ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΕΞΟΡΥΞΗ ΚΟΙΝΩΝΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΗ ΣΧΕΔΙΑΣΗ ΚΑΙ ΑΝΑΠΤΥΞΗ ΜΟΝΑΔΩΝ E-COMMERCE

Φοιτητής: Λιγνός Ιωάννης

Επιβλέπων καθηγητής: Σταμούλης Γεώργιος

Λαμία, Σεπτέμβριος 2019

UNIVERSITY OF THESSALY

FACULTY OF SCIENCES

DEPARTMENT OF COMPUTER SCIENCE

BACHELOR THESIS

SOCIAL DATA MINING IN PLANNING AND DEVELOPMENT OF E-COMMERCE UNITS

Student: Lignos Ioannis

Supervising Professor: Stamoulis Georgios

Lamia, Greece, September 2019

Υπεύθυνη Δήλωση

«Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις [1], που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

*1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάσθηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.*

*2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.*

3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια

4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.»

Περίληψη

Η συγκεκριμένη εργασία πραγματοποιήθηκε στο πλαίσιο του προπτυχιακού προγράμματος σπουδών του τμήματος πληροφορικής και τηλεπικοινωνιών του Πανεπιστημίου Θεσσαλίας και ως σκοπό έχει να αναδείξει τις έννοιες του ηλεκτρονικού εμπορίου, τρόπους με τους οποίους μπορούμε να εξορύξουμε κοινωνικά δεδομένα καθώς και την ανάγκη να εκμεταλλευτούμε τα τελευταία προς όφελος της eCommerce επιχείρησης μας.

Αναλυτικότερα στο πρώτο κεφάλαιο δίνουμε τον ορισμό του ηλεκτρονικού εμπορίου και μελετάμε εκτενώς όλες τις έννοιες που εμπλέκονται με αυτόν τον όρο. Έπειτα αναλύουμε τρόπους και τεχνικές με τις οποίες μπορούμε να εξορύξουμε κοινωνικά δεδομένα είτε με αλγόριθμους, είτε με έτοιμα «εργαλεία» που μας παρέχονται. Τέλος διαπιστώσαμε πόσο σημαντική είναι η εξόρυξη κοινωνικών δεδομένων για μια eCommerce επιχείρηση σε πεδία όπως η διαφήμιση, η δημιουργία της ιστοσελίδας και η επικοινωνία με τους πελάτες.

Abstract

This thesis was carried out as part of the undergraduate degree program Department of Computer science, University of Thessaly and its purpose is to show off the meaning of eCommerce, find ways to mine social data, and lastly to point out the need of social data mining in an eCommerce business.

In more detail in the first chapter we give the definition of eCommerce and we mention extensively all the suspects of this subject. After that, we analyzed ways and techniques for social data mining such as algorithms or already existed tools found on the web. Finally, we found out the importance of the social data mining for an eCommerce business in fields like advertising, webpage creation and communication with costumers.

Περιεχόμενα

Εισαγωγή στο e-Commerce (Ηλεκτρονικό Εμπόριο)

Οι ταχύτατες αλλαγές στον τομέα της τεχνολογίας τις τελευταίες δεκαετίες, η δραματική έξαρση της χρήσης του διαδικτύου από όλο και περισσότερους ανθρώπους καθώς και η αύξηση της παγκοσμιοποίησης προσέφεραν πρόσφορο έδαφος στην ανάπτυξη ενός νέου είδους εμπορίας και προώθησης προϊόντων και υπηρεσιών. Μιλάμε για το επονομαζόμενο Ηλεκτρονικό Εμπόριο (e-Commerce) που μέρα με τη μέρα εισέρχεται όλο και περισσότερο στις ζωές όλων μας.

Τι είναι το e-Commerce;

Ο πιο απλός και ευρέως διαδεδομένος ορισμός του ηλεκτρονικού εμπορίου είναι ότι: *«Το e-Commerce (ηλεκτρονικό εμπόριο) αποτελεί τρόπο αγοροπωλησίας αγαθών και υπηρεσιών μεταξύ ατόμων ή επιχειρήσεων οι οποίες εκτελούνται με ηλεκτρονικά μέσα. »*

Έναν άλλο ορισμό μας τον έδωσε ο γνωστός γκουρού του management Peter Drucker: *«Το Ηλεκτρονικό Εμπόριο ορίζεται ως η αστραπιαία εμφάνιση του διαδικτύου σαν έναν απ' τους σημαντικότερους, ίσως τον σημαντικότερο, παγκόσμιο διανεμητή για αγαθά, υπηρεσίες και παραδόξως για διοικητικές και επαγγελματικές θέσεις εργασίας.. Αυτό έχει επιφέρει σημαντικές αλλαγές στην οικονομία, στις αγορές ακόμα και στην ίδια τη δομή των επιχειρήσεων καθώς και στην αγοραστική συμπεριφορά των καταναλωτών.»* (Drucker 2002)

Πάλι ο Drucker αρκετά χρόνια πριν με τη φράση *«η επανάσταση του διαδικτύου μόλις άρχισε να γίνεται εμφανής»* υπονόησε αυτή τη ραγδαία αλλαγή στον τρόπο που πραγματοποιούνται οι εμπορικές συναλλαγές και ως επακόλουθο τα άτομα και οι επιχειρήσεις που εναρμονίζονται και επενδύουν στις νέες τεχνολογίες έχουν σοβαρό πλεονέκτημα σε σχέση με τους ανταγωνιστές τους που επιμένουν στις παραδοσιακές μεθόδους. Θα μελετήσουμε εκτενέστερα τα πλεονεκτήματα του ηλεκτρονικού εμπορίου στα επόμενα κεφάλαια της εργασίας.

Ένας τελευταίος ορισμός διαχωρίζει το ηλεκτρονικό εμπόριο μέσα από τέσσερις οπτικές γωνίες:

«

1. *Στον τομέα της επικοινωνίας: ως τη δυνατότητα παροχής πληροφοριών και πληρωμών.*
2. *Στον τομέα των επιχειρήσεων: ως τη χρήση της νέας τεχνολογίας που επιφέρουν αυτόματες συναλλαγές μεταξύ των επιχειρήσεων.*
3. *Στον τομέα των υπηρεσιών: ως τη μείωση του κόστους και βελτίωση της ποιότητας των υπηρεσιών που προσφέρονται.*
4. *Στον τομέα του διαδικτύου: ως τις ηλεκτρονικές αγοροπωλησίες*

»

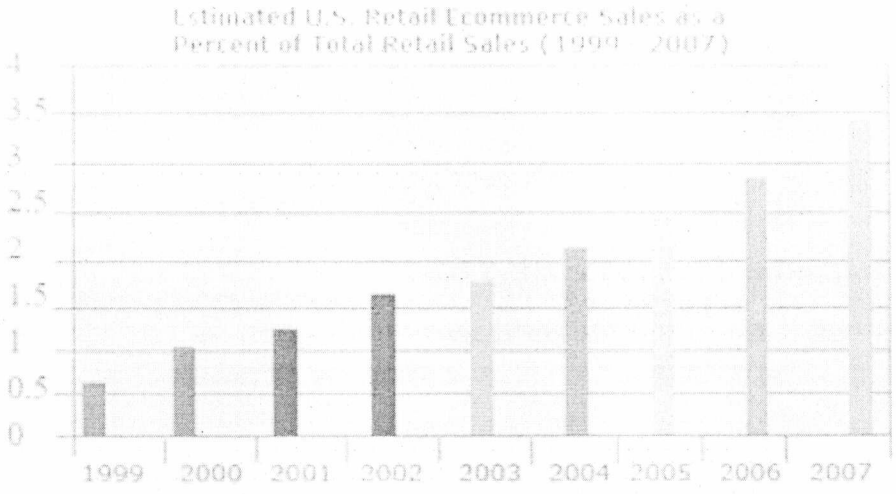
(Kalakota & Whinston)

Ιστορική αναδρομή

Το ηλεκτρονικό εμπόριο πρωτοεμφανίστηκε στα τέλη του έτους 1970 με την εμφάνιση τεχνολογιών όπως ανταλλαγή δεδομένων με ηλεκτρονικό τρόπο (EDI) και ηλεκτρονικής μεταφοράς χρημάτων (EFT). Οι συγκεκριμένες τεχνολογίες επέφεραν στις επιχειρήσεις μείωση του κόστους αφού δεν είναι απαραίτητη πια η χρήση του χαρτιού καθώς και ταχύτερη διεκπεραίωση των συναλλαγών και μείωση της γραφειοκρατίας.

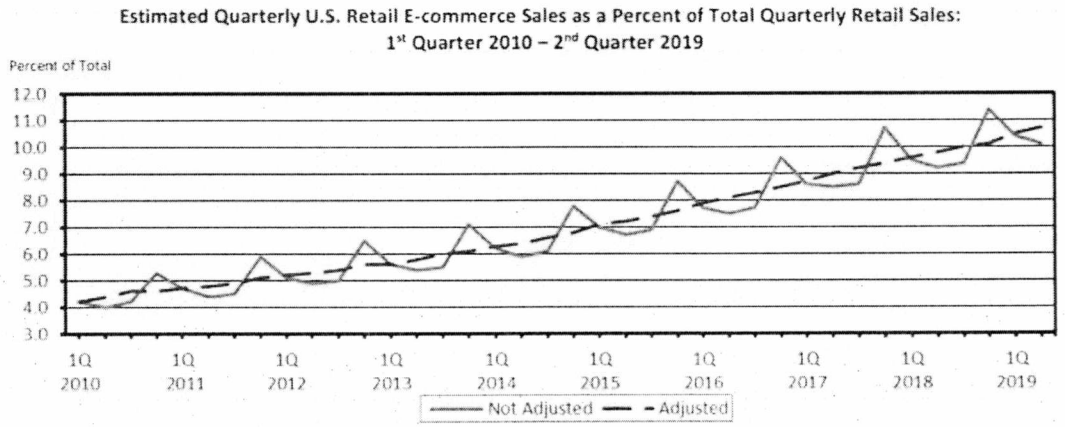
Απ' το 1994 και μετά απήλθε η επανάσταση του διαδικτύου το οποίο άρχισε να μετρά όλο και περισσότερους χρήστες. Σε αυτό συνετέλεσε ιδιαίτερα η εμφάνιση πρωτοκόλλων ασφαλείας (HTTP) καθώς και το DSL τα οποία παρείχαν μια πολύ γρηγορότερη πρόσβαση και σταθερή σύνδεση χωρίς περιορισμούς. Το 2000 πληθώρα επιχειρήσεων στις ΗΠΑ και στη Δυτική Ευρώπη άρχισαν να εισέρχονται και να παρουσιάζουν την ύπαρξη τους στον παγκόσμιο ιστό. Τότε άρχισε να αναπτύσσεται και η σύγχρονη έννοια του e-Commerce, δηλαδή της διαδικασία απόκτησης αγαθών και υπηρεσιών μέσα απ' το διαδίκτυο. Με τα χρόνια οι πωλήσεις μέσω διαδικτύου αυξάνονταν και υπολογίζεται ότι μέχρι το τέλος του 2007 οι «e-Commerce πωλήσεις»

ήταν το 3.4% των συνολικών πωλήσεων. Παρακάτω παραθέτω και ένα διάγραμμα που δείχνει την αύξηση των πωλήσεων κατά τα έτη 1999-2007 στις Ηνωμένες Πολιτείες .



Διάγραμμα 1 (Πηγή: ecommerce-land.com)

Από τότε και μέχρι σήμερα ,όπως ήταν αναμενόμενο , αυτό το ποσοστό έχει ανέβει στα ύψη. Αυξάνεται περίπου 1% κάθε χρόνο και σήμερα εν έτη 2019 έχει φτάσει στο 10,7%! Εν συνεχεία προσθέτω ένα ακόμη διάγραμμα που δείχνει την πορεία αύξησης αυτού του ποσοστού απ' το έτος 2010 μέχρι και σήμερα.

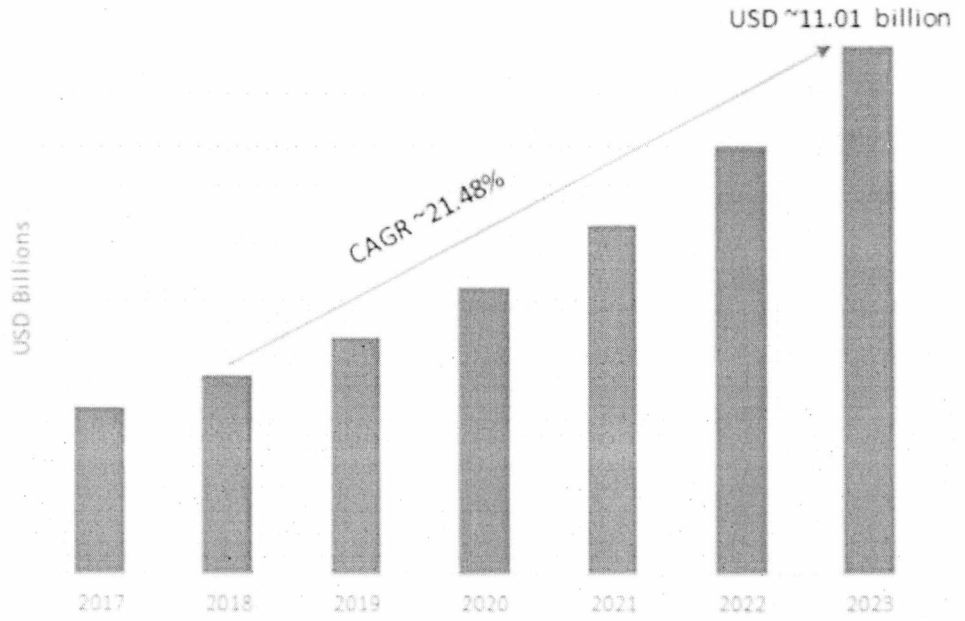


Διάγραμμα 2 (Πηγή U.S. Department of Commerce)

Σε αυτό το σημείο αξίζει να δώσουμε ξεχωριστή αναφορά στην επιχείρηση-μεγαθήριο του ηλεκτρονικού εμπορίου η οποία άλλαξε μια για πάντα τον χάρτη των ηλεκτρονικών αγορών. Μιλάμε για το Amazon, μια απ' τις πρώτες και μεγαλύτερες, ίσως η μεγαλύτερη, εταιρίες στο χώρο. Ιδρύθηκε το 1994 απ' τον Jeff Bezos στο Seattle των Ηνωμένων Πολιτειών. Είχε μεγάλη επιτυχία μέχρι το επονομαζόμενο «dot-com collapse» το 2001 όπου απήλθε μεγάλη συρρίκνωση της και έχανε συνεχώς έδαφος με πολλούς να θεωρούν ότι ήρθε και το τέλος, Προς έκπληξη όλων όμως κατάφερε να

ανακάμψει αρκετά σύντομα και μόλις το 2003 άρχισε και πάλι να βγάζει κέρδος. Σύμφωνα με στατιστικές μελέτες το 2008 το amazon προσέλκυσε πάνω από 615 εκατομμύρια χρήστες ετησίως και αυτός ο αριθμός μέχρι και σήμερα συνεχίζει να αυξάνεται. Δεν είναι τυχαίο άλλωστε ότι ο Bezos είναι ο πλουσιότερος άνθρωπος στον κόσμο με περιουσία που αγγίζει τα 131 δισεκατομμύρια δολάρια.

Η ιστορία του e-Commerce είναι μια ιστορία ενός νέου ψηφιακού κόσμου που συνεχώς μεταβάλλεται και αναπτύσσεται όσο και η ίδια η τεχνολογία κάνει σπουδαία άλματα εμπρός. Ποιο περιμένουμε όμως να είναι το μέλλον του ηλεκτρονικού εμπορίου; Θα συνεχιστεί αυτή η εντυπωσιακή άνοδος στο ποσοστιαίο μερίδιο της αγοράς; Η απάντηση είναι ναι και όλες οι ενδείξεις που έχουμε ακολουθούν αυτήν την υπόθεση. Ο τελευταίος πίνακας που παρατίθεται παρακάτω δείχνει αυτό που μόλις αναφέραμε. Την αύξηση των ποσοστιαίων μονάδων μέχρι το 2023 όπου θα φτάσει το 21,48%.



Διάγραμμα 3 (Πηγή Reuters.com)

Τα 4 διαφορετικά είδη

Με βάση τη φύση του πωλητή και την αντίστοιχη του αγοραστή στην διεξαγωγή των αγοροπωλησιών μπορούμε να διακρίνουμε το ηλεκτρονικό εμπόριο σε τέσσερις επιμέρους κατηγορίες. Το ηλεκτρονικό εμπόριο από επιχειρήσεις προς καταναλωτές (B2C- Business To Customers), από επιχειρήσεις προς επιχειρήσεις (B2B-Business To Business), από καταναλωτές προς επιχειρήσεις (C2B-Costumers To Business), από καταναλωτές προς καταναλωτές (C2C-Costumers To Costumers).

B2C(Επιχειρήσεις προς Καταναλωτές)

Όπως λέει και το όνομα του διαδραματίζονται αγορές προϊόντων και υπηρεσιών από καταναλωτές με πωλητές τις επιχειρήσεις. Πρόκειται για την πιο διαδεδομένη μορφή ηλεκτρονικού εμπορίου η οποία άρχισε διευρύνεται στα τέλη του 1995.Οι επιχειρήσεις αναζητούν τους υποψήφιους πελάτες τους με τη χρήση κοινωνικών δεδομένων στα οποία εκτελούν διάφορες τεχνικές μάρκετινγκ. (Θα μιλήσουμε για τρόπους και τεχνικές εξόρυξης κοινωνικών δεδομένων στο επόμενο κεφάλαιο της εργασίας.) Στη συνέχεια θα δώσουμε μερικά χαρακτηριστικά παραδείγματα για την κατηγορία B2C .

- Amazon.com, Wish.com, eShop
- .gr κλπ είναι παραδείγματα που επιτρέπουν το παραδοσιακό λιανεμπόριο μέσα από μια οθόνη. Και το βασικότερο; Δεν κλείνουν ποτέ! Η πηγή εσόδων τους φυσικά είναι οι πωλήσεις προϊόντων.
- Παραδείγματα όπως Netflix.com, Spotify.com, Twitch.com κλπ. παρέχουν υπηρεσίες αναπαραγωγής ταινιών, μουσικής, live streaming με κύρια πηγή εσόδων τις μηνιαίες συνδρομές.
- Παρόμοια με την πάνω κατηγορία παραδείγματα όπως McAfee, Photoshop.com, Microsoft Office παρέχουν υπηρεσίες όπως ασφάλεια υπολογιστή, επεξεργασία εικόνων και κειμένου με κύρια πηγή εσόδων τις ετήσιες συνδρομές.
- Στην κατηγορία B2C μπαίνουν επίσης sites όπως BBC.com, CNN.COM, in.gr, iefimerida.gr κλπ τα οποία δεν πωλούν άμεσα προϊόντα ή υπηρεσίες. Για την

ακρίβεια παρέχουν υπηρεσίες όπως ενημέρωσης των πολιτών δωρεάν και ως κύρια πηγή εσόδων τους είναι η διαφήμιση.

- Φυσικά δε θα μπορούσαν να λείπουν από αυτήν τη λίστα τα ευρέως διαδεδομένα στις μέρες μας κοινωνικά δίκτυα-social medias. Χαρακτηριστικά παραδείγματα Facebook, Twitter, Instagram και πολλά άλλα όπου και εδώ η κύρια πηγή εσόδων τους είναι οι διαφημίσεις.
- Τέλος sites που παρέχουν τρόπους συναλλάγματος όπως e-Trade.com με κύρια πηγή εσόδων τα τέλη συναλλαγής.

B2B(Επιχειρήσεις προς Επιχειρήσεις)

Λίγο πιο πάνω στην ιστορική αναδρομή μιλήσαμε για κάποιες τεχνολογίες εκείνης της εποχής πριν την έλευση του παγκόσμιου ιστού όπως η ανταλλαγή δεδομένων με ηλεκτρονικό τρόπο (EDI). Το EDI είναι ο πρώτος εκφραστής του B2B μοντέλου όπου επέτρεπε την άμεση σύνδεση μεταξύ επιχειρήσεων. Στις μέρες μας το ίντερνετ αποτελεί ένα πολύ σημαντικό μέσο για την εμπορία οποιασδήποτε μορφής μεταξύ επιχειρήσεων με σημαντικά πλεονεκτήματα. Να τονίσουμε εδώ ότι το B2B κατέχει το 75% όλων των συναλλαγών ηλεκτρονικού εμπορίου. Αυτό το καθιστά πολύ μεγαλύτερο σε σχέση με το πολύ διαδεδομένο B2C σε θέματα κερδών. Στη συνέχεια θα δώσουμε μερικά χαρακτηριστικά παραδείγματα για την κατηγορία B2B

- Alibaba.com παρόμοιο με το amazon.com αλλά αναφέρετε σε επιχειρήσεις καθώς πρέπει να αγοράσεις μεγάλη ποσότητα από κάθε προϊόν. Πηγή εσόδων οι πωλήσεις.
- Tetra Pak μια πολύ γνωστή σουηδική εταιρία που ασχολείται με την τυποποίηση προϊόντων άλλων επιχειρήσεων με πηγή εσόδων πάλι τις πωλήσεις προϊόντων.
- Παραδείγματα όπως salesforce.com και EmployeeMatters.com ενοικιάζουν/πωλούν εφαρμογές για επιχειρήσεις με κύρια πηγή εσόδων τις πωλήσεις υπηρεσιών.
- Πάλι τα social medias Facebook, Instagram, Twitter κλπ καθώς και εταιρίες όπως η Google,YouTube, Bing εκμεταλλεύονται τη πληθώρα χρηστών τους και τις μοναδικές πληροφορίες που έχουν συλλέξει για καθέναν απ' αυτούς και τις

πουλάνε σε επιχειρήσεις που ψάχνουν τους επόμενους πελάτες τους. Πηγή εσόδων η πώληση κοινωνικών δεδομένων.

- Skroutz.gr ένα ελληνικό παράδειγμα όπου εκμεταλλεύεται το όνομα του ώστε επιχειρήσεις να βάλουν τα προϊόντα τους στις λίστες του. Πηγή εισόδων είναι κάποιο ποσοστό επί του κέρδους που βγάζει η άλλη επιχείρηση σε μια ενδεχόμενη πώληση ενός προϊόντος μέσα από την ιστοσελίδα του.

C2B(πελάτες προς επιχειρήσεις)

Το C2B είναι ένας σχετικά πρόσφατος κλάδος ηλεκτρονικού εμπορίου και επεκτείνεται με ταχύτατους ρυθμούς. Η φιλοσοφία του είναι οι πελάτες να θέτουν συγκεκριμένους όρους στις επιχειρήσεις για μια συγκεκριμένη υπηρεσία ή προϊόν. Παρακάτω κάποια παραδείγματα.

- Χαρακτηριστικό παράδειγμα είναι η κράτηση ενός συγκεκριμένου δωματίου μια συγκεκριμένη χρονική περίοδο σε ένα ξενοδοχείο τα οποία τα θέτει ο πελάτης. Το ίδιο γίνεται και με μια πτήση με αεροπλάνο, τρένο, πλοίο κλπ.
- Ένα δεύτερο πολύ σημαντικό παράδειγμα είναι sites όπως Priceline.com στο οποίο οι πελάτες θέτουν το ποσό που σκοπεύουν να δώσουν για ένα προϊόν ή μια υπηρεσία.

C2C(πελάτες προς πελάτες)

Το C2C αναφέρετε στην διαδικτυακή εμπορική αλληλεπίδραση μεταξύ ατόμων που δεν έχουν σχέση με επιχείρηση. Μέσω λοιπόν μιας ειδικά διαμορφωμένης πλατφόρμας πραγματοποιούνται συναλλαγές προϊόντων/ υπηρεσιών. Η ύπαρξη της πλατφόρμας είναι απαραίτητη ώστε να γίνει η συναλλαγή με αντάλλαγμα ίσως κάποιο ποσοστό της πώλησης. Στη συνέχεια θα δώσουμε μερικά χαρακτηριστικά παραδείγματα για την κατηγορία C2C.

- Σαν πρώτο παράδειγμα είναι οι δημοπρασίες που γίνονται σε sites όπως το eBay . Ένα άτομο έχει το προϊόν και θέτει μια αρχική τιμή με τους ενδιαφερόμενους

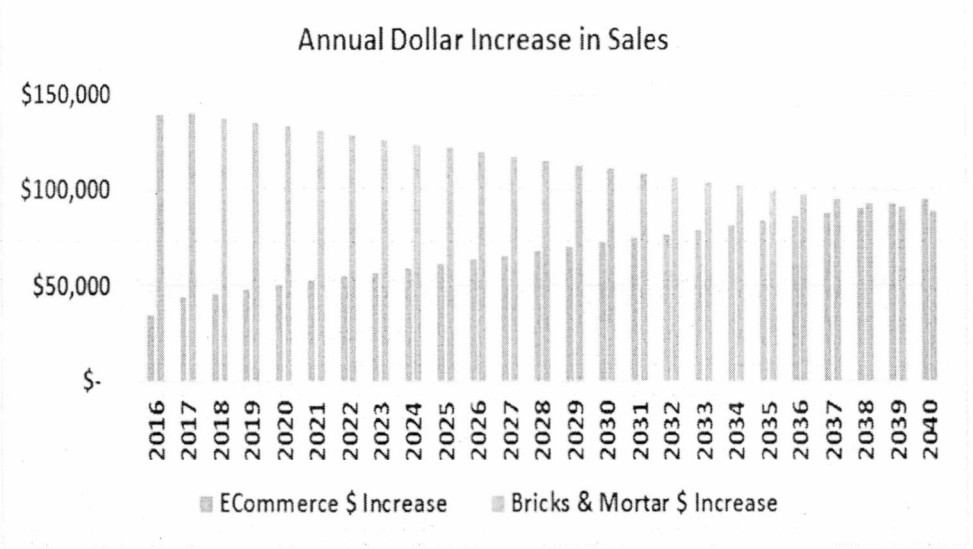
να διεκδικούν το συγκεκριμένο προϊόν μέχρι «εκεί που τραβάει η τσέπη τους». Το eBay παίρνει κάποιο ποσοστό από κάθε επιτυχημένη δημοπρασία.

- Xo.gr, car.gr ελληνικά παραδείγματα όπου άτομα μπορούν να βάζουν τα προϊόντα που θέλουν να πουλήσουν σε αυτές τις πλατφόρμες και να έρχονται σε επικοινωνία με άλλα άτομα που ενδιαφέρονται να τα αγοράσουν. Πρέπει να πληρώσεις κάποιο ποσό για να βάλεις τα προϊόντα σου στη λίστα.
- Ο συνήθης ύποπτος το Facebook έρχεται και στο C2C μοντέλο ως πλατφόρμα αυτή τη φορά και με ειδικά διαμορφωμένα γκρουπς διευκολύνει την αγοροπωλησία μεταξύ απλών ανθρώπων. Η διαδικασία είναι δωρεάν.

Πλεονεκτήματα Ηλεκτρονικού Εμπορίου

Αναμφίβολα το ηλεκτρονικό εμπόριο ήρθε για να αλλάξει τη ζωή μας. Οι επιχειρήσεις που εναρμονίζονται με αυτό έχουν σαφές πλεονέκτημα σε σχέση με τις επιχειρήσεις "brick and mortal", δηλαδή τα καταστήματα με υλική υπόσταση που δεν εκτελούν ενέργειες ώστε να μπουν στον κόσμο του διαδικτύου και του ηλεκτρονικού εμπορίου.

Στο παρακάτω διάγραμμα γίνεται αντιληπτό πως με την πάροδο του χρόνου τα παραδοσιακά καταστήματα χάνουν έδαφος έναντι τον ηλεκτρονικών καταστημάτων και πως οι ηλεκτρονικές πωλήσεις σε λίγα χρόνια θα υπερτερούν. Ήδη απ' το 2039 προβλέπεται αυτή η ενδεχόμενη «νίκη».



Διάγραμμα 4 (Πηγή bigcommerce.com)

Πάμε τώρα να αναλύσουμε γιατί συμβαίνει αυτό; Γιατί το ηλεκτρονικό εμπόριο αρχίζει και κυριαρχεί; Ποια είναι τα πλεονεκτήματα του έναντι του παραδοσιακού εμπορίου;

Για τους καταναλωτές

Το ηλεκτρονικό εμπόριο γίνεται όλο και πιο δημοφιλή στους καταναλωτές και αυτό συμβαίνει για τους παρακάτω λόγους:

- Οι καταναλωτές έχουν ένα πολύ μεγαλύτερο εύρος επιλογών το οποίο δεν προέρχεται μόνο απ' τη γειτονιά τους, την πόλη τους, τη χώρα τους αλλά από όλον τον κόσμο (με εξαίρεση κάποιους περιορισμούς εισαγωγής ή εξαγωγής που μπορούν να θέσουν οι χώρες). Οπότε το ηλεκτρονικό εμπόριο δεν έχει σύνορα.
- Υπάρχει η δυνατότητα μέσα από πληθώρα εργαλείων και επιλογών που παρέχει ένα ηλεκτρονικό κατάστημα ο πελάτης να διαμορφώσει και να εξατομικεύσει το προϊόν ή την υπηρεσία με βάση τις αρεσκείες του. Για παράδειγμα αν κάποιος θέλει να αγοράσει ένα τζιν παντελόνι συγκεκριμένου μεγέθους , το μέγεθος του μπορεί να μετρηθεί με ειδικά εργαλεία μέσω ίντερνετ.
- Στο θέμα που κάποιος έχει εντοπίσει το προϊόν που θέλει να αγοράσει και επιθυμεί να κάνει μια έρευνα αγοράς για να βρει το οικονομικότερο δε χρειάζεται να πάει από κατάστημα σε κατάστημα. Μπορεί να κάτσει σπίτι του να ανοίξει την ηλεκτρονική του συσκευή και μέσα σε δευτερόλεπτα να βρει όλες τις πληροφορίες και λεπτομέρειες που ψάχνει.
- Ένα απ' τα σημαντικότερα πλεονεκτήματα είναι το ευέλικτο ωράριο των ηλεκτρονικών καταστημάτων. Τι εννοώ με τη λέξη ευέλικτο; Ότι παραμένουν πάντα ανοιχτά 24 ώρες το 24ωρο έτοιμα να εξυπηρετήσουν όλους τους επίδοξους πελάτες ακόμα και αν αυτοί πάσχουν από αυπνία.
- Οι ηλεκτρονικές επιχειρήσεις δε χρειάζεται να πληρώνουν ΕΝΦΙΑ, νερό, ρεύμα, ενοίκια για την επιχείρηση τους. Όλα λαμβάνουν χώρα στον φανταστικό κόσμο του διαδικτύου. Ως αποτέλεσμα είναι να μπορούν να θέτουν τα προϊόντα τους σε πολύ οικονομικότερες τιμές. Οπότε οι καταναλωτές βρίσκουν τα ίδια προϊόντα με τον «έξω κόσμο» αλλά πολύ πιο φτηνά.
- Ακόμα και μετά την αγορά ενός προϊόντος ο αγοραστής δε χρειάζεται να ταξιδέψει χιλιόμετρα για να παραλάβει το προϊόν του. Μέσα σε μικρό χρονικό διάστημα θα είναι στην πόρτα του.
- Η ιδιαίτερα ενοχλητική «μόδα» της μη έκδοσης αποδείξεων που έχει καθιερωθεί στην Ελλάδα και άλλες χώρες από αρκετούς καταστηματάρχες δεν είναι εφικτή μέσα από ηλεκτρονικές αγορές και την πληρωμή online(αν φυσικά το ηλεκτρονικό κατάστημα είναι νόμιμο). Οπότε ο ίδιος ο πελάτης αισθάνεται καλύτερα που δε γίνεται άθελα του κομμάτι σε ένα κύκλωμα φοροδιαφυγής και τρέφει περισσότερη εκτίμηση και σεβασμό για τον πωλητή.

Για τις επιχειρήσεις

Απ' την άλλη πλευρά και οι επιχειρήσεις έχουν αρκετά πλεονεκτήματα και κάποια απ' αυτά τα αναφέραμε και στο επάνω κομμάτι που αφορά τους καταναλωτές.

- Πρώτο και κύριο είναι η μείωση του κόστους σε σχέση με τις επιχειρήσεις με υλική υπόσταση λόγω μείωσης φόρων. Μαζί με αυτό έρχεται και η αύξηση του κέρδους καθώς προσφέρουν χαμηλότερες τιμές και προσελκύουν όλο και περισσότερους καταναλωτές.
- Όλα γίνονται ηλεκτρονικά και όλα καταγράφονται. Οι ηλεκτρονικές επιχειρήσεις έχουν τη δυνατότητα να κρατήσουν ηλεκτρονικά αρχεία-κοινωνικά δεδομένα για την αγοραστική συμπεριφορά των καταναλωτών και να τα εκμεταλλευτούν για την περαιτέρω αύξηση των κερδών τους.
- Όπως είπαμε και παραπάνω οι επιχειρήσεις δεν έχουν ωράριο λειτουργίες. Παραμένουν ανοιχτές μέρα νύχτα και αυξάνουν τις πωλήσεις και το κέρδος τους.
- Οι επιχειρήσεις διατηρούν μεγαλύτερο πελατολόγιο καθώς δεν υπάρχουν σύνορα για τους πελάτες τους. Παράλληλα με αυτόν τον τρόπο βελτιώνουν το προφίλ τους σε άλλες χώρες διευκολύνοντας έτσι μια ενδεχόμενη επέκταση-επένδυση ως προς αυτές τις χώρες.
- Το ίδιο ισχύει και για το προσωπικό. Μπορούν να επιλέξουν το προσωπικό τους από οποιαδήποτε χώρα στον κόσμο με τα χαρακτηριστικά και τις ικανότητες που επιθυμούν χωρίς να περιορίζονται σε συγκεκριμένες επιλογές ατόμων απ' την τοπική κοινωνία.
- Η γρήγορη εισαγωγή στην αγορά είναι ένα ακόμα πλεονέκτημα για τις επιχειρήσεις καθώς επιταχύνονται οι διαδικασίες αδειοδότησης τους ακόμα και σε χώρες όπως η Ελλάδα στις οποίες κυριαρχεί η γραφειοκρατία.
- Ένα τελευταίο πλεονέκτημα είναι η ασφάλεια που παρέχει το διαδίκτυο στις επιχειρήσεις. Είναι πλέον ασφαλείς από επιθέσεις, , ληστές και «πορτοφολάδες». Μπορεί φυσικά κάποιος να σκεφτεί τους χάκερς του διαδικτύου ως ένα παρόμοιο παράδειγμα. Το απορρίπτουμε όμως καθώς είναι ελάχιστοι σε αριθμό (για να γίνει κάποιος χρειάζεται εξαιρετικές γνώσεις) και

υπάρχουν πολλοί περισσότεροι τρόποι αντιμετώπισης τους από συναγερμούς και κάμερες που υπάρχουν στα μη ηλεκτρονικά καταστήματα.

Δε θα είμασταν δίκαιοι όμως αν δεν αναφέραμε και κάποια μειονεκτήματα του ηλεκτρονικού εμπορίου:

- Πρώτο και κύριο αφορά τα προϊόντα τα οποία ο πελάτης δεν έχει τη δυνατότητα να τα δει ζωντανά, να τα αγγίξει, να τα επεξεργαστεί Ακόμα και γι' αυτό όμως η τεχνολογία έχει μεριμνήσει. Λύσεις όπως εικονική πραγματικότητα και επαυξημένη πραγματικότητα με τη χρήση ειδικών μέσων (πχ VR) είναι η απάντηση. Πάραυτα πρέπει να περάσουν κάποια χρόνια ώστε οι συγκεκριμένες τεχνολογίες να γίνουν πιο προσιτές και να διαδοθούν σε περισσότερες ηλεκτρονικές επιχειρήσεις.
- Ένα άλλο μειονέκτημα είναι ότι πολλοί άνθρωποι στον κόσμο δεν χρησιμοποιούν το διαδίκτυο για τις ηλεκτρονικές τους αγορές. Και αυτό επειδή πολλοί ζουν σε υποβαθμισμένες, τριτοκοσμικές χώρες και κοινωνίες που δεν υπάρχει πρόσβαση στο ίντερνετ. Άλλοι, κυρίως άτομα τρίτης ηλικίας, δεν έχουν τις απαραίτητες γνώσεις για να πραγματοποιήσουν ηλεκτρονικές αγορές. Στα συγκεκριμένα άτομα πολλές φορές συναντάμε και μια προκατάληψη σχετικά με το ίντερνετ. Φράσεις όπως «το κακό ίντερνετ», «το επικίνδυνο ίντερνετ», «το σατανικό ίντερνετ» σίγουρα τις έχουμε ακούσει οι περισσότεροι από μεγαλύτερους σε ηλικία ανθρώπους.
- Ένα τελευταίο και πολύ σημαντικό μειονέκτημα είναι η μειωμένη ιδιωτικότητα που έχουμε μέσα απ' το διαδίκτυο. Συνεχώς ακούμε για καταγγελίες και δίκες περί διαρροής, εκμετάλλευσης και καταρράκωσης των προσωπικών μας δεδομένων και πληροφοριών προς όφελος λίγων. Πρόσφατα μάλιστα έγινε η γνωστή δίκη του δισεκατομμυριούχου ιδρυτή του Facebook, Marc Zuckerberg για παραβίαση της ιδιωτικότητας των χρηστών του η οποία κατέληξε στην καταβολή τεράστιων χρηματικών ποσών απ' την εταιρία στο δημόσιο.

Συμπερασματικά το ηλεκτρονικό εμπόριο είναι το μέλλον και ήρθε για να μείνει. Υπάρχουν πάρα πολλοί λόγοι που το καθιστούν ελκυστικότερο απ' το παραδοσιακό εμπόριο και για τους πελάτες αλλά και για τις επιχειρήσεις. Υπάρχουν λίγα

μειονεκτήματα για τα οποία όσο η τεχνολογία προοδεύει θα εξαλειφθούν πολύ γρήγορα.

Στρατηγικός σχεδιασμός και η σημασία της συλλογής κοινωνικών δεδομένων

Είδαμε τους λόγους που καθιστούν πολύ ελκυστικό το ηλεκτρονικό εμπόριο για μια επιχείρηση που θέλει να αυξήσει την αποτελεσματικότητα και το κέρδος της. Πως θα τα καταφέρει όμως; Τι κινήσεις πρέπει να κάνει ώστε να επιβιώσει σε αυτό το νέο πολύ ανταγωνιστικό περιβάλλον; Τι στρατηγική πρέπει να ακολουθήσει και ποια είναι η σημασία της συλλογής κοινωνικών δεδομένων; Παρακάτω δίνουμε κάποια βασικά σημεία για τον στρατηγικό σχεδιασμό μιας eCommerce επιχείρησης.

- Η σχέση της με τους πελάτες είναι ένας από τους σημαντικότερους παράγοντες επιτυχίας μιας online επιχείρησης. Πρέπει να αναπτυχθούν τρόποι και τεχνικές ώστε η επιχείρηση να φαίνεται ελκυστική για την «άφιξη» νέων πελατών αλλά και για τη διατήρηση των ήδη υπαρχόντων. Πως θα γίνει όμως αυτό;
    - Με την διαφοροποίηση της από άλλες και ανάδειξης της στο χώρο της συγκεκριμένης αγοράς μέσω της διαφήμισης, των δημοσίων σχέσεων κλπ
    - Επιβράβευση των ήδη υπάρχοντων πελατών με ειδικά εκπτωτικά κουπόνια ή δώρα.
    - Η ιστοσελίδα είναι η βιτρίνα όλης της επιχείρησης. Μια όμορφη, γρήγορη και ευέλικτη ιστοσελίδα μπορεί να ωθήσει έναν επίδοξο αγοραστή την προτίμηση της συγκεκριμένης επιχείρησης και όχι τους ανταγωνιστές.
    - Εκμετάλλευση των κοινωνικών δικτύων όπως Facebook με την κατασκευή σελίδων (pages) ή ομάδων(groups) για πιο άμεση επαφή με τον κόσμο.
    - Η διάθεση ποιοτικών προϊόντων και υπηρεσιών πρὸς τους καταναλωτές μπορεί να φαίνεται αυτονόητο σε πολλούς αλλά δεν είναι για όλες τις επιχειρήσεις. Πολλοί επιλέγουν τα προϊόντα τους να έχουν χαμηλή

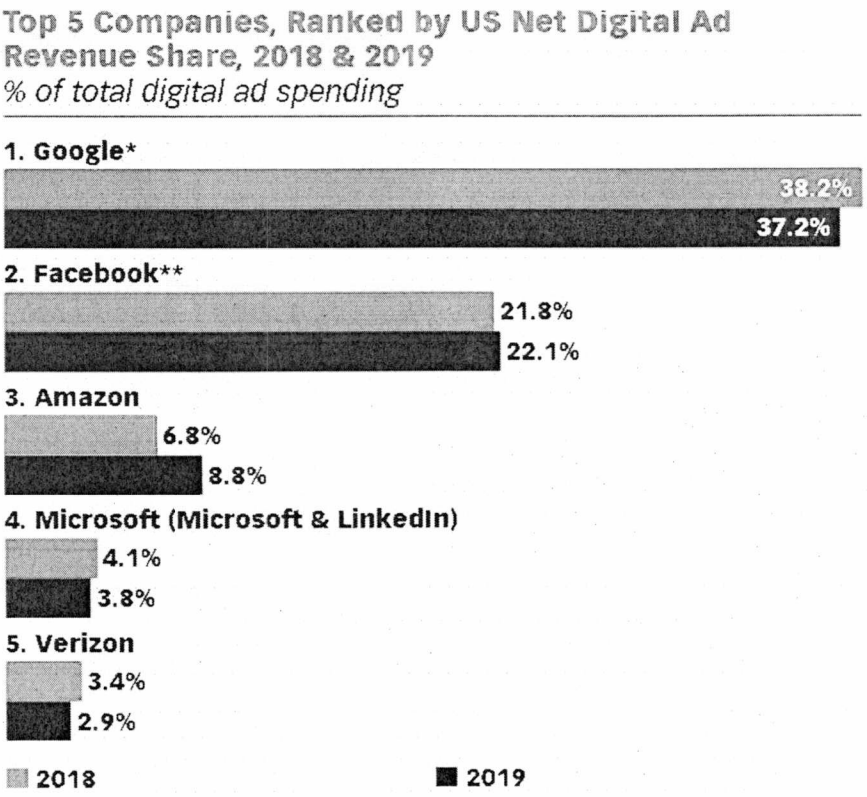
ποιότητα για να αυξήσουν το κέρδος αλλά με αυτόν τον τρόπο χάνουν σταδιακά το κοινό που τους υποστήριζε.

- Το κόστος πάντα παίζει ρόλο. Οι περισσότεροι καταναλωτές θα εξαφανιστούν από ένα κατάστημα αν θεωρήσουν ότι οι τιμές δεν ανταποκρίνονται στο προϊόν/υπηρεσία που τους παρουσιάζεται. Γιαυτό πρέπει ή να μειώσουμε τις τιμές ή όπως αναφέραμε μόλις πριν να προσφέρουμε ποιοτικότερα προϊόντα. (*Μερικές φορές η συνεχής και στοχευμένη διαφήμιση κάνει το προϊόν μας να φαίνεται ποιοτικό. Τα λεγόμενα brands(μάρκες) μπορεί να μην είναι ποιοτικά ανώτερα από τα άλλα στην αγορά, αλλά εκμεταλλεύονται το όνομα που έχουν δημιουργήσει. Να σημειωθεί πως η δημιουργία ενός απλού προϊόντος σε brand είναι μια επίπονη και χρονοβόρα διαδικασία και προϋποθέτει και την ύπαρξη μεγάλου μπάτζετ χωρίς όμως το τελευταίο να είναι δεσμευτικό)
- Αν πρόκειται για προϊόν με υλική υπόσταση και όχι υπηρεσία ένα άλλο αντικείμενο μελέτης είναι το κόστος μεταφοράς. Δε θα πρέπει να το παρακάνουμε με το κόστος καθώς κινδυνεύουμε να χάσουμε πελάτες.
- Σημαντικό ρόλο διαδραματίζει και ο τομέας της εξυπηρέτησης πελατών μιας επιχείρησης. Όσους περισσότερους τρόπους επικοινωνίας όπως τηλεφωνική επικοινωνία , επικοινωνία μέσω email, επικοινωνία μέσω live chat διαθέτει μια επιχείρηση τόσο πιο γρήγορα εξυπηρετεί και παράλληλα τόσο πιο κοντά έρχεται και με τον ίδιο τον πελάτη.

- Ο τρόπος λειτουργίας της επιχείρησης σε σχέση με το προσωπικό είναι ένα ακόμα σημείο του στρατηγικού σχεδιασμού που πρέπει να σταθούμε. Μπορούμε να διακρίνουμε δύο κατηγορίες: το προσωπικό να μην έρχεται σε άμεση επαφή δηλαδή να δουλεύουν όλοι απ' το σπίτι τους και η δεύτερη κατηγορία είναι η επιχείρηση να διαθέτει γραφεία Είναι σημαντικό οι εργαζόμενοι μιας επιχείρησης να αλληλοεπιδρούν μεταξύ τους στον πραγματικό κόσμο παρά μέσα από μια οθόνη. Αυτό μπορεί να βελτιώσει την επικοινωνία τους και επακολούθως την αποδοτικότητα τους. Στον αντίποδα η συγκεκριμένη στρατηγική αυξάνει το κόστος και περιορίζει την επιλογή του προσωπικού καθώς πρέπει να ανήκουν στο μέρος που εδρεύει η επιχείρηση. Η πρώτη κατηγορία χάνει το πλεονέκτημα της άμεσης αλληλεπίδρασης αλλά κερδίζει στο κόστος και στο διαθέσιμο προσωπικό προς πρόσληψη. Πέρα απ'

τα παραπάνω θα μπορούσε μια επιχείρηση να υιοθετήσει και μια υβριδική λύση συνδυάζοντας και τις δύο κατηγορίες.

- Ένα ακόμα σημείο αφορά τους εξωτερικούς συνεργάτες, όπως τράπεζες και μεταφορικές Οι εξωτερικοί συνεργάτες που επιλέγουμε πρέπει να λειτουργούν αποτελεσματικά και να μην υποβαθμίζουν την εικόνα και λειτουργία της επιχείρησης.
- Στο πρώτο σημείο στρατηγικού σχεδιασμού μιλήσαμε για σχέση με τους πελάτες. Πως βρίσκουν όμως αυτούς τους πελάτες οι επιχειρήσεις; Αυτό γίνεται κατά κύριο λόγο μέσω της διαφήμισης. Η διαφήμιση στις περισσότερες περιπτώσεις πραγματοποιείται μέσα από το διαδίκτυο σε ιστοσελίδες όπως google και Facebook όπου δίνουν τη δυνατότητα (ανάλογα και με το ποσό που δαπανείται) σε μια επιχείρηση να απευθυνθεί σε χιλιάδες ανθρώπους από οποιοδήποτε μέρος στον κόσμο προς αναζήτηση κέρδους. Παρακάτω δίνουμε ένα διάγραμμα με τα 5 κορυφαία sites για διαφημίσεις τα έτη 2018 και 2019



Διάγραμμα 5 (Πηγή emarketer.com)

Υπάρχουν και άλλοι τρόποι να προσεγγίσει πελάτες όπως με τη χρήση των κοινωνικών δικτύων, blogs, YouTube videos, emails και άλλα. Πολλές φορές οι επιχειρήσεις επιλέγουν τα αγοράσουν έτοιμες λίστες με πελάτες και να

προωθούν έτσι τα προϊόντα και τις υπηρεσίες τους, Κάτι που προϋποθέτει μεγάλη προσοχή και γνώση της νομοθεσίας κάθε χώρας ώστε αυτή η τεχνική να μην έρχεται σε σύγκρουση με το νόμο περί προσωπικών δικαιωμάτων. Η διαφήμιση λοιπόν αποτελεί ίσως το σημαντικότερο σημείο στρατηγικού σχεδιασμού που πρέπει να απασχολήσει τους ιθύνοντες μιας ecommerce επιχείρησης.

Εν κατακλείδι, προηγουμένως δώσαμε κάποια απ' τα βασικά σημεία στρατηγικού σχεδιασμού μιας eCommerce επιχείρησης, κάθε ένα από τα οποία είναι σημαντικό για μια επιτυχημένη πορεία. Η διαφήμιση ίσως αποτελεί το σημαντικότερο από αυτά τα τέσσερα σημεία που οι ιθύνοντες της επιχείρησης επιβάλλεται να δώσουν ιδιαίτερη έμφαση βραχυπρόθεσμα ώστε να προσελκύσουν περισσότερους πελάτες και συνεπώς να μεγιστοποιήσουν τα κέρδη τους. Με αυτόν τον τρόπο θα μπορέσουν με μεγαλύτερη άνεση στο μέλλον να ασχοληθούν και με τις υπόλοιπες στρατηγικές που αναφέραμε.

Ωραία η ιδέα της διαφήμισης αλλά δεν είναι τόσο απλό όσο ακούγεται. Δεν υφίσταται αποτελεσματική διαφήμιση χωρίς την προώθηση στο σωστό κοινό το οποίο ενδιαφέρεται για το προϊόν μας και μπορεί να μπει στη διαδικασία να το αγοράσει. Για παράδειγμα δεν μπορεί το προϊόν μας να αφορά κρέας και να το προωθήσουμε σε χορτοφάγους. Οπότε το κρίσιμο ερώτημα είναι «Που ξέρουν οι eCommerce επιχειρήσεις που θα στοχεύσουν;»

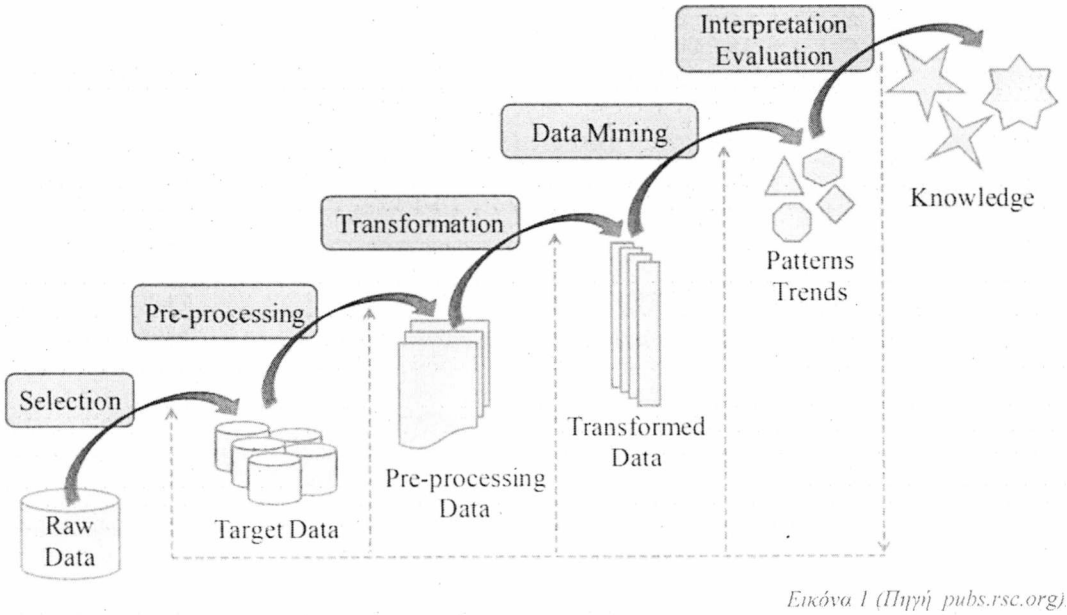
Για την απάντηση αυτής της ερώτησης απαιτείτε η εξόρυξη κοινωνικών δεδομένων.
Η συλλογή κοινωνικών δεδομένων αποτελεί το άλφα και το ωμέγα μιας επιχείρησης (για τον τομέα της διαφήμισης) διότι μπορεί να κατανοήσει καλύτερα την αγοραστική συμπεριφορά των καταναλωτών η οποία πηγάζει απ' τις συνήθειες τους, τα θέλω, τα πιστεύω τους, το υπόβαθρο τους, το επάγγελμα τους, την περιουσία τους και άλλα πολλά. Όλα αυτά μπορούν να εντοπιστούν με διάφορους τρόπους που θα αναλύσουμε στο κεφάλαιο που ακολουθεί.

Εξόρυξη κοινωνικών δεδομένων

Ως εξόρυξη κοινωνικών δεδομένων νοείται η εξαγωγή πληροφοριών που αφορούν τον κοινωνικό περίγυρο είτε μέσα από έτοιμα προγράμματα που παρέχουν αυτή τη δυνατότητα όπως Facebook Audience Insights ή Google Trends είτε με «χειροκίνητο τρόπο» με τη χρήση αλγορίθμων εξόρυξης γνώσης όπως αλγόριθμοι συσταδοποίησης (clustering), κατηγοριοποίησης (classification) και κανόνων συσχέτισης (association rules). Τα δύο έτοιμα προγράμματα χρησιμοποιούν παρόμοιους αλγόριθμους εξόρυξης , οπότε θα ήταν σωστό να τα αναλύσουμε μετά τους αλγορίθμους για να αποκτήσουμε μια πιο σφαιρική εικόνα του αντικειμένου.

Κατηγορίες αλγορίθμων εξόρυξης

Οι αλγόριθμοι γνώσης είναι ευρέως διαδεδομένοι σε επιχειρήσεις ηλεκτρονικού εμπορίου διότι τους δίνει τη δυνατότητα να εξάγουν αποτελέσματα και συμπεράσματα για την αγοραστική συμπεριφορά των καταναλωτών. Να εντοπίσουν ευκαιρίες στη συγκεκριμένη αγορά και να αυξήσουν τις πωλήσεις και τα κέρδη τους. Αξίζει απλά να αναφέρουμε ότι η εξόρυξη γνώσης μέσω αλγορίθμων είναι ένα βήμα της συνολικής διαδικασίας εύρεσης-ανακάλυψης γνώσης μέσα από κοινωνικά δεδομένα η οποία περιλαμβάνει 5 βήματα όπως φαίνεται στην παρακάτω εικόνα.



Εικόνα 1 (Πηγή pubs.rsc.org)

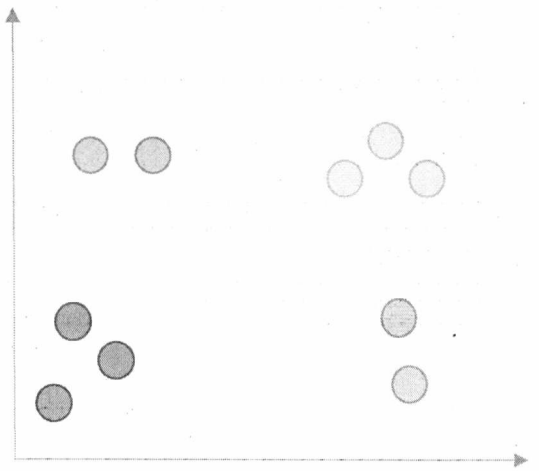
Σε αυτήν την ενότητα θα αναλύσουμε 3 κατηγορίες αλγορίθμων εξόρυξης κοινωνικών δεδομένων που σχετίζονται με την συσταδοποίηση (clustering), την κατηγοριοποίηση (classification) και τους κανόνες συσχέτισης (association rules). Στην πρώτη κατηγορία και την τρίτη κατηγορία ανήκουν οι περιγραφικοί αλγόριθμοι που στόχο έχουν να βρουν τρόπους για να περιγράψουν τα δεδομένα ενώ στη δεύτερη κατηγορία ανήκουν οι λεγόμενοι «προβλεπτικοί αλγόριθμοι» όπου επιχειρούν να δουν στο μέλλον(πχ να βρουν τη μελλοντική τιμή μιας μεταβλητής)

Συσταδοποίηση (Clustering)

Η κατηγορία αλγορίθμων συσταδοποίησης χωρίζει τα δεδομένα μας σε δύο ή περισσότερες κατηγορίες-συστάδες. Οι δύο αρχές που μας ενδιαφέρουν για να μπορούμε να πούμε ότι είναι επιτυχημένη μια συσταδοποίηση είναι:

1. Τα δεδομένα (αντικείμενα) που ανήκουν σε μια συστάδα πρέπει να είναι όμοια μεταξύ τους.
2. Τα δεδομένα που ανήκουν σε μια συστάδα πρέπει να είναι διαφορετικά από τα δεδομένα άλλων συστάδων.

Για παράδειγμα έχουμε μια eCommerce επιχείρηση που πουλάει κιθάρες. Έχουμε μια λίστα δεδομένων με υποψήφιους πελάτες. Θέλουμε να ομαδοποιήσουμε τους πελάτες με βάση δύο παράγοντες: την ηλικία τους και το ετήσιο εισόδημα τους. Οπότε έστω ότι στον άξονα των χ έχουμε την ηλικία τους και στον άξονα των y το εισόδημα. Παρακάτω έχω δημιουργήσει ένα απλό διάγραμμα για τις ανάγκες του παραδείγματος..



Διάγραμμα 6

Κάθε κύκλος αναπαριστά έναν πελάτη. Όπως γίνεται αντιληπτό αν δει κανείς τα χρώματα των κύκλων έχουμε τέσσερις ομάδες πελατών(clusters). Η μπλε ομάδα αφορά μικρές ηλικίες με μικρό εισόδημα, η πορτοκαλί πάλι μικρές ηλικίες αλλά υψηλό εισόδημα, η πράσινη ομάδα μεγαλύτερες ηλικίες και χαμηλό εισόδημα και τέλος η κίτρινη μεγαλύτερες ηλικίες με υψηλό εισόδημα. Που μας χρησιμεύει αυτό; Έστω ότι έχουμε 4 είδη κιθάρας, 2 οικονομικά το ένα έχει νεανικό στυλ και το άλλο κλασσικό και 2 ακριβά το ένα με νεανικό στυλ, το άλλο κλασσικό. Με βάση τις συστάδες που φτιάξαμε μπορούμε να διαφημίσουμε την ακριβή κιθάρα με κλασσικό στυλ μόνο στην κίτρινη ομάδα κοκ. Φανταστείτε να μην κάναμε την ομαδοποίηση και να επιλέγαμε να διαφημίσουμε την ακριβή κιθάρα σε όλους. Και θα πληρώναμε περισσότερα χρήματα και το ποσοστό αγοράς/προβολής θα ήταν πολύ μικρό διότι η διαφήμιση μας δε θα ήταν καθόλου στοχευμένη.

Το παραπάνω παράδειγμα μας έδωσε μια ιδέα του πώς χρησιμοποιούν συσταδοποίηση οι eCommerce επιχειρήσεις. Στη συνέχεια θα αναφέρουμε και αναλύσουμε τρεις αλγορίθμους συσταδοποίησης:

- K-means
- Συσσωρευτική Ιεραρχική Συσταδοποίηση
- DBSCAN

K-means

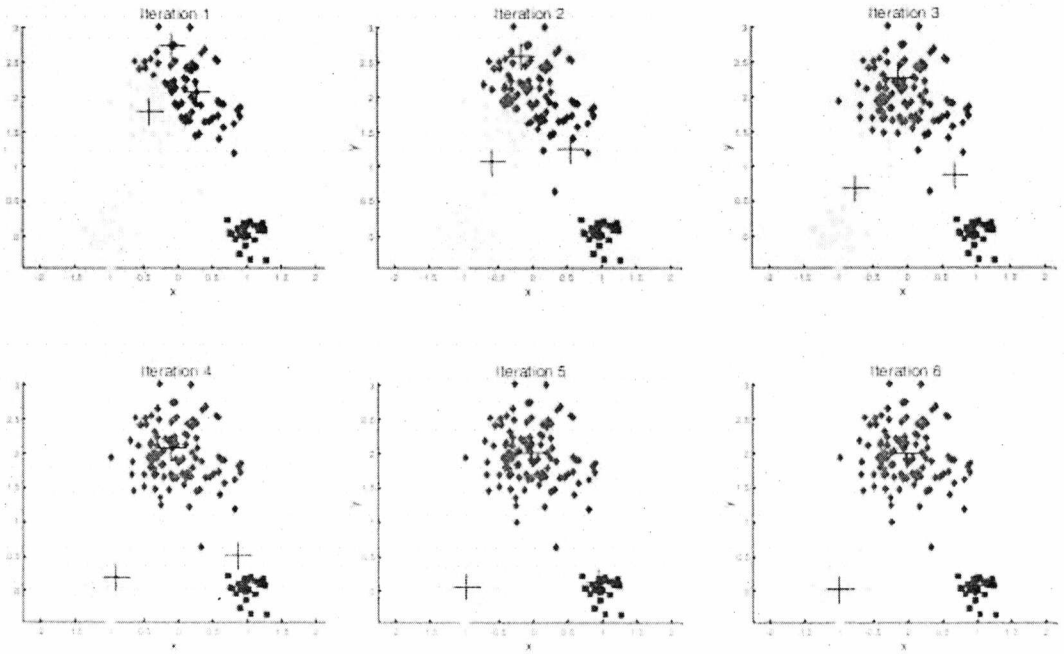
Ο πρώτος και διασημότερος αλγόριθμος συσταδοποίησης είναι ο K-means. Η βασική αρχή του είναι ότι υπάρχουν K κεντρικά σημεία κάθε ένα από τα οποία ανήκει σε μόνο μία συστάδα . Συνεπώς υπάρχουν K συστάδες. Κάθε αντικείμενο συσχετίζεται με το κοντινότερο του κεντρικό σημείο και με αυτόν τον τρόπο διαμορφώνονται οι συστάδες. Ακολουθεί ο βασικός αλγόριθμος K-means με χρήση ψευδογλώσσας

1: Επιλογή K σημείων ως τα αρχικά κεντρικά σημεία
2: ***Repeat***
3: Ανάθεση όλων των αρχικών σημείων στο κοντινότερο τους από τα K κεντρικά σημεία
4: Επαναυπολογισμός του κεντρικού σημείου κάθε συστάδας

*5: **Until** τα κεντρικά σημεία να μην αλλάζουν*

(από Introduction to Data Mining ελληνική έκδοση)

Παρατηρούμε πως ο αλγόριθμος έχει μια είσοδο Κ όπου προσδιορίζουμε πόσες συστάδες χρειαζόμαστε. Όσο περισσότερες τόσο μεγαλύτερη και η λεπτομέρεια της κάθε συστάδας. (το οποίο δεν είναι πάντα χρήσιμο διότι αυξάνει την πολυπλοκότητα και την κατανόηση των αποτελεσμάτων. Φανταστείτε στο προηγούμενο παράδειγμα μας να είχαμε δύο πελάτες ηλικίας 50 ετών με ένα ευρώ διαφορά στο εισόδημα τους αλλά παρόλαυτα να μην ανήκαν στην ίδια ομάδα λόγω ύπαρξης υπερβολικά πολλών συστάδων.) Αυτό που κάνει στη συνέχεια ο αλγόριθμος αφού αναθέσουμε τιμή στο Κ είναι να αναθέσει όλα τα σημεία στο κοντινότερο γιαυτά κεντρικό σημείο και έτσι να δημιουργηθούν οι συστάδες. Ακολουθεί ο επαναυπολογισμός του κεντρικού σημείου κάθε συστάδας. Αυτό σημαίνει ότι τα κεντρικά σημεία δεν είναι σταθερά αλλά μετακινούνται στο χώρο. Συνήθως μετακινείται στο κέντρο του cluster του χωρίς να είναι απαραίτητο να συμπίπτει με ένα απ' τα σημεία. Μπορούμε να υπολογίσουμε αυτό το «κέντρο» βρίσκοντας την μέση τιμή για τις συντεταγμένες όλων των σημείων της συστάδας. Ένα πολύ στοχευμένο παράδειγμα από το βιβλίο "Introduction to Data Mining"(δες βιβλιογραφία) που δείχνει την μετακίνηση των κεντρικών σημείων σε κάθε επανάληψη του αλγόριθμου. (τα κεντρικά σημεία συμβολίζονται με σταυρό)



Εικόνα 2 (Πηγή: Introduction to Data Mining)

Μια παραλλαγή του K-means που αξίζει να αναφέρουμε είναι ο διχοτομικός K-means. Η λογική εδώ είναι βάζουμε αρχικά όλα τα σημεία σε μια συστάδα και τη χωρίζουμε στη μέση. Επαναλαμβάνουμε το ίδιο μέχρι να παράξουμε K συστάδες. Ακολουθεί ο αλγόριθμος.

1: Αρχικοποίησε τη λίστα συστάδων ώστε να περιέχει τη συστάδα που αποτελείται από όλα τα σημεία

2: ***Repeat***

3: *Επιλογή μιας συστάδας απ' τη λίστα των συστάδων*

4: ***for*** *i=1 to num_of_trials* ***do***

5: *διχοτόμησε την επιλεγμένη συστάδα χρησιμοποιώντας τον βασικό αλγόριθμο k-means*

6: *Πρόσθεσε στη λίστα απ' τις δύο συστάδες αυτή με τη μικρότερη διασπορά*

7: ***Until*** *η λίστα να περιέχει K συστάδες*

(από Introduction to Data Mining ελληνική έκδοση)

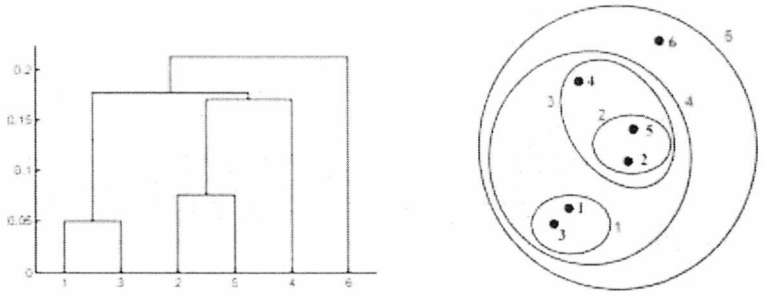
Αυτή η παραλλαγή του αρχικού αλγορίθμου μας επιλύει το πρόβλημα των «άδειων συστάδων» που συναντάτε συχνά. Αυτό σημαίνει ότι κάποιες απ' τις αρχικές συστάδες μας να είναι άδειες καθώς κανένα σημείο δεν είναι αρκετά κοντά στα κεντρικά τους σημεία.

Αφού ολοκληρώσαμε την ανάλυση μας για τον K-means αλγόριθμο θα δώσουμε κάποια μειονεκτήματα του:

- Όταν τα clusters μας έχουν διαφορετικά μεγέθη δεν μπορεί να τα διακρίνει.(πχ αν έχουμε τρεις συστάδες και η μία είναι πολύ μεγαλύτερη δεν μπορεί να την εντοπίσει)
- Όταν τα clusters μας έχουν διαφορετικές πυκνότητες πάλι δυσκολεύεται να τα ξεχωρίσει.
- Όταν τα clusters μας δεν έχουν κυκλικά σχήματα δεν μπορεί να τα εντοπίσει.

Συσσωρευτική Ιεραρχική Συσταδοποίηση

Η λογική εδώ είναι ότι έχεις κάποιες συστάδες- εμφολευμένες συστάδες η οποίες απεικονίζονται με μορφή που θυμίζει δέντρο. Αυτός ο αλγόριθμος δεν απαιτεί να του ορίσουμε τον αριθμό των συστάδων όπως ο K-means. Αντιθέτως μπορούμε να πάρουμε όσες συστάδες θέλουμε κόβοντας το δέντρο στο επιθυμητό σημείο. Στην Εικόνα 3 βλέπουμε τη μορφή του δέντρου καθώς και τις εμφολευμένες συστάδες.



Εικόνα 3 (Πηγή: Introduction to Data Mining)

Ο αλγόριθμος της Συσσωρευτικής Ιεραρχικής Συσταδοποίησης είναι ο παρακάτω:

1: Υπολόγισε τη μήτρα εγγύτητας

2: ***Repeat***

3: Συγχώνευσε τις πλησιέστερες δύο συστάδες

4: Ενημέρωσε τη μήτρα εγγύτητας

5: ***Until*** *τα κεντρικά σημεία να μην αλλάζουν*

(από Introduction to Data Mining ελληνική έκδοση)

Μήτρα εγγύτητας ή αλλιώς πίνακας γειτνίασης είναι ένας δυσδιάστατος πίνακας (εκτός και αν δουλεύουμε σε τρισδιάστατο χώρο κλπ) όπου οι γραμμές του=στήλες του=σύνολο των συστάδων. Η διαδικασία δημιουργία του πίνακα είναι η εξής:

- Αρχικά κάθε σημείο είναι και μια συστάδα στον πίνακα.
- Έπειτα από τις συγχωνεύσεις σημείων-συστάδων ενημερώνεται ο πίνακας με τις νέες συστάδες (στο παράδειγμα της εικόνας 3 διακρίνονται 5 συστάδες).

- Σύμφωνα με τον αλγόριθμο πρέπει να συγχωνεύσουμε τις δύο κοντινότερες συστάδες.
- Οπότε ο πίνακας αν είχε ν συστάδες τώρα θα έχει ν-1 μετά και απ' την συγχώνευση.
- Η ενημέρωση του πίνακα γειτνίασης πραγματοποιείται με διάφορους τρόπους:
    - Μικρότερη απόσταση: Δηλαδή με βάση τη μικρότερη απόσταση μεταξύ των συστάδων το οποίο σημαίνει ότι για δύο συστάδες παίρνουμε την απόσταση απ' τα δύο πιο κοντινά τους σημεία χωρίς να παίζουν ρόλο τα υπόλοιπα.
    - Μεγαλύτερη απόσταση: Δηλαδή με βάση τη μεγαλύτερη απόσταση μεταξύ δύο συστάδων, δηλαδή παίρνουμε την απόσταση από τα πιο μακρινά σημεία τους.
    - Μέση απόσταση: Όπως λέει και το όνομα της είναι η μέση τιμή της απόστασης όλων των σημείων.
    - Κεντρική απόσταση: Δηλαδή η απόσταση μεταξύ των κεντρικών σημείων των συστάδων.

Το μειονέκτημα του συγκεκριμένου αλγόριθμου είναι ότι οι αποφάσεις που παίρνουμε είναι και τερματικές. Για παράδειγμα αφού συγχωνεύσουμε δύο συστάδες δεν μπορούμε να τις ξαναχωρίσουμε.

DBSCAN

Ο συγκεκριμένος αλγόριθμος συσταδοποίησης αναζητεί και βρίσκει περιοχές με υψηλή πυκνότητα με τη βοήθεια περιοχών χαμηλής πυκνότητας. Αναλυτικότερα παίρνουμε κάθε σημείο ξεχωριστά και βρίσκουμε το σύνολο των σημείων που εντοπίζονται σε μια ακτίνα Eps απ' αυτό. Με βάση το αποτέλεσμα το σημείο θα είναι σημείο πυρήνα, σημείο ορίου ή σημείο θορύβου

- Σημείο πυρήνα: Αν μέσα στην ακτίνα του Eps υπάρχει συγκεκριμένος αριθμός σημείων που συμπίπτει ή είναι μεγαλύτερος από ένα προκαθορισμένο κατώφλι MinPts τότε είναι σημείο πυρήνα.

- Σημείο ορίου: Ένα σημείο που βρίσκεται εντός της ακτίνας Eps ενός σημείου πυρήνα τότε λέγεται σημείο ορίου
- Σημείο θορύβου: Ένα σημείο που βρίσκεται εκτός της ακτίνας ενός σημείου πυρήνα (και δεν είναι το ίδιο πυρήνας) ονομάζεται σημείο θορύβου.

Ο αλγόριθμος DBSCAN δίνεται παρακάτω:

1: Χαρακτήρισε όλα τα σημεία ως πυρήνα, ορίου, θορύβου.
2: Διέγραψε τα σημεία θορύβου.
3: Τοποθέτησε μια ακμή ανάμεσα σε όλα τα σημεία πυρήνα που βρίσκονται εντός απόστασης Eps μεταξύ τους.
4: Όρισε κάθε ομάδα συνδεδεμένων σημείων πυρήνα ως χωριστή συστάδα.
5: Εκχώρησε κάθε σημείο ορίου σε μια από τις συστάδες των σημείων πυρήνα.

(από Introduction to Data Mining ελληνική έκδοση)

Τα αρνητικά του DBSCAN αφορούν αρχικά τις πυκνότητες που ποικίλουν σε μέγεθος καθώς δεν μπορεί να τις διαχειριστεί με επιτυχία. Επίσης αν έχουμε δεδομένα πολλών διαστάσεων οι πυκνότητες θα είναι πιο δύσκολο να οριστούν και συνεπώς ο αλγόριθμος πάλι θα παρουσιάσει πρόβλημα.

Κατηγοριοποίηση (classification)

Άλλη μια ευρέως διαδεδομένη κατηγορία αλγορίθμων εξόρυξης κοινωνικών δεδομένων είναι η κατηγοριοποίηση. Όπως λέει και το όνομα της θέλουμε να τοποθετήσουμε τα δεδομένα σε κάποια κατηγορία. Πως γίνεται αυτό; Αρχικά κάθε δεδομένο έχει ένα σύνολο από χαρακτηριστικά. (Δε μας περιορίζει ο αριθμός και το είδος των χαρακτηριστικών.) Έπειτα με τη χρήση ενός αλγόριθμου-μοντέλου κατηγοριοποίησης τα επεξεργαζόμαστε και καταλήγουμε σε ένα αποτέλεσμα που αφορά την κατηγορία στην οποία ανήκει το δεδομένο. Όλη αυτή η διαδικασία συνήθως περιγράφεται με μια συνάρτηση f(x,y) όπου "x" είναι το σύνολο των ιδιοτήτων και "y" η κατηγορία στην οποία ανήκει. Η συνάρτηση είναι το μοντέλο κατηγοριοποίησης που χρησιμοποιούμε.

Εικόνα 4

Για παράδειγμα ας θεωρήσουμε τον παρακάτω πίνακα με κάποια τυχαία στοιχεία πελατών που συλλέξαμε.

ID	Ηλικία	Χώρα	Εισόδημα	Αγοραστική Συμπεριφορά
1	57	Ελλάδα	1200	Τσιγκούνης
2	31	ΗΠΑ	3000	Ανοιχτοχέρης
3	19	ΗΠΑ	900	Τσιγκούνης
4	21	Ελλάδα	500	Ανοιχτοχέρης
5	34	ΗΠΑ	2000	Ανοιχτοχέρης

Πίνακας 1

Παρατηρούμε πως υπάρχουν 3 ιδιότητες Ηλικία, Χώρα, Εισόδημα και η κατηγορία Αγοραστική Συμπεριφορά που προσδιορίζει πόσο εύκολα αγοράζει ο καθένας τους κάθε προϊόν.(το παράδειγμα δημιουργήθηκε για τους σκοπούς της πτυχιακής μου εργασίας και δεν γνωρίζω αν ανταποκρίνεται στην πραγματικότητα)
Η διαδικασία όπως είπαμε και πριν είναι να πάρουμε το σύνολο των ιδιοτήτων, να βάλουμε σε μια συνάρτηση ή μοντέλο κατηγοριοποίησης και να πάρουμε μια έξοδο (Τσιγκούνης ή Ανοιχτοχέρης) που στην προκειμένη περίπτωση θα αφορά την αγοραστική συμπεριφορά των πελατών.

Διακρίνουμε το μοντέλο κατηγοριοποίησης σε δύο κατηγορίες:

1. Περιγραφικό μοντέλο: Με τη βοήθεια του μπορούμε να κατανοήσουμε τι χαρακτηριστικά ανήκουν σε μια συγκεκριμένη κατηγορία. Να πούμε για παράδειγμα ότι αν η χώρα που μένει ο πελάτης είναι η Αμερική τότε η αγοραστική του συμπεριφορά είναι ανοιχτοχέρης.
2. Προβλεπτικό μοντέλο: Όταν δεν έχουμε στοιχεία για κάποιες τιμές ιδιοτήτων όπως για παράδειγμα χώρα Γερμανία μπορούμε να βάλουμε το μοντέλο μας να προβλέψει την κατηγορία στην οποία ανήκει ο συγκεκριμένος πελάτης.

Τα πιο γνωστά μοντέλα κατηγοριοποίησης δίνονται παρακάτω:

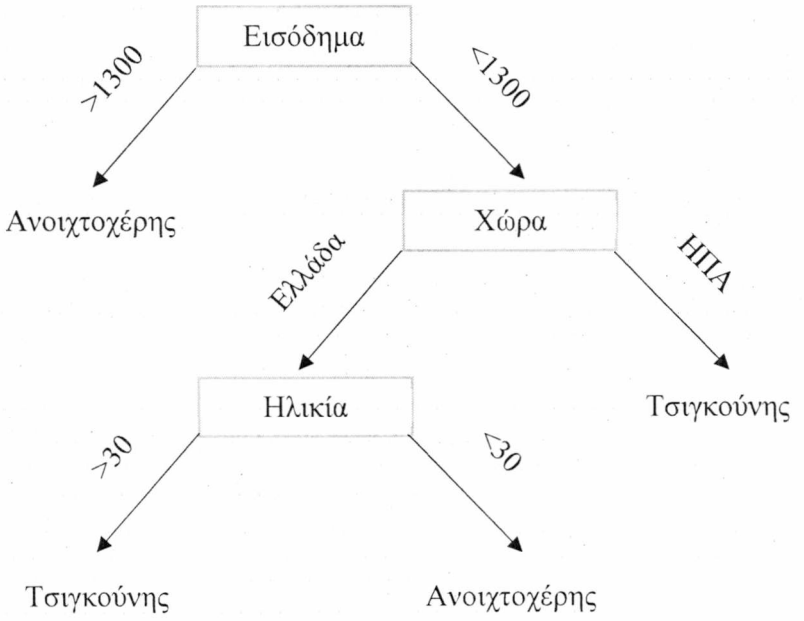
- Δέντρα απόφασης
- Μέθοδος Κανόνων
- Κοντινότεροι γείτονες
- Κατηγοριοποιητές Bayes

- Νευρωνικά δίκτυα
- Διανύσματα υποστήριξης

Στο κεφάλαιο αυτό θα ασχοληθούμε τρία από αυτά τα μοντέλα: τα δέντρα απόφασης, τη μέθοδο των κανόνων και τους κατηγοριοποιητές Bayes.

Δέντρα απόφασης

Τα δέντρα απόφασης είναι απ' τα απλότερα μοντέλα κατηγοριοποίησης που υπάρχουν αλλά είναι ευρέως διαδεδομένα και χρησιμοποιούνται από πολλά άτομα/επιχειρήσεις. Βασίζεται στη λογική ενός δέντρου στου οποίου τη ρίζα υπάρχει μια ιδιότητα (μπορούμε να επιλέξουμε όποια ιδιότητα επιθυμούμε) και στα φύλλα υπάρχουν τα χαρακτηριστικά μας. Για να φτάσουμε απ' τη ρίζα στα φύλλα χρησιμοποιούμε μια σειρά από ερωτήσεις με βάση τις τιμές των ιδιοτήτων που έχουμε. Στο παράδειγμα με τους πελάτες που δώσαμε πριν η ρίζα θα μπορούσε για να είναι το εισόδημα. Τα φύλλα θα ήταν τσιγκούνης ή ανοιχτοχέρης στα οποία θα φτάναμε μέσω ερωτήσεων . Δηλαδή θα ήταν κάπως έτσι:



Διάγραμμα 7

Αρχικά λοιπόν ρωτάμε αν το εισόδημα είναι μεγαλύτερο των 1300 ευρώ. Αν ισχύει τότε μπορούμε να θέσουμε ως την κατηγορία του συγκεκριμένου ατόμου ως

Ανοιχτοχέρη χωρίς να κοιτάξουμε άλλες ιδιότητας. Αν δεν ισχύει οδηγούμαστε στην επόμενη ερώτηση της χώρας. Αν είναι από ΗΠΑ έχουμε τελειώσει και τον κατηγοριοποιούμε ως Τσιγκούνη. Αν είναι από Ελλάδα όμως οδηγούμαστε στην επόμενη ερώτηση της ηλικίας. Αν είναι κάτω των 30 είναι Ανοιχτοχέρης , αν είναι πάνω από 30 είναι Τσιγκούνης και κάπου εδώ φτάσαμε στο τέλος.

Αναφέραμε πριν ότι ως ρίζα μπορούμε να επιλέξουμε μια ιδιότητα της αρεσκείας μας. Το ίδιο ισχύει και με τους κόμβους παιδιά. Πχ στο προηγούμενο παράδειγμα δεν μας εμπόδιζε κανείς να αλλάξουμε θέσεις στη χώρα και στην ηλικία στο δέντρο. Αυτό σημαίνει πως θα έχουμε εκθετικό αριθμό δέντρων. Κάποια από αυτά είναι πιο δύσχρηστα-αργά και κάποια άλλα πιο αποτελεσματικά. Η διαδικασία του να επιλέξουμε το πιο αποτελεσματικό δέντρο είναι συχνά πολύ δύσκολη γιατί φανταστείτε αντί για 3 ιδιότητες να είχαμε 300; Πόσες διαφορετικές περιπτώσεις θα υπήρχαν. Παρόλαυτα έχουν αναπτυχθεί κάποιοι αλγόριθμοι που προσπαθούν να επιλέξουν το πιο αποτελεσματικό δέντρο με τον πιο γνωστό από αυτούς τον αλγόριθμο του Hunt να παρουσιάζεται παρακάτω.

- Αλγόριθμος του Hunt

  Είναι ένας άπληστος (greedy) αλγόριθμος ο οποίος χτίζει το δέντρο αναδρομικά. Δηλαδή αρχίζει από μία ρίζα και θέτει και τα δύο φύλλα της σε μια συγκεκριμένη κατηγορία χωρίς παιδιά έπειτα βάζει κόμβους-παιδιά και με αυτόν τον τρόπο συνεχίζει τη διαδικασία αναδρομικά

Μέθοδοι κανόνων

Το επόμενο μοντέλο κατηγοριοποίησης που θα αναλύσουμε είναι αυτό της δημιουργίας κανόνων. Οι κανόνες έχουν την προγραμματιστική λογική "if-then-else" που μεταφράζεται ότι αν μια εγγραφή (στο γνωστό παράδειγμα μας ένας πελάτης) καλύπτεται από έναν κανόνα τότε της προσδίδουμε το χαρακτηριστικό του κανόνα και αν όχι οδηγούμαστε σε έναν επόμενο.

Οι κανόνες για τους οποίους μιλάμε έχουν την μορφή r_i : (Condition$_i$) -> y_i με το μέρος του κανόνα πριν το τοξάκι να ονομάζεται συνθήκη εισόδου και το μέρος μετά το τοξάκι συνθήκη εξόδου ή επακόλουθο χαρακτηριστικό. Ακολουθεί ένα παράδειγμα:

Ας κατασκευάσουμε κανόνες που προκύπτουν απ' το γνωστό μας παράδειγμα της αγοραστικής συμπεριφοράς κάποιων πελατών. Για λόγους διευκόλυνσης επαναλαμβάνεται εδώ ο πίνακας.

ID	Ηλικία	Χώρα	Εισόδημα	Αγοραστική Συμπεριφορά
1	57	Ελλάδα	1200	Τσιγκούνης
2	31	ΗΠΑ	3000	Ανοιχτοχέρης
3	19	ΗΠΑ	900	Τσιγκούνης
4	21	Ελλάδα	500	Ανοιχτοχέρης
5	34	ΗΠΑ	2000	Ανοιχτοχέρης

r_1 : (Εισόδημα>1300) -> ανοιχτοχέρης

r_2 : (Εισόδημα<1300) ∩ (Χώρα=ΗΠΑ) -> τσιγκούνης

r_3 : (Εισόδημα<1300) ∩ (Χώρα=Ελλάδα) ∩ (Ηλικία >30) -> τσιγκούνης

r_4 : (Εισόδημα<1300) ∩ (Χώρα=Ελλάδα) ∩ (Ηλικία <30) -> ανοιχτοχέρης

Παρατηρήστε πως οι παραπάνω κανόνες πέρα από τα δεδομένα του πίνακα μπορούν να βγουν εύκολα και απ' το δέντρο που είχαμε δημιουργήσει στο προηγούμενο μοντέλο κατηγοριοποίησης με δέντρα απόφασης. Συνεπώς η μέθοδος εξαγωγής κανόνων θα ανήκει σε μία απ' τις δύο κατηγορίες:

- Άμεση: κάθε κανόνας απορρέει από τα δεδομένα που έχουμε όπως παράδειγμα ο πίνακας με τους υποψήφιους πελάτες. Ένας αλγόριθμος που κάνει αυτόματα αυτή τη διαδικασία είναι ο Learn-One-Rule.

- Έμμεση : κάθε κανόνας απορρέει με την βοήθεια άλλων μοντέλων κατηγοριοποίησης όπως τα δέντρα απόφασης ή τα Νευρωνικά Δίκτυα τα οποία θα αναλύσουμε αργότερα στην εργασία.

Κριτήρια που ορίζουν την «ποιότητα » ενός κανόνα είναι η κάλυψη και η ακρίβεια. Το πρώτο υπολογίζεται ως οι εγγραφές που καλύπτονται απ' τον κανόνα προς όλες τις εγγραφές. Για παράδειγμα ο κανόνας r_1 καλύπτει δύο πελάτες οπότε κάλυψη 2/5= 40%. Το δεύτερο υπολογίζεται ως οι εγγραφές με έξοδο y που καλύπτονται απ' τον κανόνα προς όλες τις εγγραφές με έξοδο y. Για παράδειγμα ο κανόνας r_1 πάλι έχει ακρίβεια 2/3=66%.

Φανταστείτε τέλος μια εγγραφή στον πίνακα των πελατών μας να καλυπτόταν από δύο κανόνες οι οποίοι έβγαζαν διαφορετική έξοδο y. Τι θα κάναμε σε αυτό το είδος σύγκρουσης;

- Μια λύση είναι οι διατεταγμένοι κανόνες. Θα καταγράφαμε τους κανόνες σε μια φθίνουσα σειρά με βάση μια προτεραιότητα. Οπότε η εγγραφή θα επέλεγε τον κανόνα με τη μεγαλύτερη προτεραιότητα και δε θα μπορούσε ποτέ να έχει δυο εξόδους.
- Οι μη διατεταγμένοι κανόνες είναι η δεύτερη λύση. Με αυτήν την τεχνική μια εγγραφή μπορεί να ανήκει σε περισσότερους από έναν κανόνες. Σε μια λίστα καταγράφει τις εξόδους του κάθε κανόνα με τη μορφή «ψήφων». Κάθε φορά που ένας κανόνας έχει την ίδια έξοδο με έναν άλλο τότε ενημερώνεται ο πίνακας. Τέλος κερδίζει ο κανόνας με τους περισσότερους ψήφους.

Κατηγοριοποιητές Bayes

Έστω ότι έχουμε έναν πελάτη στην επιχείρηση και θέλουμε να τον κατατάξουμε σε μια κατηγορία. Μια εναλλακτική λύση πέρα απ' τα δέντρα απόφασης και τους κανόνες που έχουμε μελετήσει είναι να χρησιμοποιήσουμε πιθανότητες. Ο Bayes μας διευκολύνει σε αυτό με τον παρακάτω τύπο του

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Ο οποίος υπολογίζει μια υπό συνθήκη πιθανότητα P(Y|X).

Υπάρχουν δύο υλοποιήσεις των κατηγοριοποιητών του Bayes.

- Απλοϊκός Bayes
- Δίκτυο πεποίθησης του Bayes

Εμείς θα ασχοληθούμε μόνο με ην πρώτη υλοποίηση, τον Απλοϊκό Bayes. του οποίου ο τύπος είναι:

$$P(X|Y = y) = \prod_{i=1}^{d} P(X_i|Y = y)$$

όπου κάθε σύνολο χαρακτηριστικών X = {X1,X2,X3,…,Xd} αποτελείται από d χαρακτηριστικά. Πάμε όμως να πάρουμε τα πράγματα με τη σειρά με ένα παράδειγμα ώστε να γίνει ευκολότερη η κατανόηση του συγκεκριμένου τρόπου κατηγοριοποίησης.

Έστω ότι πάλι έχουμε το αγαπημένο μας παράδειγμα με την αγοραστική συμπεριφορά των πελατών ο πίνακας του οποίου επαναλαμβάνεται παρακάτω:

ID	Ηλικία	Χώρα	Εισόδημα	Αγοραστική Συμπεριφορά
1	57	Ελλάδα	1200	Τσιγκούνης
2	31	ΗΠΑ	3000	Ανοιχτοχέρης
3	19	ΗΠΑ	900	Τσιγκούνης
4	21	Ελλάδα	500	Ανοιχτοχέρης
5	34	ΗΠΑ	2000	Ανοιχτοχέρης
6	26	Ελλάδα	700	?

Όπως παρατηρήσατε στον πίνακα έχουμε προσθέσει μια νέα σειρά. Είναι ένας νέος πελάτης που θα προσπαθήσουμε να τον κατηγοριοποιήσουμε με την υλοποίηση του απλού Bayes σε τσιγκούνη ή ανοιχτοχέρη.

Η πρώτη μας κίνηση είναι στις μέχρι τώρα εγγραφές μας να βρούμε τις πιθανότητες για κάποιον πελάτη να είναι τσιγκούνης και τις πιθανότητες να είναι ανοιχτοχέρης. Με το γράμμα «Α» θα συμβολίζουμε τον ανοιχτοχέρη και με το γράμμα «Τ» τον τσιγκούνη. Έχουμε:

$$P(A) = \frac{3}{5} = 0.6$$

$$P(T) = \frac{2}{5} = 0.4$$

Ο τύπος του απλοϊκού Bayes για το παράδειγμα μας θα έχει τις παρακάτω δύο μορφές για την κάθε κατηγορία που έχουμε:

$$P(X|A) = P(\textit{εισόδημα} < 1300|A) \times P(\textit{χώρα} = \textit{Ελλάδα}|A) \times P(\textit{ηλικία} < 30|A)$$

$$P(X|T) = P(\textit{εισόδημα} < 1300|T) \times P(\textit{χώρα} = \textit{Ελλάδα}|T) \times P(\textit{ηλικία} < 30|T)$$

Πριν λοιπόν μπορέσουμε να υπολογίσουμε αυτές τις πιθανότητες πρέπει να βρούμε τις έξι επιμέρους πιθανότητες. Οπότε θα έχουμε σύμφωνα με τον κανόνα του Bayes:

$$P(\text{εισόδημα} < 1300|A) = \frac{P(A|\text{εισόδημα} < 1300) \times P(A)}{P(\text{εισόδημα} < 1300)} = \frac{\frac{1}{3} \times 0.6}{\frac{3}{5}} = 0.33$$

$$P(\text{εισόδημα} < 1300|T) = \frac{P(T|\text{εισόδημα} < 1300) \times P(T)}{P(\text{εισόδημα} < 1300)} = \frac{\frac{2}{2} \times 0.4}{\frac{3}{5}} = 0.66$$

$$P(\chi\acute{\omega}\rho\alpha = Ελλάδα|A) = \frac{P(A|\chi\acute{\omega}\rho\alpha = Ελλάδα) \times P(A)}{P(\chi\acute{\omega}\rho\alpha = Ελλάδα)} = \frac{\frac{1}{3} \times 0.6}{\frac{2}{5}} = 0.495$$

$$P(\chi\acute{\omega}\rho\alpha = Ελλάδα|T) = \frac{P(T|\chi\acute{\omega}\rho\alpha = Ελλάδα) \times P(T)}{P(\chi\acute{\omega}\rho\alpha = Ελλάδα)} = \frac{\frac{1}{2} \times 0.4}{\frac{2}{5}} = 0.5$$

$$P(\text{ηλικία} < 30|A) = \frac{P(A|\text{ηλικία} < 30) \times P(A)}{P(\text{ηλικία} < 30)} = \frac{\frac{1}{3} \times 0.6}{\frac{2}{5}} = 0.495$$

$$P(\text{ηλικία} < 30|T) = \frac{P(T|\text{ηλικία} < 30) \times P(T)}{P(\text{ηλικία} < 30)} = \frac{\frac{1}{2} \times 0.4}{\frac{2}{5}} = 0.5$$

Οπότε αφού υπολογίσαμε αυτές τις 6 τιμές είμαστε έτοιμη να χρησιμοποιήσουμε τον τύπο του απλοϊκού Bayes. Έχουμε:

$$P(X|A) = 0.33 \times 0.495 \times 0.495 = 0.0808$$

και

$$P(X|T) = 0.66 \times 0.5 \times 0.5 = 0.165$$

Τέλος πρέπει να βρούμε την εκ των υστέρων πιθανότητα για κάθε μια απ' τις δυο κατηγορίες. Οπότε:

$$P(A|X) = P(X|A) \times \frac{3}{5} = 0.0808 \times 0.6 = 0.048$$

και

$$P(T|X) = P(X|T) \times \frac{2}{5} = 0.165 \times 0.3 = 0.049$$

Συνεπώς ο νέος πελάτης μας κατηγοριοποιείται ως τσιγκούνης.

Κανόνες Συσχέτισης (association rules)

Η τελευταία κατηγορία εξόρυξης κοινωνικών δεδομένων που θα μελετήσουμε είναι οι κανόνες συσχέτισης. Είναι ιδιαίτερα χρήσιμη σε μια επιχείρηση ηλεκτρονικού εμπορίου καθώς μπορεί να υποδείξει συμπεριφορές των καταναλωτών-πελατών που δε φαίνονται με γυμνό μάτι. Αυτή η συμπεριφορά μετά μπορεί να χρησιμοποιηθεί για πολλούς σκοπούς όπως cross sales (διασταυρωμένες πωλήσεις), κουπόνια εκπτώσεων κτλ.

Θα επεκταθούμε στο θέμα των cross sales με ένα παράδειγμα. Όσοι δεν είστε εξοικειωμένοι με τον όρο "Cross Sales" ορίζονται ως η προσπάθεια των παραγωγών να μεγιστοποιήσουν το κέρδος προτείνοντας στους αγοραστές και άλλα παρεμφερή προϊόντα ή υπηρεσίες. Ας υποθέσουμε λοιπόν πως έχουμε ένα ηλεκτρονικό κατάστημα το οποίο πουλάει σχολικά προϊόντα (e-βιβλιοπωλείο). Πρόσφατα κάναμε πέντε πωλήσεις και κρατήσαμε ένα αρχείο με τα προϊόντα που περιέχονταν στην κάθε πώληση:

Πώληση 1	{μολύβι ,γόμα, τετράδιο}
Πώληση 2	{στυλό, μολύβι}
Πώληση 3	{στυλό, τετράδιο}
Πώληση 4	{στυλό, μολύβι, γόμα, τετράδιο}
Πώληση 5	{μολύβι, γόμα}

Πίνακας 2

Παρατηρούμε ότι υπάρχει άμεση και ισχυρή σχέση μεταξύ του μολυβιού και της γόμας στις πωλήσεις. Αυτό σημαίνει πως θα ήταν σοφό σε κάποιον που αγόραζε μολύβι να του προτείναμε να αγοράσει και μια γόμα. Αυτή είναι η έννοια των cross sales. Αυτή η σχέση που μόλις αναφέραμε θα μπορούσε να παρασταθεί και με έναν κανόνα συσχέτισης όπως φαίνεται παρακάτω:

{μολύβι -> γόμα}

Μια διαφορετική αναπαράσταση του παραπάνω πίνακα θα μπορούσε να γίνει με τη βοήθεια διάδικων αριθμών για τα δεδομένα.

Πώληση	στυλό	μολύβι	γόμα	τετράδιο
1	0	1	1	1
2	1	1	0	0
3	1	0	0	1
4	1	1	1	1

5	0	1	1	0

Πίνακας 3

Είναι εύκολο να καταλάβει κανείς ότι όπου υπάρχει 1 σημαίνει ότι το προϊόν συμμετείχε στη συγκεκριμένη αγορά και όπου υπάρχει 0 όχι.

Αν τώρα δημιουργήσουμε δύο σύνολα, το σύνολο Π={π1,π2,π3,π4,π5} που περιέχει όλες τις πωλήσεις και το Α={στυλό, μολύβι, γόμα, τετράδιο} που περιέχει όλα τα αντικείμενα τότε κάθε πώληση ξεχωριστά απ' το σύνολο των πωλήσεων θα περιέχει μια συλλογή από αντικείμενα. Αυτή η συλλογή και κάθε τέτοια συλλογή (όχι απαραίτητα των πωλήσεων, μπορεί να είναι και μια τυχαία με βάση τα δεδομένα) ονομάζεται *στοιχειοσύνολο*. Αυτό χρησιμεύει στην δημιουργία κανόνων συσχέτισης με στοιχειοσύνολα τα οποία είναι ξένα μεταξύ τους.

Υπάρχουν δύο παράμετροι για να κρίνουν το πόσο ισχυρός είναι ένας κανόνας συσχέτισης.

- Υποστήριξη(support): Αυτή η παράμετρος αν είναι αρκετά χαμηλή μπορεί να υποδηλώνει ότι ο κανόνας μας δε σημαίνει απολύτως τίποτα για την συμπεριφορά των πελατών και ήταν απλά μια τυχαία εξαίρεση. Μπορούμε να την υπολογίσουμε με τον τύπο:

$$Support, s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N}$$

(Βιβλίο: Εισαγωγή στην εξόρυξη δεδομένων)

- Εμπιστοσύνη(confidence): Αυτή η παράμετρος υποδηλώνει πως με την αγορά προϊόντων ενός στοιχειοσύνολου είναι πολύ συχνό να αγοραστούν και τα προϊόντα του άλλου στοιχειοσύνολου στον κανόνα. Ο τύπος της συγκεκριμένης παραμέτρου είναι:

$$Confidence, c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)}$$

(Βιβλίο: Εισαγωγή στην εξόρυξη δεδομένων)

Το πρόβλημα που συναντάμε είναι ότι όταν έχουμε πάρα πολλούς πελάτες, πάρα πολλά δεδομένα πως θα εξάγουμε συσχετίσεις-κανόνες από αυτά. Και ακόμα χειρότερα πως θα αξιολογήσουμε τους κανόνες με βάση τους δύο παραμέτρους που αναφέραμε πριν; Θα παίρνουμε για κάθε κανόνα και θα κάνουμε όλες τις πράξεις; Ας πάρουμε για παράδειγμα την πώληση 4 στον πίνακα 2 και ας προσπαθήσουμε να φτιάξουμε αρκετούς απ' τους κανόνες της.

{στυλό, μολύβι, γόμα} -> {τετράδιο}
{στυλό, μολύβι} -> {γόμα, τετράδιο}
{στυλό} -> {μολύβι, γόμα, τετράδιο}
{στυλό, γόμα} -> {μολύβι, τετράδιο}
{στυλό, τετράδιο} -> (μολύβι, γόμα}
{στυλό, μολύβι, τετράδιο} -> {γόμα}
{στυλό, γόμα, τετράδιο} -> {μολύβι}

και αυτοί είναι απλά κάποιοι κανόνες. Όλοι αυτοί οι κανόνες ανήκουν στο ίδιο στοιχειοσύνολο {στυλό, μολύβι, γόμα, τετράδιο} Πρέπει συνεπώς με κάποιον τρόπο να απαλείφουμε τα στοιχειοσύνολα που δεν εμφανίζονται συχνά ώστε να μειωθεί η πολυπλοκότητα. Οπότε η διαδικασία που πρέπει να ακολουθήσουμε περιλαμβάνει δύο βήματα

1. Εύρεση Συχνών Στοιχειοσυνόλων
2. Δημιουργία Κανόνων

Εύρεση Συχνών Στοιχειοσυνόλων

Σύμφωνα με το βιβλίο «Εισαγωγή στην εξόρυξη δεδομένων»(δες βιβλιογραφία) «*αν ένα σύνολο δεδομένων περιέχει κ αντικείμενα τότε υπάρχουν $2^κ$ στοιχειοσύνολα που μπορεί να παράγει.*» Αυτός είναι τεράστιος αριθμός.

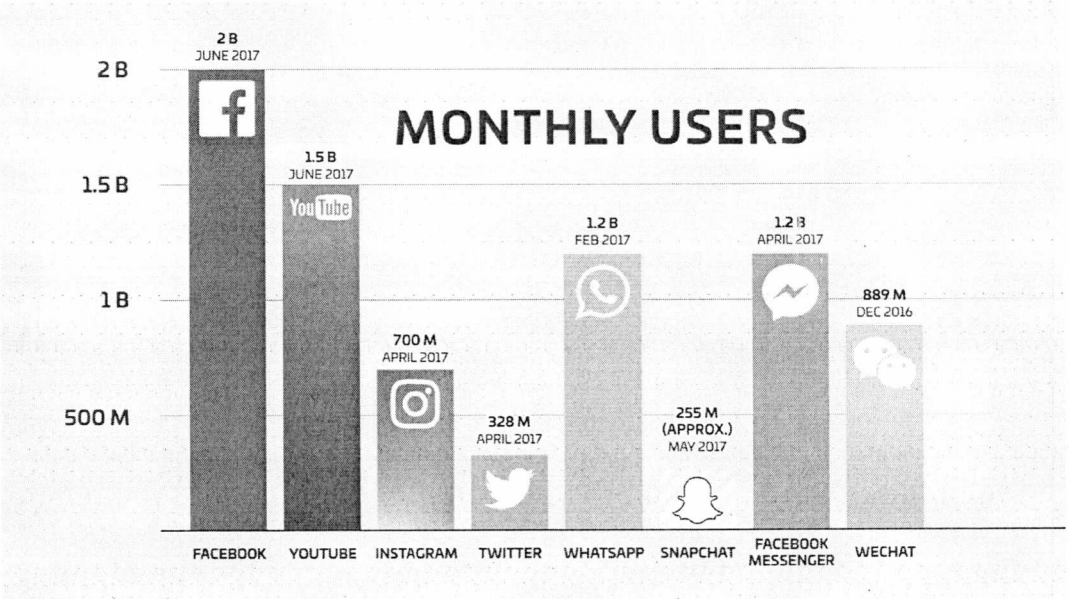
Για να μικρύνει αυτός ο αριθμός χρησιμοποιούνται αλγόριθμοι όπως ο αλγόριθμος του Apriori, ο αλγόριθμος FPGROWTH. τους οποίους δε θα αναλύσουμε στη συγκεκριμένη πτυχιακή εργασία.

Δημιουργία Κανόνων

Με τα στοιχειοσύνολα λοιπόν που ανακαλύψαμε με κάποιον απ' τους παραπάνω αλγόριθμους θα δημιουργήσουμε κανόνες συσχέτισης ώστε να μπορέσουμε να εξάγουμε συμπεράσματα για την αγοραστική συμπεριφορά των πελατών της eCommerce επιχείρησης μας. Οι κανόνες που θα δημιουργήσουμε πρέπει να έχουν μεγάλη υποστήριξη ώστε να σημαίνει ότι δεν είναι τυχαίος. Για την δημιουργία κανόνων έχουν δημιουργηθεί αλγόριθμοι με τον σημαντικότερο πάλι να είναι του Apriori αυτήν τη φορά όμως για παραγωγή κανόνων.

Έτοιμα εργαλεία εξόρυξης

Πέρα απ' τις μεθόδους και αλγορίθμους εξόρυξης που είδαμε προηγουμένως στην εργασία ,υπάρχουν έτοιμα εργαλεία εξόρυξης κοινωνικών δεδομένων που μπορεί να φανούν χρήσιμα σε μια eCommerce επιχείρηση. Κυκλοφορούν πολλά τέτοια εργαλεία στο διαδίκτυο με τα δύο δημοφιλέστερα εξ αυτών να είναι το Audience Insights του Facebook και το Google trends τα οποία θα μελετήσουμε εκτενώς. Οι δύο αυτοί κολοσσοί που κατέχουν τα συγκεκριμένα εργαλεία και τα προσφέρουν δωρεάν στις επιχειρήσεις εκτιμάται πως ξεπερνούν τα τρία δισεκατομμύρια μηνιαίους χρήστες παγκοσμίως.



Διάγραμμα 8 (Πηγή techbranch.com)

Απ' το παραπάνω διάγραμμα το YouTube ανήκει στη Google και το Instagram, What's up, Messenger ανήκουν στο Facebook. Αυτή η μεγάλη βάση δεδομένων με χρήστες που διαθέτουν είναι ένας σημαντικός λόγος για να πείσει μια επιχείρηση να τα χρησιμοποιήσει. Διότι όλοι αυτοί οι χρήστες από την πλευρά μιας επιχείρησης και ιδιαίτερα μιας eCommerce επιχείρησης μεταφράζονται σε επίδοξους πελάτες.

Facebook Audience Insights

Το πρώτο εργαλείο εξόρυξης κοινωνικών δεδομένων που θα μελετήσουμε είναι το Facebook Audience Insights. Μέσω της μεγάλης βάσης χρηστών που διαθέτει με τη χρήση αλγόριθμων παρόμοιων με αυτόν που μελετήσαμε, εξάγει αποτελέσματα με τα οποία eCommerce επιχειρήσεις μπορούν να βγάλουν σημαντικά συμπεράσματα για την αγοραστική συμπεριφορά των πελατών. Πάμε να πάρουμε τα πράγματα με τη σειρά.

Η διαδικασία για να εισέλθει κάποιος μέσα στο Facebook Audience Insights είναι να κατέχει έναν λογαριασμό είτε στο Business Manager είτε στο Ads Manager. Όταν δημιουργήσει τον λογαριασμό αρκεί να ακολουθήσει το μονοπάτι Menu -> Plan -> Audience Insights. Η πρώτη σελίδα που αντικρίζουμε όταν ολοκληρώσουμε την διαδικασία φαίνεται παρακάτω:

Εικόνα 5

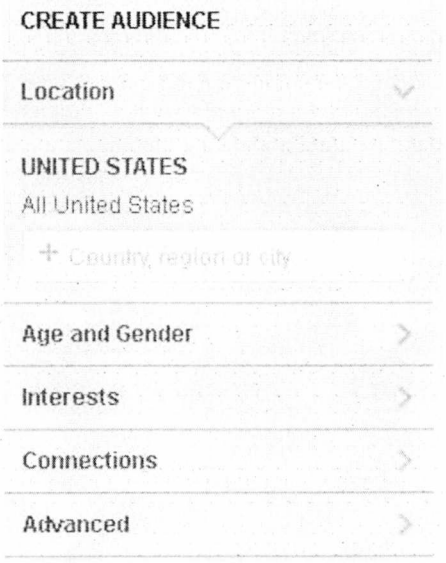
Με μια πρώτη ματιά καταλαβαίνει κανείς ότι πρόκειται για ένα γραφικά διαμορφωμένο εργαλείο με διαγράμματα και στατιστικές μελέτες των χρηστών του (όλα αυτά προέκυψαν από αλγόριθμους εξόρυξης δεδομένων). Περιλαμβάνει δύο μέρη. Τη δημιουργία του κοινού που θέλουμε να αναλύσουμε (αριστερή στήλη) και τα δεδομένα/πληροφορίες που προκύπτουν απ' αυτό (δεξιά μεγάλη στήλη) Εν συνεχεία θα αναλύσουμε τα περιεχόμενα κάθε μέρους ώστε να αποκτήσουμε μια πιο σφαιρική άποψη του αντικειμένου.

Δημιουργία κοινού

Μέσω αυτής της στήλης μπορούμε να προσαρμόσουμε το κοινό που μας ενδιαφέρει και να εξάγουμε κοινωνικά δεδομένα για την επιχείρηση μας. Μεγάλο ρόλο διαδραματίζει το προϊόν/υπηρεσία που θέλουμε να προωθήσουμε, ποια χώρα στοχεύουμε, ποια ηλικιακή ομάδα, ποιο κοινωνικό γκρουπ και άλλα πολλά. Η δημιουργία κοινού αναλυτικότερα έχει πέντε επιλογές όπως φαίνεται στην εικόνα 6.

Εικόνα 6

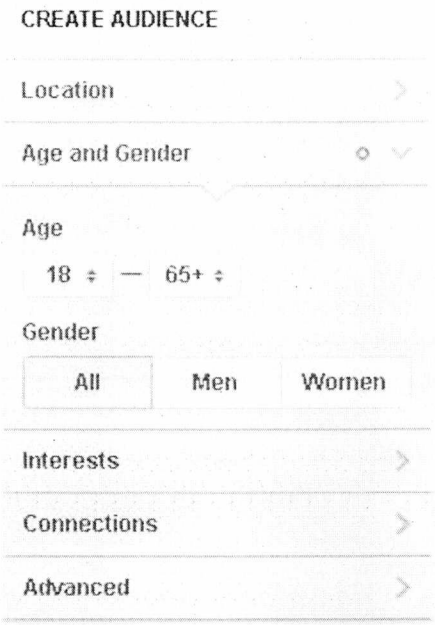
1. Περιοχή (Location): Όπως μαρτυρά και το όνομα της ορίζουμε την περιοχή, γεωγραφικό μέρος που στοχεύουμε πελάτες. Η περιοχή αυτή μπορεί να είναι χώρα(country), νομός(region) ή πόλη(city). Δυστυχώς δεν παρέχει τη δυνατότητα να μπορείς να αναζητήσεις με βάση την ήπειρο π.χ. Ευρώπη, Ασία . (δες εικόνα 7)



Εικόνα 7

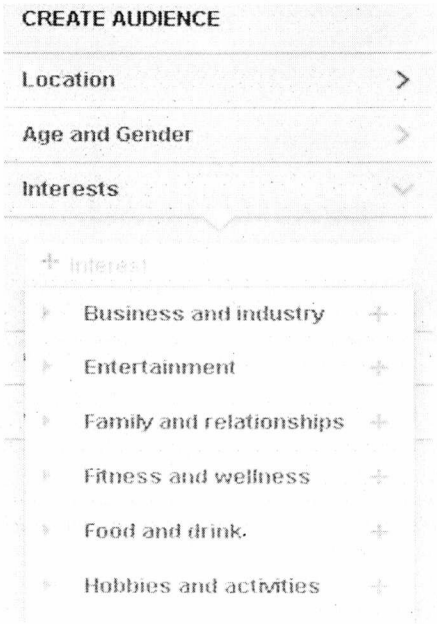
2. Φύλλο και Ηλικία (Age and Gender): Ορίζουμε το φύλλο και την ηλικία του κοινού που θέλουμε να στοχεύσουμε. Για το φύλλο το Facebook μας δίνει δύο επιλογές, γυναίκα ή άντρας. Για την ηλικία μας επιτρέπει να αναζητήσουμε μόνο ενήλικες από 18 μέχρι 65+. Δηλαδή δεν μπορούμε να διαχωρίσουμε

ηλικιακές ομάδες ατόμων άνω των 65 ετών. Από 18 μέχρι 65 πάντως έχουμε μεγάλη ευχέρεια. (εικόνα 8)



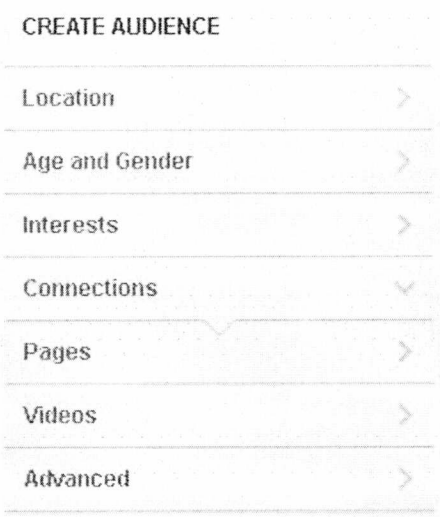
Εικόνα 8

3. Ενδιαφέροντα (Interests): Υπάρχουν τόσες πολλές και διαφορετικές επιλογές που μας προσφέρονται για να προσδιορίσουμε τα ενδιαφέροντα των υποψήφιων πελατών μας. Τρόποι διασκέδασης, οικογένεια και φίλοι, υγεία και ευεξία, τρόποι και συνήθειες διατροφής, δραστηριότητες, ενδιαφέροντα κλπ, είναι όλα εκεί, ότι χρειάζεται μια επιχείρηση eCommerce. (εικόνα 9)



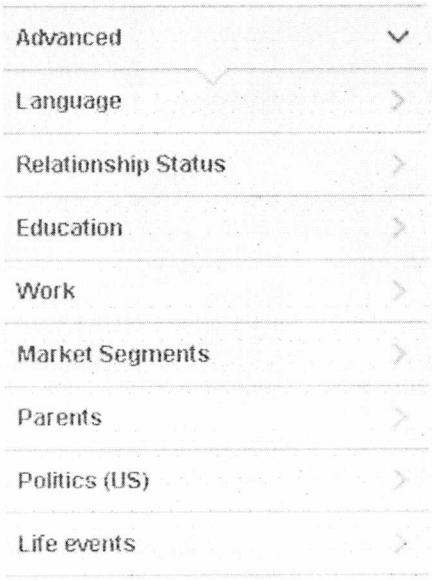
Εικόνα 9

4. Αλληλεπίδραση (Connection): Η επιλογή αυτή προϋποθέτει να κατέχουμε μια σελίδα με την επιχείρηση μας στο κοινωνικό δίκτυο Facebook ή να έχουμε αναρτήσει ένα βίντεο. Μπορούμε να δούμε κάποια χαρακτηριστικά με τα άτομα που αλληλοεπιδρούν στη σελίδα της επιχείρησης μας και να διαπιστώσουμε το προφίλ των πελατών μας. (εικόνα 10)



Εικόνα 10

5. Περισσότερα (Advanced): Τέλος το Audience Insights έχει τοποθετήσει κάποιες ακόμα πληροφορίες σχετικά με τη γλώσσα, την οικογενειακή κατάσταση, τη μόρφωση, το επάγγελμα, πολίτικές πεποιθήσεις και άλλα σε μια κατηγορία ονόματι περισσότερα (advanced). (εικόνα 11)

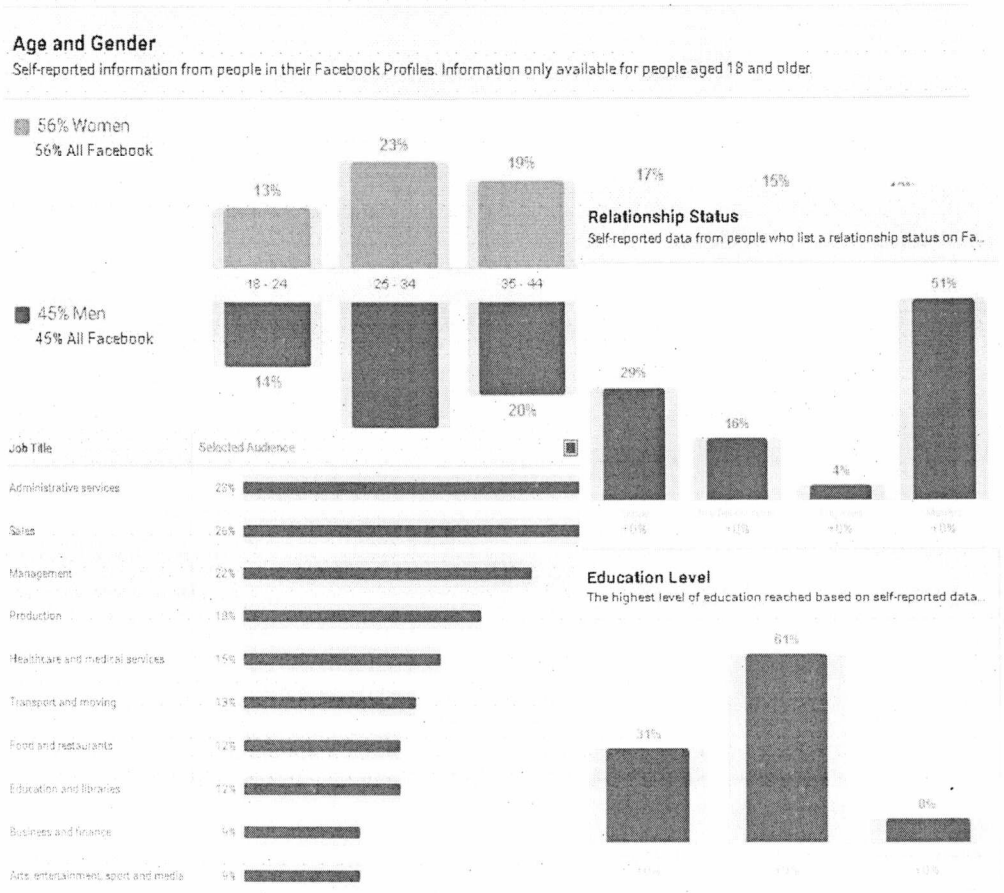


Εικόνα 11

Πληροφορίες για το κοινό που δημιουργήσαμε

Όταν ολοκληρώσουμε τον ορισμό του κοινού το οποίο αναζητούμε απομένει να πάρουμε τις κατάλληλες πληροφορίες ώστε να προωθήσουμε κατάλληλα προϊόντα και υπηρεσίες στο κατάλληλο κοινό. Η περιοχή δεξιά της στήλης δημιουργίας κοινού όπως μπορεί να παρατηρήσει κανείς στην Εικόνα 5 (σελίδα 43) μας παρέχει αυτές τις πληροφορίες. Παρατηρούμε ότι έχει και αυτή τέσσερις διαφορετικές επιλογές: Δημογραφικά Στοιχεία (Demographics), Σελίδες που τους αρέσουν (Page likes), Περιοχή (Location), Δραστηριότητα (Activity).
Πάμε όμως να τα αναλύσουμε με τη σειρά:

1. Δημογραφικά στοιχεία: Παρέχουν πληροφορίες όπως η ηλικία και το φύλλο (Age and Gender) του κοινού μας καθώς και το ποσοστό που ανήκει στην εκάστοτε ηλικιακή ομάδα, για την οικογενειακή κατάσταση (Relationship Status), το επίπεδο μόρφωσης(Education Level), και τον τίτλο εργασίας τους(Job Title). (εικόνα 12)



Εικόνα 12

2. Σελίδες που τους αρέσουν: Οι σελίδες είναι μια πολύ διαδεδομένη λειτουργία στο Facebook που επιτρέπουν στα άτομα να αλληλοεπιδρούν με πράγματα που τους ενδιαφέρουν όπως χόμπι, φαγητό, αθλήματα, προϊόντα κλπ. Αυτή η επιλογή μας παρέχει μια λίστα με τις αυθεντικές σελίδες και όχι απλά το είδος της σελίδας. Είναι μια απ' τις βασικότερες δυνατότητες που παρέχει το Audience Insights καθώς μας δίνεται η δυνατότητα να «κατασκοπεύσουμε» ανταγωνιστές. Βλέποντας τις σελίδες επιτυχημένων ανταγωνιστών μπορούμε να μάθουμε απ' αυτούς και να ανακαλύψουμε τρόπους ώστε να κάνουμε καλύτερη την επιχείρηση μας. (Εικόνα 13)

Top Categories

1	Brewery	Budweiser • Bud L
2	Coffee Shop	Dunkin'
3	Shopping & retail	Groupon • T-Mobi
4	Pizza Place	Pizza Hut
5	Food & drink	Coca-Cola
6	Food and drink company	OREO • Gatorade
7	Computer company	Samsung
8	Furniture	IKEA
9	Politician	Paul Ryan • Mitt F
10	Non-profit organisation	NRA - National Rifl

Page Likes

Facebook Pages that are likely to be relevant to your audience based on Facebook Page likes

Page	Relevance	Audience	Facebook	Affinity
Budweiser	1	3.8m	3.8m	10x
Pepsi	2	4.9m	4.9m	10x
Dunkin'	3	6.3m	6.3m	10x
Groupon	4	6.1m	6.2m	10x
Pizza Hut	5	4.8m	4.8m	10x
Coca-Cola	6	8.9m	8.9m	10x
OREO	7	8.1m	8.1m	10x
Samsung	8	14.6m	14.7m	10x

Εικόνα 13

Πριν πάμε στην επόμενη επιλογή οφείλουμε να τονίσουμε μια πολύ σημαντική παράμετρο που παρατηρούμε στην Εικόνα 13. Λόγος για τη *Σχέση(Affinity)*. Αυτή η παράμετρος μας δείχνει την πιθανότητα του κοινού μας να του αρέσει η αντίστοιχη σελίδα σε σχέση με όλο το κοινό του Facebook. Για παράδειγμα η σελίδα της Pepsi αρέσει στο κοινό μας 10 φορές περισσότερο απ' ότι στο υπόλοιπο σχέδιο. Αυτό πολλές φορές μας βοηθάει να καταλαβαίνουμε αν το κοινό που έχουμε επιλέξει είναι σχετικό με τις σελίδες των ανταγωνιστών ή είναι πολύ ευρύ. Μας ενδιαφέρει ως επιχείρηση eCommerce να μην είναι ευρύ και η παράμετρος *Affinity* να είναι όσο μεγαλύτερη γίνεται. Με αυτό τον τρόπο θα ξέρουμε ότι το κοινό που δημιουργήσαμε είναι το σωστό. Στο παράδειγμα με την Pepsi το 10x *Affinity* υποδηλώνει ότι το κοινό μας είναι πάρα πολύ ευρύ σχετικά με τα προϊόντα αναψυκτικού "cola" . Αν η eCommerce επιχείρηση μας πουλούσε αναψυκτικά αυτής της μορφής τότε θα σήμαινε πως το κοινό που έχουμε δημιουργήσει δεν είναι

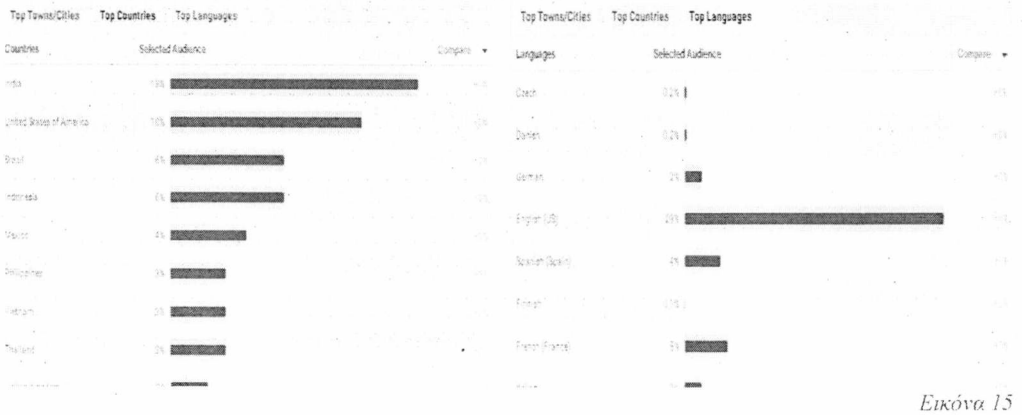
τόσο ακριβές. Στο συγκεκριμένο παράδειγμα δεν έχουμε δημιουργήσει καν κοινό, είναι όλες οι επιλογές προκαθορισμένες γιαυτό συνέβη αυτό. Για τις ανάγκες της πτυχιακής ας επιχειρήσουμε να δημιουργήσουμε ένα κοινό και να ανεβάσουμε αυτήν την παράμετρο της Pepsi. Παρακάτω τα βήματα που εκτελώ:

- Απ' τη στήλη της δημιουργίας κοινού πάμε στα ενδιαφέροντα πληκτρολογούμε Pepsi και έπειτα την επιλέγουμε. Η παράμετρος μας για τη σελίδα Pepsi έχει αυξηθεί στο 150x
- Βλέπουμε τα Demographics και παρατηρούμε αρχικά στην ηλικία και το φύλο ότι η Pepsi είναι πιο διαδεδομένη σε άτομα ηλικίας 25-34 ετών. Το επιλέγουμε και παρατηρούμε ότι το Affinity έχει πάει 160x
- Με παρόμοια λογική θέτουμε την οικογενειακή κατάσταση σε ελεύθερος, και τον τίτλο εργασίας σε εστιατόρια ή μεταφορές. Η παράμετρος μας έχει εκτοξευθεί στο 203x όπου και θα ολοκληρώσουμε το παράδειγμα μας. (εικόνα 14).

Page	Relevance	Audience	Facebook	Affinity
Pepsi	1	13.6K	4.8m	203x
All That	2	2.7K	1m	188x
NickRewind	4	3.2K	1.7m	133x

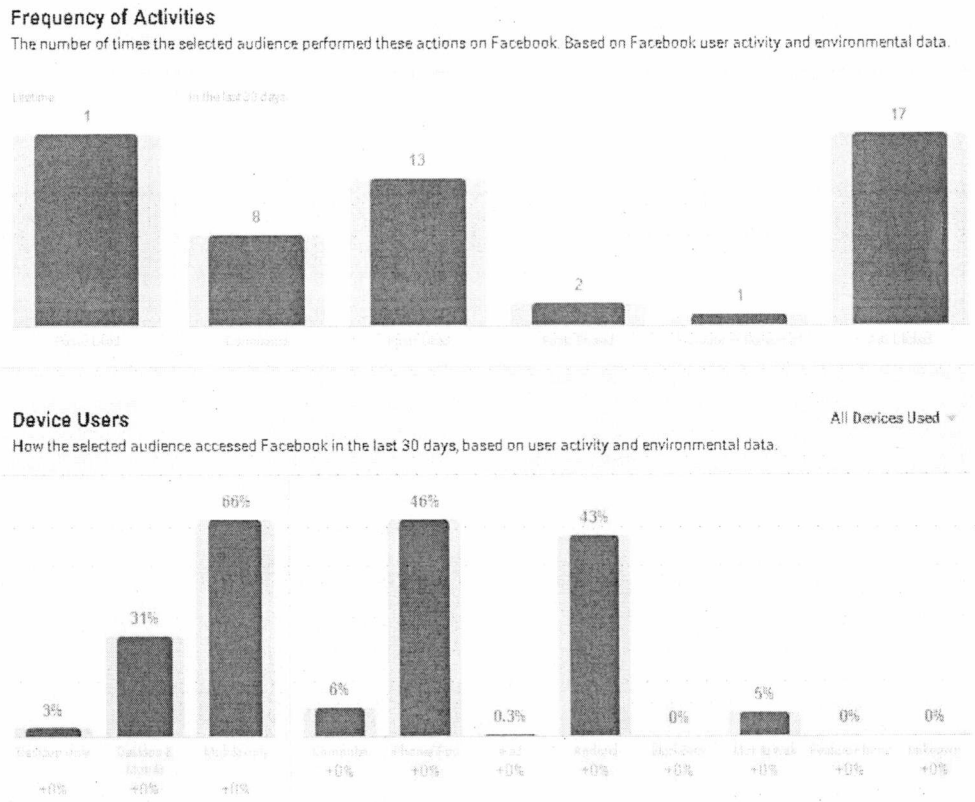
Εικόνα 14

3. Περιοχή: Η περιοχή όπως υποδηλώνει και το όνομα της μας υποδεικνύει τις περιοχές στις οποίες μένει το κοινό που δημιουργήσαμε. Έχει τρεις επιπλέον επιλογές: Οι κορυφαίες πόλεις, πολιτείες(ΗΠΑ) στις οποίες ανήκει το κοινό μας, τις κορυφαίες χώρες, και τις κορυφαίες γλώσσες τις οποίες μιλάνε.(εικόνα 15)



Εικόνα 15

4. Δραστηριότητα: Η δραστηριότητα μας δίνει πληροφορίες όπως ο τρόπος αλληλεπίδρασης του κοινού σε σελίδες (πχ συνηθίζει να πατάει μου αρέσει, να γράφει σχόλιο, να κάνει κοινοποίηση κλπ) και οι συσκευές που χρησιμοποιεί για να συνδέεται με το Facebook. Αν είμαστε μια επιχείρηση eCommerce που αναπτύσσει εφαρμογές desktop δε θα ήταν έξυπνο να προωθήσουμε το προϊόν μας σε άτομα που χρησιμοποιούν κατά κόρον κινητά για την πλοήγηση τους. (Εικόνα 16)

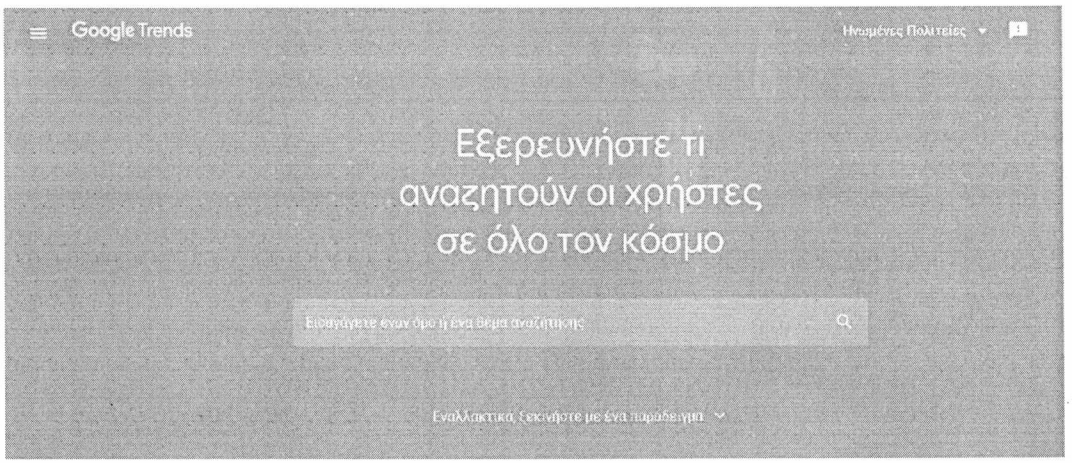


Εικόνα 16

Google Trends

Το δεύτερο εργαλείο εξόρυξης που θα μελετήσουμε είναι το Google Trends. Θα μπορούσε κάποιος να πει ότι είναι ένα πιο εύκολο Audience Insights καθώς δεν χρειάζεται να δημιουργήσεις ένα κοινό. Το μόνο που χρειάζεται να κάνεις είναι να πληκτρολογήσεις το αντικείμενο που σε ενδιαφέρει και θα σου παρέχει πληροφορίες σχετικά με τους χρήστες της Google που αναζήτησαν ή αλληλοεπίδρασαν με το συγκεκριμένο ή κάποιο άλλο παρόμοιο αντικείμενο. Επιπλέον δεν απαιτεί σύνδεση ή δημιουργία κάποιου λογαριασμού όπως το Audience Insights. Αντιθέτως είναι ελεύθερο για όλους.

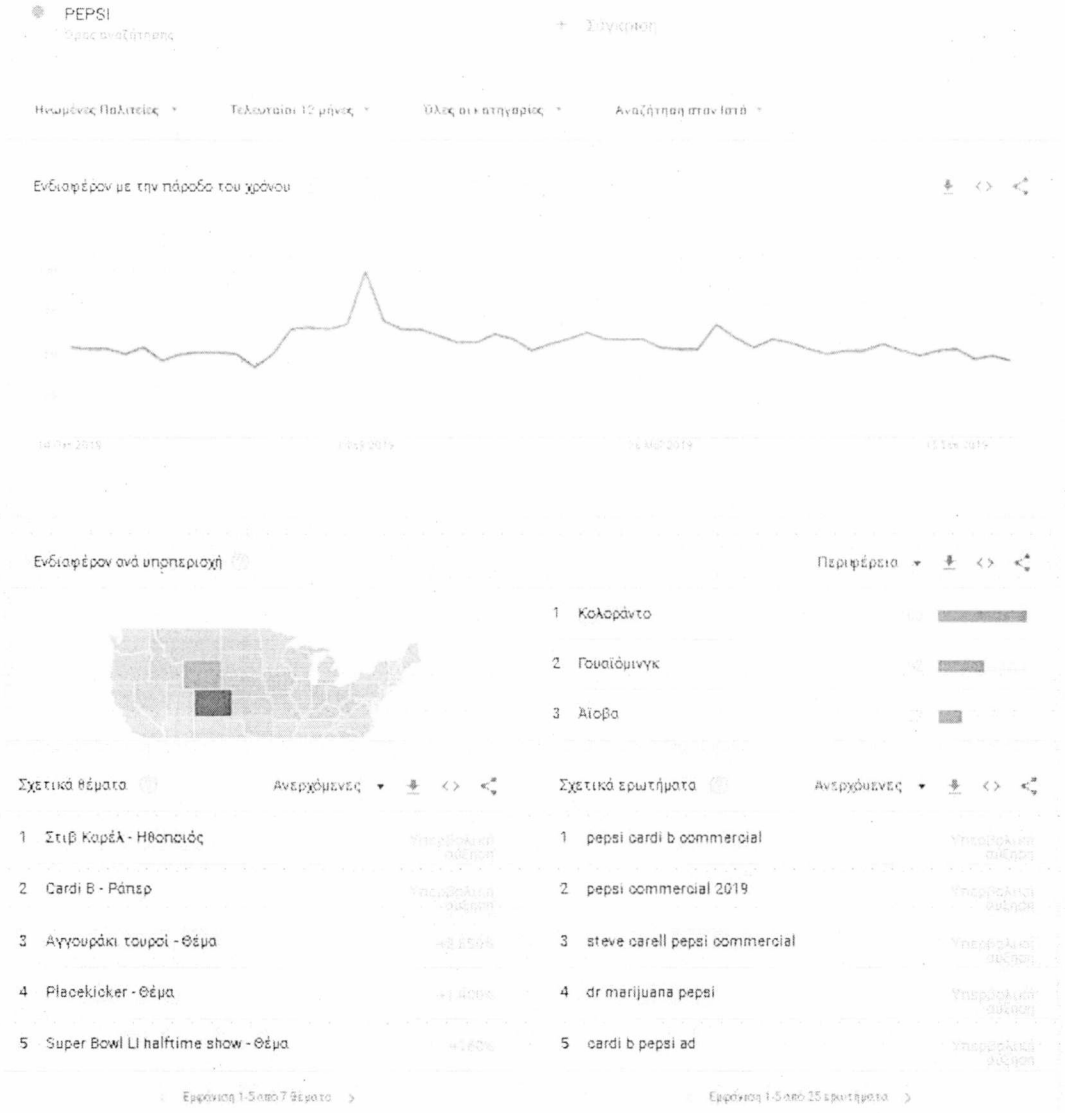
Η απλοϊκότητα που αναφέραμε πριν γίνεται αντιληπτή απ' την αρχική του κιόλας σελίδα.



Εικόνα 17

Ας επιχειρήσουμε σε αυτό το σημείο να εισάγουμε έναν όρο για να συνεχίσουμε την ανάλυση του συγκεκριμένου εργαλείου. Θα χρησιμοποιήσω και εδώ τον όρο «PEPSI» ως παράδειγμα όπως είχαμε κάνει και με το Audience Insightsς. (στη σελίδα 49)

Μετά την εισαγωγή ενός όρου το Google Trends μας μεταφέρει σε μια σελίδα (Εικόνα 18) όπου μπορούμε να διακρίνουμε χρήσιμες πληροφορίες όπως το ενδιαφέρον των χρηστών για τον συγκεκριμένο όρο με την πάροδο του χρόνου, το ενδιαφέρον ανά γεωγραφική περιοχή και σχετικά θέματα και ερωτήματα που επίσης αναζήτησαν οι χρήστες και έχουν άμεση σχέση με τον συγκεκριμένο όρο, στην περίπτωση μας PEPSI.

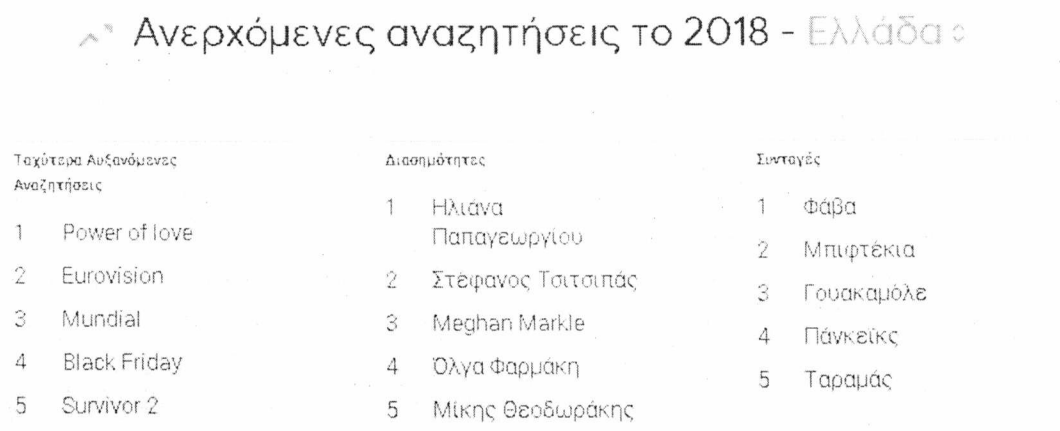


Εικόνα 18

Όπως φαίνεται στο πάνω μέρος της εικόνας πρέπει εμείς να ορίσουμε τη χώρα ή την περιοχή για την οποία θα δοθούν πληροφορίες (μας παρέχεται και η επιλογή Παγκοσμίως) καθώς και το χρονικό διάστημα την κατηγορία και το είδος της αναζήτησης.

Τέλος και ανεξάρτητα απ' τον όρο που έχουμε πληκτρολογήσει το Google Trends μας παρέχει και άλλες δυνατότητες.

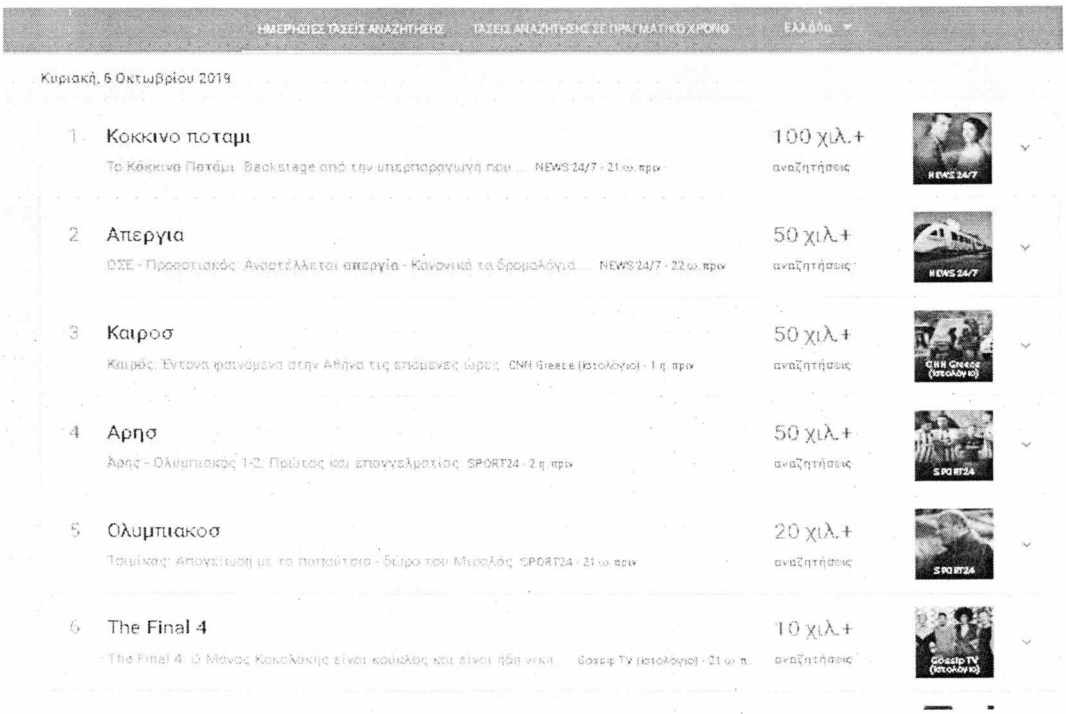
Η μία είναι οι ανερχόμενες αναζητήσεις για το 2018 (κάθε φορά παρουσιάζει το προηγούμενο έτος, όταν μπούμε στο 2020 τότε θα δείξει του 2019) για κάποια συγκεκριμένη χώρα το οποίο φαίνεται στην εικόνα 19.



Εικόνα 19

(επέλεξα Ελλάδα για να τιμήσω τη χώρα μου και μετά την εμφάνιση των αποτελεσμάτων το μετάνιωσα.)

Η δεύτερη είναι οι δημοφιλείς αναζητήσεις για κάθε μέρα ξεχωριστά καθώς και σε πραγματικό χρόνο. Αυτό είναι ιδιαίτερα χρήσιμο για μια eCommerce επιχείρηση καθώς μπορεί να ενημερώνεται και να συμβαδίζει με τις εξελίξεις και τις απαιτήσεις κάθε περιόδου και γιατί όχι, κάθε μέρας ξεχωριστά. (Εικόνα 20)



Εικόνα 20

Κοινωνικά δεδομένα και eCommerce

Αρχικά ορίσαμε τι σημαίνει eCommerce και είδαμε όλες τις έννοιες που εμπλέκονται με αυτόν τον όρο όπως τα είδη του, τα πλεονεκτήματα του, στρατηγικό σχεδιασμό των επιχειρήσεων όπου και θέσαμε ένα σημαντικό ερώτημα. Πως ξέρουν οι επιχειρήσεις ηλεκτρονικού εμπορίου σε τι κοινό θα απευθυνθούν για να προωθήσουν τα προϊόντα τους; Και φυσικά απαντήσαμε με την εξόρυξη κοινωνικών δεδομένων. Οπότε η επόμενη ενότητα αφορούσε ακριβώς αυτό: τρόπους με τους οποίους μια επιχείρηση ηλεκτρονικού εμπορίου μπορεί να εξορύξει κοινωνικά δεδομένα. Η τελευταία ενότητα της πτυχιακής εργασίας ξαναορίζει τη σχέση αυτών των δύο εννοιών και επισημαίνει τη σημασία των κοινωνικών δεδομένων για τις eCommerce επιχειρήσεις.

Τα κοινωνικά δεδομένα που αποκτούμε μέσω της εξόρυξης δεν χρησιμεύουν μόνο στη διαφήμιση και στην αύξηση των πωλήσεων. Βέβαια αυτός είναι ο απώτερος σκοπός αλλά για να επιτευχθεί πρέπει να προσέξουμε και άλλες πτυχές.

Μία απ' αυτές είναι η σχεδίαση της ιστοσελίδας μας με βάση αυτά τα δεδομένα. Η ιστοσελίδα είναι το α και το ω μιας eCommerce επιχείρησης. Είναι η ψηφιακή βιτρίνα της στην οποία παρουσιάζει τα προϊόντα και τις υπηρεσίες της. Επιβάλλεται να είναι σωστά σχεδιασμένη αν θέλουμε να κάνουμε εντύπωση στους επισκέπτες μας, Πολλοί επιχειρηματίες υποστηρίζουν ότι ο σωστός σχεδιασμός είναι μια αφηρημένη έννοια και αν μια ιστοσελίδα είναι σωστή ή λάθος είναι κάτι υποκειμενικό. Οι ίδιοι που το υποστηρίζουν αυτό για την κατασκευή της ιστοσελίδας τους δρουν με την κοινή λογική χωρίς να διαθέτουν κάποιο στοιχείο-απόδειξη ότι αυτό που παρουσιάζουν όντως θα λειτουργήσει. Θα αποδείξω γιατί δεν είναι μια αφηρημένη έννοια με ένα παράδειγμα.

- Στο παράδειγμα έστω ότι η eCommerce επιχείρηση μας προωθεί ιατρικό εξοπλισμό σε ιδιώτες ιατρούς στην Ελλάδα. Πρέπει να χρησιμοποιήσουμε κάποια μέθοδο εξόρυξης για να διαπιστώσουμε αν οι γιατροί στην Ελλάδα ενδιαφέρονται για το εντυπωσιακό στο μάτι προϊόν ή γιαυτό που θα κάνει πιο εύκολη την εξέταση του ασθενούς. Έστω ότι με τα κοινωνικά δεδομένα που συλλέγουμε συμπεραίνουμε ότι ανήκουν στη δεύτερη κατηγορία Αυτό που θα ήθελε συνεπώς, να αντικρίσει ένας επίδοξος πελάτης ιατρός δεν είναι τόσο τα εφέ, τα χρώματα και οι εντυπωσιακές εικόνες της ιστοσελίδας μας αλλά η

μεγάλη λεπτομέρεια και πληροφορία που παρέχεται σχετικά με τον εξοπλισμό όπως τα χαρακτηριστικά του και ρεαλιστικές διακριτικές εικόνες.

Αυτό που γίνεται αντιληπτό με το παραπάνω παράδειγμα είναι ότι η σωστή σχεδίαση μιας ιστοσελίδας δεν είναι αφηρημένη έννοια και έχει άμεση σχέση με τη συμπεριφορά των ατόμων στα οποία απευθυνόμαστε. Δραματικό ρόλο διαδραματίζει λοιπόν η εξόρυξη κοινωνικών δεδομένων στην σχεδίαση της ιστοσελίδας μιας επιχείρησης που ασχολείται με το ηλεκτρονικό εμπόριο.

Μια δεύτερη πτυχή διαφαίνεται στην ανάπτυξη και παρουσίαση τον προϊόντων ή υπηρεσιών που προωθούμε. Για παράδειγμα κάποια άτομα θέλουν να τους παρουσιαστεί το προϊόν τους με ένα στιλάτο δέμα και άλλοι δεν τους νοιάζει το δέμα και θέλουν απλά το προϊόν τους. Με τη πλειοψηφία πελατών έχουμε να κάνουμε πάλι θα φανεί μέσω των κοινωνικών δεδομένων που θα εξορύξουμε.

Η τελευταία πτυχή αναφέρεται στο πως αντιμετωπίζουμε τους πελάτες στο επικοινωνιακό κομμάτι. Οι πελάτες μας στη συγκεκριμένη περιοχή μιας συγκεκριμένης χώρας που στοχεύουμε είναι μορφωμένοι ή όχι; Αν ναι πρέπει να τους προσεγγίσουμε με πιο έξυπνο τρόπο. Είναι υπομονετικοί ή όχι; Αν ναι μπορούμε να τους «ζαλίζουμε» με την αποστολή email χωρίς ιδιαίτερες επιπτώσεις. Πολλά τέτοια παραδείγματα που πρέπει μια επιχείρηση να απαντήσει μέσα από την εξόρυξη κοινωνικών δεδομένων.

Συμπερασματικά σε αυτό το κεφάλαιο αναφέραμε τη σχέση μεταξύ eCommerce επιχείρησης και της εξόρυξης κοινωνικών δεδομένων η οποία είναι μεγάλη και ο σωστός συνδυασμός και οι σωστές αποφάσεις που παίρνουμε για αυτά τα δύο είναι καθοριστικές για την επιτυχημένη πορεία μας στον κόσμο του διαδικτύου.

Βιβλιογραφία

1. Electronic Commerce. A manager's Guide, Kalakota R. and Whinston A. 1997

2. Εισαγωγή στο ηλεκτρονικό εμπόριο, Πομπόρτσης Ανδρέας Σ, Τσουλφάς Ανέστης Γ, Εκδόσεις Τζιόλα, 2002

3. Managing in the Next Society. New York: Truman Talley Books, Drucker P, 2002

4. Εισαγωγή στην εξόρυξη δεδομένων, Pang-Ning Tan, Michael Steinback, Vipin Kumar, Εκδόσεις Τζιόλα, 2010

5. History of ecommerce, Διαδικτυακός τόπος https://www.ecommerce-land.com/history_ecommerce.html

6. Quarterly retail e-commerce sales, US Department of Commerce. Διαδικτυακός τόπος: https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf

7. Omnichannel Retail Commerce Platform Market 2019 Size, Share, Industry Segmentation, Key Strategies, Competitive Landscape, With Regional Forecast To 2023, Reuters, Διαδικτυακός τόπος: https://www.reuters.com/brandfeatures/venture-capital/article?id=84658

8. Ecommerce 101 + The History of Online Shopping: What The Past Says About Tomorrow's Retail Challenges Διαδικτυακός τόπος: https://www.bigcommerce.com/blog/ecommerce/#the-future-of-ecommerce

9. Types of E-Commerce Models, Eyerys Διαδικτυακός τόπος: https://www.eyerys.com/articles/types-e-commerce-models

10. Top 5 companies ranked by us net digital ad revenue share 2018-2019 of total digital ad spending, Emarketer, Διαδικτυακός Τόπος: https://www.emarketer.com/chart/226372/top-5-companies-ranked-by-us-net-digital-ad-revenue-share-2018-2019-of-total-digital-ad-spending

11. Introduction to Data Mining, thecrazyprogrammer, Διαδικτυακός Τόπος: https://www.thecrazyprogrammer.com/2018/02/introduction-to-data-mining.html

12. Joint probability, conditional probability and Bayes' theorem Διαδικτυακός Τόπος: https://www.ling.upenn.edu/courses/ling052/Bayes1.html

13. Facebook now has 2 Billion Monthly users .. and responsibility, techbranch, Διαδικτυακός Τόπος: https://techcrunch.com/2017/06/27/facebook-2-billion-users/