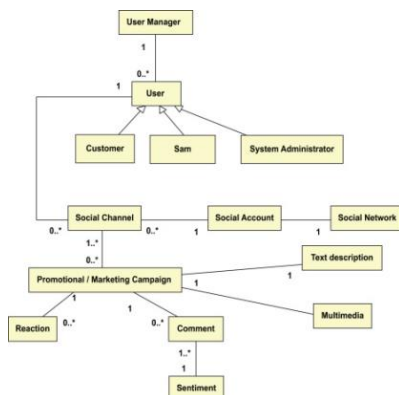
Ταυτοποίηση

Οι χρήστες που αναγνωρίζονται από το σύστημα DSS αφορούν τον πελάτη, ο οποίος διαχειρίζεται τους λογαριασμούς του σε κοινωνικά δίκτυα και έχει πρόσβαση σε διάφορες λειτουργίες. Επιπλέον τον διαχειριστή λογαριασμού κοινωνικής δικτύωσης ο οποίος είναι υπεύθυνος για την διαχείριση των δραστηριοτήτων των πελατών στα κοινωνικά δίκτυα και εκτελεί παρόμοιες λειτουργίες. Τέλος ο

διαχειριστής του συστήματος, ο οποίος χειρίζεται τα προφίλ όλων των χρηστών και έχει πλήρη πρόσβαση σε όλες τις λειτουργίες του συστήματος.

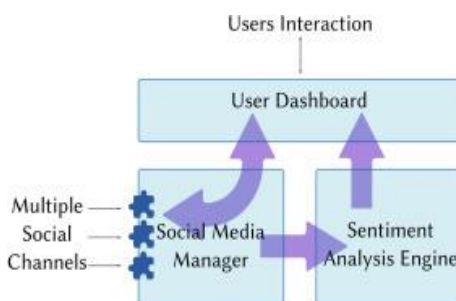
Διαγράμματα τάξης

Χρησιμοποιείται ο φορμαλισμός UML 2 προκειμένου να περιγραφούν οι σχέσεις μεταξύ των τάξεων μετά την ανάλυση των απαιτήσεων. Οι τάξεις αυτές χρησιμοποιούνται για την υλοποίηση των δομών δεδομένων του λογισμικού και των βάσεων δεδομένων.



Εικόνα 35. UML διάγραμμα κύριων τάξεων

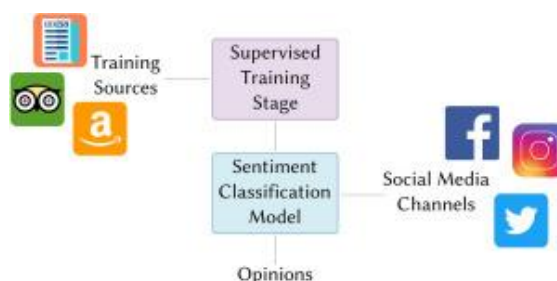
Τελευταίο βήμα για τον σχεδιασμό του DSS είναι ο καθορισμός των κύριων λειτουργικών μονάδων. Τον πίνακα εργαλείων χρήστη, τον υπεύθυνο κοινωνική δικτύωση (SMM) και την ανάλυση συναισθήματος (SAE)



Εικόνα 36. Βασικές λειτουργικές ομάδες DSS

Η ανάλυση συναισθήματος στο σύστημα εκφράζεται μέσω της αναπαράστασης τους κειμένου σε αριθμητικό διανυσματικό χώρο και της ταξινόμησης με χρήση μοντέλων ταξινόμησης μηχανικής μάθησης. Το DSS υιοθετεί μια προσέγγιση βασισμένη σε αναπαράσταση κειμένου Bag-Of-Words και μοντέλα ταξινόμησης. Για να εκπαιδευτούν τα μοντέλα χρειάζεται ένα σύνολο δεδομένων με ετικέτες συναισθήματος, το οποίο πρέπει να προσαρμοστεί στο πλαίσιο εφαρμογής. Έτσι κάθε σενάριο απαιτεί ξεχωριστό μοντέλο ταξινόμησης, που πρέπει να προσαρμοστεί ανάλογα με τον τομέα και το πλαίσιο της εφαρμογής.

Για να αντιμετωπιστεί το ζήτημα της επισημάνσης κειμένων για την εκπαίδευση μοντέλων, χρησιμοποιείται μια μέθοδος για την αυτόματη επιλογή κειμένων με ετικέτα από τις κριτικές των χρηστών που περιέχουν κείμενο και βαθμολογία. Για κάθε σενάριο, επιλέγονται κριτικές που ταιριάζουν με το συγκεκριμένο περιβάλλον και χρησιμοποιούνται ως δεδομένα εκπαίδευσης.



Εικόνα 37. Σχέδιο ανάλυσης μεταξύ πηγών

Για την αναπαράσταση του κειμένου χρησιμοποιείται η τεχνική tokenization, έπειτα υλοποιείται ένα φιλτράρισμα λέξεων και αφαιρούνται οι λέξεις "στοπ" που δεν παρέχουν πληροφορίες στην ανάλυση του κειμένου. Στην συνέχεια, πραγματοποιείται η διαδικασία αναγωγής λέξεων stemming και ένα φιλτράρισμα Stem (στελεχών) και τέλος ακολουθεί η αναπαράσταση των χαρακτηριστικών για να μπορεί να εφαρμοστεί ένας αλγόριθμος ταξινόμησης. Το κείμενο μετατρέπεται σε αριθμητικό διάνυσμα και επισημαίνεται ως θετικό, αρνητικό ή ουδέτερο ανάλογα με το συναίσθημα. Το SAE παράγει ένα σύνολο ταξινομημένων κειμένων από τα σχόλια των χρηστών. Για τον προσδιορισμό των παραμέτρων ενός μοντέλου ταξινόμησης χρησιμοποιούνται οι αλγόριθμοι SVM, Naive Bayes, C4.5 decision trees, K-Nearest Neighbor.

Εκπαίδευση

Για την εκπαίδευση των ταξινομητών σε διάφορα σενάρια, χρησιμοποιούνται ξεχωριστά σετ εκπαίδευσης για κάθε σενάριο. Τα σετ εκπαίδευσης περιλαμβάνουν κριτικές από μέσα κοινωνικής δικτύωσης και ιστότοπους. Για παράδειγμα, για το σενάριο εστιατορίων, συλλέγονται κριτικές από το Trip Advisor, ενώ για το σενάριο ηλεκτρονικών καταστημάτων ευρείας κατανάλωσης συλλέγονται κριτικές από την Amazon. Κάθε σετ περιλαμβάνει θετικές και αρνητικές κριτικές για την εκπαίδευση του μοντέλου, με την ουδέτερη κριτική να προέρχεται από διάφορες πηγές ιστού για να εξασφαλιστεί η αντικειμενικότητα.

Για κάθε σενάριο συγκρίνονται τα προτεινόμενα μοντέλα ταξινόμησης. Τα πειράματα εκτελούνται χρησιμοποιώντας διασταυρούμενη επικύρωση πέντε φορές, όπου το σύνολο χωρίζεται σε πέντε τμήματα. Τα αποτελέσματα της εκτέλεσης αξιολογούνται για τη συνολική απόδοση του κάθε μοντέλου στα συγκεκριμένα σενάρια χρησιμοποιώντας τις κλασσικές μετρήσεις.

	Σενάριο	Accuracy	Precision			Recall			F1score		
			Θετικό	Αρνητικό	Ουδέτερο	Θετικό	Αρνητικό	Ουδέτερο	Θετικό	Αρνητικό	Ουδέτερο
SVM	Εστιατόριο	91.95	90.6	88.1	97.2	88.9	89.3	97.6	89.7	88.7	97.4
SVM	Online κατάστημα	93.01	92.2	91.4	95.5	92	91	96	92.1	91.2	95.7

Πίνακας 8. Ανάλυση καλύτερης απόδοσης μοντέλου σε δύο διαφορετικά σενάρια

Παρατηρείται ότι ο αλγόριθμος ταξινόμησης SVM επιτυγχάνει μεγαλύτερη μέση ακρίβεια 91,95% στο σενάριο του εστιατορίου και 93,01% στο σενάριο ηλεκτρικών ειδών. Συνεπώς τα μοντέλα SVM επιλέγονται για να ενσωματωθούν στην τελική υλοποίηση του συστημάτων DSS ανάλυσης συναισθήματος. Για να δοκιμαστεί το προτεινόμενο σύστημα, πραγματοποιείται μια προσομοιωμένη καμπάνια ακρόασης μέσω κοινωνικών δικτύων. Αρχικά δημιουργούνται ψεύτικες σελίδες σε Facebook, Twitter και Instagram και στη συνέχεια εξετάζονται δημόσιες σελίδες και κανάλια επιχειρήσεων. Η δεύτερη προσέγγιση περιλαμβάνει τη σύνδεση με δημόσια κοινωνικά κανάλια, τη λήψη επιλεγμένων αναρτήσεων και σχολίων και την ταξινόμηση τους από το σύστημα αναγνώρισης συναισθήματος. Για την ταξινόμηση των σχολίων, χρησιμοποιούνται τα μοντέλα ταξινόμησης SVM για τα δύο σενάρια. Το σύστημα καταγράφοντας τα σχόλια από τα κοινωνικά κανάλια πρέπει να μπορεί να ταξινομεί σχόλια σχετικά με διάφορες πτυχές όπως το φαγητό, την εξυπηρέτηση, την αποστολή προϊόντων κ.λπ. Οι αρχικές λέξεις-κλειδιά που αξιοποιούνται για την αναζήτηση των σχολίων συγκεντρώνονται σε σύνολα δεδομένων, μετατρέποντας τις σειρές λέξεων σε διανύσματα. Τα δεδομένα εκπαιδεύονται χρησιμοποιώντας αλγορίθμους μηχανικής μάθησης και τα αποτελέσματα αξιολογούνται από τις μετρήσεις.

Σχόλια	Πηγή	Εκτίμηση Αξιολόγησης	Πραγματική Αξιολόγηση
Αγόρασα ένα S7 και μου το στέλνουν μετά από 3 εβδομάδες! Πολύ κακή εξυπηρέτηση ... αποφύγετε αυτό το κατάστημα		Αρνητικός	Αρνητικός
Παρόλο που δεν μου αρέσουν τα τηλέφωνα Samsung, βρήκα εξαιρετικές προσφορές και καλή υπηρεσία αποστολής		Αρνητικός	Θετικός
Έχετε επιλογές χωρίς γλουτένη;		Ουδέτερος	Ουδέτερος
Αχχχχ νοστιμιο		Θετικός	Θετικός

Πίνακας 9. Παραδείγματα ταξινόμησης σχολίων.

Με πάνω από 100.000 επιλεγμένα σχόλια, πραγματοποιείται επαλήθευση για την ορθότητα της ταξινόμησης σε περισσότερα από 500 τυχαία σχόλια για κάθε σενάριο. Ανά σενάριο επιλέγονται τυχαία τρεις ομάδες που περιέχουν 170 σχόλια χωρισμένα σε θετικά, αρνητικά και ουδέτερα. Τα αποτελέσματα επιβεβαιώνουν τη σταθερότητα του συστήματος από την προηγούμενη αξιολόγηση. Παρόλο που η απόδοση μπορεί να μειωθεί λόγω μεταβολών στα χαρακτηριστικά των δεδομένων, η μελέτη δείχνει ότι το σύστημα παραμένει σταθερό και αποδοτικό.

4.8 Δέσμευση πελατών

Οι Pansari & Kumar ^[32] θεωρούν ότι η σχέση μεταξύ του πελάτη και της επιχείρησης εξαρτάται από την ικανοποίηση, την εμπιστοσύνη, και τη δέσμευση που έχει ο κάθε πελάτης. Για να δημιουργηθεί μια σχέση μεταξύ της επιχείρησης και του πελάτη πρέπει να υπάρχει άμεση ή έμμεση αλληλεπίδραση μεταξύ τους.

Αυτό μπορεί να γίνει είτε μέσω συζητήσεων ή αντιδράσεων στα μέσα κοινωνικής δικτύωσης, συζητήσεις άλλοτε με την επιχείρηση και άλλοτε μεταξύ των πελατών, είτε μέσω των κριτικών που αναρτά ο κάθε πελάτης για κάποιο προϊόν.

Αυτές οι δραστηριότητες σύμφωνα με την ικανοποίηση και την σύνδεση του πελάτη με την επιχείρηση βοηθούν στην δημιουργία μιας δυνατής σχέσης που βασίζεται στην εμπιστοσύνη. Η εμπιστοσύνη περιγράφεται ως η πεποίθηση και η πίστη που έχει ένας πελάτης να ενεργεί με ηθικό τρόπο και να δημιουργεί σχέσεις ανταλλαγής που εκτιμώνται ιδιαίτερα (Morgan and SD Hunt. (1994)). ^[33]

Όσο περισσότερο αλληλεπιδρά ένας πελάτης, τόσο πιο επωφελής γίνεται η σχέση για την επιχείρηση, καθώς αναπτύσσει πιο αποτελεσματικές στρατηγικές μάρκετινγκ και αποκτά πρόσβαση σε περισσότερες προσωπικές πληροφορίες. Η παρουσία μιας επιχείρησης στα κοινωνικά δίκτυα κάνει ακόμη πιο εύκολη την δημιουργία αυτής της σχέσης καθώς οι επιχειρήσεις αλληλεπιδρούν άμεσα με τους πελάτες και με την σειρά τους οι πελάτες μοιράζονται πληροφορίες μεταξύ τους στα σχόλια. Έτσι δεν χτίζουν σχέση μόνο με τους υπάρχοντες πελάτες όμως και με πιθανούς μελλοντικούς πελάτες. Αυτή η διαδραστικότητα δίνει στον πελάτη έναν νέο ρόλο καθώς μέσω της αλληλεπίδρασης του βοηθά στην προώθηση και την αύξηση της αξίας μιας παροχής ή ενός προϊόντος.

Πέρα όμως από τη θετική δέσμευση του πελάτη μπορεί να δημιουργηθεί αρνητική δέσμευση ανάλογα με το σθένος της συμπεριφοράς η οποία μπορεί να επηρεάσει αρνητικά την φήμη και την ανάπτυξη της επιχείρησης και να μεταδοθεί ραγδαία στο διαδίκτυο. Do et al., 2020 ^[34]

Με τη χρήση της μηχανικής μάθησης μπορεί να αναλυθεί γρήγορα και αποτελεσματικά ένα μεγάλο σύνολο δεδομένων από σχόλια και κριτικές, βελτιώνοντας τη σχέση της επιχείρησης με τον πελάτη, μεγιστοποιώντας τις θετικές επιπτώσεις και προλαμβάνοντας τις ζημιές που μπορεί να επιφέρει μία αρνητική δέσμευση.

Η Dinda Thalia Andariesta ^[35] προτείνει την ανάπτυξη μοντέλων πρόβλεψης που μπορούν να προβλέπουν το επίπεδο αλληλεπίδρασης των αναρτήσεων στο Twitter για επιχειρήσεις, με βάση τις δραστηριότητες τους στα μέσα κοινωνικής δικτύωσης βελτιώνοντας τη στρατηγική τους.

Σύνολο δεδομένων

Για την συλλογή των δεδομένων έγινε ανάκτηση πληροφοριών από τρεις λογαριασμούς ηλεκτρονικού εμπορίου @tokopedia, @bukalapak, @bliblidotcom στο Twitter. Κατά την διάρκεια επτά μηνών συλλέχτηκαν συνολικά 12,786 αναρτήσεις οι οποίες συγκέντρωσαν 11,870,254 likes και 2.735.886 retweets.

Προετοιμασία δεδομένων

Σε αυτό το στάδιο εξετάζονται τα χαρακτηριστικά που υπάρχουν σε μία ανάρτηση,

συμπεριλαμβανομένων κάποιων μεταβλητών πρόβλεψης, όπως τη χρήση συνδέσμων, hashtags, εικόνων, βίντεο καθώς και την ημέρα ανάρτησης. Η ημέρα ανάρτησης θεωρείται αριθμητική μεταβλητή καθώς αναφέρεται στον χρόνο που πέρασε από την τελευταία ανάρτηση δηλαδή στην συχνότητα ανάρτησης. Ενώ οι υπόλοιπες μεταβλητές, θεωρούνται δυαδικές οπότε καταγράφονται ως "Ναι" ή "Όχι".

Αξιολόγηση δεδομένων

Αναλύονται οι αναρτήσεις για να μετρηθούν τα ποσοστά αλληλεπίδρασης στο Twitter. Τα ποσοστά κατηγοριοποιούνται σε τέσσερα επίπεδα, χαμηλό 0-0,02%, καλό (0,02-0,09%) , υψηλό (0,09-0,33%) και υψηλό (> 0,33%) με βάση το ποσοστό αλληλεπίδρασης, για να αξιολογηθούν ως προς την αποτελεσματικότητά τους. Στην συνέχεια, υπολογίζονται τα συνολικά ποσοστά αλληλεπίδρασης από το σύνολο των retweets και likes σε κάθε ανάρτηση, και κατηγοριοποιούνται βάσει του αριθμού των ακολούθων, για να προσδιοριστεί η δυναμική της αλληλεπίδρασης. Τα αποτελέσματα της ανάλυσης παρουσιάζουν ανισορροπία οπότε χωρίζεται το σύνολο δεδομένων σε 80% για εκπαίδευση και 20% για δοκιμές. Χρησιμοποιείται μια μέθοδος δειγματοληψίας, η οποία διαμορφώνει τον πληθυσμό σε στρώματα και επιλέγει δείγματα από κάθε στρώμα συνδυάζοντάς τα για να εκτιμήσει τις παραμέτρους του πληθυσμού διασφαλίζοντας την ισορροπία των κλάσεων.

Τα μοντέλα πρόβλεψης που αξιολογούνται είναι τα ακόλουθα: Logistic regression, Decision Tree, K-nearest Neighbor, Random Forest

	Class	LR	DT	KNN	RF
Tokopedia	Μέσος όρος Τάξεων	Precision: 50.2% Recall: 50.6% F1score: 50.4%	Precision: 57.7% Recall: 59.4% F1score: 58.6%	Precision: 53.7% Recall: 55.5% F1score: 67.3%	Precision: 63.7% Recall: 77% F1score: 71.8%
Bukalapak	Μέσος όρος Τάξεων	Precision: 31% Recall: 37.5% F1score: 34%	Precision: 30.3% Recall: 45.5% F1score: 36%	Precision: 34.4% Recall: 71.5% F1score: 46.4%	Precision: 46.1% Recall: 78.6% F1score: 58.1%
Blibli	Μέσος όρος Τάξεων	Precision: 23% Recall: 43% F1score: 33.4%	Precision: 28.3% Recall: 59.55% F1score: 38.4%	Precision: 30.5% Recall: 79.4% F1score: 44%	Precision: 35.3% Recall: 84.3% F1score: 49.8%

Πίνακας 10. Αξιολόγηση απόδοσης μοντέλων για δέσμευση πελατών.

Αποτελέσματα

Παρατηρείται ότι το μοντέλο DT απαιτεί λιγότερο χρόνο επεξεργασίας και στους τρεις λογαριασμούς, όμως η απόδοση του είναι χαμηλότερη από τα μοντέλα RF και το kNN. Το RF εμφανίζει καλύτερο AUC, παρόλα αυτά η ανάπτυξη του απαιτεί περισσότερο χρόνο. Επίσης παρέχει την καλύτερη τιμή ανάκλησης των δεδομένων υποδηλώνοντας ότι παρουσιάζει χαμηλό ποσοστό λανθασμένων αρνητικών προβλέψεων. Τις χαμηλότερες αποδόσεις παρατηρείται να τις σημειώνει το μοντέλο λογιστικής παλινδρόμησης LR. Συμπεραίνουμε πως το RF αν και απαιτεί περισσότερο χρόνο από τα υπόλοιπα μοντέλα, προσφέρει υψηλότερη απόδοση στην ταξινόμηση ,ανθεκτικότητα στην προσαρμογή και αποτελεσματική αντιμετώπιση προβλημάτων υψηλής διάστασης. Τα μοντέλα δείχνουν ισορροπία στην ακρίβεια και την ανάκληση για την εταιρεία @tokopedia αλλά υψηλό ποσοστό λανθασμένων θετικών προβλέψεων στα άλλες δυο εταιρείες. Η κάθε μία διαφοροποιείται στα μοτίβα των αναρτήσεων, με την tokopedia να χρησιμοποιεί περισσότερα διαδραστικά στοιχεία, ενώ η Blibli χρησιμοποιεί περισσότερες εικόνες και hashtags.

4.9 Συγκεντρωτικός Πίνακας Αποτελεσμάτων

	Μοντέλο πρόβλεψης	Αποτέλεσμα
Πρόβλεψη συμπεριφοράς	XGBoost	Μέσα από δύο σει πειραμάτων αντίθεσης, αποδεικνύεται ότι ο αλγόριθμος που επιλέχθηκε σε αυτή την εργασία μπορεί να φιλτράρει αποτελεσματικά τα χαρακτηριστικά, γεγονός που απλοποιεί την πολυπλοκότητα του μοντέλου σε κάποιο βαθμό και βελτιώνει την ακρίβεια ταξινόμησης της μηχανικής μάθησης. AUC > 0.76 F1 > 0.90
Πρόβλεψη αποχώρησης	Stochastic Gradient Boost	Η καμπύλες ROC και AUC εμφάνισαν την καλύτερη απόδοση ίση με 0.84 και ακρίβεια 0.83 συμβάλλοντας σε καλύτερες προβλέψεις και μειώνοντας το κόστος.
Τμηματοποίηση αγοράς	Ιεραρχική ομαδοποίηση	Παρέχεται μια προσέγγιση βάσει δεδομένων ομαδοποιώντας τα προφίλ επισκεπτών με βάση ένα εκτεταμένο σύνολο χαρακτηριστικών. Τα αποτελέσματα συμβάλλουν στην προσαρμοστικότητα των τμημάτων μάρκετινγκ ώστε να λαμβάνουν αποφάσεις βάσει δεδομένων και τελικά να οδηγούν σε κέρδος.
Click Through Rate	Τυχαίο Δάσος	Αναπτύχθηκε ο αλγόριθμος με python και χρησιμοποιώντας ένα σύνολο δεδομένων η ακρίβεια απόδοσης έφτασε στο 95.2%
Απάτη κλικ	Τυχαίο Δάσος	Ο αλγόριθμος εμφανίζει μια ικανοποιητική απόδοση ακρίβειας 84% και μπορεί να διαπιστώσει εάν ο χρήστης είναι άνθρωπος ή bot. Ο εντοπισμός των bots είναι πιο ακριβείς από τον εντοπισμό των ανθρώπων.
Chatbot	Λογιστική παλινδρόμηση	Χρησιμοποιώντας τον αλγόριθμο παλινδρόμησης, η προσωπικότητα του καταναλωτή μπορεί να προβλεφθεί μέσω των αλληλεπιδράσεων, επιτρέποντας στα chatbot να προσομοιώσουν μια προσωπικότητα. Επηρεάζοντας θετικά την αφοσίωση των καταναλωτών και τα αποτελέσματα αγορών.
Ανάλυση συναισθήματος	Μηχανές διανυσμάτων υποστήριξης	Εξετάζοντας δυο διαφορετικά σενάρια και αξιοποιώντας πάνω από 100.000 σχόλια ο αλγόριθμος SVM έδειξε την καλύτερη απόδοση ταξινόμησης και στις δύο περιπτώσεις με 90,1% και 93,01 %.
Δέσμευση πελατών	Τυχαίο δάσος	Η απόδοση της ακρίβειας με 83.9% δείχνει ότι παίζει καθοριστικό ρόλο στην πρόβλεψη της δέσμευσης και το μοντέλο μπορεί να υποστηρίξει τους διαχειριστές μάρκετινγκ και κοινωνικών μέσων στη βελτίωση της στρατηγικής τους στα κοινωνικά μέσα.

Κεφάλαιο 5 - Συμπεράσματα

Το μέλλον της μηχανικής μάθησης (ML) στο μάρκετινγκ είναι έτοιμο για μετασχηματικές εξελίξεις, με αρκετές βασικές τάσεις και εξελίξεις να επαναπροσδιορίζουν το τοπίο. Η προγνωστική ανάλυση αναδύεται ως βασικός άξονας, όπου οι αλγόριθμοι ML αξιοποιούν τη δύναμη των ιστορικών δεδομένων για την πρόβλεψη και την κατανόηση της συμπεριφοράς των πελατών με πρωτοφανή ακρίβεια. Αυτό δίνει τη δυνατότητα στους επαγγελματίες του μάρκετινγκ να εφαρμόζουν εξαιρετικά στοχευμένες καμπάνιες, εξατομικευμένες προτάσεις και εξελιγμένα συστήματα βαθμολόγησης δυνητικών πελατών, οδηγώντας τα ποσοστά αφοσίωσης και μετατροπών σε νέα ύψη. Η υπερ-εξατομίκευση γίνεται κεντρικό θέμα στις στρατηγικές μάρκετινγκ χάρη στην ικανότητα της ML να προσαρμόζει δυναμικά τις προτάσεις περιεχομένου, τις προτάσεις προϊόντων και τις διαφημίσεις με βάση περίπλοκα μοτίβα στη συμπεριφορά των μεμονωμένων χρηστών.

Αυτό το επίπεδο εξατομίκευσης υπερβαίνει την τμηματοποίηση, επιτρέποντας στους επαγγελματίες του μάρκετινγκ να δημιουργούν προσαρμοσμένες εμπειρίες που έχουν μεγαλύτερη απήχηση στους καταναλωτές, ενισχύοντας την αφοσίωση της επωνυμίας και την ικανοποίηση των πελατών. Ο αυτοματισμός φτάνει σε νέα ύψη στις προσπάθειες μάρκετινγκ. Οι αλγόριθμοι ML δεν αυτοματοποιούν μόνο εργασίες ρουτίνας, αλλά εμβαθύνουν επίσης στη δημιουργία περιεχομένου, από τη δημιουργία αντιγράφων διαφήμισης έως τη παραγωγή οπτικά συναρπαστικών στοιχείων. Αυτό όχι μόνο εξορθολογεί τη δημιουργική διαδικασία, αλλά ενισχύει επίσης τη συνολική αποτελεσματικότητα, απελευθερώνοντας τους επαγγελματίες του μάρκετινγκ να επικεντρωθούν σε στρατηγικές πτυχές των καμπανιών τους.

Η δυναμική βελτιστοποίηση τιμολόγησης είναι ένας άλλος τομέας όπου η ML έχει σημαντικό αντίκτυπο. Αναλύοντας δεδομένα σε πραγματικό χρόνο, όπως η ζήτηση, η τιμολόγηση των ανταγωνιστών και οι συνθήκες της αγοράς, οι αλγόριθμοι ML δίνουν τη δυνατότητα στις επιχειρήσεις να προσαρμόζουν τις στρατηγικές τιμολόγησης αμέσως. Αυτό διασφαλίζει τη βέλτιστη τιμολόγηση για μέγιστη κερδοφορία και ανταγωνιστικότητα σε μια ταχέως μεταβαλλόμενη αγορά. Το μάρκετινγκ συνομιλίας βιώνει μια μεγάλη επανάσταση λόγω των chatbots που υποστηρίζονται από ML. Αυτά τα εξελιγμένα ρομπότ μπορούν να προσελκύσουν χρήστες σε πραγματικό χρόνο, παρέχοντας αλληλεπιδράσεις με επίγνωση περιβάλλοντος και εξατομικευμένες αλληλεπιδράσεις.

Καθώς τα όρια μεταξύ του παραδοσιακού μάρκετινγκ και της εξυπηρέτησης πελατών θολώνουν, το μάρκετινγκ συνομιλίας που βασίζεται σε ML γίνεται ισχυρό εργαλείο για τις επωνυμίες να συνδέονται με το κοινό τους σε βαθύτερο επίπεδο. Η ικανότητα της ML στην τμηματοποίηση και τη στόχευση πελατών ωθεί τα όρια της ακρίβειας του μάρκετινγκ. Οι επαγγελματίες μπορούν πλέον να εντοπίσουν διαφορετικά μοτίβα και συμπεριφορές εντός του κοινού-στόχου τους, επιτρέποντας πιο αναλυτική και αποτελεσματική προσαρμογή καμπάνιας. Αυτό έχει ως αποτέλεσμα υψηλότερα ποσοστά μετατροπής και πιο σημαντική απόδοση στις επενδύσεις μάρκετινγκ.

Ο εντοπισμός και η πρόληψη της απάτης αποτελούν κρίσιμα ζητήματα στο ψηφιακό μάρκετινγκ και η μηχανική μάθηση διαδραματίζει καθοριστικό ρόλο στη διαφύλαξη των προϋπολογισμών από σπάταλες δαπάνες. Οι προηγμένοι αλγόριθμοι μπορούν να αναλύουν μοτίβα και ανωμαλίες σε πραγματικό χρόνο, εντοπίζοντας και μετριάζοντας πιθανές δόλιες δραστηριότητες, όπως απάτη κλικ ή δημιουργία ψεύτικου λογαριασμού.

Τα εργαλεία ανάλυσης συναισθήματος, που υποστηρίζονται από ML προσφέρουν στους επαγγελματίες του μάρκετινγκ τη δυνατότητα να μετρούν την κοινή γνώμη και το συναίσθημα της επωνυμίας σε πολλά κανάλια. Αυτός ο βρόχος ανάδρασης σε πραγματικό χρόνο δίνει τη δυνατότητα στους επαγγελματίες του μάρκετινγκ να προσαρμόζουν τις καμπάνιες τους άμεσα αντιμετωπίζοντας ανησυχίες ή αξιοποιώντας το θετικό συναίσθημα για να ενισχύσουν τη φήμη της επωνυμίας.

Η βελτιστοποίηση μάρκετινγκ μεταξύ καναλιών επωφελείται ουσιαστικά από τις αναλυτικές δυνατότητες της μηχανικής μάθησης. Αναλύοντας τα δεδομένα απόδοσης σε διάφορα κανάλια, οι έμποροι μπορούν να προσδιορίσουν τις πιο αποτελεσματικές πλατφόρμες για συγκεκριμένα τμήματα κοινού. Αυτή η προσέγγιση βάσει δεδομένων επιτρέπει πιο στρατηγική κατανομή πόρων και προϋπολογισμού για τη μεγιστοποίηση του αντίκτυπου. Η επαυξημένη ανάλυση, που καθοδηγείται από την ML, εγκαινιάζει μια νέα εποχή λήψης αποφάσεων με γνώμονα τα δεδομένα.

Αυτά τα εργαλεία αυτοματοποιούν τη δημιουργία πληροφοριών και προτάσεων, παρέχοντας στους επαγγελματίες του μάρκετινγκ ενεργή νοημοσύνη που μπορεί να λειτουργήσει ώστε να βελτιώνουν και να βελτιστοποιούν συνεχώς τις στρατηγικές τους.

Βιβλιογραφία

- [1] Jeffrey W. Seifert. (2004) Data mining: An overview. National security issues, Chapter 11. ([Google Book](#))
- [2] Chon Ho, Yu. (2010). Exploratory data analysis in the context of data mining and resampling. International Journal of Psychological Research, 3(1), 9-22. ([PDF](#))
- [3] E. Triantaphyllou. (2010) Data mining and knowledge discovery via logic-based methods: theory, algorithms and applications-Part I: Algorithmic Issues. ([Google Book](#))
- [4] Indranil Bose, Radha K. Mahapatra: Business data mining - a machine learning perspective. Information & Management Volume 39, Issue 2, Pages 211-225 [https://doi.org/10.1016/S0378-7206\(01\)00091-X](https://doi.org/10.1016/S0378-7206(01)00091-X)
- [5] P. Harrington (2012). Machine learning in action ([Google Book](#))
- [6] Kristina P. Sinaga, Miin-Shen Yang (2020) Unsupervised K-means clustering algorithm IEEE Access, Volume: 8 Page(s): 80716 – 80727 [doi: 10.1109/ACCESS.2020.2988796](https://doi.org/10.1109/ACCESS.2020.2988796)
- [7] Batta Mahesh "Machine Learning Algorithms - A Review", International Journal of Science and Research (2018) ([PDF](#))
- [8] [32] Harsh h. Patel, Purvi Prajapati "Study and analysis of Decision tree Based Classifications algorithms" International Journal of Computer Sciences and Engineering, Vol. 6, Issue 10 (2018) ([PDF](#))
- [9] Daniel Berrar "Bayes' Theorem and Naive Bayes Classifier" Tokyo Institute of Technology 2019 Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, Pages 403-412 ([Google Book](#))
- [10] Jerome H. Friedman, "Stochastic Gradient Boosting", Computational Statistics & Data Analysis, Volume 38, Issue 4, Pages 367-378 (2002)
- [11] Anett Erdmann, Ramón Arilla, José M. Ponzoa "Search engine optimization: The long-term strategy of keyword choice" Journal of Business Research Volume 144, (2022) Pages 650-662 <https://doi.org/10.1016/j.jbusres.2022.01.065>
- [12] Eva Lahuerta Otero , Pablo A. Muñoz Gallego , Renée M.E. Pratt "Click-and-Mortar SMEs: Attracting customers to your website" Business Horizons Volume 57, Issue 6, November–December 2014, Pages 729-736 <https://doi.org/10.1016/j.bushor.2014.07.006>
- [13] Zofija Tupikovskaja-Omovie, David Tyler "Eye tracking technology to audit google analytics: Analysing digital consumer shopping journey in fashion m-retail" International Journal of Information Management Volume 59, August (2021) 102294 <https://doi.org/10.1016/j.ijinfomgt.2020.102294>

- [14] Matthew D. Dean, Dinah M. Payne, Brett J.L Landry "Data mining: an ethical baseline for online privacy policies" *Journal of Enterprise Information Management* (2016). <https://doi.org/10.1108/JEIM-04-2014-0040>
- [15] Anthony J. Rhem "Ethical Use of Data in AI Applications" *Ethics - Scientific Research, Ethical Issues, Artificial Intelligence and Education* Miroslav Radenkovic (2023) DOI: 10.5772/intechopen.1001597
- [16] Fiona Ellis-Chadwick , Neil F. Doherty "Web advertising: The role of e-mail marketing *Journal of Business Research* Volume 65, Issue 6, (2012) Pages 843-848 <https://doi.org/10.1016/j.jbusres.2011.01.005>
- [17] Camelia Budac, Lucian Blaga "Theoretical Approaches on Successful Email Marketing Campaigns", University of Sibiu, Romania ([PDF](#))
- [18] Bilal Aslam, Heikki Karjaluo "Digital advertising around paid spaces, E-advertising industry's revenue engine: A review and research agenda", *Telematics and Informatics* Volume 34, Issue 8, (2017) Pages 1650-1662 <https://doi.org/10.1016/j.tele.2017.07.011>
- [19] Michel van der Borgh, Edwin J. Nijssen, Jeroen J.L. Schepers "Unleash the power of the installed base: Identifying cross-selling opportunities from solution offerings" *Industrial Marketing Management* Volume 108, (2023), Pages 122-133 <https://doi.org/10.1016/j.indmarman.2022.11.010>
- [20] Sima Magatef , Manaf Al-Okaily , Laila Ashour , Tala Abuhussein "The impact of electronic customer relationship management strategies on customer loyalty: A mediated model" *Journal of Open Innovation: Technology, Market, and Complexity* Volume 9, Issue 4, December 2023, 100149 <https://doi.org/10.1016/j.joitmc.2023.100149>
- [21] [58] Maryam I.Y. Suleiman, Dr. Norhilmi Bin Muhammad, Ibrahim Suleiman Yahaya, Dr. Magaji Abubakar Adamu, Amina Usman Sabo "Benefits And Challenges: For Marketing Strategies On-Line" *European Journal of Molecular & Clinical Medicine* ISSN 2515-8260 Volume 07, Issue 03, (2020)
- [22] Peter S.H. Leeftang, Peter C. Verhoef, Peter Dahlström , Tjark Freundt "Challenges and solutions for marketing in a digital era " *European Management Journal* Volume 32, Issue 1, (2014) Pages 1-12 <https://doi.org/10.1016/j.emj.2013.12.001>
- [23] K Uday Kiran and Thangaraja Arumugam "Role of programmatic advertising on effective digital promotion strategy: A conceptual framework" *Journal of Physics: Conference Series*, Volume 1716, National Science, Engineering and Technology Conference (NCSET) (2020) Vellore Institute of Technology, Chennai, India [doi:10.1088/1742-6596/1716/1/012032](https://doi.org/10.1088/1742-6596/1716/1/012032)
- [24] Zhengyi Hu "Consumer Behavior Prediction Based on Machine Learning Scenarios", SHU Business School, Sheffield Hallam University, Sheffield, UK ([PDF](#))
- [25] Cheng-Ju Liu,Tien-Shou Huang,Ping-Tsan Ho ,Jui-Chan Huang,Ching-Tang Hsieh "Machine learning-based e-commerce platform repurchase customer prediction model" (2020) <https://doi.org/10.1371/journal.pone.0243105>

- [26] "Customer churning analysis using machine learning algorithms" B. Prabadevi, R. Shalini, B.R. Kavitha, *International Journal of Intelligent Networks* Volume 4, (2023) Pages 145-154 <https://doi.org/10.1016/j.ijin.2023.05.005>
- [27] McIlroy A., Barnett S. Building customer relationship: Do discount cards work? *Managing Service Quality*, 10 (6) (2000), pp. 347-355
- [28] Rik van Leeuwen, Ger Koole "Data-driven market segmentation in hospitality using unsupervised machine learning", *Machine Learning with Applications* Volume 10, 15 December 2022, 100414 <https://doi.org/10.1016/j.mlwa.2022.100414>
- [29] [91] Malak Aljabri , Rami Mustafa A. Mohammad "Click fraud detection for online advertising using machine learning" *Egyptian Informatics Journal* Volume 24, Issue 2, (2023) Pages 341-350 <https://doi.org/10.1016/j.eij.2023.05.006>
- [30] Eleni Adamopoulou, Lefteris Moussiades "Chatbots: History, technology, and applications" *Machine Learning with Applications* Volume 2 (2020) 100006 <https://doi.org/10.1016/j.mlwa.2020.100006>
- [31] Pietro Ducange, Michela Fazzolari, Marinella Petrocchi, Massimo Vecchio "An effective Decision Support System for social media listening based on cross-source sentiment analysis models" *Engineering Applications of Artificial Intelligence* Volume 78 (2019) Pages 71-85 <https://doi.org/10.1016/j.engappai.2018.10.014>
- [32] Pansari, A., & Kumar, V. (2017). Customer engagement: the construct, antecedents, and consequences. *Journal of the Academy of Marketing Science*, 45(3), 294–311. <https://doi.org/10.1007/s11747-016-0485-6>
- [33] Archana Nayak Kini, Basri Savitha, Iqbal Thonse Hawaldar "Brand loyalty in FinTech services: The role of self-concept, customer engagement behavior and self-brand connection", *Journal of Open Innovation: Technology, Market, and Complexity* Volume 10, Issue 1, (2024) 100240 <https://doi.org/10.1016/j.joitmc.2024.100240>
- [34] Hengyun Li a, Hongbo Liu, Hyejo Hailey Shin, Haipeng Ji "Impacts of user-generated images in online reviews on customer engagement: A panel data analysis", *Tourism Management* Volume 101, (2024) 104855 <https://doi.org/10.1016/j.tourman.2023.104855>
- [35] Dinda Thalia Andariesta, Meditya Wasesa "Machine learning models to predict the engagement level of Twitter posts: Indonesian e-commerce case study" *Procedia Computer Science* Volume 227, (2023) Pages 823-832 <https://doi.org/10.1016/j.procs.2023.10.588>

Διαδικτυακές πηγές

- (1) Data mining, Wikipedia
- (2) Statical Classification, Wikipedia
- (3) Regression analysis, Wikipedia
- (4) Cluster Analysis, Wikipedia
- (5) Summet Kumar Agrawal, Metrics To evaluate your Classification Model to take the right decisions, Analytics Vidhya (2024)
- (6) Evaluation Metrics to Evaluate Classification Models, Enjoy algorithms
- (7) Weka, Wikipedia
- (8) Weka Data Mining, Javapoint
- (9) Python, Wikipedia
- (10) Python libraries, Great learning (2024)
- (11) What is tokenization in NLP? Here's All you need to know, Aravindpai Pai, Analytics Vidhya (2024)
- (12) Text pre-processing: Stop words removal using different libraries, Chetna Khanna, Towards Data Science (2021)
- (13) Lemmatization in Natural Language Processing (NLP) and Machine Learning, Sunny Srinidhi, Builtin (2023)
- (14) Stemming and Lemmatization, The Stanford Natural Language Processing Group
- (15) POS(Parts-Of-Speech) Tagging in NLP, GeeksForGeeks
- (16) Named Entity Recognition, GeeksforGeeks
- (17) Text Cleaning Methods in NLP, Kashish Rastogi, Analytics Vidhya (2022)
- (18) Machine learning, Wikipedia
- (19) Supervised Learning: Definition and examples, Big Blue Data Academy (2023)
- (20) A Comprehensive Guide to Ensemble Learning (with Python codes), Aishwarya Singh, Analytics Vidhya (2023)
- (21) Unsupervised Learning, Google Cloud
- (22) What is Reinforcement Learning? Aws Amazon
- (23) K-means clustering, Wikipedia

- (24) Decision Tree learning, Wikipedia
- (25) Decision Tree pruning, Wikipedia
- (26) Νευρωνικά Δίκτυα (Neural Networks): Ορισμός & Εφαρμογές, Big Blue Data Academy (2023)
- (27) Bayes' Theorem, Wikipedia
- (28) Bayesian Network, Wikipedia
- (29) What is the k-nearest neighbors (KNN) algorithm? , IMB
- (30) Support Vector Machine, Wikipedia
- (31) Logistic Regression, Wikipedia
- (32) When to use it: Logistic Regression, Avitevet
- (33) XGBoost, GeeksForGeeks
- (34) K-mean Clustering: Getting the Optimal Number of Clusters, Ankita Banerji, Analytics Vidhya (2023)
- (35) What is Random Forest?, IBM
- (36) Search Engine Optimization (SEO) Starter Guide, Google For Developers
- (37) Google Search Console Tools, Google
- (38) Data analysis and measurement, Shubhinagpuri, Medium (2023)
- (39) Digital marketing: what are the challenges for 2023? , Doxee
- (40) Marketing Globalisation: Challenges and Opportunities, Content Nation
- (41) “Programmatic advertising, Definition, benefits, examples, and how it works” Amazon Ads
- (42) Interactive Advertising, Wikipedia
- (43) Barbara von der Osten, “Why Should You Include Interactive Ads in Your Digital Marketing Strategy?” RockContent (2022)
- (44) Native Advertising, Wikipedia
- (45) Roshan Paswan, “The Ultimate Guide To Start Generating AI-Powered Ad Creatives At Scale” , Medium (2023)
- (46) Mike Kaput, “AI in advertising: Everything you need to know” , Marketing Artificial Intelligence Institute (2024)

- (47) “What is shoppable content? Why it matters for eCommerce?” Storyly
- (48) Paul Slaney, “Influencer Marketing and Shoppable Instagram: A winning combination” LinkedIn (2023)
- (49) Ad Fraud, Wikipedia
- (50) “Ad Fraud, what it is, How It works & 13 common types” , Datadome
- (51) Ad blocking, Wikipedia
- (52) 14 Effective Strategies To Overcome Ad Blocking, Forbes
- (53) Krishan Arora, The impact of privacy regulations on digital marketing, Forbes (2023)
- (54) General Data Protection Regulation, Wikipedia
- (55) General Data Protection Regulation, Intersoft Consulting
- (56) California Consumer Privacy Act, Wikipedia
- (57) California Consumer Privacy Act, State of California Department of Justice
- (58) What is Ad Viewability, Macrometa
- (59) How to Build a Churn Prediction Model to Predict Customer Churn, UserPilot
- (60) Hanson Cheng, “The Importance of Machine Learning for Customer Segmentation” Freedomtoascend (2023)
- (61) Aparna Dhinakaran, Best Practices In ML Observability for Click-Through Rate Models, Medium (2021)
- (62) Kshitij Kutumbe, Ads Click Through Rate(CTR) Analysis And Prediction, Medium (2023)
- (63) Click Fraud, Wikipedia
- (64) How machine learning works for payment fraud detection and prevention, Stripe
- (65) What is Chatbot, Oracle
- (66) The Uses and Benefits of AI Chatbots, Helpshift
- (67) Sentiment Analysis: Definition and Importance, Big Blue Data Academy

Αναφορές Εικόνων και Πινάκων

<https://www.ml4devs.com/articles/machine-learning-intro-for-developers/>

<https://www.ml4devs.com/articles/machine-learning-intro-for-developers/>

https://www.researchgate.net/figure/Data-classification-clustering-and-regression-for-healthcare-data-analytics-ML_fig1_339847881

<https://www.edureka.co/blog/introduction-to-machine-learning/>

<https://www.statisticshowto.com/hierarchical-clustering/>

<https://medium.datadriveninvestor.com/k-means-clustering-4a700d4a4720>

<https://machinelearningtheory.org/docs/Random-Forest/tree/>

https://en.wikipedia.org/wiki/Neural_network_%28machine_learning%29

<https://innovationyourself.com/naive-bayes-classifier/>

<https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>

<https://www.geeksforgeeks.org/introduction-to-support-vector-machines-svm/>

<http://avitevet.com/machine-learning/when-to-use-it-logistic-regression/>

<https://www.mdpi.com/2076-3417/12/15/7739>

https://www.researchgate.net/figure/A-general-architecture-of-XGBoost_fig3_335483097

<https://www.mdpi.com/2076-3417/12/15/7739>

<https://machinelearningmastery.com/what-is-the-weka-machine-learning-workbench/>

<https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3755824/>

<https://dl.acm.org/doi/full/10.1145/3554729>

<https://corporatefinanceinstitute.com/resources/accounting/return-on-investment-roi-formula/>

<https://www.slidegeeks.com/email-marketing-process-flow-chart-for-business-promotion-ppt-styles-sample-pdf>

<https://whatagraph.com/templates/google-search-console-dashboard>

<https://www.superoffice.com/blog/crm-software-statistics/>

<https://business.instagram.com/blog/ar-ad-facebook-instagram>

<https://creately.com/diagram/example/iutwrw2j1/chatbot-flowchart-classic>