



Πολυτεχνείο Κρήτης
Σχολή Μηχανικών Παραγωγής και Διοίκησης

Τμηματοποίηση Αγοράς με χρήση μεθόδων
Εξόρυξης Δεδομένων και Πολυκριτήριας
Ανάλυσης

Διπλωματική Εργασία
Παρούσης Γιώργος

Επιβλέπων
Ματσατσίνης Νικόλαος, Καθηγητής

Χανιά, Δεκέμβριος 2023

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω θερμά τον καθηγητή μου κ. Νικόλαο Ματσατσίνη για την ανάθεση της παρούσας διπλωματικής εργασίας καθώς και για την συνεχή καθοδήγηση που μου παρείχε κατά την διάρκεια εκπόνησης της.

Επίσης θέλω να ευχαριστήσω την υποψήφια διδάκτωρ κ. Φωτεινή Καλαφάτη για τις συμβουλές και την καθοδήγηση της κατά τη συγγραφή του κειμένου της διπλωματικής εργασίας.

Τέλος ευχαριστώ πολύ την οικογένεια μου για την συνεχή στήριξη της καθ' όλη την διάρκεια των σπουδών μου.

Περίληψη

Στις επιχειρήσεις δίνεται ιδιαίτερη σημασία στην δύναμη των πληροφοριών. Το τμήμα που είναι υπεύθυνο με την ανάλυση δεδομένων πρέπει να εκμεταλλεύεται κάθε πληροφορία που μπορεί να προκύψει από τα δεδομένα που συλλέγονται. Στόχος είναι η ανακάλυψη νέων ευκαιριών, η αντιμετώπιση προβλημάτων και ο εντοπισμός των περιοχών που χρήζουν βελτίωση. Αντίστοιχα το τμήμα του μάρκετινγκ είναι υπεύθυνο για την επιλογή της καλύτερης δυνατής στρατηγικής για την εισαγωγή νέων προϊόντων και υπηρεσιών και την τροποποίηση ή την απόσυρση υπαρχόντων προϊόντων και υπηρεσιών.

Η τμηματοποίηση της αγοράς αποτελεί πολύ σημαντική πρακτική για τον στρατηγικό σχεδιασμό μιας ομάδας μάρκετινγκ, με σκοπό τη βελτιστοποίηση μιας στοχευμένης διαφημιστικής καμπάνιας σε ένα τμήμα του καταναλωτικού κοινού με παρόμοια χαρακτηριστικά συμπεριφοράς. Η επιτυχία μιας επιχείρησης έρχεται και από την ικανότητα της να προσαρμόζει τα προϊόντα της σε αυτά τα ιδιαίτερα χαρακτηριστικά τους ώστε να μπορεί να ικανοποιήσει τις αντίστοιχες ομάδες των πελατών που απευθύνεται.

Στην παρούσα εργασία παρουσιάζεται μια μεθοδολογία για την ανάλυση και τμηματοποίηση της αγοράς. Η μεθοδολογία συνδυάζει μια πολυκριτήρια μέθοδο και έναν αλγόριθμο Νευρωνικών Δικτύων. Πιο συγκεκριμένα χρησιμοποιείται η μεθοδολογία UTASTAR σε συνδυασμό με τα Self Organizing Maps.

Πίνακας Περιεχομένων

Κεφάλαιο 1ο: Εισαγωγή	1
1.1 Σκοπός της Εργασίας.....	1
1.2 Καθορισμός του Προβλήματος.....	1
Κεφάλαιο 2ο: Υφιστάμενη Κατάσταση	2
2.1 Εισαγωγή	2
2.2 Συνδυασμός Πολυκριτήριας Ανάλυσης Και Νευρωνικών Δικτύων- Περιπτώσεις Εφαρμογής	2
Κεφάλαιο 3ο: Θεωρητικό Υπόβαθρο.....	7
3.1 Συστήματα Υποστήριξης Αποφάσεων.....	7
3.2 Πολυκριτήρια Ανάλυση.....	8
3.2.1 Μεθοδολογία Μοντελοποίησης Ενός Πολυκριτήριου Προβλήματος	9
3.3 Μάρκετινγκ.....	11
3.3.1 Εισαγωγή	11
3.3.2 Μίγμα Μάρκετινγκ.....	11
3.3.3 Ψηφιακό Μάρκετινγκ.....	14
3.3.4 Στρατηγική Ψηφιακού Μάρκετινγκ.....	16
3.4 Τμηματοποίηση της Αγοράς.....	16
3.5 Τεχνητή Νοημοσύνη(Artificial Intelligence)	18
3.6 Μηχανική Μάθηση	18
3.6.1 Εισαγωγή	18
3.6.2 Είδη Μηχανικής Μάθησης	18
3.7 Συσταδοποίηση(Clustering)	19
3.8 Προεπεξεργασία Δεδομένων	20
3.8.1 Ελλιπείς Τιμές.....	20
3.8.2 Θόρυβος στα Δεδομένα	20
3.8.3 Ασυμπεπλή Δεδομένα	21
3.8.4 Ενοποίηση Των Δεδομένων.....	21
3.8.5 Μετασχηματισμός Δεδομένων	21
3.9 Νευρωνικά Δίκτυα.....	23
3.9.1 Δομή Τεχνητού Νευρωνικού Δικτύου	23
3.9.2 Συνάρτηση Ενεργοποίησης	24
3.9.3 Αισθητήρας (Perceptron)	29
3.9.4 Πολύ-Επίπεδοι Αισθητήρες	30
3.9.5 Αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων	31
3.9.6 Δίκτυα Πρόσθιας Τροφοδότησης Ενός Επιπέδου.....	32
3.9.7 Πολυεπίπεδα Δίκτυα Πρόσθιας Τροφοδότησης (Multilayer Feed-Forward Neural Networks)	32
3.9.8 Αναδρομικά Νευρωνικά Δίκτυα.....	33
Κεφάλαιο 4ο: Η Πολυκριτήρια Μέθοδος UTASTAR και το Νευρωνικό Δίκτυο SOM	35
4.1 Εισαγωγή	35
4.2 Πολυκριτήρια Μέθοδος - UTASTAR	35

4.2.1	Αλγόριθμος Utastar	36
4.3	Νευρωνικά Δίκτυα: Self-Organizing Maps.....	38
4.3.1	Δομή Και Λειτουργία.....	39
4.3.2	Κώδικας Som	40
4.3.3	U-Matrix.....	42
4.3.4	Αλγόριθμος K-Means	43
4.3.5	Συντελεστής Silhouette	44
4.3.6	Κανόνας του αγκώνα	45
Κεφάλαιο 5ο: Παρουσίαση και Εφαρμογή Μεθοδολογίας: Τμηματοποίηση και Ανάλυση Αγοράς με SOM και K-Means		46
5.1	Εισαγωγή	46
5.2	Δεδομένα.....	46
5.3	Μεθοδολογία	47
5.4	Εφαρμογή UTASTAR.....	51
5.5	Καθαρισμός Δεδομένων – Αφαίρεση ακραίων τιμών	51
5.6	Τμηματοποίηση των Καταναλωτών Βάσει των Ολικών Χρησιμοτήτων	52
5.7	Περιπτώσεις Ακραίων Τιμών	57
5.8	Τμηματοποίηση Βάσει των Βαρών των Χαρακτηριστικών	57
5.9	Συνδυασμός Συμπερασμάτων με τα Αποτελέσματα από το MARKEX.....	64
5.9.1	Βάρη Των Κριτηρίων.....	64
5.9.2	Ολικές Χρησιμότητες Καταναλωτών	65
Κεφάλαιο 6ο: Συμπεράσματα		68
Βιβλιογραφία		69

Κεφάλαιο 1^ο: Εισαγωγή

1.1 Σκοπός της Εργασίας

Στο πλαίσιο των επιχειρήσεων, το τμήμα που ασχολείται με την ανάλυση δεδομένων είναι υπεύθυνο για τη χρήση κάθε πληροφορίας που προκύπτει από τα δεδομένα, με σκοπό την ανακάλυψη νέων ευκαιριών, την αντιμετώπιση προκλήσεων και την εντοπισμό περιοχών για βελτίωση. Συγκεκριμένα, στον τομέα του μάρκετινγκ, οι υπεύθυνοι πρέπει να επιλέγουν την καλύτερη στρατηγική για τα προβλήματα που σχετίζονται με την εισαγωγή νέων προϊόντων, την αλλαγή ή απόσυρση υπαρχόντων προϊόντων.

Η διαδικασία λήψης αποφάσεων σε αυτό το πεδίο είναι πολύπλοκη, λόγω των πολλαπλών κριτηρίων και των διαφορετικών απόψεων των ανθρώπων που συμμετέχουν. Ο ανταγωνισμός στην αγορά απαιτεί τη λήψη αποφάσεων με όσο το δυνατόν μεγαλύτερη ακρίβεια. Επομένως, είναι αναγκαίο να ελαχιστοποιηθούν τα σφάλματα προκειμένου να ληφθεί η βέλτιστη απόφαση.

Στην παρούσα εργασία παρουσιάζεται η ανάλυση της Γαλλικής αγοράς λαδιού από την έρευνα που δημοσιεύτηκε από τους «Υ. Siskos, Ν.Ε. Matsatsinis και Γ. Baourakis» το 1998 με τίτλο «*Multicriteria analysis in agricultural marketing: The case of French olive oil market*». Η ανάλυση πραγματοποιείται με συνδυασμό μιας πολυκριτήριας μεθοδολογίας και ενός αλγορίθμου Νευρωνικών Δικτύων. Πιο συγκεκριμένα χρησιμοποιείται η μεθοδολογία UTA STAR σε συνδυασμό με τα Self Organizing Maps.

1.2 Καθορισμός του Προβλήματος

Το πρόβλημα που επιλύεται είναι η τμηματοποίηση και η ανάλυση της αγοράς και πραγματοποιείται με συνδυασμό μιας πολυκριτήριας μεθόδου και τα νευρωνικά δίκτυα. Το νευρωνικό δίκτυο, Self Organizing Maps, που απαιτεί εκπαίδευση και κατάλληλη επιλογή δεδομένων για την ορθότερη λειτουργία του, συνδυάζεται με την πολυκριτήρια μέθοδο ανάλυσης αποφάσεων UTA STAR. Με την εφαρμογή του παραπάνω συνδυασμού μεθόδων ανάλυσης, πραγματοποιείται η τμηματοποίηση, η ανάλυση της αγοράς και η περιγραφή των ομάδων που προκύπτουν.

Κεφάλαιο 2ο: Υφιστάμενη Κατάσταση

2.1 Εισαγωγή

Η χρήση των νευρωνικών δικτύων σε συνδυασμό με τις μεθόδους πολυκριτήριας ανάλυσης για την επίλυση προβλημάτων είναι ένας αναπτυσσόμενος κλάδος. Τα τελευταία χρόνια ερευνάται αν ο συνδυασμός των μεθόδων βελτιώνει την απόδοση των αποτελεσμάτων στα διάφορα προβλήματα πολυκριτήριας ανάλυσης. Η προσθήκη των νευρωνικών δικτύων χρησιμοποιείται κυρίως για να βοηθήσει στα προβλήματα που προκύπτουν λόγω της τυχαιότητας στην διαδικασία της λήψης μιας απόφασης.

Στη συνέχεια παρουσιάζονται έρευνες που παρουσιάζουν την υφιστάμενη κατάσταση στην χρήση των συνδυασμών των νευρωνικών δικτύων και των πολυκριτήριων μεθόδων. Πιο συγκεκριμένα χρησιμοποιούν Analytic Hierarchy Process-AHP και TOPSIS, Technique for Order Preference by Similarity to the Ideal Solution ως μεθόδους πολυκριτήριας ανάλυσης σε συνδυασμό με νευρωνικά δίκτυα όπως Back Propagation Neural Network-(BPNN), Self Organizing Maps.

2.2 Συνδυασμός Πολυκριτήριας Ανάλυσης Και Νευρωνικών Δικτύων-Περιπτώσεις Εφαρμογής

Η λύση πολυκριτήριων προβλημάτων γίνεται με μεθόδους πολυκριτήριας ανάλυσης. Τα τελευταία χρόνια όμως γίνεται συνδυασμός των μεθόδων Πολυκριτήριας ανάλυσης με νευρωνικά δίκτυα.

Στην ερευνά από τον Tueeno Kohonen παρουσιάζει μια θεωρητική μελέτη και υπολογιστικές προσομοιώσεις μιας νέας διαδικασίας αυτοοργάνωσης (Self Organizing Process). Ανακαλύπτει μια διαδικασία που δημιουργεί ένα απλό δίκτυο φυσικών στοιχείων που λαμβάνει σήματα από ένα πρωτογενή χώρο συμβάντων και οι αναπαραστάσεις των σημάτων χαρτογραφούνται αυτόματα σε ένα σύνολο εξόδων. Το βασικό αυτοοργανωμένο σύστημα είναι ένα μονοδιάστατος ή δυσδιάστατος πίνακας μονάδων επεξεργασίας που μοιάζει με ένα δίκτυο μονάδων λογικής κατωφλιού και χαρακτηρίζεται από μικρής εμβέλειας ανατροφοδότησης μεταξύ γειτονικών μονάδων. Χρησιμοποιούνται διάφοροι τύποι υπολογιστικών προσημειώσεων για να επιδείξουν τη διαδικασία ταξινόμησης καθώς και τις συνθήκες υπό τις οποίες αποτυγχάνει.

Στην έρευνα των Hongjun Yang, Juanjuan Xue (2016) γίνεται αξιολόγηση της διδασκαλίας των καθηγητών. Συνδυάζονται η πολυκριτήρια μέθοδος Analytic Hierarchy Process-AHP και Back Propagation Neural Network- (BPNN) δημιουργώντας μια καινούργια μέθοδο αξιολόγησης. Από την χρήση του συνδυασμού των δύο μεθόδων υπολογίζονται οι συντελεστές αξιολόγησης της

απόδοσης, οι αντικειμενικοί και υποκειμενικοί συντελεστές προτίμησης και από τον συνδυασμό προκύπτουν τα συνολικά βάρη των κριτηρίων. Τα αποτελέσματα της έρευνας έδειξαν ότι η αξιολόγηση με τον συνδυασμό των δύο μεθόδων είναι πιο ακριβής από τον ήδη υπάρχον τρόπο.

Ο συνδυασμός των μεθόδων Analytic Hierarchy Process AHP και fuzzy νευρωνικού δικτύου χρησιμοποιήθηκε στην έρευνα: του Arpan Kumar Kar (2009) για την λύση του προβλήματος πολυκριτήριας ανάλυσης για επιλογή προμηθευτών. Σκοπό είχε την κατηγοριοποίηση των προμηθευτών σε κατάλληλους και ακατάλληλους. Τα βάρη για το νευρωνικό προκύπτουν από την μέθοδο Analytic Hierarchy Process AHP και η κατηγοριοποίηση των προμηθευτών από το fuzzy νευρωνικό δίκτυο.

Στην έρευνα των Su-Li Yan, Ying Wang, Ji-Cheng Liu (2012) γίνεται ανάλυση της αξιολόγησης του συστήματος Business Intelligence BIS. Για την επίτευξη του στόχου χρησιμοποιείται η μέθοδος Analytic Hierarchy Process AHP για τον καθορισμό των συντελεστών των βαρών, στη συνέχεια η fuzzy TOPSIS καθορίζει την βαθμολογία του συστήματος της κάθε εταιρίας και τέλος χρησιμοποιείται το νευρωνικό δίκτυο BP για την αξιολόγηση του συστήματος

Σκοπός της έρευνας των Carlos Pestana Barros και Peter Wanke (2015) είναι η αξιολόγηση της αποδοτικότητας των αφρικανικών αεροπορικών εταιριών μέσω του συνδυασμού της μεθόδου TOPSIS, Technique for Order Preference by Similarity to the Ideal Solution, και νευρωνικού δικτύου. Αρχικά στο πρώτο στάδιο με την μέθοδο TOPSIS γίνεται καθορισμός της σχετικής αποδοτικότητας των αεροπορικών εταιριών με τους πιο διαδεδομένους δείκτες που χρησιμοποιούνται. Στο δεύτερο στάδιο τα αποτελέσματα από την TOPSIS δηλαδή η κατάταξη των εναλλακτικών, συνδυάζονται με το νευρωνικό δίκτυο με σκοπό την δημιουργία μοντέλου πρόβλεψης της αποδοτικότητας των αεροπορικών εταιριών.

Στην έρευνά των Nilashi, M., Mardani, A., Liao, H., Ahmadi, H., Abdul Manaf, A., & Almukadi, W. (2019) εξετάζεται η επίδραση των πράσινων ξενοδοχείων με υπηρεσίες σπά στην ικανοποίηση των ταξιδιωτών. Προτείνεται μια μέθοδος για ανάλυση διαδικτυακών σχολίων με μεθόδους πολυκριτήριας λήψης αποφάσεων, εξόρυξης κειμένου και τεχνικών προβλεπτικής μάθησης για τον εντοπισμό των παραγόντων που επηρεάζουν τις αποφάσεις των ταξιδιωτών. Γίνεται χρήση των Self Organizing Maps για την κατηγοριοποίηση των πελατών βάση των προτιμήσεων τους και της συμπεριφοράς τους. Εφαρμόζει το LDA (Blei et al. 2003; Liu, 2015; Ματσατσίνης, 2021) για την ανάλυση των κριτηρίων κειμένου, παρέχοντας πληροφορίες για τα θέματα και τα συναισθήματα που εκφράζουν οι πελάτες. Χρησιμοποιεί το TOPSIS για την κατάταξη των χαρακτηριστικών του ξενοδοχείου βάσει πολλαπλών κριτηρίων, εντοπίζοντας τις

πιο ευνοϊκές πτυχές. Χρησιμοποιεί την προβλεπτική μάθηση με μια νευρο-ασαφή (fuzzy) τεχνική για την αποκάλυψη επιπέδων ικανοποίησης του πελάτη, προσφέροντας μια λεπτομερή κατανόηση των παραγόντων ικανοποίησης. Οι ομάδες που εξετάζει είναι: οι μόνοι ταξιδιώτες, οι οικογένειες, ζευγάρια και οι παρέες φίλων. Χρησιμοποιεί τις κριτικές από 152 ξενοδοχεία στην Μαλαισία. Τα ευρήματα της έρευνας παρέχουν μια προσέγγιση για τις αποφάσεις των ταξιδιωτών στην επιλογή του ξενοδοχείου και παρέχουν επίσης πληροφορίες στους διαχειριστές των ξενοδοχείων για να τους βοηθήσουν να βελτιώσουν την ποιότητα των υπηρεσιών και να αναπτύξουν στρατηγικές Μάρκετινγκ.

Στη έρευνα των *Melody Y. Kiang, Michael Y. Hu και Dorothy M. Fisher (2004)* εφαρμόζονται τα νευρωνικά δίκτυα Self Organizing Maps σε δεδομένα πελατών από την AT&T. Συγκρίνονται με μια μεθοδολογία δύο σταδίων που συνδυάζει ανάλυση παραγόντων και συσταδοποίηση k-means στον εντοπισμό των αγορών. Τα αποτελέσματα της σύγκρισης έδειξαν ότι τα Self Organizing Maps ξεπερνούν την διαδικασία των δύο σταδίων και έχουν ενισχυμένη αποτελεσματικότητα στον εντοπισμό αγορών.

Η έρευνα των *Manojit Chattopadhyay, Pranab K Dan, Sitanath Majumdar και Partha Sarathi Chakraborty*, αναφέρεται στο ρόλο του αλγορίθμου Τεχνητών Νευρωνικών Δικτύων (ANN) στην αγοροπωλησία και επισημαίνει την ανάγκη για μια συλλογική ανασκόπηση της βιβλιογραφίας και ενός συστήματος ταξινόμησης προς την αναγνώριση των μελλοντικών τάσεων στην έρευνα αγοροπωλησίας. Η έρευνα περιλαμβάνει την δημιουργία μιας βάσης δεδομένων που καλύπτει περιέχει την βιβλιογραφία από το 2000 έως το 2010 και την πρόταση ενός σχήματος ταξινόμησης για τα εντοπισμένα άρθρα. Εντοπίστηκαν συνολικά χίλια άρθρα, και ένα υποσύνολο περίπου εκατό (100) σχετικών άρθρων ελέγχθηκε και ταξινομήθηκε στη συνέχεια βάσει του κύριου προσανατολισμού του καθενός. Τα βασικά συμπεράσματα της μελέτης δείχνουν ότι οι εφαρμογές βασισμένες σε ANN δέχονται περισσότερη προσοχή, από τις εφαρμογές που βασίζονται σε SOM και λαμβάνουν την δεύτερη θέση στην αγοροπωλησία.

Στη έρευνα του *Chenguang Wang* χρησιμοποιείται ένα μοντέλο μη εποπτευόμενης μάθησης το SOM. Το μοντέλο αναλύει τα δεδομένα των πελατών με μια διαδικασία που ονομάζεται μηχανική χαρακτηριστικών, επιλέγοντας σημαντικά συμπεριφορικά χαρακτηριστικά. Στη συνέχεια οι πελάτες ομαδοποιούνται χρησιμοποιώντας το SONN Νευρωνικό Δίκτυο Αυτοοργάνωσης. Τα αποτελέσματα επιβεβαιώνουν την απόδοση του προτεινόμενου μοντέλου με υψηλή ικανότητα ομαδοποίησης και διακριτής ικανότητας για την βελτίωση των κερδών της επιχείρησης στον τομέα του Μάρκετινγκ.

Η έρευνα των Vannessa Duarte , Sergio Zuniga-Jara και Sergio Contreras εξετάζει την υιοθέτηση των μεθόδων Μηχανικής Μάθησης στις εφαρμογές του Μάρκετινγκ την περίοδο 2008-2022. Από την έρευνα προκύπτει ότι κατά την διάρκεια αυτής της περιόδου η υιοθέτηση των μεθόδων Μηχανικής Μάθησης στο Μάρκετινγκ αυξήθηκε σημαντικά. Η υιοθέτηση περιλαμβάνει πέρα από την χρήση των κλασσικών μεθόδων όπως τα τεχνητά νευρωνικά δίκτυα έως υβριδικές μεθόδους που συνδυάζουν διάφορες τεχνικές για την βελτίωση των αποτελεσμάτων. Οι εφαρμογές που χρησιμοποιούνται ποικίλουν ευρέως, συμπεριλαμβανομένων της βαθιάς μάθησης, της εποπτευόμενης μάθησης, της ενισχυμένης μάθησης, της μη εποπτευόμενης μάθησης και των υβριδικών μεθόδων. Διαπιστώθηκε ότι τα κύρια προβλήματα Μάρκετινγκ που επιλύονται με την χρήση της μηχανικής μάθησης σχετίζονται με την συμπεριφορά του καταναλωτή, τα συστήματα προτάσεων, την πρόβλεψη, την αγοροπωλησία και την ανάλυση κειμένου και περιεχομένου.

Στην έρευνα των Jiapeng Liua, Xiuyu Liaoa, Wei Huangα και Xianzhao Liaob προτείνετε μια νέα προσέγγιση για την τμηματοποίηση της αγοράς μέσω υποστήριξης αποφάσεων με πολλαπλά κριτήρια που ενσωματώνει ανάλυση προτίμησης και επιλογή τμηματοποίησης. Η προσέγγιση χρησιμοποιεί μια συναρτησιακή τιμή ως μοντέλο προτίμησης και απαιτεί από τους καταναλωτές να παρέχουν συγκριτικές αξιολογήσεις ορισμένων προϊόντων ως πληροφορίες προτίμησης. Για την ανάλυση των προτιμήσεων κάθε καταναλωτή, η προσέγγιση εφαρμόζει το παράδειγμα διακριτοποίησης και τη στοχαστική ανάλυση πολυκριτήριας αποδοχής για την παραγωγή ενός συνόλου συναρτήσεων τιμής σύμφωνα με τις πληροφορίες προτιμήσεις που παρέχει κάθε καταναλωτής. Με βάση την ανάλυση προτίμησης, προτείνετε ένα νέο μέτρο για την μέτρηση της ομοιότητας μεταξύ των προτιμήσεων διάφορων καταναλωτών και αναπτύσσετε ένας αλγόριθμος ιεραρχικής συσταδοποίησης για την πραγματοποίηση καταμερισμού της αγοράς. Με σκοπό να βοηθήσει τις εταιρίες να εξυπηρετούν καταναλωτές από διάφορα τμήματα με στοχευμένες στρατηγικές Μάρκετινγκ και τα κατάλληλα προϊόντα, η προσέγγιση προτείνει να δημιουργηθεί μια αντιπροσωπευτική συνάρτηση τιμής και μια μοναδική κατάταξη των προϊόντων για κάθε καταναλωτή, ώστε τα προϊόντα που βρίσκονται στην πρώτη θέση της λίστας να τους παρουσιάζονται. Τέλος παρουσιάζεται ένα ενδεικτικό παράδειγμα ενός προβλήματος καταμερισμού της αγοράς και αναλύει την εφαρμογή της προτεινόμενης προσέγγισης.

Ο σκοπός της έρευνας, του Jonathan Z. Bloom, είναι να εξεταστεί ένα αυτοοργανωμένο νευρωνικό δίκτυο για την τμηματοποίηση της διεθνούς

τουριστικής αγοράς του Κέιπ Τάουν, στην Νότια Αφρική. Χρησιμοποιείται ένα νευρωνικό δίκτυο με ανάδραση για να συμπληρώσει την τμηματοποίηση, παρέχοντας περισσότερες πληροφορίες με βάση τη σχέση εισόδου-εξόδου και αναλύσεις ευαισθησίας. Η έρευνα έδειξε ότι η τουριστική Μητρόπολη του Κέιπ θα μπορούσε να χρησιμοποιεί τα μοντέλα νευρωνικού δικτύου για να παρακολουθεί την μεταβαλλόμενη συμπεριφορά των τουριστών εντός και μεταξύ των τμημάτων. Το άρθρο προτείνει επίσης εφαρμογές στον τομέα του Μάρκετινγκ για το Κέιπ Τάουν.

Η έρευνα των Amir Hossein Azadniaa, Muhamad Zameri Mat Samana, Kuan Yew Wonga, Pezhman Ghadimib και Norhayati Zakuanc παρουσιάζει μια μεθοδολογία επιλογής βιώσιμων δραστηριοτήτων και προμηθευτών για την εφοδιαστική τους αλυσίδα. Η μεθοδολογία περιλαμβάνει τμηματοποίηση και πολυκριτήρια λήψη αποφάσεων για την επίλυση του προβλήματος. Γίνεται χρήση του Self-Organizing Maps για την τμηματοποίηση των προμηθευτών και την προκαταρκτική κατάταξή τους με βάση την προτίμηση των καταναλωτών και χαρακτηριστικά βιωσιμότητας. Έπειτα γίνεται χρήση των μεθόδων πολυκριτήριας ανάλυσης για να καταταχθούν οι προμηθευτές με σκοπό να πραγματοποιηθεί συντονισμός μεταξύ των προμηθευτών και των πελατών. Έχει πραγματοποιηθεί μια μελέτη προκειμένου να επιδείξει την αποτελεσματικότητα της προτεινόμενης προσέγγισης.

Από τις παραπάνω έρευνες προκύπτει ότι οι εφαρμογές στο Μάρκετινγκ με χρήση Τεχνίτης Νοημοσύνης είναι πολλαπλές. Πιο ειδικά η χρήση των Self-Organizing Maps σε συνδυασμό με πολυκριτήριες μεθόδους αποδεικνύονται αποδοτικές για την τμηματοποίηση μιας αγοράς από τις έρευνες που παρουσιάστηκαν παραπάνω. Η ίδια λογική ακολουθείται και στην συγκεκριμένη διπλωματική εργασία.

Στην συγκεκριμένη εργασία ακολουθείται η λογική συνδυασμού των μεθόδων τμηματοποίησης για την αγορά λαδιού στη Γαλλία. Χρησιμοποιείται η πολυκριτήρια μέθοδος Utastar στο λογισμικό Markex σε συνδυασμό με το νευρωνικό δίκτυο Self Organizing Maps.

Κεφάλαιο 3ο: Θεωρητικό Υπόβαθρο

3.1 Συστήματα Υποστήριξης Αποφάσεων

Τα Συστήματα Υποστήριξης Αποφάσεων (ΣΥΑ) αποσκοπούν στην παροχή στους χρήστες τους της βέλτιστης δυνατής υποστήριξης για τη λήψη αποφάσεων. Με τη χρήση αναλυτικών μεθόδων, αλγορίθμων βελτιστοποίησης και επεξεργασίας δεδομένων, τα ΣΥΑ βοηθούν τους αποφασίζοντες στην ανάλυση των εναλλακτικών και διευκολύνουν την επιλογή της καταλληλότερης εναλλακτικής.

Ο Little όρισε (1970) ότι ένα σύστημα υποστήριξης αποφάσεων είναι ένα σύνολο διαδικασιών για την επεξεργασία δεδομένων και κρίσεων με στόχο να υποβοηθήσουν τους μάνατζερ στην διαδικασία λήψης μιας απόφασης. Τα ΣΥΑ βοηθούν τον αποφασίζοντα στην τελική απόφαση του σε μη δομημένα ή ημιδομημένα προβλήματα απόφασης, που οι επιλογές του είναι ταχέως μεταβαλλόμενες.

Τα ΣΥΑ εμφανίστηκαν για πρώτη φορά προς το τέλος της δεκαετίας του '60, ταυτόχρονα με την εμφάνιση μιας νέας κατηγορίας Πληροφοριακών Συστημάτων που επικεντρώνονταν στα ΣΥΑ με βάση τα μοντέλα (model-driven DSS) ή Συστήματα Αποφάσεων Διοίκησης (Management Decision Systems). Στη συνέχεια, στη δεκαετία του '70 ακολούθησε η θεωρητική ανάπτυξη και η ενσωμάτωση συστημάτων οικονομικού προγραμματισμού-λογιστικών φύλλων (spreadsheet DSS) και ΣΥΑ για την υποστήριξη της ομαδικής λήψης αποφάσεων (Group DSS). Τη δεκαετία του '90, εμφανίστηκαν τα Διοικητικά Πληροφοριακά Συστήματα (Executive Information Systems), τα Συστήματα OLAP και Επιχειρηματικής Ευφυΐας (Business Intelligence). Στα μέσα της δεκαετίας του '90, παρουσιάστηκαν τα ΣΥΑ που βασίζονται στη γνώση (Knowledge-Based DSS) και τα ΣΥΑ που βασίζονται στον ιστό (Web-based DSS).

Ένα ΣΥΑ είναι ικανό να επιλύσει προβλήματα που δεν επιλύονται με προβλήματα βελτιστοποίησης και να οδηγήσει τον αποφασίζοντα σε υλοποιήσιμες αποφάσεις. Από ένα ΣΥΑ παρέχεται πολλαπλός αριθμός εναλλακτικών λύσεων και με την χρήση του βελτιώνεται η ποιότητα της πληροφόρησης που βασίζεται η απόφαση. Η κατανόηση του προβλήματος από τον αποφασίζοντα είναι καλύτερη και οδηγείται σε βελτιωμένες αποφάσεις.

3.2 Πολυκριτήρια Ανάλυση

Η Πολυκριτήρια Λήψη Αποφάσεων αναπτύχθηκε επειδή υπήρχε ανάγκη επίλυσης των πολυδιάστατων προβλημάτων που αντιμετώπιζαν οι επιχειρήσεις και τα μοντέλα που προϋπήρχαν αδυνατούσαν να τα επιλύσουν με την χρήση ενός κριτηρίου. Επειδή τα κριτήρια αξιολόγησης των εναλλακτικών αποφάσεων είναι πολλαπλά και αντικρουόμενα στα πολυκριτήρια προβλήματα, τα πολυκριτήρια προβλήματα χαρακτηρίζονται από χαμηλό βαθμό απόδοσης (Ματσατσίνης, 2010).

Οι διαφορετικές προσεγγίσεις της διαδικασίας απόφασης δημιούργησαν δυο σχολές. Τη Πολυκριτήρια Λήψη Αποφάσεων (MultiCriteria Decision Making – MCDM) και την Πολυκριτήρια Υποβοήθηση της Λήψης Αποφάσεων (MultiCriteria Decision Aid – MCDA).

Η Πολυκριτήρια Λήψη Αποφάσεων (MCDM), αναπαριστά τις προτιμήσεις των αποφασίζοντων σε ένα σύνολο εναλλακτικών επιλογών ως ένα σύστημα αξιών. Το σύστημα αξιών αφορά την διαμόρφωση μιας συνάρτησης χρησιμότητας και των σχετικών βαρών των προτεραιοτήτων. Σύμφωνα με αυτή οι προτιμήσεις του αποφασίζοντα σε ένα σύνολο εναλλακτικών ενεργειών λαμβάνονται υπόψη στη διαμόρφωση ενός συστήματος αξιών, το οποίο ικανοποιεί ένα σύνολο συνθηκών, μέσα από το οποίο ο αποφασίζων θα οδηγηθεί στην επιλογή της σωστότερης λύσης. Βασικοί στόχοι της Πολυκριτήριας Λήψης Αποφάσεων είναι:

- Ο καθορισμός των συνθηκών που απαιτούνται ώστε να υπάρχει το σύστημα αξιών.
- Η ύπαρξη υποστήριξης του αποφασίζοντα για να οδηγείται στην εύρεση, μέσα από μια διαδικασία, ενός συστήματος αξιών και να καταλήγει στη σωστή απόφαση.

Στην Πολυκριτήρια Υποβοήθηση της Λήψης Αποφάσεων (MCDA) σύμφωνα με τον Roy (1990) κατά τη διαδικασία λήψης μιας απόφασης:

- Συμμετέχουν περισσότεροι του ενός αποφασίζοντες με διαφορετικές απόψεις και πιστεύω.
- Στην περίπτωση του ενός αποφασίζοντα οι προτιμήσεις του έχουν συνήθως μεγάλο βαθμό ασάφειας με ασυνέπειες και συγκρούσεις.
- Ο τρόπος με τον οποίο γίνεται ο καθορισμός των εναλλακτικών επιλογών, των κριτηρίων, της αξιολόγησης των εναλλακτικών επιλογών και των προτιμήσεων των αποφασίζοντων είναι αυθαίρετος

Σκοπός της Πολυκριτήριας Ανάλυσης Υποστήριξης της λήψης αποφάσεων είναι να υποδείξει στον αποφασίζοντα, μέσω της σταδιακής κατανόησης και βελτίωσης των αποδόσεων και των γνώσεων του, στην καλύτερη δυνατή απόφαση για το συγκεκριμένο πρόβλημα.

3.2.1 Μεθοδολογία Μοντελοποίησης Ενός Πολυκριτήριου Προβλήματος

Σύμφωνα με τον Roy (1976, 1990) ο προτεινόμενος τρόπος μοντελοποίησης ενός προβλήματος Πολυκριτήριας Ανάλυσης αποτελείται από τέσσερα στάδια.

Αρχικά γίνεται αντιληπτό το αντικείμενο της απόφασης και αναλύεται κάθε απόφαση σε ένα πεπερασμένο ή συνεχές σύνολο εναλλακτικών επιλογών $\{A = a_i, i = 1, 2, \dots, n\}$. Υπάρχουν τέσσερα είδη προβληματικών όσο αφορά το σύνολο των εναλλακτικών επιλογών.

- Προβληματική α: Επιλογή μιας και μόνο εναλλακτική μέσα από ένα σύνολο.
- Προβληματική β: Αφορά την ταξινόμηση (sorting) όλων των εναλλακτικών ενεργειών του συνόλου σε ομάδες με συγκεκριμένες ιδιότητες.
- Προβληματική γ: Αφορά την κατάταξη (ranking) όλων των εναλλακτικών ενεργειών του συνόλου A από την προτιμότερη στην λιγότερο προτιμητέα.
- Προβληματική δ: Αφορά την απλή περιγραφή όλων των εναλλακτικών επιλογών του συνόλου A καθώς και των συνεπειών τους ώστε να είναι κατανοητές για τους αποφασίζοντες.

Τα προβλήματα συνήθως είναι σύνθετα και για την επίλυση τους μπορεί να χρειαστούν περισσότερες από μια προβληματική στα διάφορα στάδια λήψης μιας απόφασης.

Στην συνέχεια γίνεται καθορισμός των κριτηρίων. Σαν κριτήριο ορίζεται κάθε μονότονη μεταβλητή η οποία δηλώνει τις προτιμήσεις του αποφασίζοντα. Τα κριτήρια χωρίζονται σε ποσοτικά που εκφράζονται από μια συνεχή κλίμακα και σε ποιοτικά-διάταξης που περιγράφονται από κλίμακες διακριτών τιμών. Κάθε εναλλακτική επιλογή ενός συνόλου A μπορεί να εκτιμηθεί από τους αποφασίζοντες ως ένα σύνολο στοιχειωδών επιπτώσεων αποτελούμενα από ιδιότητες, χαρακτηριστικά, μειονεκτήματα, πλεονεκτήματα κ.α. Roy (1985). Για τον καθορισμό της κατάλληλης οικογένειας κριτηρίων ο Roy (1985) πρότείνει τα ακόλουθα βήματα:

1. Καθορισμός του συνόλου των εναλλακτικών επιλογών
2. Ανάλυση των στοιχειωδών επιπτώσεων των εναλλακτικών επιλογών
3. Καθορισμός των αξόνων προτίμησης
4. Επιλογή των διαστάσεων
5. Καθορισμός της συνεπούς οικογένειας κριτηρίων

Τα κριτήρια συμβολίζονται με :

$$g_j(a_i), i = 1, 2, \dots, n \text{ και } j = 1, 2, \dots, m$$

Ορίζουμε με, $g: A \rightarrow R$ και $a \rightarrow g(a)$

Όπου,

$g(a)$: η εκτίμηση της εναλλακτικής επιλογής a του συνόλου A , πάνω στο κριτήριο g .

Η συνάρτηση g θα πρέπει να πληροί την ιδιότητα της μονοτονίας σύμφωνα με την οποία για δυο εναλλακτικές a και b του συνόλου A θα ισχύει:

- Η a προτιμάται από την b όταν $g(a) > g(b)$
- Η a είναι ισοδύναμη με την b όταν : $g(a) = g(b)$

Συνεπής οικογένεια κριτηρίων ονομάζεται το σύνολο των κριτηρίων τα οποία χρησιμοποιούνται για την λήψη μιας απόφασης και πρέπει να τηρούν κάποιες ιδιότητες:

- Πρέπει να πληρούν την ιδιότητα της μονοτονίας ή συνέπειας, δηλαδή στην περίπτωση δυο εναλλακτικών επιλογών a και b , η a προτιμάται από την b όταν ισχύει:

$g_i(a) = g_j(b) \forall i = j \text{ και } g_i(a) > g_j(b)$, τότε η εναλλακτική a προτιμάται από την b .

- Πρέπει να έχουν επάρκεια στα πλαίσια της διαθέσιμης πληροφορίας, δηλαδή αν για ένα ζεύγος εναλλακτικών επιλογών a και b ισχύει $g_i(a) = g_j(b) \forall i = 1, 2, \dots, n$ τότε η εναλλακτική a είναι ισοδύναμη της b δηλαδή δεν απουσιάζει κανένα κριτήριο απόφασης από το σύνολο των χρησιμοποιούμενων κριτηρίων.

- Να μην είναι πλεοναστικά, δηλαδή η αφαίρεση ενός κριτηρίου από το σύνολο των κριτηρίων απόφασης που χρησιμοποιείται να είναι ικανή να αναιρέσει μια από τις προηγούμενες δύο συνθήκες για κάποιο ζευγάρι εναλλακτικών επιλογών.

3.3 Μάρκετινγκ

3.3.1 Εισαγωγή

Οι ορισμοί για το Μάρκετινγκ είναι πολλαπλοί και έχουν διατυπωθεί είτε από θεωρητικές είτε επαγγελματικές ενώσεις. Σύμφωνα με τον Philip Kotler, ένας από τους πιο γνωστούς θεωρητικούς και καθηγητές στον χώρο του Μάρκετινγκ, «Το Μάρκετινγκ είναι μια κοινωνική και διοικητική διαδικασία με την οποία άτομα και ομάδες αποκτούν ότι χρειάζονται και επιθυμούν, μέσω παραγωγής, της προσφοράς και της ανταλλαγής προϊόντων που έχουν αξία για αυτούς με άλλα προϊόντα»(Kotler,2000).

Το Μάρκετινγκ μπορεί να περιγραφεί ως οι διαδικασίες που χρησιμοποιεί μια επιχείρηση για να προωθήσει την πώληση ενός προϊόντος ή μιας υπηρεσίας. Περιλαμβάνει τη διαδικασία της δημιουργίας, επικοινωνίας, παράδοσης και ανταλλαγής προσφορών που έχουν ως σκοπό να ικανοποιήσουν τις ανάγκες αλλά και τις επιθυμίες των καταναλωτών μιας επιχείρησης ή ενός οργανισμού. Η αντιστοίχιση προϊόντων με πελάτες εξασφαλίζει τελικά την κερδοφορία.

3.3.2 Μίγμα Μάρκετινγκ

Βασικός παράγοντας επιτυχίας κάθε οργανισμού ή επιχείρησης είναι το Μάρκετινγκ, καθώς αποτελεί το κύριο μέσω επικοινωνίας μεταξύ της επιχείρησης και των πελατών. Η πιο γνωστή στρατηγική του Μάρκετινγκ είναι το μίγμα Μάρκετινγκ (Thabit H. & Manaf B. Raewf, 2018). Το μίγμα Μάρκετινγκ πηγάζει από την θεωρία της μικροοικονομίας (Chong, 2003). Ορίζεται ως ένα σύνολο εργαλείων και παραγόντων που καθοδηγούν τις βασικές αποφάσεις των επιχειρήσεων, με σκοπό τη. Αύξηση της ελκυστικότητας των προϊόντων και υπηρεσιών, καθώς και την προσαρμογή στις ανάγκες των καταναλωτών (Palmer,2004).

Το μίγμα Μάρκετινγκ αποτελείται από τέσσερις βασικούς συντελεστές, οι οποίοι με την σωστή χρήση και τον κατάλληλο συνδυασμό καθορίζουν μια στρατηγική Μάρκετινγκ επιτυχή. Αυτοί οι συντελεστές ονομάζονται 4Ps από τα αρχικά των αγγλικών όρων Product, Price, Place, Promotion. Το ευρύ μείγμα Μάρκετινγκ αποτελείται και από άλλους τρεις συντελεστές: Άνθρωποι (People), Διαδικασίες (Process), Φυσικές Αποδείξεις (Physical Evidence).

Το βασικό μείγμα του Μάρκετινγκ αποτελούν(4Ps):

Προϊόν (Product)

Το προϊόν αναφέρεται στο αντικείμενο ή αντικείμενα η εταιρία έχει σκοπό να προσφέρει στους πελάτες. Ένα προϊόν ορίζεται ως ένα αγαθό ή υπηρεσία που μπορεί να προσφερθεί σε μια αγορά για την προσέλκυση της προσοχής, χρήση ή κατανάλωση(Singh, 2012). Το προϊόν έχει σκοπό να καλύψει ένα καινού στην αγορά ή να ανταποκριθεί στις μεγάλες ανάγκες ενός ήδη υπάρχοντος προϊόντος. Για να προετοιμαστεί κατάλληλα για μια καλή στρατηγική μάρκετινγκ το ανάλογο τμήμα μάρκετινγκ πρέπει να καταλάβει πλήρως και σε βάθος, το προϊόν που πωλείται, τον ανταγωνισμό και τα χαρακτηριστικά που το ξεχωρίζουν, αν μπορεί να συνδυαστεί με ένα ήδη υπάρχον προϊόν ή αν μπορούν να χρησιμοποιηθούν οι διαδικασίες παραγωγής του από κάποια ήδη υπάρχουσα γραμμή παραγωγής της εταιρίας και τέλος αν υπάρχουν υποκατάστατα προϊόντα στην αγορά.

Τιμή (Price)

Δεύτερός βασικός πυλώνας του μίγματος Μάρκετινγκ αποτελεί η τιμή. Τιμή θεωρείται οποιαδήποτε χρέωση υπάρχει σε ένα προϊόν ή μια υπηρεσία (Borden&Marshall,1959). Ο καθορισμός της τιμής ενός προϊόντος ή μιας υπηρεσίας αποτελεί μια πολύπλοκη διαδικασία. Για την τιμή οι εταιρίες πρέπει να υπολογίσουν το κόστος παραγωγής κάθε προϊόντος, τα κόστη μάρκετινγκ και τα κόστη διανομής. Επίσης πρέπει να υπολογίσουν την τιμή των ανταγωνιστικών προϊόντων στην αγορά και πόσο πρόθυμοι είναι οι πελάτες να αγοράσουν με την συγκεκριμένη τιμή το προϊόν. Όλοι αυτοί οι παράγοντες επηρεάζουν την κερδοφορία του οργανισμού (Singh, 2012).

Δίκτυο Διανομής (Place)

Το Δίκτυο διανομής είναι ο τρίτος πυλώνας του μίγματος Μάρκετινγκ και αναφέρεται στην στρατηγική διανομής, η οποία περιλαμβάνει τον τρόπο με τον οποία θα γίνει το προϊόν ή η υπηρεσία διαθέσιμη στο κατάλληλο μέρος της αγοράς. Ο στόχος είναι να διασφαλιστεί ότι το προϊόν φτάνει στους σωστούς πελάτες την κατάλληλη στιγμή και το κατάλληλο μέρος, είτε μέσα από μεσάζοντες, είτε απευθείας στους πελάτες/καταναλωτές (Thabit & Younus, 2015). Επιπρόσθετα, ένας άλλος σημαντικός παράγοντας για την ομαλή λειτουργία της αλυσίδας του δικτύου διανομής είναι τα logistics, τα οποία είναι υπεύθυνα για την παράδοση του προϊόντος στον τελικό πελάτη, αλλά και της μεταφοράς και αποθήκευσης των προϊόντων.

Προώθηση (Promotion)

Από τους ισχυρότερους πυλώνες του μείγματος μάρκετινγκ είναι η προώθηση. Η προώθηση είναι ο τρόπος με τον οποίο μια επιχείρηση προβάλλει τα προϊόντα ή τις υπηρεσίες της στην αγορά. Περιλαμβάνει την διαφήμιση, τις δημόσιες σχέσεις, μάρκετινγκ περιεχομένου, το email μάρκετινγκ, διαδικτυακή παρουσία, affiliate μάρκετινγκ, μάρκετινγκ κοινωνικών δικτύων κ.α. Η προώθηση συνιστά σημαντικό μέρος του συνολικού σχεδιασμού μάρκετινγκ, καθώς βοηθάει στην κατανόηση, αναγνώριση και ενίσχυση του επιχειρησιακού μηνύματος που η επιχείρηση επιθυμεί να μεταδώσει στο κοινό της.

Και σε το πιο ευρύ μείγμα (7P) περιλαμβάνει ακόμα:

Άνθρωποι (People)

Στον όρο Άνθρωποι περιλαμβάνονται όλοι όσοι εργάζονται σε μια επιχείρηση καθώς και τους πελάτες της επιχείρησης. Ο τρόπος με τον οποίο οι εργαζόμενοι επικοινωνούν και εξυπηρετούν τους πελάτες διαδραματίζει σημαντικό ρόλο στην εμπειρία του καταναλωτή με το προϊόν, στην αξιολόγηση του και στην πρόθεσή του να ξαναγοράσει το προϊόν. Πρέπει οι άνθρωποι να είναι κατάλληλα εκπαιδευμένοι ώστε να είναι ικανοί να δημιουργούν αξιόπιστες σχέσεις με τους πελάτες.

Διαδικασίες (Process)

Οι διαδικασίες αναφέρονται στις σειρές βημάτων και διαδικασιών που χρησιμοποιούνται για την υλοποίηση των στρατηγικών Μάρκετινγκ και την παροχή των προϊόντων και υπηρεσιών στους πελάτες. Στόχος κάθε επιχείρησης είναι να δημιουργήσει μια ομαλή, αποτελεσματική και φιλική προς τον πελάτη εμπειρία με τα προϊόντα και τις υπηρεσίες τους. Οι διαδικασίες κατατάσσονται στις παρακάτω κατηγορίες ανάλογα με την φύση τους:

- α. Διαχείριση Αλυσίδας Εφοδιαστικής (Supply Chain Management): Οι διαδικασίες αναφορικά με την αποτελεσματική διαχείριση της αλυσίδας εφοδιαστικής είναι ουσιώδεις για την αποτελεσματική παροχή προϊόντων ή υπηρεσιών στον πελάτη.
- β. Συστήματα Πληροφορικής (Information Systems): Η χρήση σύγχρονων τεχνολογικών συστημάτων πληροφορικής είναι κρίσιμη για τη συλλογή, ανάλυση, και ερμηνεία δεδομένων πελατών για την καλύτερη κατανόηση των αναγκών τους.
- γ. Διαδικασίες Καινοτομίας (Innovation Processes): Η διαδικασία καινοτομίας επιτρέπει στις επιχειρήσεις να παραμένουν ενημερωμένες και ανταγωνιστικές μέσω συνεχούς ανάπτυξης και ενσωμάτωσης νέων ιδεών και τεχνολογιών.

- δ. Διαδικασίες Παραγωγής (Production Processes): Η οργάνωση και βελτιστοποίηση των διαδικασιών παραγωγής είναι κρίσιμη για την εξασφάλιση ποιοτικών προϊόντων και την αποτελεσματική παραγωγή.
- ε. Διαδικασίες Εξυπηρέτησης Πελατών (Customer Service Processes): Οι διαδικασίες που αφορούν την εξυπηρέτηση των πελατών συμβάλλουν στη δημιουργία θετικών εμπειριών και τη διατήρηση της εμπιστοσύνης πελατών.

Φυσικές Αποδείξεις (Physical Evidence)

Τα φυσικά στοιχεία παρέχουν απτές αποδείξεις για την ποιότητα και την εμπειρία που προσφέρει μια επιχείρηση. Τα φυσικά στοιχεία είναι ιδιαίτερα χρήσιμα εργαλεία για τον καταναλωτή που έχει την πρόθεση να αγοράσει ένα προϊόν, αλλά πρώτα χρειάζεται κάποια αποδεικτικά στοιχεία προκειμένου να αντιληφθεί την ποιότητα του προϊόντος ή της υπηρεσίας που πρόκειται να προμηθευτεί.

3.3.3 Ψηφιακό Μάρκετινγκ

Οι προηγμένες τεχνολογίες έχουν προκαλέσει επανάσταση σε σχεδόν κάθε πτυχή της ανθρώπινης ζωής, συμπεριλαμβανομένων των τρόπων με τους οποίους οι επιχειρήσεις προωθούν προϊόντα και υπηρεσίες στους καταναλωτές. Το διαδίκτυο, τα ψηφιακά μέσα και τα κοινωνικά δίκτυα χρησιμοποιούνται πλέον ευρέως στον επιχειρηματικό τομέα, με σκοπό την εκσυγχρονισμένη προσέγγιση του παραδοσιακού μάρκετινγκ. Εταιρίες σε διάφορους κλάδους, όπως το λιανικό εμπόριο, η μεταποίηση και η υγειονομική περίθαλψη, επενδύουν σημαντικά σε τεχνολογικές λύσεις για τη βελτιστοποίηση των δραστηριοτήτων τους. Οι καινοτομίες που προκύπτουν στον τομέα της τεχνολογίας παρέχουν στους υπεύθυνους μάρκετινγκ τη δυνατότητα να δημιουργήσουν, να επικοινωνήσουν και να προσφέρουν αξία στους πελάτες τους (Grewal, D., Hultand, J., Kopalle, P.K. κ.ά., 2020).

Το ψηφιακό μάρκετινγκ (digital marketing) είναι ένας όρος «ομπρέλα» που περιλαμβάνει πολλά κανάλια και τακτικές στο διαδίκτυο με στόχο την διαφήμιση και προώθηση προϊόντων και υπηρεσιών. Είναι μια από τις πιο αποτελεσματικές στρατηγικές για την ανάπτυξη των επιχειρήσεων (business development) και την αναγνωρισιμότητα της επωνυμίας (brand awareness). Οι επιχειρήσεις εκμεταλλεύονται ψηφιακά κανάλια, όπως μηχανές αναζήτησης (SEO), μέσα κοινωνικής δικτύωσης (social media), ηλεκτρονικό ταχυδρομείο (e-mail) και άλλες ιστοσελίδες προκειμένου να συνδεθούν με τρέχοντες και υποψήφιους πελάτες.

- Μηχανές Αναζήτησης (SEO)

Σκοπός της βελτιστοποίησης των Μηχανών Αναζήτησης (SEO) είναι μέσα από στρατηγικές διαδικασίες να βελτιστοποιήσει την βαθμολογία μιας ιστοσελίδας ή ενός διαδικτυακού περιεχομένου, προκυμμένου να είναι πιο εμφανής, δηλαδή να «ανέβει» πιο ψηλά, στα αποτελέσματα των μηχανών αναζήτησης. Η βελτιστοποίηση επιτυγχάνεται με την βοήθεια συγκεκριμένων λέξεων-κλειδιά, ώστε να ενισχυθεί η ορατότητα του και να βελτιώνει την εμφάνιση του σε θέσεις.

- Κοινωνικά Δίκτυα (Social Media)

Ο συγκεκριμένος τύπος Μάρκετινγκ περιλαμβάνει την δημιουργία διαδικτυακής παρουσίας σε συγκεκριμένες πλατφόρμες κοινωνικών μέσων, όπως για παράδειγμα Facebook, Instagram, Twitter, YouTube, LinkedIn κ.α. Οι εταιρίες έχουν την δυνατότητα να τοποθετήσουν πληρωμένες διαφημίσεις για να προωθηθούν από τους αλγορίθμους και να έχουν μεγαλύτερη πιθανότητα να τις δουν οι θεατές. Σχεδόν όλες οι εταιρίες πλέον χρησιμοποιούν συνδυασμό μέσων κοινωνικής δικτύωσης. Η επιλογή προκύπτει με βάση σε ποια κοινωνική πλατφόρμα το κοινό-στόχος είναι πιο ενεργό και το είδος της επιχείρησης που εκπροσωπείται.

- Διαδικτυακή Παρουσία

Η Διαδικτυακή παρουσία στο Μάρκετινγκ είναι μια από τις πιο σημαντικές στρατηγικές και χρησιμοποιεί το διαδίκτυο για την δημιουργία ή την αύξηση της επισκεψιμότητας σε έναν ιστότοπο και να μεταδώσει μηνύματα μάρκετινγκ σε στοχευμένο κοινό. Έχει ως στόχο την προσέλκυση του στόχου-κοινού που θα προβεί στις ανάλογες ενέργειες, όπως την πραγματοποίηση μιας αγοράς. Η εξέλιξη της διαδικτυακής διαφήμισης έχει καταφέρει να είναι πολύ πιο στοχευμένη στο κοινό που απευθύνετε, και καταφέρνει λόγω αυτού να έχει σημαντικό προβάδισμα έναντι της παραδοσιακής διαφήμισης.

- Email:

Το email Μάρκετινγκ αποτελεί ένα ευρέως διαδεδομένο εργαλείο Μάρκετινγκ που χρησιμοποιούν πολλές επιχειρήσεις. Οι εταιρίες αποκτούν τις διευθύνσεις email των πελατών ή των πιθανών πελατών και διανείμουν μηνύματα. Αυτά τα μηνύματα περιέχουν κουπόνι, ευκαιρίες έκπτωσης ή ενημέρωση για επερχόμενες εκπτώσεις. Με αυτή την προσέγγιση, οι επιχειρήσεις στοχεύουν στην δημιουργία εμπιστοσύνης με τους πελάτες τους και στην ενίσχυση της επωνυμίας τους.

- Affiliate Μάρκετινγκ

Το Affiliate Μάρκετινγκ είναι ένας τρόπος διαδικτυακής προώθησης μέσω συνεργασίας. Ο όρος Affiliate Μάρκετινγκ περιγράφει την σχέση μεταξύ του διαφημιζόμενου και του συνεργάτη (Affiliate), όπου ο συνεργάτης προωθεί

επισκέπτες προς την ιστοσελίδα του διαφημιζόμενου και αμείβεται μόνο εάν οι επισκέπτες πραγματοποιήσουν μια συγκεκριμένη δράση, συνήθως αγορά. Αυτός ο τρόπος αμοιβής είναι γνωστός ως Cost Per Action (CPA).

- Περιεχομένου

Το Μάρκετινγκ περιεχομένου αφορά στην δημιουργία και προώθηση περιεχομένου με σκοπό την δημιουργία αναγνωσιμότητας, την αύξηση επισκεψιμότητας, την εκδήλωση ενδιαφέροντος και τελικά την μετατροπή τους σε πελάτες. Οι πιο συνηθισμένοι τρόποι για τις επιχειρήσεις για να πετύχουν το Μάρκετινγκ Περιεχομένου είναι τα βίντεο και τα podcasts.

3.3.4 Στρατηγική Ψηφιακού Μάρκετινγκ

Οι στρατηγικές Μάρκετινγκ είναι ένα ολοκληρωμένο σχέδιο που περιλαμβάνει την απαραίτητη έρευνα, τα απαραίτητα πλάνα και τακτικές, ακολουθεί τις αξίες της εταιρίας και στοχεύει στην δημιουργία ενός βιώσιμου ανταγωνιστικού πλεονεκτήματος.

Τα κοινωνικά μέσα έχουν τροποποιήσει τον τρόπο εφαρμογής των στρατηγικών μάρκετινγκ και τις βασικές στρατηγικές όπως η 4P. Οι έννοιες που αναφέρθηκαν προηγούμενος Προϊόν, Τιμή, Τοποθεσία και Προώθηση έχουν εξελιχθεί ενσωματώνοντας ψηφιακά χαρακτηριστικά. Το προϊόν πρέπει να ικανοποιεί τις προσδοκίες των χρηστών στο διαδίκτυο. Η Τιμή ορίζεται λαμβάνοντας υπόψη ότι οι χρήστες στο διαδίκτυο είναι ικανοί να πραγματοποιήσουν γρήγορες συγκρίσεις τιμών, να πληρώσουν διαδικτυακά, να χρησιμοποιήσουν κουπόνια και εκπτώσεις και έτσι η τιμολόγηση είναι δυναμική. Η Τοποθεσία αναφέρεται στο διαδικτυακό περιβάλλον και στην ευκολία εύρεσης και περιήγησης στο διαδικτυακό κατάστημα. Η Προώθηση πραγματοποιείται χρησιμοποιώντας πολλαπλά διαδικτυακά κανάλια, καθιστώντας πολύ εύκολη την ανάπτυξη προσωπικών σχέσεων.

3.4 Τμηματοποίηση της Αγοράς

Η τμηματοποίηση της αγοράς ανήκει στις σημαντικότερες πρακτικές του στρατηγικού σχεδιασμού στον τομέα του μάρκετινγκ και συμβάλει στην κατανόηση της καταναλωτικής συμπεριφοράς, αποτελώντας κρίσιμο παράγοντα για την επιτυχία της επιχείρησης στην αγορά (Morgan, Whitley, Feng, & Chari, 2019). Αφενός μελετώντας την καταναλωτική συμπεριφορά, δημιουργούνται τα θεμέλια για αποδοτική τμηματοποίηση, αφετέρου στοχεύοντας τη κατάλληλη ομάδα καταναλωτών, κρίνεται επιτυχής μια ενέργεια μάρκετινγκ (Peter & Olson, 2010). Η λογική της τμηματοποίησης της αγοράς στηρίζεται στο ότι ένα προϊόν δεν μπορεί να είναι αρεστό για ολόκληρο το καταναλωτικό κοινό, απαιτώντας έτσι τον σχεδιασμό μιας στρατηγικής μάρκετινγκ. Ο τρόπος ζωής, η γνώση για το προϊόν και η αγοραστική συμπεριφορά διαφέρουν από άτομο σε άτομο. Συνεπώς, οι επιτυχημένες

επιχειρήσεις προσαρμόζουν τα προϊόντα τους για να ικανοποιούν κάθε κατηγορία καταναλωτών.

Η τμηματοποίηση της αγοράς χωρίζει τους καταναλωτές σε ομάδες, όπου τα μέλη της κάθε ομάδας έχουν κοινά χαρακτηριστικά σε σχέση με τα δημογραφικά τους χαρακτηριστικά (π.χ. ηλικία, φύλλο, εισόδημα), γεωγραφικά χαρακτηριστικά, ψυχομετρικά χαρακτηριστικά και η συμπεριφορά τους. Η τμηματοποίηση των καταναλωτών σε έρευνα Μάρκετινγκ αποσκοπεί στην βελτίωση της κατανόησης της βάσης πελατών και στο να προσαρμόζει τις υπηρεσίες της για να ικανοποιεί τις συγκεκριμένες ανάγκες και επιθυμίες κάθε ομάδας. Αυτή η προσέγγιση επιτρέπει στις επιχειρήσεις να διαμορφώνουν πιο στενευμένες και αποτελεσματικές στρατηγικές Μάρκετινγκ βελτιώνοντας τελικά την ικανοποίηση των συμμετεχόντων.

Το μειονέκτημα της τμηματοποίησης της αγοράς είναι η επένδυση χρόνου και ανθρώπινου δυναμικού από την επιχείρηση. Μπορεί η στρατηγική τμηματοποίησης να είναι αποτυχημένη και να αποτελεί σημαντική σπατάλη πόρων για την εταιρία. Συνεπώς, μια επιχείρηση πρέπει να συγκεντρώσει όλες τις πληροφορίες που χρειάζεται για να κρίνει αν η στρατηγική τμηματοποίησης είναι συμφέρουσα επιλογή.

Για την επιτυχή ανάλυση της τμηματοποίησης της αγοράς απαιτείται η χρήση των κατάλληλων μεθόδων. Από την χρήση των καταλληλότερων μεθόδων από τον αναλυτή θα προκύψουν και τα ανάλογα αποτελέσματα. Για αυτό το λόγο μια επιχείρηση πέρα από την συλλογή των σωστών δεδομένων χρειάζεται και ένας ικανός αναλυτής.

Για την διασφάλιση της ποιότητας της ανάλυσης απαιτείται μια σειρά από προϋποθέσεις, όπως η συλλογή χρήσιμων δεδομένων, η καλή διερεύνηση αυτών των δεδομένων, η σωστή τμηματοποίηση του καταναλωτικού κοινού και η περιγραφή αυτών των τμημάτων.

Το στάδιο της συλλογής των δεδομένων αποτελεί σημαντικό παράγοντα για την καλή ποιότητα της τμηματοποίησης, καθώς όσο καλή και να είναι η στατιστική μέθοδος που χρησιμοποιείται δεν μπορεί να διορθώσει τα κακής ποιότητας δεδομένα. Δηλαδή, η τμηματοποίηση της αγοράς θα μπορεί να είναι τόσο καλή όσο είναι και τα δεδομένα που έχουν συλλεχθεί.

Τελικός σκοπός είναι η διερεύνηση των ομάδων που προέκυψαν από την τμηματοποίηση και να επιτευχθεί ένα λεπτομερέστερο προφίλ των ομάδων. Η τμηματοποίηση των πελατών δημιουργεί τις ομάδες και καθορίζει ποιες μπορούν να επιλεγθούν για τις διαφημιστικές καμπάνιες. Η επιχείρηση με αυτά τα στοιχεία, πρέπει να αναπτύξει ένα μείγματος μάρκετινγκ που να εξυπηρετεί κάθε ομάδα.

3.5 Τεχνητή Νοημοσύνη(Artificial Intelligence)

Η Τεχνητή Νοημοσύνη (ΤΝ) αναφέρεται σε έναν κλάδο της επιστήμης των υπολογιστών που επιδιώκει την ανάπτυξη και την εφαρμογή νέων στρατηγικών έρευνας και μεθόδων αναπαράστασης της γνώσης σε προγράμματα ηλεκτρονικών υπολογιστών. Κύριος σκοπός της ΤΝ είναι να προσομοιώσει τη διαδικασία που ακολουθεί το ανθρώπινο μυαλό για την επίλυση διαφόρων προβλημάτων. Σε αντίθεση με τους συμβατικούς υπολογιστές, η ΤΝ αναζητά νέες προσεγγίσεις για να αντιμετωπίσει προβλήματα που ξεπερνούν τις περιορισμένες υπολογιστικές δυνατότητες των συμβατικών συστημάτων. Παράλληλα, η συμβάλλει στη βαθύτερη κατανόηση της ανθρώπινης νοημοσύνης και στην ανάπτυξη προηγμένων λειτουργιών που εμπνέονται από αυτήν.

3.6 Μηχανική Μάθηση

3.6.1 Εισαγωγή

Η Μηχανική Μάθηση αναφέρεται σε ένα φαινόμενο όπου ένα σύστημα ή ένα πρόγραμμα υπολογιστή βελτιώνει την απόδοσή του σε μια συγκεκριμένη εργασία, χωρίς να χρειάζεται να προγραμματιστεί εκ νέου. Στόχος της Μηχανικής Μάθησης είναι να δημιουργήσει προγράμματα τα οποία μπορούν να βελτιώσουν την απόδοσή τους μέσω της αξιοποίησης προηγούμενης γνώσης και εμπειρίας.

Ένας πιο επίσημος ορισμός για τη Μηχανική Μάθηση έχει προταθεί από τον Tom M. Mitchell (1997):

- "Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P , αν η επίδοσή του σε εργασίες της κλάσης T , όπως αποτιμάται από το μέτρο P , βελτιώνεται με την εμπειρία E ."

Αυτός ο ορισμός επισημαίνει ότι η Μηχανική Μάθηση επιτυγχάνεται μέσω της εμπειρίας, όπου το πρόγραμμα μαθαίνει νέες πληροφορίες και προσαρμόζει την απόδοσή του για να βελτιώσει την εκτέλεση συγκεκριμένων εργασιών. Οι εργασίες αυτές αξιολογούνται με ένα μέτρο επίδοσης, το οποίο μπορεί να μετρήσει την απόδοση του συστήματος σε σχέση με τις εκάστοτε εργασίες.

3.6.2 Είδη Μηχανικής Μάθησης

Στην μηχανική μάθηση υπάρχουν τρεις τρόποι μάθησης, όπως και οι τρόποι που μαθαίνει ο άνθρωπος.

1. Επιβλεπόμενη Μάθηση (Supervised Learning) είναι η διαδικασία που ο αλγόριθμος κατασκευάζει μια συνάρτηση που απεικονίζει δεδομένες εισόδους (σύνολο εκπαίδευσης) σε γνωστές επιθυμητές εξόδους, με

στόχο τη γενίκευση στην συνάρτησης αυτής και για εισόδους με άγνωστη έξοδο. Χρησιμοποιείται σε προβλήματα:

- ο Ταξινόμησης (Classification)
 - ο Πρόγνωσης (Prediction)
 - ο Διερμηνείας (Interpretation)
2. Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning) είναι η διαδικασία που ο αλγόριθμος κατασκευάζει ένα μοντέλο για κάποιο σύνολο εισόδων υπό την μορφή παρατηρήσεων χωρίς να γνωρίζει τις επιθυμητές εξόδους. Χρησιμοποιείται σε προβλήματα
- ο Ανάλυσης Συσχετισμών (Association Analysis) :
 - ο Μείωση της διάστασης των δεδομένων (Dimensionality reduction)
 - ο Ομαδοποίησης (Clustering)
3. Ενισχυτική μάθηση (Reinforcement Learning) ο αλγόριθμος μαθαίνει μια στρατηγική ενεργειών μέσα από άμεση αλληλεπίδραση με το περιβάλλον. Χρησιμοποιείται σε προβλήματα Σχεδιασμού (Planning) όπως ο έλεγχος ρομπότ και η βελτιστοποίηση εργασιών σε εργοστασιακούς χώρους.

3.7 Συσταδοποίηση (Clustering)

Η συσταδοποίηση είναι μια ευρέως διαδεδομένη διαδικασία που χρησιμοποιείται για τη στατιστική ανάλυση δεδομένων σε πολλούς τομείς, όπως η Μηχανική Μάθηση, η εξόρυξη δεδομένων και η ανάλυση εικόνας. Ο στόχος της συσταδοποίησης είναι να διαχωρίσει ένα σύνολο δεδομένων σε ομάδες (συστάδες) με κοινά χαρακτηριστικά, δηλαδή τα δεδομένα κάθε ομάδας είναι όμοια μεταξύ τους και διαφορετικά από τις άλλες ομάδες.

Η συσταδοποίηση ανήκει στην κατηγορία της Μη Επιβλεπόμενης Μάθησης, καθώς δεν απαιτεί ετικέτες ή κλάσεις για τα δεδομένα. Αντίθετα, αναζητά ομοιότητες και μοτίβα μεταξύ τους. Οι τεχνικές συσταδοποίησης χωρίζονται συνήθως σε:

1. Ιεραρχική συσταδοποίηση (Hierarchical Clustering): Προσεγγίζει τη συσταδοποίηση με ιεραρχικό τρόπο, δημιουργώντας ένα δέντρο συστάδων (δενδρογράμμο). Μπορεί να αρχίζει με ένα δέντρο που περιέχει όλα τα δεδομένα και στη συνέχεια διαχωρίζεται ή άνω-προς-κάτω δηλαδή αρχίζει με ξεχωριστές συστάδες και στη συνέχεια συνδέονται. Η επιλογή του σημείου διαχωρισμού γίνεται με βάση την απόσταση μεταξύ των στοιχείων των συστάδων.
2. Διαμεριστική συσταδοποίηση (Distribution-based Clustering): Βασίζεται στην πιθανό-θεωρητική μοντελοποίηση των δεδομένων και επιχειρεί να περιγράψει την κατανομή τους. Στόχος είναι να βρεθούν δεσμευμένα (συστάδες) που αντιστοιχούν σε πιθανο-θεωρητικές κατανομές.

3. Συσταδοποίηση που βασίζεται στην πυκνότητα (Density-based Clustering): Εντοπίζει περιοχές υψηλής πυκνότητας στα δεδομένα και θεωρεί τα δεδομένα εκεί ως μια συστάδα. Αυτή η τεχνική είναι κατάλληλη για δεδομένα με πολλές συστάδες διαφορετικού μεγέθους και πυκνότητας.
4. Συσταδοποίηση που βασίζεται σε κέντρα (Centroid-based Clustering): Οι συστάδες καθορίζονται από τη θέση ενός κέντρου, που μπορεί να είναι ένα σημείο που εκπροσωπεί την ομάδα ή το κέντρο μάζας των δεδομένων. Η συσταδοποίηση K-means είναι ένα παράδειγμα τέτοιας τεχνικής.

3.8 Προεπεξεργασία Δεδομένων

Η διαδικασία της προεπεξεργασίας των δεδομένων είναι απαραίτητη για την εξασφάλιση της καλής ποιότητας των δεδομένων. Τα βασικότερα προβλήματα που συναντώνται στα δεδομένα και χρειάζεται η προεπεξεργασία είναι:

- Ελλιπείς τιμές
- Θόρυβος στα δεδομένα
- Ασυνέπειες στα δεδομένα

3.8.1 Ελλιπείς Τιμές

Για την αντιμετώπιση του προβλήματος των ελλιπών τιμών ακολουθείται η διαδικασία της αναγνώρισης των γραμμών που περιλαμβάνουν τις ελλιπείς τιμές και στη συνέχεια η διαγραφή ή η συμπλήρωση τους. Για την διαδικασία της συμπλήρωσης των δεδομένων γίνεται χρήση :

- Της καθολικής σταθεράς
- Της μέσης τιμής του χαρακτηριστικού
- Της μέσης τιμής των δειγμάτων της ίδιας κλάσης
- Της πιο πιθανής τιμής που παράγεται με μεθόδους όπως παλινδρόμηση δέντρα απόφασης κ.α.

3.8.2 Θόρυβος στα Δεδομένα

Το πρόβλημα του θορύβου στα δεδομένα λύνεται με την μέθοδο της ενδοχείας (binning) και την μέθοδο της συσταδοποίησης (clustering). Κατά τη διαδικασία της ενδοχείας γίνεται σε πρώτη φάση η ταξινόμηση των

δεδομένων και στην συνέχεια ο διαχωρισμός τους σε δοχεία. Ο διαμερισμός γίνεται με δύο τρόπους

1. Διαμερισμός ίσου πλάτους(απόσταση): Το διάστημα των δεδομένων χωρίζεται σε N διαστήματα ίσου μεγέθους. Το πρόβλημα είναι ότι ο διαμερισμός ίσου πλάτους είναι επιρρεπής σε ακραίες τιμές
2. Διαμερισμός ίσου βάθους(συχνότητα): Το διάστημα των δεδομένων χωρίζεται σε N διαστήματα με τον ίδιο αριθμό δειγμάτων. Σε αυτή την περίπτωση υπάρχει καλύτερη κλιμάκωση των δεδομένων.

Οι μέθοδοι που χρησιμοποιούνται στην ενδοχείωση είναι ίδιοι και στην διακριτοποίηση και είναι :

- Εξομάλυνση με βάση την μέση τιμή του κάθε δοχείου: Αντικατάσταση των τιμών με την μέση τιμή του δοχείου.
- Εξομάλυνση με χρήση του μέσου κάθε δοχείου: Αντικατάσταση των τιμών με την μεσαία τιμή κάθε δοχείου.
- Εξομάλυνση με χρήση των ορίων κάθε δοχείου: Αντικατάσταση των τιμών με την κοντινότερη τιμή των ορίων.

3.8.3 Ασυνεπή Δεδομένα

Ασυνέπεια δεδομένων σημαίνει ότι δεδομένα από διαφορετικές πηγές έχουν διαφορετικές εκδόσεις αποθηκευμένων δεδομένων. Η επίλυση του προβλήματος επιτυγχάνεται είτε με χειρωνακτική διόρθωση είτε με ημιαυτόματα, δηλαδή με εργαλεία καθαρισμού δεδομένων (data scrubbing tools) ή εργαλεία λογιστικού ελέγχου(data auditing tools).

3.8.4 Ενοποίηση Των Δεδομένων

Η ενοποίηση είναι η διαδικασία συλλογής των δεδομένων από πολλαπλές πηγές με σκοπό την δημιουργία μιας ολοκληρωμένης έκδοσης. Κατά την διαδικασία της ενοποίησης εντοπίζονται πιθανές συγκρούσεις ή ασυνέπειες στις τιμές των δεδομένων και αναλύονται. Επίσης πλεονάσματα στα δεδομένα εντοπίζονται με την ανάλυση συσχετίσεων. Η ενοποίηση σαν διαδικασία βελτιώνει την ταχύτητα της εξόρυξης των δεδομένων και βελτιώνει την ποιότητα των αποτελεσμάτων.

3.8.5 Μετασχηματισμός Δεδομένων

Ο μετασχηματισμός των δεδομένων έχει ως στόχο την σύγκριση δεδομένων που αρχικά δεν είναι συγκρίσιμα. Έχει ως αποτέλεσμα τη βελτίωση της ακρίβειας των αποτελεσμάτων των αλγορίθμων εξόρυξης ή τη μείωση του όγκου των δεδομένων. Γίνεται χρήση του μετασχηματισμού για τους παρακάτω λόγους:

- Ομαλοποίηση των δεδομένων και εξάλειψη του θορύβου
- Κανονικοποίηση των δεδομένων, δηλαδή την μετατροπή του συνόλου σε ένα συγκεκριμένο εύρος τιμών.
- Την δημιουργία νέων χαρακτηριστικών από τα ήδη υπάρχοντα

Η κανονικοποίηση είναι απαραίτητη για την αντιμετώπιση προβλημάτων που προκύπτουν κατά την εξόρυξη δεδομένων. Χρησιμοποιείται ώστε οι τιμές να ανήκουν σε ένα συγκεκριμένο εύρος τιμών. Οι πιο διαδεδομένες μέθοδοι κανονικοποίησης συνοψίζονται παρακάτω:

- Κανονικοποίηση min-max: Σκοπός της κανονικοποίησης min-max είναι οι τιμές να ανήκουν σε ένα νέο περιορισμένο εύρος. Γίνεται χρήση του τύπου:

$$V_{new} = \frac{V - \min}{\max - \min} (\max_{new} - \min_{new}) + \min_{new}$$

- Κανονικοποίηση z-score: Γίνεται χρήση της μέσης τιμής και της τυπικής απόκλισης ώστε τα δεδομένα να έχουν μέση τιμή 0 και τυπική απόκλιση 1. Ο τύπος που χρησιμοποιείται είναι :

$$V_{new} = \frac{v - \mu}{\sigma}$$

Όπου,

- ο μ: μέση τιμή του χαρακτηριστικού
- ο σ: τυπική απόκλιση

3.9 Νευρωνικά Δίκτυα

Τα Τεχνητά Νευρωνικά Δίκτυα – ΤΝΔ (Artificial Neural Networks-ANN) αποτελούνται από υπολογιστικούς κόμβους, νευρώνες που συνδέονται μεταξύ τους με σκοπό να εντοπίσει τις σχέσεις που υπάρχουν στα δεδομένα εισόδου. Μιμείται την λειτουργία του κεντρικού νευρικού συστήματος και του ανθρώπινου εγκεφάλου. Τα Τεχνητά Νευρωνικά Δίκτυα αποτελούνται από πολλές εισόδους x_i και μόνο μια έξοδο y . Κάθε είσοδος x_i συνδέεται με ένα βάρος w_i και τα αποτελέσματα αθροίζονται μέσω τη συνάρτησης αθροίσματος.

$$F = \sum_i^n x_i w_i$$

Ο τεχνητός νευρώνας δίνει έξοδο μέσω της συνάρτησης μετάβασης (transfer function) όταν το ζυγισμένο άθροισμα των εισόδων είναι μεγαλύτερο μιας ορισμένης τιμής κατωφλίου (threshold value) θ .

$$\sum_i^n x_i w_i - \theta > 0$$

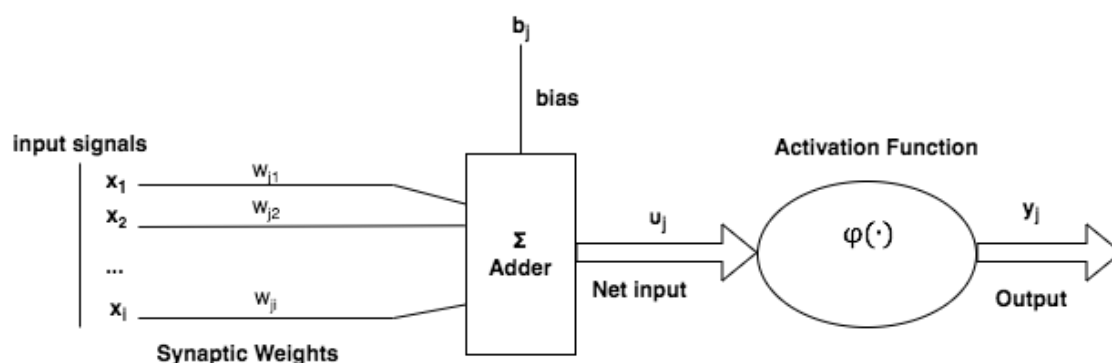
Ο στοιχειώδης Perceptron (basic Perceptron) είναι ένα από τα πιο απλά ΤΝΔ και αποτελείται από μόνο ένα νευρώνα. Η έξοδος για ένα διάνυσμα εισόδου $x = (x_1, x_2, x_3, \dots, x_n)$ δίνεται μέσω της συνάρτησης g ως:

$$\alpha = g\left(\sum_i^n x_i w_i\right)$$

3.9.1 Δομή Τεχνητού Νευρωνικού Δικτύου

Η δομή ενός τεχνητού νευρωνικού δικτύου (ΤΝΔ) βασίζεται στους τεχνητούς νευρώνες που αποτελούν τις βασικές μονάδες επεξεργασίας. Κάθε νευρώνας λαμβάνει είσοδο από άλλους νευρώνες ή από την αρχική είσοδο του δικτύου. Το βασικό μοντέλο του νευρώνα που χρησιμοποιείται συχνά σε τεχνητά νευρωνικά δίκτυα παρουσιάζεται στο σχήμα 3-1.

Το βασικό μοντέλο του νευρώνα που χρησιμοποιείται κυρίως σε υλοποιήσεις τεχνητών νευρωνικών δικτύων φαίνεται στο σχήμα 3-1.



Σχήμα 3-1: Βασικό Μοντέλο Νευρώνα

Πηγή: Li, S. (n.d.). *Illustration of neuron structure of the neural network*. From https://researchgate.net/figure/Illustration-of-neuron-structure-of-the-neural-network_fig2_329990010

Αρχικά, κάθε στοιχείο του εισαγόμενου διανύσματος πολλαπλασιάζεται με το αντίστοιχο βάρος. Στη συνέχεια, οι γινόμενες εισόδους συνυπολογίζονται μαζί με έναν εξωτερικό παράγοντα, τη μεροληψία ή κατώφλι (bias, threshold), και παράγεται το τοπικό πεδίο (net input, induced local field, activation potential).

Μέχρι αυτό το σημείο, ο νευρώνας παρέχει μια γραμμική συνδυαστική λειτουργία των εισόδων με βάρη προσαρμοσμένα στο δίκτυο. Πριν φτάσει στη συνάρτηση ενεργοποίησης, ο νευρώνας χαρακτηρίζεται ως γραμμικός και μπορεί να θεωρηθεί ως ένα γραμμικό προσαρμοσμένο φίλτρο (linear adaptive filter). Ένα νευρωνικό δίκτυο που αποτελείται μόνο από γραμμικούς νευρώνες χαρακτηρίζεται ως γραμμικό. Στη συνέχεια, εφαρμόζεται η συνάρτηση ενεργοποίησης ή συνάρτηση μεταφοράς (activation function ή squashing function), και παράγεται το τελικό αποτέλεσμα στην έξοδο του νευρώνα.

Αυτή είναι η βασική δομή ενός τεχνητού νευρωνικού δικτύου, και η επεξεργασία γίνεται μέσω αλληλεπίδρασης των νευρώνων και των βαρών τους. Με την εκπαίδευση του δικτύου, τα βάρη προσαρμόζονται έτσι ώστε να επιτυγχάνεται η επιθυμητή εξόδος για κάθε είσοδο.

3.9.2 Συνάρτηση Ενεργοποίησης

Η συνάρτηση ενεργοποίησης είναι βασικό κομμάτι για την επεξεργασία του σήματος. Χρησιμοποιούνται διάφορες συναρτήσεις ενεργοποίησης και κάθε μία εξάγει διαφορετική έξοδο. Η φύση της συνάρτησης ενεργοποίησης έχει επιρροή στην ταχύτητα εκμάθησης του μοντέλου. Μερικές από τις πιο διαδεδομένες είναι οι παρακάτω:

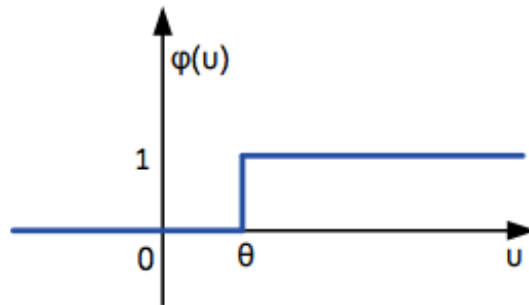
Όσον αφορά τις γραμμικές συναρτήσεις ενεργοποίησης:

- Βηματική συνάρτηση: Η βηματική συνάρτηση παίρνει δύο τιμές και θεωρείται ότι είναι η συνάρτηση ενεργοποίησης στους βιολογικούς νευρώνες και αυτή που χρησιμοποιήθηκε στο πρώτο μοντέλο τεχνητού νευρώνα. Στην περίπτωση που ο νευρώνας χρησιμοποιεί συνάρτηση κατωφλίου δηλαδή:

$$v_k = \sum_{i=0}^n x_i w_i - \theta_k$$

Ακολουθεί το μοντέλο McCulloch – Pitts:

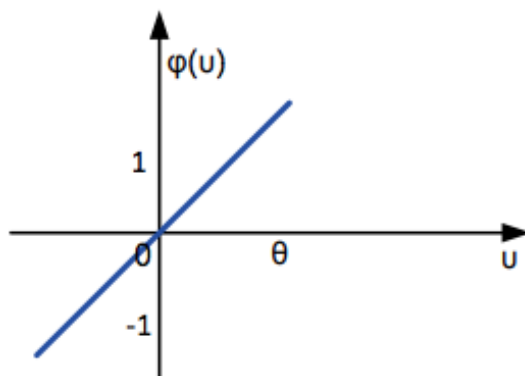
$$\varphi(v) = \begin{cases} 1, & v \geq \theta \\ 0, & v < \theta \end{cases}$$



Σχήμα 3-2: Απεικόνιση Βηματικής Συνάρτησης

Πηγή: Κυρίτσης Κ. (2014,Μάιος). *Νευρωνικά Δίκτυα και Μηχανές Διανυσματικής Υποστήριξης* (dissertation).

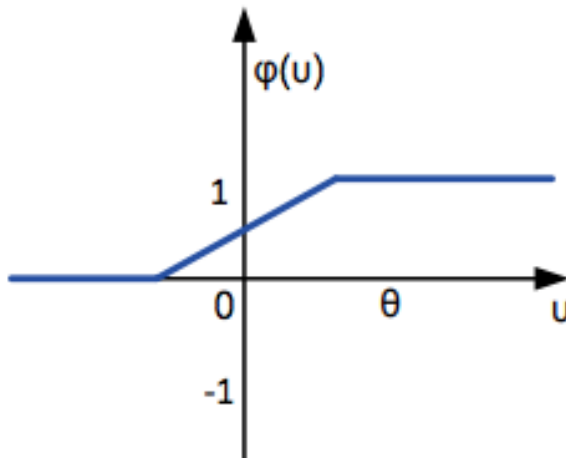
- Γραμμική συνάρτηση: Υπάρχουν φορές που χρησιμοποιείται ως συνάρτηση ενεργοποίησης η γραμμική συνάρτηση $\varphi(v) = v$ όταν δεν υπάρχει κάποια μεταβολή από την είσοδο του νευρώνα στην έξοδο του.



Σχήμα 3-3: Απεικόνιση Γραμμικής Συνάρτησης

(Πηγή: Κυρίτσης Κ. (2014,Μάιος). *Νευρωνικά Δίκτυα και Μηχανές Διανυσματικής Υποστήριξης* (dissertation).)

- Τμηματικά γραμμική συνάρτηση: Η βηματική συνάρτηση επειδή η πρώτη παράγωγος της είναι ίση με μηδέν δεν προτιμάται σαν συνάρτηση ενεργοποίησης στα Τεχνητά Νευρωνικά Δίκτυα. Η μάθηση στα νευρωνικά δίκτυα είναι η μεταβολή των τιμών των βαρών και στα μαθηματικά, η μεταβολή σημαίνει παράγωγος οπότε είναι λογικό να μην προτιμάται η γραμμική συνάρτηση ενεργοποίησης.



Σχήμα 3-4: Απεικόνιση Τμηματικά Γραμμικής Συνάρτησης

Πηγή: Κυρίτσης Κ. (2014,Μάιος). *Νευρωνικά Δίκτυα και Μηχανές Διανυσματικής Υποστήριξης* (dissertation).

Όσο αφορά τις μη- γραμμικές συναρτήσεις ενεργοποίησης:

- Σιγμοειδής Συνάρτηση (Sigmoid Function): Η σιγμοειδής συνάρτηση είναι μη γραμμική συνάρτηση και η γραφική παράστασή της είναι σχήματος S-καμπύλης. Ο τύπος είναι της μορφής:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

Όπου,

- x_0 : το κεντρικό σημείο της καμπύλης
- L : η μέγιστη τιμή της καμπύλης και
- k : η παράμετρος ρυθμού ανάπτυξης της καμπύλης.

Το πεδίο ορισμού της είναι όλο το σύνολο των πραγματικών αριθμών και το σύνολο τιμών της το σύνολο $(0, L)$. Ανάλογα με τις τιμές των παραμέτρων της

σιγμοειδής συνάρτησης παίρνει και διάφορες μορφές. Οι πιο χαρακτηριστικές είναι η λογιστική συνάρτηση (Logistic Function) και η συνάρτηση υπερβολικής εφαπτομένης (Hyperbolic Tangent).

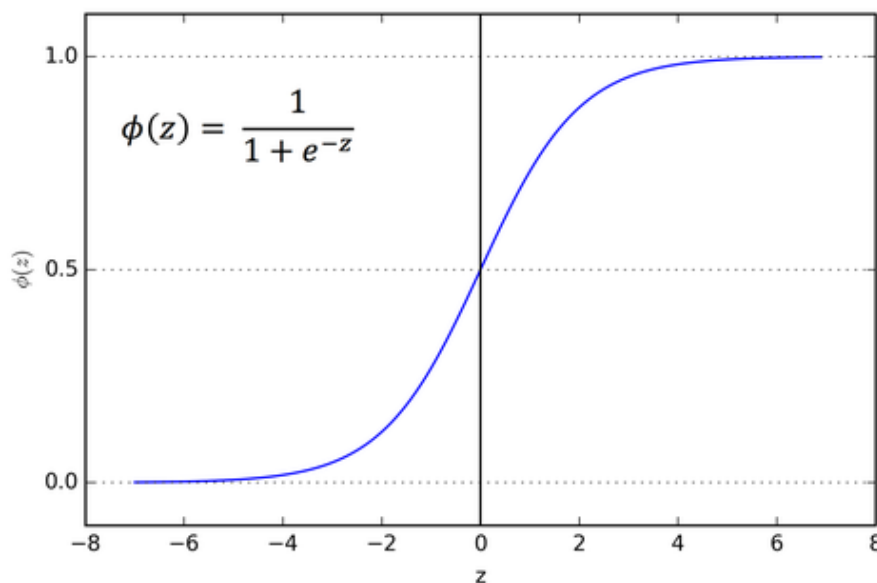
- Λογιστική Συνάρτηση (Logistic Function):

Για $L=1$, $k=1$ και $\mathbf{x}_0=0$ ο γενικός τύπος της σιγμοειδής συνάρτησης γίνεται

$$f(x) = \frac{1}{1 + e^{-x}}$$

Η Λογιστική συνάρτηση όπως και η Σιγμοειδής είναι μη γραμμική. Παρόμοια με την Σιγμοειδή το πεδίο ορισμού είναι το σύνολο των πραγματικών αριθμών και το σύνολο τιμών της είναι το $(0,1)$ αφού το $L=1$.

Η Λογιστική συνάρτηση προτιμάται ως συνάρτηση ενεργοποίησης διότι παράγει κανονικοποιημένες τιμές, δηλαδή τιμές ανάμεσα στο $(0,1)$. Όμως τα αρνητικά που την χαρακτηρίζουν είναι ότι είναι υπολογιστικά σύνθετη και ότι δεν είναι συμμετρική κοντά στο μηδέν που σημαίνει ότι οι τιμές των εξόδων των νευρώνων που παράγονται είναι ίδιες. [16]



3-5: Απεικόνιση Λογιστικής Συνάρτησης

Πηγή: Hamhochoi. (2020, April 24). *Batch normalization*. Hamhochoi. From <https://dothanhblog.wordpress.com/2019/12/29/batch-normalization/>

- Συνάρτηση Υπερβολικής Εφαπτομένης (Hyperbolic Tangent, tanh) :

$$\varphi(u) = \tanh(a \cdot u)$$

Η συνάρτηση Υπερβολικής Εφαπτομένης χρησιμοποιείται σε προβλήματα μάθησης γιατί μπορεί να δέχεται δεδομένα με αρνητικές τιμές. Το πεδίο τιμών είναι $(-1, 1)$. Όσο αυξάνεται το a τόσο η κλίση γίνεται μεγαλύτερη. Σε μεγάλες τιμές η γραφική παράσταση παίρνει τη μορφή της βηματικής συνάρτησης και της συνάρτησης προσήμου αλλά έχουν την ιδιότητα της συνέχειας και της παραγωγισιμότητας. Οι περισσότεροι αλγόριθμοι εκπαίδευσης Τεχνητών Νευρωνικών Δικτύων χρειάζονται τις μερικές παραγώγους των εξόδων των νευρώνων ως προς τα βάρη.

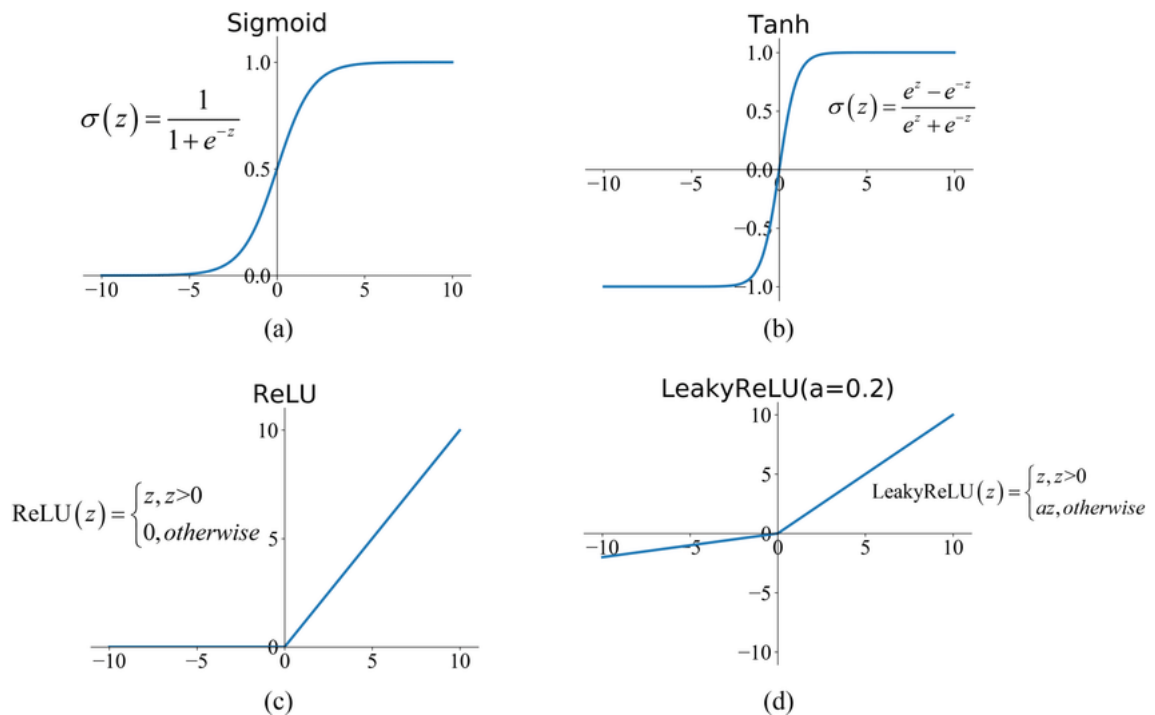
- Συνάρτηση Ράμπας ή Ανορθωμένη Γραμμική Μονάδα (Rectified Linear Unit, ReLU)

$$\sigma(z) = \max(0, z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

Η συνάρτηση ράμπας ή Ανορθωμένη Γραμμική Μονάδα χρησιμοποιείται πολύ στα Τεχνητά Νευρωνικά Δίκτυα καθώς επιταχύνει πολύ την διαδικασία της εκπαίδευσής. Λόγω της φύσης της συνάρτησης, ένα σημαντικό μειονέκτημα είναι ότι οι νευρώνες, των οποίων οι εισοδοί έχουν αρνητικές τιμές δίνουν ως έξοδο μηδέν. Αυτό μπορεί να οδηγήσει μεγάλο μέρος του δικτύου να μείνει ανενεργό μειώνοντας έτσι σημαντικά την αποτελεσματικότητά του.

- Διαρρέουσα Ανορθωμένη Γραμμική Μονάδα (Leaky ReLU & Parametric ReLU PReLU): Λόγω του προβλήματος που έχει η ReLU δημιουργήθηκε η Διαρρέουσα Ανορθωμένη Γραμμική Μονάδα. Το μέρος της συνάρτησης που γινόταν μηδέν αντικαταστάθηκε με μια μικρή θετική κλίση στα αρνητικά, αποτρέποντας έτσι την απενεργοποίηση των νευρώνων.

$$\sigma(z) = \max(0.01z, z) = \begin{cases} 0.01, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

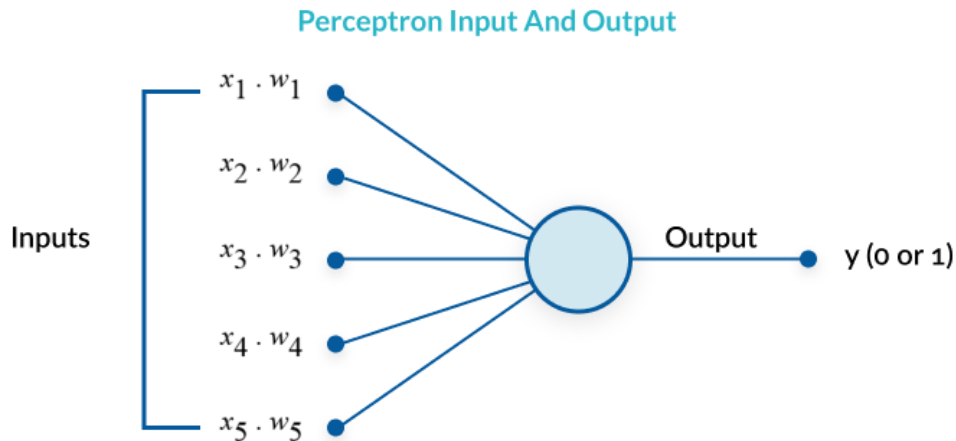


Σχήμα 3-6: Μη Γραμμικές Συναρτήσεις Ενεργοποίησης

Πηγή: Feng, J. (n.d.). *Commonly used activation functions: (A) sigmoid, (b) tanh, (c) relu ... researchgate*. Retrieved from https://www.researchgate.net/figure/Commonly-used-activation-functions-a-Sigmoid-b-Tanh-c-ReLU-and-d-LReLU_fig3_335845675

3.9.3 Αισθητήρας (Perceptron)

Ένα από τα πρώτα τεχνητά νευρωνικά δίκτυα που αναπτύχθηκαν είναι το μοντέλο του αισθητήρα, που είναι γνωστό και ως Perceptron. Διατυπώθηκε για πρώτη φορά από τον Rosenblatt (1958) και θεωρείται το πιο απλό τεχνητό νευρωνικό δίκτυο αφού αποτελείται από δύο επίπεδα. Το πρώτο επίπεδο είναι οι είσοδοι στο δίκτυο και δεν γίνεται κάποια επεξεργασία διότι δεν έχει νευρώνες. Το δεύτερο επίπεδο αποτελείται από υπολογιστικούς νευρώνες και αποτελούν το επίπεδο εξόδου του δικτύου.



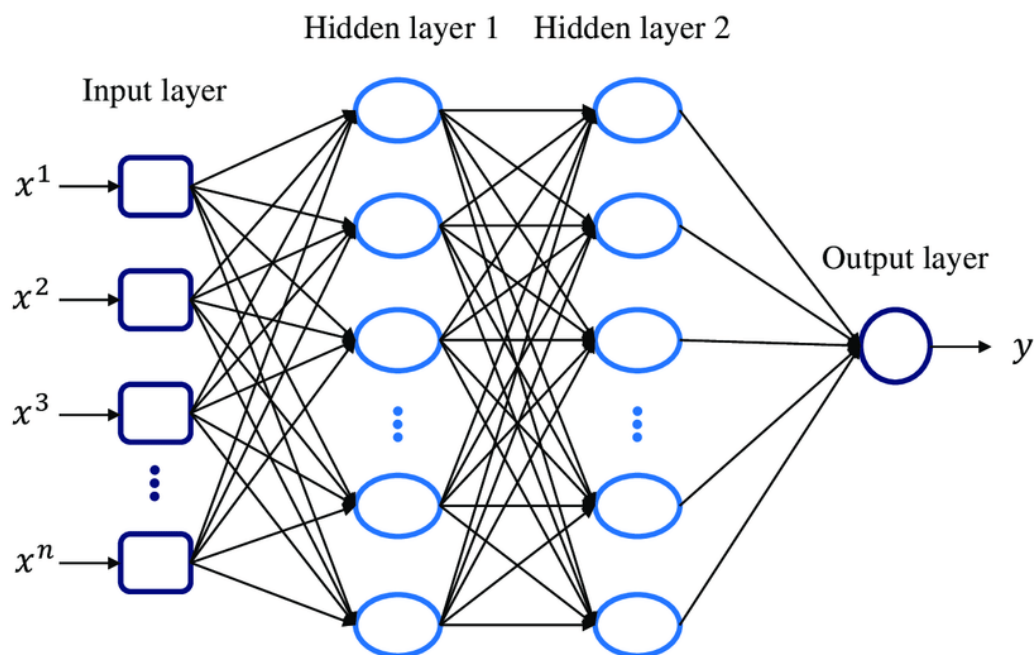
Σχήμα 3-7: Μοντέλο του Αισθητήρα, Είσοδος – Έξοδος

Πηγή : *Perceptrons and multi-layer perceptrons*. IndianTechWarrior. (2022, June 5). Retrieved from <https://indiantechwarrior.com/perceptrons-and-multi-layer-perceptrons/>

3.9.4 Πολύ-Επίπεδοι Αισθητήρες

Οι πολυεπίπεδοι αισθητήρες (MultiLayer Perceptron (MLP)) αναπτύχθηκαν λόγω της αδυναμίας του απλού αισθητήρα να λύσει απλά μη γραμμικά προβλήματα. Τα πολυεπίπεδα δίκτυα Perceptron είναι δίκτυα πρόσθιας τροφοδότησης (Feed-Forward Neural Networks). Οι πολυεπίπεδοι αισθητήρες βασίζονται στην δομή και την λειτουργία του επίπεδου αισθητήρα και έχουν την προσθήκη κρυφών επιπέδων ανάμεσα στην είσοδο και έξοδο. Τα κρυφά επίπεδα δέχονται πληροφορία από την είσοδο σε μορφή διανύσματος. Αποτελούνται από υπολογιστικούς νευρώνες στους οποίους γίνεται η επεξεργασία, ίδια με την επεξεργασία στον απλό αισθητήρα. Η έξοδος ενός κρυφού επιπέδου γίνεται είσοδος στο επόμενο κρυφό επίπεδο και επαναλαμβάνεται η ίδια διαδικασία μέχρι την τελική έξοδο.

Όσο αφορά τις συναρτήσεις ενεργοποίησης τα πολυεπίπεδα δίκτυα MLP χρησιμοποιούν μη γραμμικές συναρτήσεις, οι οποίες είναι συνεχείς και παραγωγίσιμες στο πεδίο ορισμού τους. Διαφορά με τα μονοεπίπεδα δίκτυα αποτελεί ότι τα μονοεπίπεδα είναι ικανά να μαθαίνουν αποκλειστικά γραμμικός διαχωρίσιμα προβλήματα. Με την ανακάλυψη των πολυεπίπεδων δικτύων πλέον μπορούν να αντιμετωπίσουν πιο σύνθετα μη-γραμμικώς διαχωρίσιμα προβλήματα.



Σχήμα 3-8: Πολυεπίπεδοι Αισθητήρες

Πηγή: *Multilayer perceptron (MLP) architecture with two hidden layers and one prediction output*. Ανακτήθηκε από

https://www.researchgate.net/figure/Multilayer-perceptron-MLP-architecture-with-two-hidden-layers-and-one-prediction-output_fig1_349630467

3.9.5 Αρχιτεκτονικές Τεχνητών Νευρωνικών Δικτύων

Τα τεχνητά νευρωνικά δίκτυα δημιουργήθηκαν με στόχο την προσομοίωση της λειτουργίας του ανθρώπινου εγκεφάλου. Υπάρχουν δυο βασικές κατηγορίες τεχνητών νευρωνικών δικτύων ανάλογα με το πως συνδέονται οι κόμβοι μεταξύ τους:

- Πρόσθιας Τροφοδότησης (Feed Forward)
- Οπίσθιας Τροφοδότησης (Feed Backward ή Recurrent)

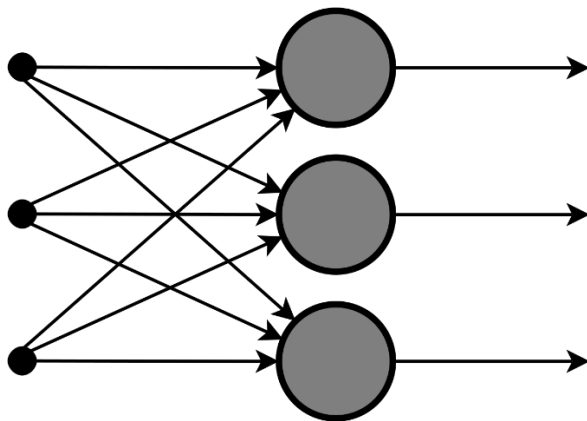
Όσο αφορά τα νευρωνικά δίκτυα πρόσθιας τροφοδότησης οι κόμβοι δίνουν είσοδο σε κόμβους μόνο επόμενου επιπέδου. Δηλαδή δεν υπάρχει έξοδος κάποιου κόμβου που αποτελεί είσοδο ίδιου ή προηγούμενου επιπέδου. Δίκτυα που η έξοδος κάποιου κόμβου που αποτελεί είσοδο ίδιου ή προηγούμενου επιπέδου ονομάζονται οπισθοδιάδοσης (backpropagation).

Τα νευρωνικά δίκτυα οπίσθιας τροφοδότησης ονομάζονται και ανατροφοδοτούμενα ΤΝΔ (Recurrent ANN) και έχουν την δυνατότητα οι κόμβοι να τροφοδοτούν με την έξοδο τους και κόμβους ίδιου ή προηγούμενων

επιπέδων. Τα δίκτυα που η ανατροφοδότηση αφορά κόμβους στο ίδιο επίπεδο ονομάζονται αυτοσυσχετιζόμενες μνήμες (autoassociated memories) αλλιώς καλούνται ετεροσυσχετιζόμενες μνήμες (heteroassociated memories).

3.9.6 Δίκτυα Πρόσθιας Τροφοδότησης Ενός Επιπέδου

Τα δίκτυα πρόσθιας τροφοδότησης ενός επιπέδου (One Layer Feed-forward Neural Networks) αποτελούνται από δύο επίπεδα. Η δομή τους είναι η πιο απλή δομή ενός νευρωνικού δικτύου καθώς έχει το επίπεδο εισόδου και το επίπεδο εξόδου. Το επίπεδο εισόδου αποτελείται από νευρώνες εισόδου, οι οποίοι συλλέγουν και μεταφέρουν τα δεδομένα εισόδου, δεν κάνουν κάποιο υπολογισμό. Στο επίπεδο εξόδου υπάρχουν υπολογιστικοί νευρώνες και υπολογίζουν την έξοδο του δικτύου.



Σχήμα 3-9: Απεικόνιση Δικτύου Πρόσθιας Τροφοδότησης Ενός Επιπέδου

Πηγή: *Single Layer Feed Forward Neural Network, HD png download, Transparent png image*. Ανακτήθηκε στις November 8, 2022, από https://www.pngitem.com/middle/hbbhobT_single-layer-feed-forward-neural-network-hd-png/

3.9.7 Πολυεπίπεδα Δίκτυα Πρόσθιας Τροφοδότησης (Multilayer Feed-Forward Neural Networks)

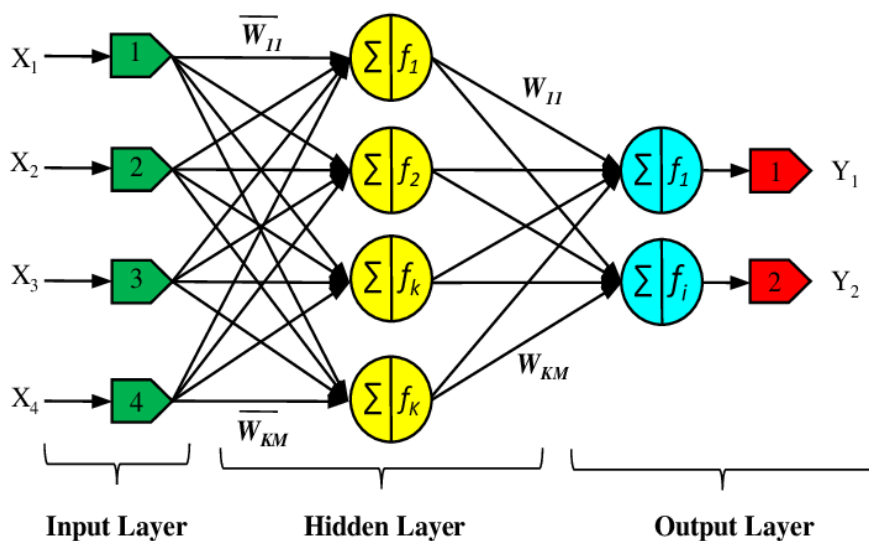
Τα πολυεπίπεδα δίκτυα πρόσθιας τροφοδότησης (Multilayer Feed-forward Neural Networks) είναι μια εξέλιξη των δικτύων πρόσθιας τροφοδότησης ενός επιπέδου, με την προσθήκη ενός ή περισσοτέρων κρυφών επιπέδων ανάμεσά τους. Αντίθετα με τα δίκτυα ενός επιπέδου που έχουν μόνο ένα επίπεδο κρυφών νευρώνων, τα πολυεπίπεδα δίκτυα έχουν πολλά κρυφά επίπεδα που περιέχουν υπολογιστικούς νευρώνες.

Το επίπεδο εισόδου αποτελείται από μη-υπολογιστικούς νευρώνες οι οποίοι λαμβάνουν και μεταφέρουν απλά την πληροφορία προς τα κρυφά επίπεδα. Τα κρυφά επίπεδα περιέχουν υπολογιστικούς νευρώνες και είναι υπεύθυνα για τον υπολογισμό των ενδιάμεσων αναπαραστάσεων των δεδομένων. Μετά τον

υπολογισμό στα κρυφά επίπεδα, οι τιμές πολλαπλασιάζονται με τα αντίστοιχα βάρη και χρησιμοποιούνται ως είσοδος για τους νευρώνες του επιπέδου εξόδου. Οι νευρώνες του επιπέδου εξόδου υπολογίζουν τις τελικές τιμές του δικτύου.

Η προσθήκη κρυφών επιπέδων επιτρέπει στα πολυεπίπεδα δίκτυα προσθιας τροφοδότησης να υπολογίζουν πιο σύνθετες μη γραμμικές συναρτήσεις και να αντιμετωπίζουν πιο πολύπλοκα δεδομένα. Ο κάθε νευρώνας σε ένα κρυφό επίπεδο υπολογίζει μια γραμμική συνδυαστική σχέση των εισόδων του και των αντίστοιχων βαρών, ενώ η συνάρτηση ενεργοποίησης του επιπέδου του προσδίδει μη γραμμικότητα. Με την χρήση πολλών κρυφών επιπέδων, το δίκτυο μπορεί να εκτελέσει πιο περίπλοκες υπολογιστικές διεργασίες και να αναπαραστήσει πιο πολύπλοκες συναρτήσεις.

Τα πολυεπίπεδα δίκτυα προσθιας τροφοδότησης έχουν ευρεία εφαρμογή σε πολλά προβλήματα μηχανικής μάθησης, όπως η αναγνώριση προτύπων, η εκτίμηση και πρόβλεψη, και η ταξινόμηση δεδομένων.



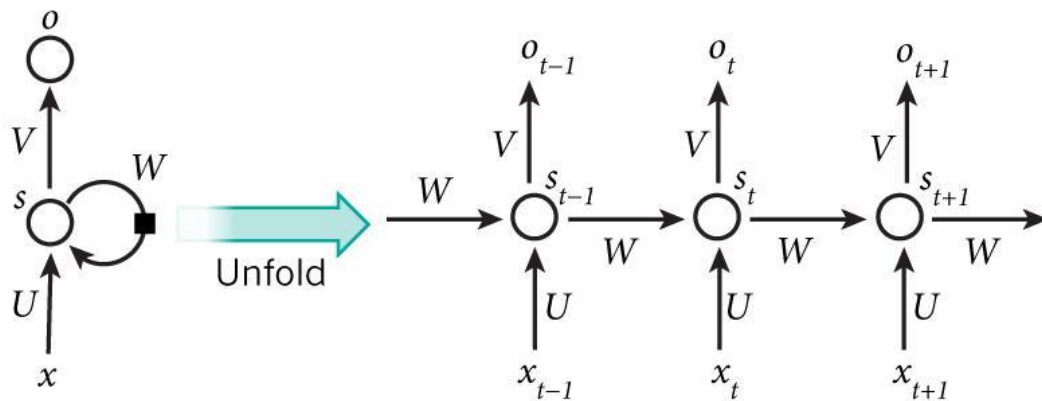
Σχήμα 3-10: Απεικόνιση Πολυεπίπεδου Δικτύου Πρόσθιας Τροφοδότησης

Πηγή: *Structure of the multilayer feed-forward neural network*. (n.d.). Retrieved November 8, 2022, from https://www.researchgate.net/figure/Structure-of-the-multilayer-feed-forward-neural-network_fig2_321281271

3.9.8 Αναδρομικά Νευρωνικά Δίκτυα

Τα Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks) χρησιμοποιούν για την παραγωγή μιας εξόδου πέρα από τα δεδομένα εισόδου και τις προηγούμενες εξόδους. Τα Αναδρομικά Νευρωνικά Δίκτυα (RNN) δεν έχουν νευρώνες προς μόνο μια κατεύθυνση όπως τα Πολυεπίπεδα Δίκτυα Πρόσθιας

Τροφοδότησης αλλά μπορούν να μεταφέρουν πληροφορία και προς το προηγούμενο και το ίδιο επίπεδο. Λόγο της ικανότητας επικοινωνίας ανάμεσα σε προηγούμενα και ίδια επίπεδα τα Αναδρομικά Νευρωνικά Δίκτυα έχουν την ικανότητα αντίληψης του χρόνου. Μπορούν δηλαδή να προβλέψουν την έξοδο της $t+1$ στιγμής συναρτήσει των δεδομένων των στιγμών t και $t-1$.



3-11 : Απεικόνιση Αναδρομικού Νευρωνικού Δικτύου

Πηγή: RNN. RNN - programmer sought. (n.d.). Retrieved from <https://programmersought.com/article/4868807525/>

Κεφάλαιο 4ο: Η Πολυκριτήρια Μέθοδος UTASTAR και το Νευρωνικό Δίκτυο SOM

4.1 Εισαγωγή

Η πολυκριτήρια μέθοδος UTASTAR αναλύεται λεπτομερώς, παρέχοντας μια κατανόηση των βασικών αρχών και των σταδίων της μεθόδου. Επιπλέον, αναφέρονται οι τεχνικές που χρησιμοποιούνται για την αξιολόγηση των κριτηρίων και την καθορισμό των βαρών τους.

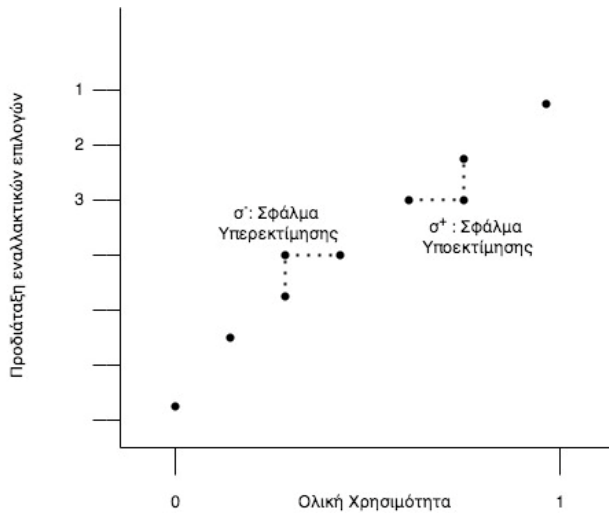
Στη συνέχεια, παρουσιάζεται το νευρωνικό δίκτυο αυτοοργανωμένων χαρτών (SOM), το οποίο αποτελεί ένα είδος τεχνητού νευρωνικού δικτύου. Εξηγείται η αρχιτεκτονική και η λειτουργία του SOM, καθώς και ο τρόπος με τον οποίο εφαρμόζεται για την ανάλυση και την ομαδοποίηση δεδομένων.

4.2 Πολυκριτήρια Μέθοδος - UTASTAR

Η μέθοδος UTASTAR προτάθηκε από τους Siskos και Yiannacopoulos (1985) αποτελεί μέθοδο πολυκριτήριας ανάλυσης αποφάσεων και είναι μια βελτίωση της μεθόδου UTA των Jacquet-lagrece and Siskos (1982) (Siskos et al., 2016).

Η μέθοδος UTA όπως και η μέθοδος UTASTAR έχει ως στόχο την ανάλυση των προτιμήσεων του αποφασίζοντα. Σκοπός των μεθόδων είναι το μοντέλο να ανταποκρίνεται όσο το δυνατόν περισσότερο στην αρχική προδιάταξη του αποφασίζοντα. Υπολογίζονται δηλαδή οι μερικές χρησιμότητες συναρτήσεων κάποιων πολυκριτηριακών εκτιμήσεων και μιας διάταξης εναλλακτικών επιλογών ώστε το μοντέλο να ανταποκρίνεται όσο περισσότερο γίνεται στην πραγματικότητα.

Η διαφορά της UTA με την UTASTAR βρίσκεται στην συνάρτηση σφάλματος. Στην UTA υπήρχε μια μοναδική συνάρτηση σφάλματος $\sigma: A \rightarrow [0,1]$, όπου $\sigma(a)$ εκφράζει την διαφορά της τιμής της χρησιμότητας της αρχικής προδιάταξης και της υπολογισμένης χρησιμότητας $u[g(a)]$ της εναλλακτικής επιλογής a . Όμως προκύπτει το πρόβλημα ότι η συνάρτηση αυτή δεν επαρκεί για την ελαχιστοποίηση της διασποράς όλων των σημείων γύρω από την μονότονη καμπύλη. Έτσι δημιουργήθηκε η μέθοδος UTASTAR για να δώσει λύση στο πρόβλημα των σημείων που βρίσκονται δεξιά της καμπύλης και απαιτείται η αφαίρεση ενός ποσού χρησιμότητας, όπως φαίνεται στο διάγραμμα 4-1.



Διάγραμμα 4-1 : Χρησιμότητα και προδιάταξη στην ανάλυση παλινδρόμησης
Πηγή: [Ματσατσίνης, Ν. (2010)]

4.2.1 Αλγόριθμος Utastar

Ο αλγόριθμος πολυκριτήριας ανάλυσης UTASTAR αποτελείται από τα παρακάτω βήματα :

1. Με βάση την αρχική προδιάταξη του αποφασίζονται εκφράζονται οι ολικές χρησιμότητες των εναλλακτικών επιλογών $u(g(a))$ με $a \in A$ με όρους των προσθετικών μερικών χρησιμοτήτων $u_i(g_i^j)$

Όπου:

- $A=\{a,b,c,\dots\}$ ένα σύνολο εναλλακτικών επιλογών πάνω σε μια δεδομένη δομή προτιμήσεων

και στη συνέχεια εκφράζονται τα w_{ij} όπου λαμβάνονται υπόψη οι περιορισμοί μονοτονικότητας των κριτηρίων :

$$w_{ij} = u_i(g_i^{j+1}) - u_i(g_i^j) \geq 0 \quad \forall i \text{ και } j$$

Όπου:

- w_{ij} εκφράζουν τις αποστάσεις μεταξύ των επιπέδων αξιολόγησης των κριτηρίων
- $i = 1, 2, \dots, n$ ο αύξων αριθμός κάθε κριτηρίου
- $j = 2, 3, \dots, a_i - 1$ ο αύξων αριθμός υποδιαστήματος του i κριτηρίου

2. Γράφεται κάθε ζεύγος από την κορυφή προς την βάση της προδιάταξης ως εξής:

$$\Delta(a_k, a_{k+1}) = U(g(a_k)) - \sigma^+(a_k) + \sigma^-(a_k) - U(g(a_{k+1})) + \sigma^+(a_{k+1}) + \sigma^-(a_{k+1})$$

Ο αριθμός των σχέσεων αυτών ισούται με το πλήθος των εναλλακτικών επιλογών-ενεργειών μείον ένα.

3. Σχηματίζεται το γραμμικό πρόβλημα:

$$[min]z = \sum_{k=1}^m \sigma^+(a_k) + \sigma^-(a_k)$$

Υπό τους περιορισμούς:

$$\Delta(a_k, a_{k+1}) \geq \delta \text{ εαν } a_k > a_{k+1} \forall k$$

$$\Delta(a_k, a_{k+1}) = 0 \text{ εαν } a_k \sim a_{k+1} \forall k$$

$$\sum_{i=1}^n \sum_{j=1}^{a_i-1} w_{ij} = 1$$

$$w_{ij} \geq 0 \quad \sigma^+(a_k) \geq 0, \sigma^-(a_k) \geq 0 \forall i, j, k$$

$$\sum_{k=1}^m (\sigma^+(a_k) + \sigma^-(a_k)) \leq z^* + \varepsilon$$

Όπου,

δ : μια μικρή θετική τιμή

z : μια βέλτιστη τιμή (σφάλμα) του γραμμικού προβλήματος του βήματος 3 και

ε : ένας πολύ μικρός αριθμός ή μηδέν

4. Ελέγχεται η ύπαρξη πολλαπλών βέλτιστων ή ημιβέλτιστων λύσεων (έλεγχος ευστάθειας) στο γραμμικό πρόβλημα. Σε περίπτωση που η λύση δεν είναι μοναδική, υπολογίζεται η μέση συνάρτηση προσθετικής αξίας μεταξύ των κοντινών βέλτιστων λύσεων, οι οποίες μεγιστοποιούν τις ακόλουθες αντικειμενικές συναρτήσεις

$$u_i(g_i^*) = \sum_{j=1}^{a_i-1} w_{ij} \quad \forall i = 1, 2, \dots, n$$

Σε περίπτωση που η πολυκριτήρια εκτίμηση ενός αποφασίζοντα για μια εναλλακτική επιλογή σε ένα κριτήριο, δεν είναι από τις διακριτές τιμές του κριτηρίου αυτού τότε η μερική χρησιμότητα του υπολογίζεται με γραμμική παρεμβολή σύμφωνα με τον ακόλουθο τύπο :

$$u_i(g_i^x) = u_i(g_i^{j-1}) + \frac{g_i^x - g_i^{j-1}}{g_i^j - g_i^{j-1}} [u_i(g_i^j) - u_i(g_i^{j-1})]$$

Το νόημα της προσθετικής συνάρτησης χρησιμότητας είναι ότι η ολική χρησιμότητα μιας απόφασης ισούται με το άθροισμα των μερικών χρησιμοτήτων των κριτηρίων στα οποία εκτιμάται.

Η προσθετική συνάρτηση που χρησιμοποιείται για τον υπολογισμό των ολικών χρησιμοτήτων είναι η εξής:

$$U(g) = \sum_i^n p_i u_i(g_i)$$

- p_1, p_2, \dots, p_n : οι σταθερές που υποδηλώνουν τη σημασία (βάρος) των κριτηρίων και εφαρμόζονται:

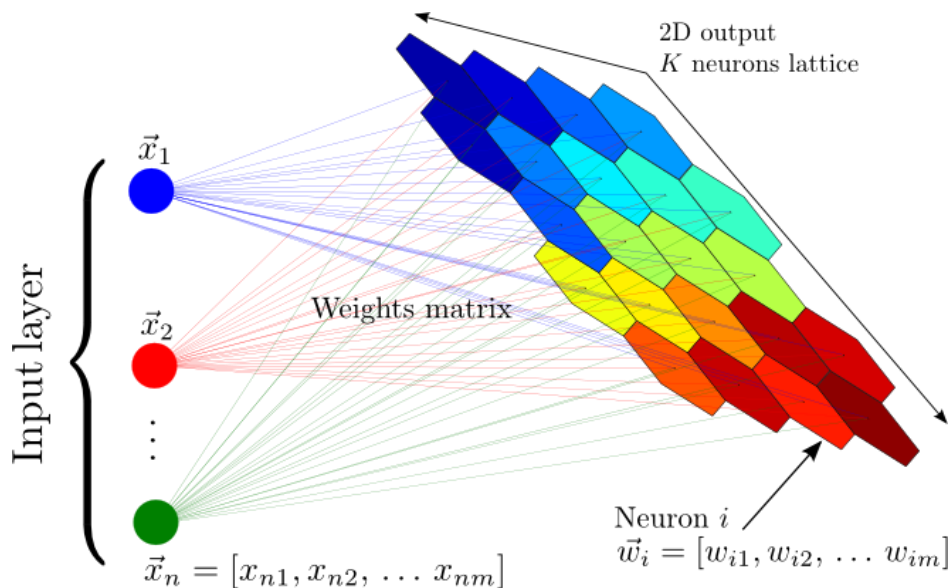
$$\sum_i^n p_i = 1$$

4.3 Νευρωνικά Δίκτυα: Self-Organizing Maps

Οι αυτοοργανωμένοι χάρτες (Self-Organizing Maps) SOM είναι ένα μοντέλο νευρωνικού δικτύου που χρησιμοποιείται κυρίως για συσταδοποίηση και την . Παρουσιάστηκε για πρώτη φορά από τον T. Kohonen (1982) και αναφέρεται επίσης ως χάρτης Kohonen. Ο αλγόριθμος προσφέρει μείωση των διαστάσεων δηλαδή αναπαριστά τον χώρο εισόδου ως ένα χώρο μειωμένης διάστασης συνήθως δυο διαστάσεων, που ονομάζεται χάρτης. Χρησιμοποιεί ανταγωνιστική μάθηση, δηλαδή μια συνάρτηση γειτονιάς για να διατηρήσει τις τοπολογικές ιδιότητες του χώρου εισόδου.

Οι αυτοοργανωμένοι χάρτες αποτελούνται από τους νευρώνες εισόδου, τους υπολογιστικούς νευρώνες και τους νευρώνες εξόδου. Το στρώμα με τους νευρώνες εξόδου μπορεί να αναπαρασταθεί ως ένας δισδιάστατος χάρτης όπως φαίνεται στο σχήμα 4-2. Το δίκτυο οργανώνει τα βάρη του ώστε να αναγνωρίζει όποια κανονικότητα μπορεί να υπάρχει στα διανύσματα εισόδου.

Αντιστοιχίζει την είσοδο με την έξοδο και γι' αυτό αποτελεί μια αυτοοργανωμένη απεικόνιση χαρακτηριστικών.



Σχήμα 4-2 : Απεικόνιση Αυτοοργανωμένου Χάρτη (Self-Organizing Maps)

Πηγή: Caparrini, F. S., & Work, W. W. (n.d.). *Self organizing maps (SOM) in NetLogo*. Self Organizing Maps (SOM) in NetLogo - Fernando Sancho Caparrini. Retrieved November 8, 2022, from <http://www.cs.us.es/~fsancho/?e=136>

4.3.1 Δομή Και Λειτουργία

Οι αυτο-οργανωμένες χάρτες είναι ένα είδος νευρωνικών δικτύων που χρησιμοποιούνται για την ανάλυση και οργάνωση πολυδιάστατων δεδομένων. Αυτά τα δίκτυα επιτρέπουν την αναπαράσταση δεδομένων σε χωρική μορφή, όπου οι παρόμοιες εισόδους τοποθετούνται κοντά ή μία στην άλλη στον χώρο του χάρτη.

Η διαδικασία της αυτο-οργάνωσης ξεκινά με την αρχικοποίηση των βαρών στο δίκτυο. Ένα τυχαίο διάνυσμα από τα δεδομένα επιλέγεται και υπολογίζεται η απόστασή του από κάθε νευρώνα στο δίκτυο. Ο νευρώνας με τη μικρότερη απόσταση θεωρείται ο "Best Matching Unit" (BMU).

Στη συνέχεια, εντοπίζεται η γειτονιά γύρω από τον BMU και τα βάρη των γειτόνων ανανεώνονται. Αυτό συνεπάγεται την προσαρμογή των βαρών ώστε να προσεγγίζουν περισσότερο το διάνυσμα εισόδου. Η ακριβής μέθοδος ανανέωσης των βαρών μπορεί να ποικίλλει ανάλογα με την υλοποίηση του αλγορίθμου αυτο-οργάνωσης.

Η διαδικασία της αυτο-οργάνωσης συνεχίζεται για όλα τα δεδομένα στο σύνολο εκπαίδευσης. Μετά από αρκετές επαναλήψεις, τα βάρη στο δίκτυο συγκλίνουν και ο χάρτης αναπαριστά την τοπολογία των δεδομένων με βάση τις παρόμοιες ιδιότητές τους.

Υπολογίζονται οι αποστάσεις μεταξύ του x και όλων των νευρώνων m .

$$\|x - m_b\| = \min_i \{\|x - m_i\|\}$$

Ο νευρώνας m_b που έχει την ελάχιστη απόσταση ονομάζεται BMU. Γίνεται ενημέρωση των βαρών του BMU και των τοπολογικών γειτόνων ως προς τα βάρη του x μέσω :

$$m_i(t+1) = m_i(t) + a(t)h_{bi}(t)[x - m_i(t)]$$

Όπου, t είναι ο τρέχων χρόνος, $a(t)$ είναι ο συντελεστής προσαρμογής και $h_{bi}(t)$ είναι η συνάρτηση γειτονιάς με κέντρο τον BMU. Η συνάρτηση γειτονιάς ορίζεται από τον τύπο:

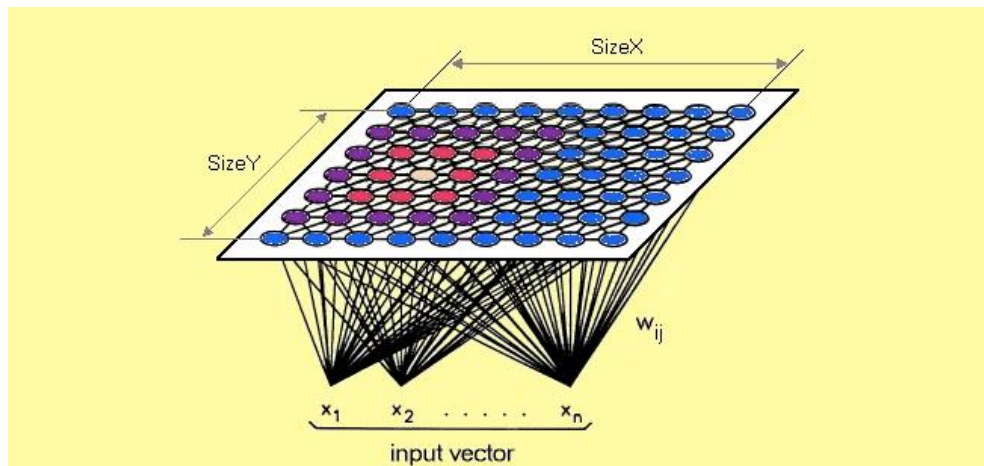
$$h_{bi}(t) = \exp\left(-\frac{\|r_b - r_i\|^2}{2\sigma^2(t)}\right)$$

Όπου, r_b και r_i είναι οι θέσεις των νευρώνων i και b στο πλέγμα του SOM.

4.3.2 Κώδικας Som

Γίνεται χρήση υφιστάμενου κώδικα αλγορίθμου SOM στο προγραμματιστικό περιβάλλον Pythοn. Εκτελείται σε λειτουργικό σύστημα macOS και αποτελείται από τα παρακάτω βήματα:

1. Αρχικοποιούνται τα βάρη κάθε κόμβου
2. Ένα διάνυσμα επιλέγεται τυχαία από το σετ των δεδομένων
3. Εξετάζεται κάθε νευρώνας για να υπολογιστεί μέσω της ευκλείδειας απόστασης συνήθως ποιανού νευρώνα τα βάρη μοιάζουν περισσότερο στο διάνυσμα εισόδου. Ο νευρώνας που κερδίζει ονομάζεται Best Matching Unit (BMU)
4. Υπολογίζεται η γειτονιά του BMU. Ο αριθμός των γειτόνων μειώνεται μέσα στον χρόνο.
5. Γίνεται ενημέρωση στο βάρος που κερδίζει. Οι γείτονες επίσης ενημερώνονται. Όσο πιο κοντά βρίσκεται ένας κόμβος στο BMU τόσο πιο πολύ μεταβάλλονται τα βάρη του και όσο πιο μακριά βρίσκεται από το BMU τόσο λιγότερο μεταβάλλονται.
6. Το βήμα 2 με 5 επαναλαμβάνεται για N φορές.



Σχήμα 4-3 : Σχηματική απεικόνιση Self-Organizing Maps

Πηγή: KHAZRI, A. (2021, October 13). Self organizing maps. Medium. Retrieved March 2, 2023, from <https://towardsdatascience.com/self-organizing-maps-1b7d2a84e065>

Best Matching Unit: είναι η τεχνική που υπολογίζει την απόσταση του κάθε βάρους στο διάνυσμα του δείγματος, ελέγχοντας όλα τα βάρη. Το βάρος με την μικρότερη απόσταση είναι ο Νικητής Νευρώνας (BMU). Υπάρχουν διάφοροι τρόποι υπολογισμού της απόστασης, ο πιο διαδεδομένος τρόπος είναι μέσω της Ευκλείδειας απόστασης.

Παράδειγμα

Πλέγμα SOM: Έχουμε ένα 2D πλέγμα SOM διαστάσεων 2x2, δηλαδή με 4 νευρώνες.

Δεδομένα εισόδου: Έχουμε ένα διάνυσμα εισόδου: (0.3, 0.4, 0.5, 0.6).

Διανύσματα βαρών των νευρώνων:

- Νευρώνας 1 με συντεταγμένες (0,0) με διάνυσμα βάρους : (0.1, 0.2, 0.3, 0.4).
- Νευρώνας 2 με συντεταγμένες (0,1) με διάνυσμα βάρους : (0.5, 0.6, 0.7, 0.8).
- Νευρώνας 3 με συντεταγμένες (1,0) με διάνυσμα βάρους : (0.9, 1.0, 1.1, 1.2).
- Νευρώνας 4 με συντεταγμένες (1,1) με διάνυσμα βάρους : (0.2, 0.3, 0.4, 0.5).

Υπολογισμός Αποστάσεων:

Υπολογίστε την Ευκλείδεια απόσταση μεταξύ των δεδομένων εισόδου (0.3, 0.4, 0.5, 0.6) και του διανύσματος βαρών κάθε νευρώνα:

Απόσταση προς το Νευρώνα 1:

$$\sqrt{((0,3 - 0,1)^2 + (0,4 - 0,2)^2 + (0,5 - 0,3)^2 + (0,6 - 0,4)^2)} = \sqrt{0,16} = 0,4$$

Απόσταση προς το Νευρώνα 2:

$$\sqrt{((0,3 - 0,5)^2 + (0,4 - 0,6)^2 + (0,5 - 0,7)^2 + (0,6 - 0,8)^2)} = \sqrt{0,64} = 0,8$$

Απόσταση προς το Νευρώνα 3:

$$\sqrt{((0,3 - 0,9)^2 + (0,4 - 1,0)^2 + (0,5 - 1,1)^2 + (0,6 - 1,2)^2)} = \sqrt{1,44} = 1,2$$

Απόσταση προς το Νευρώνα 4:

$$\sqrt{((0,3 - 0,2)^2 + (0,4 - 0,3)^2 + (0,5 - 0,4)^2 + (0,6 - 0,5)^2)} = \sqrt{0,4} = 0,2$$

Επιλογή BMU: Ο νευρώνας με τη μικρότερη Ευκλείδεια απόσταση από τα δεδομένα εισόδου είναι σε αυτήν την περίπτωση ο Νευρώνας 4.

Άρα σημαίνει ότι το διάνυσμα εισόδου θα αποτυπωθεί στο πλέγμα του SOM στον 4^ο νευρώνα.

4.3.3 U-Matrix

Η μέθοδος U-Matrix αναπτύχθηκε για την αναπαράσταση των πολυδιάστατων δεδομένων εισόδου σε ένα πλέγμα δυο διαστάσεων. Αυτό το πλέγμα ονομάζεται χάρτης και δημιουργείται με την βασική ιδέα που χρησιμοποιεί το SOM με τον υπολογισμό της ευκλείδειας απόστασης. Ο αλγόριθμος που χρησιμοποιεί το SOM έχει σχεδιαστεί για να διατηρεί τις σχέσεις που έχουν τα δεδομένα στην αρχική τους πολυδιάστατη κατάσταση πάνω στον δισδιάστατο χάρτη. Υψηλές τιμές του U-Matrix εκφράζουν ανομοιότητες μεταξύ των γειτονικών νευρώνων και είναι τα σύνορα των συστάδων, μοιάζουν σαν «όροι». Χαμηλές τιμές εκφράζουν όμοιους γειτονικούς νευρώνες, οι «κοιλάδες» στον χάρτη υποδεικνύουν τα κέντρα των συστάδων και η συγκέντρωσή τους δημιουργεί τις συστάδες. Επίσης οι περιθωριακές τιμές φαίνονται στον U-Matrix σαν «χωνί». Με αυτόν τον τρόπο ο U-Matrix μπορεί να χρησιμοποιηθεί για να αναδείξει ομοιότητες και ανομοιότητες μεταξύ των νευρώνων.

Ο αλγόριθμος SOM αποτελείται από ένα ορθογώνιο πλέγμα (X, Y). Έστω ότι $b_{x,y}$ είναι ο πίνακας των νευρώνων και $w_{i,x,y}$ ο πίνακας με τα βάρη κάθε νευρώνα μετά την εκπαίδευση. Κάθε μη ακραίος νευρώνας b έχει 8 γείτονες και έχει 3 διευθύνσεις dx, dy, dxy στον U-Matrix. Οι διευθύνσεις ορίζονται ως εξής:

$$dx(x,y) = \|b_{x,y} - b_{x+1,y}\| = \sqrt{\sum_i (w_{i,x,y} - w_{i,x+1,y})^2}$$

$$dy(x,y) = \|b_{x,y} - b_{x,y+1}\| = \sqrt{\sum_i (w_{i,x,y} - w_{i,x,y+1})^2}$$

$$dxy(x,y) = \frac{1}{2} \left(\frac{\|b_{x,y} - b_{x+1,y+1}\|}{\sqrt{2}} + \frac{\|b_{x,y+1} - b_{x+1,y}\|}{\sqrt{2}} \right)$$

4.3.4 Αλγόριθμος K-Means

Ο αλγόριθμος K-means (Aggarwal and Reddy, 2013; Tan et al., 2014; Bell, 2020; Ματσατσίνης, 2021) είναι από τους πιο γνωστούς και πολυχρησιμοποιημένους αλγόριθμους συσταδοποίησης. Στόχος του αλγορίθμου είναι η εύρεση των k σημείων στο d -διάστατο χώρο, τα οποία ελαχιστοποιούν την μέση απόσταση ελαχίστων τετραγώνων από το κοντινότερο του κέντρο. Στον k-means, ο αριθμός των συστάδων k δίνεται ως είσοδος στον αλγόριθμο και κάθε συστάδα σχετίζεται με ένα κεντρικό σημείο. Αρχίζει επιλέγοντας τυχαία έναν αριθμό κέντρων που χρησιμοποιούνται ως τα αρχικά κέντρα για κάθε cluster. Στην συνέχεια επιλέγει κάθε στοιχείο από το σύνολο δεδομένων και το αντιστοιχίζει με το κοντινότερο σε αυτό κεντρικό σημείο. Σε επόμενο βήμα υπολογίζονται ξανά τα νέα κεντρικά σημεία και ακολουθείται η ίδια διαδικασία ανάθεσης του κάθε στοιχείου των δεδομένων στο νέο κοντινότερο κεντρικό σημείο. Ο αλγόριθμος k-means υλοποιείται σε τέσσερα βήματα:

1. Διαχωρισμός σε μη κενά σύνολα.
2. Υπολογίζονται τα κεντρικά σημεία που θα είναι τα κέντρα βάρους, των συστάδων του τρέχοντος διαχωρισμού.
3. Γίνεται ανάθεση κάθε αντικειμένου στη συστάδα με το κοντινότερο κεντρικό σημείο.
4. Επανάληψη του βήματος 2.
5. Ο αλγόριθμος τερματίζει όταν η ανάθεση ολοκληρωθεί και τα κέντρα δεν μεταβάλλονται.

Η αρχικοποίηση των κέντρων βάρους των συστάδων έχει σημαντικό ρόλο για την πορεία του αλγορίθμου. Μια κακή αρχικοποίηση των κέντρων μπορεί να επηρεάσει αρνητικά την συσταδοποίηση και να οδηγήσει σε συστάδες κακής ποιότητας. Επίσης στον αλγόριθμο k-means πρέπει να γίνει επιλογή του αριθμού των συστάδων από τον χρήστη με βάση την δική του εμπειρία. Αυτό αποτελεί μειονέκτημα του αλγορίθμου καθώς δεν υπάρχει κανόνας ώστε να οδηγηθεί ο χρήστης στην σωστή επιλογή.

4.3.5 Συντελεστής Silhouette

Ο συντελεστής Silhouette είναι ένα μέγεθος υπολογισμού της ποιότητας του αλγορίθμου συσταδοποίησης. Κυμαίνεται ανάμεσα στο -1 και στο 1. Όσο μεγαλύτερη τιμή σημαίνει ότι το στοιχείο της συστάδας μοιάζει πολύ με τα υπόλοιπα στοιχεία της συστάδας και διαφοροποιείται ξεκάθαρα από τις γειτονικές συστάδες. Τιμές κοντά στο 0 υποδεικνύουν ότι η απόσταση ανάμεσα στο εξεταζόμενο σημείο και στα υπόλοιπα σημεία των γειτονικών συστάδων δεν είναι σημαντική και τιμές κοντά στο -1 υποδεικνύουν ότι οι συστάδες έχουν χωριστεί με λάθος τρόπο.

Για τον υπολογισμό του δείκτη πρέπει να υπολογιστούν πρώτα οι συντελεστές a και b για κάθε στοιχείο. Όπου a η μέση απόσταση του i στοιχείου από όλα τα υπόλοιπα στοιχεία της συστάδας στην οποία ανήκει και ο συντελεστής b είναι η ελάχιστη μέση απόσταση του στοιχείου i με τα στοιχεία κάποιας k συστάδας.

$$a(i) = \frac{1}{C_i - 1} \sum_{j \in C_i, i \neq j} d(i, j)$$
$$b(i) = \min_{k \neq i} \frac{1}{C_k} \sum_{j \in C_k} d(i, j)$$

Το C_i συμβολίζει το πλήθος των στοιχείων της i συστάδας και το $d(i, j)$ την απόσταση μεταξύ των στοιχείων i και j . Ορίζεται ο δείκτης Silhouette για το στοιχείο i σύμφωνα με τον παρακάτω τύπο:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \text{ if } |C_i| > 1$$

$$s(i) = 0, \text{ if } |C_i| = 1$$

4.3.6 Κανόνας του αγκώνα

Ο κανόνας του αγκώνα διατυπώθηκε για πρώτη φορά από τον R. Thorndike το 1953. Η επινόηση αυτού του κανόνα αποτελεί μια σημαντική συνεισφορά στον τομέα της συσταδοποίησης και έχει γίνει ευρέως διαδεδομένη τεχνική για την εύρεση του βέλτιστου αριθμού συστάδων σε πολλούς αλγόριθμους συσταδοποίησης.

Ο κανόνας του αγκώνα (Elbow method) είναι μια τεχνική που χρησιμοποιείται για τον καθορισμό του κατάλληλου αριθμού συστάδων σε έναν αλγόριθμο συσταδοποίησης, κυρίως στον αλγόριθμο k-means. Ο κανόνας αυτός παρέχει μια ευκρινή ένδειξη για τον αριθμό των συστάδων που προσφέρει τα καλύτερα αποτελέσματα, όπου η αύξηση ή μείωση του αριθμού των συστάδων δεν έχει σημαντική επίδραση στην ποιότητα της συσταδοποίησης.

Βήματα της μεθόδου:

1. Εκτέλεση του αλγορίθμου συσταδοποίησης (π.χ. k-means) για διάφορους αριθμούς συστάδων, από 1 έως ένα μέγιστο προκαθορισμένο αριθμό.
2. Υπολογισμός του αθροίσματος των τετραγώνων των αποστάσεων από τα κέντρα των συστάδων για κάθε αριθμό συστάδων.
3. Σχεδίαση ενός γραφήματος με τον αριθμό των συστάδων στον οριζόντιο άξονα και το άθροισμα των τετραγώνων των αποστάσεων στον κατακόρυφο άξονα.
4. Αναζήτηση του σημείου στο γράφημα όπου η καμπύλη του αθροίσματος των τετραγώνων των αποστάσεων αλλάζει την κατεύθυνσή της και μοιάζει με έναν "αγκώνα".
5. Επιλογή του αριθμού των συστάδων στο σημείο όπου βρίσκεται ο "αγκώνας". Αυτός ο αριθμός συστάδων θεωρείται ο βέλτιστος.

Σημειώνεται ότι, αν και μεγαλύτερος αριθμός συστάδων μπορεί να βελτιώσει την απόδοση στα δεδομένα εκπαίδευσης, η χρήση ενός πολύ μεγάλου αριθμού συστάδων μπορεί να οδηγήσει σε overfitting, όπου το μοντέλο είναι υπερβολικά προσαρμοσμένο στα δεδομένα εκπαίδευσης και χάνει την ικανότητα γενίκευσης σε νέα δεδομένα.

Κεφάλαιο 5ο: Παρουσίαση και Εφαρμογή Μεθοδολογίας: Τμηματοποίηση και Ανάλυση Αγοράς με SOM και K-Means

5.1 Εισαγωγή

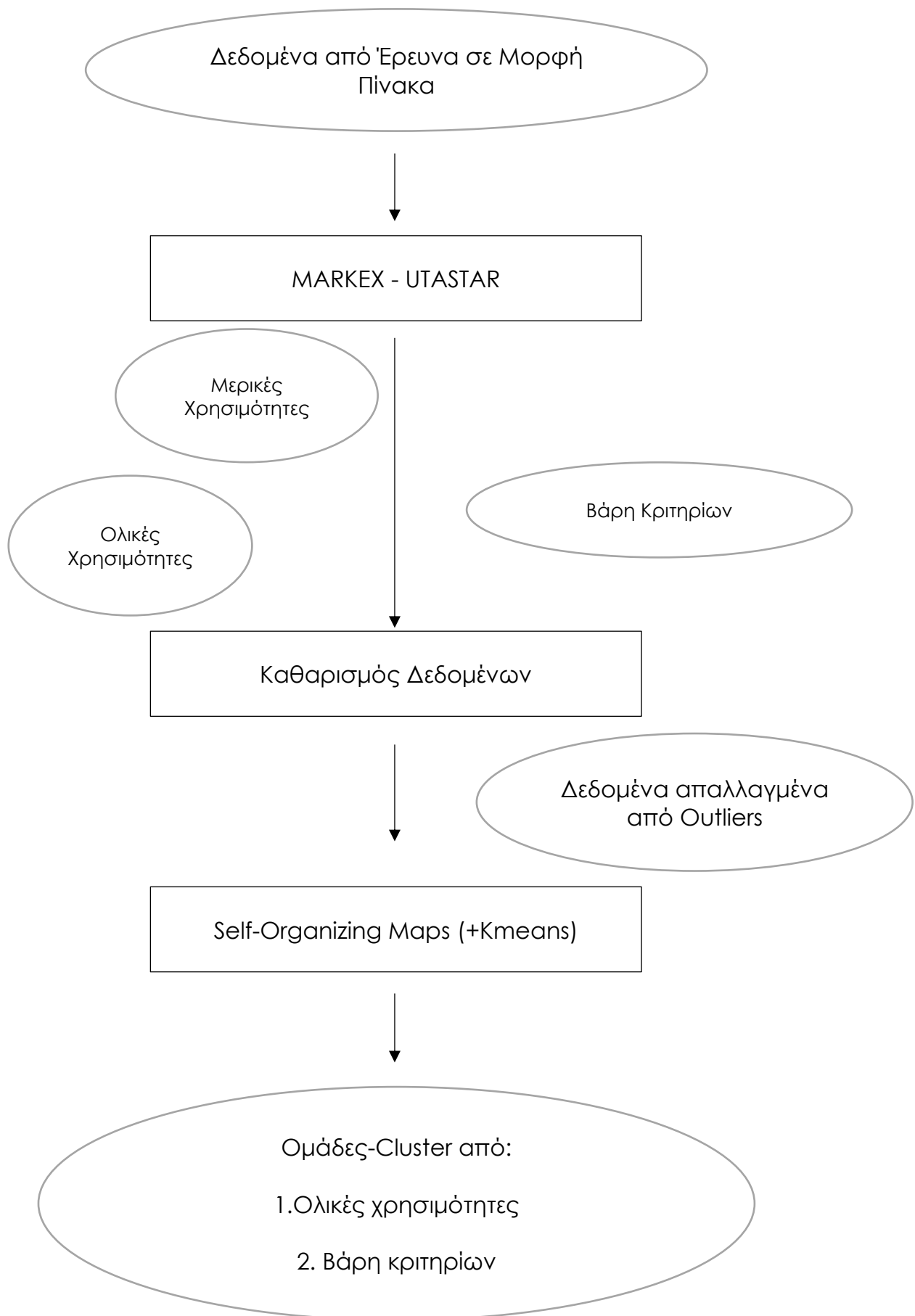
Στο παρών κεφάλαιο παρουσιάζεται και πραγματοποιείται η τμηματοποίηση και η ανάλυση της αγοράς.

5.2 Δεδομένα

Τα δεδομένα που χρησιμοποιούνται για την συσταδοποίηση των διακοσίων τεσσάρων (204) καταναλωτών είναι τα βάρη των χαρακτηριστικών καθώς και οι ολικές χρησιμότητες. Προέρχονται από την εφαρμογή της Utastar σε περιβάλλον MARKEX (Matsatsinis and Siskos, 1999), στα δεδομένα, της έρευνας που πραγματοποιήθηκε στην Γαλλία και πιο συγκεκριμένα στο Παρίσι στον τομέα του μάρκετινγκ αγροτικών προϊόντων, για τους καταναλωτές της αγοράς λαδιού (Siskos et al. 2001).

Ο στόχος της έρευνας είναι να αποτυπώσει την χρησιμότητα της πολυκριτήριας προσέγγισης στην ανάλυση δεδομένων προτίμησης του καταναλωτή και την δυνατότητα της να υποστηρίζει τις διαδικασίες ανάπτυξης νέων προϊόντων από τις γεωργικές επιχειρήσεις. Στην έρευνα παρουσιάζεται μια μεθοδολογία που βασίζεται στον καταναλωτή για την υποστήριξη αποφάσεων ανάπτυξης προϊόντων, όπου εστιάζει στον καθορισμό του μοντέλου το οποίο αναπαριστά με τον καλύτερο δυνατόν τρόπο τις προτιμήσεις κάθε μεμονωμένου καταναλωτή. Ένα σύστημα υποστήριξης αποφάσεων συνοψίζει την ανάλυση για το σύνολο των ερωτηθέντων καταναλωτών προκειμένου να καθοριστεί το ιδανικό προφίλ ενός νέου προϊόντος και να προσομοιωθεί η στρατηγική διείσδυσης του στην αγορά. Παρουσιάζονται τα αποτελέσματα από την εφαρμογή της μεθοδολογίας στη βάση δεδομένων έρευνας που προέρχεται από την αγορά ελαιολάδου του Παρισιού. Τέλος, το άρθρο καταλήγει σε κάποιες προτάσεις για την πρακτική του μάρκετινγκ στη αγροτική επιχειρηματικότητα. (Siskos et al. 2001)

5.3 Μεθοδολογία



Η μεθοδολογία που αναπτύχθηκε για την τμηματοποίηση της αγοράς των καταναλωτών χρησιμοποιεί συνδυασμό δεδομένων που έχουν παραχθεί από εφαρμογή της UTASTAR, και τον αλγόριθμο Self Organizing Maps για την τμηματοποίηση των καταναλωτών.

Αρχικά, από έρευνα, αντλούνται δεδομένα για μια συγκεκριμένη αγορά. Τα δεδομένα επεξεργάζονται κατάλληλα ώστε να είναι ιδανικά για την επεξεργασία τους από το λογισμικό MARKEX, που χρησιμοποιεί τον αλγόριθμο UTASTAR.

Το MarkeX είναι ένα σύστημα που λειτουργεί ως σύμβουλος για τους ανθρώπους στον τομέα του μάρκετινγκ καθώς δίνει την δυνατότητα για οπτική υποστήριξη η οποία βοηθά στην καλύτερη κατανόηση και να ξεπεραστεί η έλλειψη ειδίκευσης.

Ο Ματσατσίνης και Σίσκος (1993) παρουσίασαν μια πρωτότυπη βασισμένη στον πελάτη μεθοδολογία για να βοηθήσει στην δημιουργία νέων προϊόντων στις επιχειρήσεις. Στο στάδιο του ερωτηματολογίου της αγοράς, κάθε πελάτης δίνει τις βαθμολογίες του σε ένα σετ από προϊόντα που παίρνουν μέρος στην έρευνα, με βάση κάποια κριτήρια. Τέλος απαιτείται να βαθμολογήσει τα προϊόντα με βάση την σειρά προτεραιότητας.

Η τμηματοποίηση της αγοράς στο MarkeX γίνεται, με την πολυκριτήρια μέθοδο Utastar. Γίνεται δηλαδή με ανάλυση κριτηρίων με βάση την σημαντικότητα τους η οποία βασίζεται στις επιλογές των καταναλωτών.

Η πολυκριτήρια μεθοδολογία που χρησιμοποιήθηκε έχει ένα πολύ σημαντικό πλεονέκτημα που αφορά τη δυνατότητα χειρισμού τόσο ποσοτικών όσο και ποιοτικών κριτηρίων.

CONSUMER PREFERENCE ASSESSMENT

What is your opinion about the following products?

PRODUCT		CARAPELLI	LERIDA	KOLYBARI	HEDIARD	LA JARRE d'OR	PUGET
IMAGE (a)	BAD						
	UNAWARE						
	GOOD						
	VERY GOOD						
COLOUR (b)	UNNATURAL						
	NATURAL						
	ATTRACTIVE						
ODOUR (c)	UNNATURAL						
	NATURAL						
	PLEASANT						
TASTE (d)	UNNATURAL						
	NATURAL						
	DELICIOUS						
PACKAGING (e)	GOOD						
	FAIR						
	GOOD						
	VERY GOOD						

(a) What is your opinion about the quality of..... extra virgin olive oil?

(b) Have you ever heard from others (relatives, friends) about..... extra virgin olive oil?

(c) Which is your opinion about the package of..... extra virgin olive oil?

(d) Which is your opinion about the producing company of..... extra virgin olive oil?

RANKING

Which of the following extra virgin olive oils would you buy?

Repetitive question: If you could not find your oil of choice, which one would you buy?

CARAPELLI	LERIDA	KOLYBARI	HEDIARD	LA JARRE D'OR	PUGET

Ενδεικτικό ερωτηματολόγιο από το MARKEX με την κωδικοποίηση των κριτηρίων

Με βάση τα δεδομένα που αντλούνται από το ειδικό ερωτηματολόγιο, δημιουργείται από το MARKEX, ένας πολυκριτήριος πίνακας για κάθε καταναλωτή, της παρακάτω μορφής (Πίνακας : παράδειγμα από εφαρμογή στην περίπτωση ανάπτυξης ενός νέου προϊόντος ελαιόλαδου για την Γαλλική αγορά), ο οποίος θα χρησιμοποιηθεί εν συνεχεία για την είσοδο των δεδομένων στη UTASTAR και την επίλυση του αντίστοιχου γραμμικού προβλήματος. Η UTASTAR εφαρμόζεται για κάθε καταναλωτή ξεχωριστά κατασκευάζοντας με αυτό τον τρόπο τα μοντέλα συμπεριφοράς τους.

ΠΡΟΙΟΝΤΑ	ΠΡΟΔΙΑΤΑΞΗ	ΕΠΙΡΡΟΗ	ΧΡΩΜΑ	ΑΡΩΜΑ	ΓΕΥΣΗ	ΣΥΣΚΕΥΑΣΙΑ	ΤΙΜΗ (€)
CARAPELLI	3	2	2	2	2	2	31
LERIDA	4	1	2	2	2	2	65
KOLIMVARI	3	2	2	2	2	2	20
HEDIARD	2	1	2	2	3	3	48
LA JARRE D'OR	3	2	2	2	3	2	37
PUGET	1	3	1	2	3	2	18

Πίνακας: Πολυκριτήριοι πίνακας προδιατάξης των προϊόντων

Στην συνέχεια, έπειτα από την εφαρμογή του αλγορίθμου προκύπτουν τα δεδομένα των ολικών χρησιμότητων των εναλλακτικών επιλογών των καταναλωτών, καθώς και τα βάρη που δίνει κάθε καταναλωτής σε κάθε κριτήριο των προϊόντων που αξιολογεί.

Στην συνέχεια διερευνάται η διενέργεια τμηματοποιήσεων με την χρήση Τεχνητής Νοημοσύνης και πιο συγκεκριμένα Self Organizing Maps. Η τμηματοποίηση των καταναλωτών γίνεται με τα παραπάνω δεδομένα, τα οποία έχουν μορφή πίνακα. Δηλαδή πραγματοποιείται τμηματοποίηση της αγοράς με δύο διαφορετικούς τρόπους. Ο πρώτος τρόπος χρησιμοποιεί τα δεδομένα των ολικών χρησιμότητων των καταναλωτών για τα προϊόντα (Παράγραφος 5.6). Ο δεύτερος τρόπος τμηματοποίησης χρησιμοποιεί τα βάρη που δίνει κάθε καταναλωτής σε κάθε κριτήριο των προϊόντων που αξιολογεί (Παράγραφος 5.8).

Αντλείται μια βιβλιοθήκη, ανοιχτή προς όλους τους χρήστες, από ιστότοπο με αλγορίθμους (github.com), ονόματος sompy. Η συγκεκριμένη βιβλιοθήκη περιέχει τον αλγόριθμο Self Organizing Maps. Η εφαρμογή και τροποποίησή της γίνεται σε περιβάλλον Python.

Πριν την τμηματοποίηση της αγοράς πραγματοποιείται ο καθαρισμός των δεδομένων. Η διαδικασία που ακολουθείτε είναι η εύρεση και αφαίρεση των ακραίων τιμών και από τα βάρη των χαρακτηριστικών καθώς και από τις ολικές χρησιμότητες των καταναλωτών. Η εύρεση των ακραίων τιμών και στις δύο περιπτώσεις πραγματοποιείται με την διαδικασία που χρησιμοποιεί το Interquartile Range IQR, ενδοτεταρτημοριακό Εύρος.

Τα δεδομένα αποκτούν την κατάλληλη μορφή για την εισαγωγή τους στον αλγόριθμο Self Organizing Maps σε περιβάλλον Python.

Γίνονται οι απαραίτητες αλλαγές στον αλγόριθμο Self Organizing Maps για να ανταποκρίνεται καλύτερα στα δεδομένα που χρησιμοποιούνται.

Με γνώμονα τον δείκτη Silhouette αποφασίζεται ο πιο αποδοτικός αριθμός συστάδων για την τμηματοποίηση.

Τα δεδομένα εισόδου, συνδέονται με τους νευρώνες του πλέγματος SOM διαμέσου των βαρών. Τα βάρη είναι διανύσματα που έχουν τις ίδιες διαστάσεις με το διάνυσμα των δεδομένων εισόδου. Η συνάρτηση που συνδέει τα δύο μεταξύ τους, δεδομένα εισόδου και νευρώνες στο πλέγμα του SOM, είναι η Ευκλείδεια απόσταση του διανύσματος εισόδου με το διάνυσμα του βάρους που αντιστοιχεί στον κάθε νευρώνα.

Οι νευρώνες πλέον στο πλέγμα του SOM(U-Matrix) τοποθετούνται σε κοντινές αποστάσεις ανάλογα με την ομοιότητα τους. Σε αυτό το σημείο μέσα στον αλγόριθμο εφαρμόζεται ο αλγόριθμος K-means ώστε να τοποθετηθούν οι αντίστοιχες ταμπέλες στους ήδη τμηματοποιημένους καταναλωτές.

Πραγματοποιείται ανάλυση των συστάδων, ο σχολιασμός των καταναλωτών που ανήκουν σε αυτές και τα συμπεράσματα που προκύπτουν. Οι ομάδες-τμήματα που προκύπτουν πρέπει να είναι ομοιογενείς και να περιγράφονται ανάλογα με τα προτιμήσεις τους.

5.4 Εφαρμογή UTASTAR

Η εφαρμογή της UTASTAR γίνεται σε περιβάλλον MARKEX (Matsatsinis and Siskos, 1999), στο Web-MARKEX. Σκοπός της εφαρμογής είναι η έξοδος των ολικών χρησιμότητων των εναλλακτικών και των βαρών των χαρακτηριστικών για κάθε καταναλωτή.

Η μέθοδος UTASTAR επιτρέπει την κατάταξη και τον προσδιορισμό της βέλτιστης εναλλακτικής λύσης βάσει ενός συνόλου κριτηρίων. Οι αποφασίζοντες παρέχουν τις προτιμήσεις τους για τις εναλλακτικές σε σχέση με κάθε κριτήριο, και η μέθοδος χρησιμοποιεί μαθηματικά μοντέλα για την επίτευξη ενός συνολικού σκορ ολικής χρησιμότητας και την κατάταξη των εναλλακτικών.

Με την εφαρμογή του συστήματος MARKEX προκύπτουν οι μερικές και ολικές χρησιμότητες των εναλλακτικών επιλογών για τον κάθε καταναλωτή και τα βάρη σημαντικότητας των κριτηρίων για τον κάθε καταναλωτή.

5.5 Καθαρισμός Δεδομένων – Αφαίρεση ακραίων τιμών

Πριν την τμηματοποίηση της αγοράς πραγματοποιείται ο καθαρισμός των δεδομένων. Η διαδικασία που ακολουθείτε είναι η εύρεση και αφαίρεση των ακραίων τιμών και από τα βάρη των χαρακτηριστικών καθώς και από τις ολικές χρησιμότητες των καταναλωτών. Η εύρεση των ακραίων τιμών και στις δύο

περιπτώσεις πραγματοποιείται με την διαδικασία που χρησιμοποιεί το Interquartile Range IQR, ενδοτεταρτημοριακό Εύρος.

Το IQR είναι ένα μέτρο διακύμανσης που χρησιμοποιείται στη στατιστική. Υπολογίζεται ως η διαφορά μεταξύ του τρίτου τεταρτημορίου (Q3) και του πρώτου τεταρτημορίου (Q1) ενός συνόλου δεδομένων.

5.6 Τμηματοποίηση των Καταναλωτών Βάσει των Ολικών Χρησιμοτήτων

Η τμηματοποίηση στο Markex (Matsatsinis and Siskos, 1999) γίνεται, με την πολυκριτήρια μέθοδο Utastar. Γίνεται δηλαδή με ανάλυση κριτηρίων με βάση την σημαντικότητα τους η οποία βασίζεται στις επιλογές των καταναλωτών.

Στο παρών υποκεφάλαιο παρουσιάζεται η εφαρμογή της μεθοδολογίας τμηματοποίησης της αγοράς στα δεδομένα των ολικών χρησιμοτήτων των εναλλακτικών, από την UTASTAR σε περιβάλλον MARKEX (Matsatsinis and Siskos, 1999).

Αντλείται μια βιβλιοθήκη, ανοιχτή προς όλους τους χρήστες, από ιστότοπο με αλγορίθμους (github.com), ονόματος sompy. Η συγκεκριμένη βιβλιοθήκη περιέχει τον αλγόριθμο Self Organizing Maps. Η εφαρμογή και τροποποίησή της γίνεται σε περιβάλλον Python και πιο συγκεκριμένα στο Pycharm.

Κατά την τμηματοποίηση της αγοράς με γνώμονα τις ολικές χρησιμότητες των εναλλακτικών των καταναλωτών με το νευρωνικό δίκτυο Self Organizing Maps. Με την μέθοδο δοκιμής και σφάλματος σε περιβάλλον Python καθορίστηκαν:

- Το μέγεθος του πλέγματος του SOM ίσο με 20x20 νευρώνες
- Ο συντελεστής train_len_factor ίσο με 5
- Τυχαία αρχικοποίηση των βαρών

Οι υπόλοιπες παράμετροι είναι στην προκαθορισμένη επιλογή.

Οι ολικές χρησιμότητες των εναλλακτικών των καταναλωτών προέρχονται από την εφαρμογή της Utastar σε περιβάλλον MARKEX (Matsatsinis and Siskos, 1999), στα δεδομένα, της έρευνας που πραγματοποιήθηκε στην Γαλλία και πιο συγκεκριμένα στο Παρίσι στον τομέα του γεωργικού μάρκετινγκ, για τους καταναλωτές της αγοράς λαδιού.(Siskos et al. 2001)(Παράγραφος 5.2). Οι ολικές χρησιμότητες των εναλλακτικών έχουν μορφή πίνακα έξι(6) x διακόσια τέσσερα(204), αφού αποτελούνται από τα έξι(6) προϊόντα και από τους διακόσιους τέσσερεις (204) καταναλωτές.

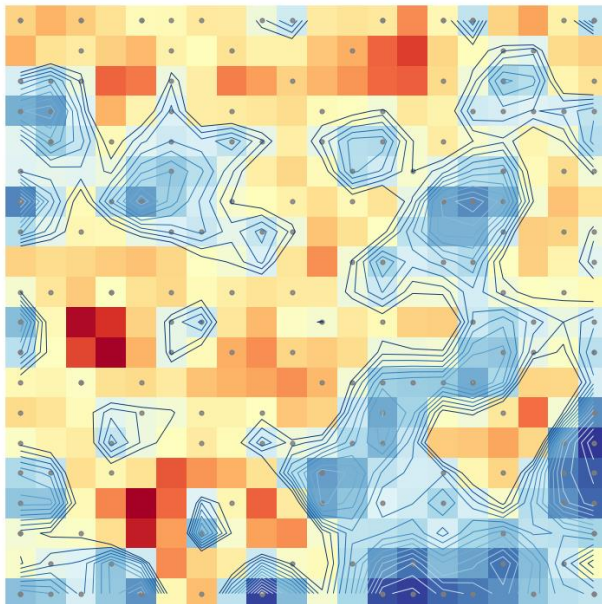
Τα δεδομένα αφού αποθηκευτήκαν σε αρχείο excel και έπειτα τροποποιήθηκαν σε μορφή csv για την χρήση τους στο προγραμματιστικό περιβάλλον python.

Οι ολικές χρησιμότητες συνδέονται με τους νευρώνες του πλέγματος SOM διαμέσου των βαρών. Τα βάρη είναι διανύσματα που έχουν τις ίδιες διαστάσεις

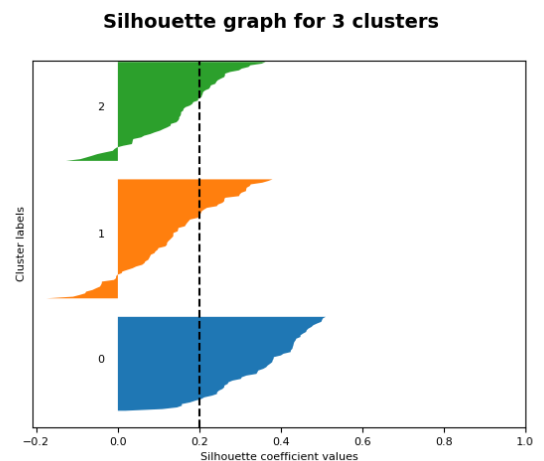
με το διάνυσμα των ολικών χρησιμότητων. Η συνάρτηση που συνδέει τα δύο μεταξύ τους, ολικές χρησιμότητες και νευρώνες στο πλέγμα του SOM, είναι η Ευκλείδεια απόσταση του διανύσματος εισόδου με το διάνυσμα του βάρους που αντιστοιχεί στον κάθε νευρώνα.

Οι νευρώνες πλέον στο πλέγμα του SOM(U-Matrix) τοποθετούνται σε κοντινές αποστάσεις ανάλογα με την ομοιότητα τους. Σε αυτό το σημείο μέσα στον αλγόριθμο εφαρμόζεται ο αλγόριθμος K-means ώστε να τοποθετηθούν οι αντίστοιχες ταμπέλες στους ήδη τμηματοποιημένους καταναλωτές.

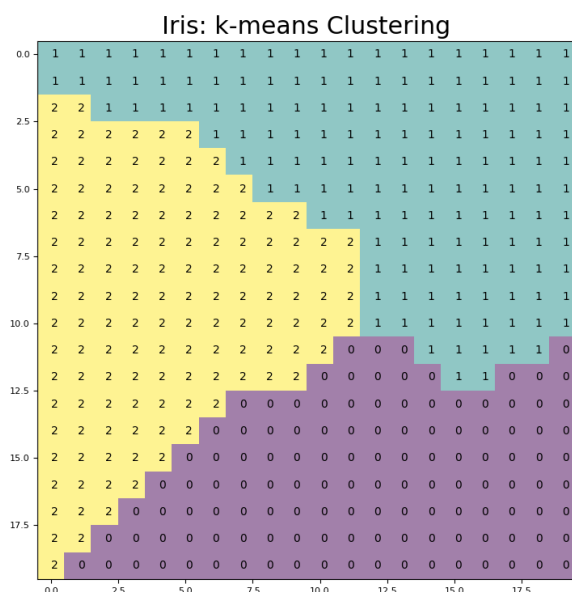
Η αγορά τμηματοποιήθηκε σε τρεις συστάδες έπειτα από την εφαρμογή του αλγορίθμου. Με γνώμονα τον δείκτη silhouette για την επιλογή του αριθμού των συστάδων, οι συστάδες που χωρίστηκαν οι καταναλωτές είναι τρεις σε αριθμό. Αυτό συνέβη διότι η μεγαλύτερη τιμή του δείκτη silhouette ήταν 0.199 κατά την επανάληψη του αλγορίθμου που τμηματοποιούσε την αγορά σε τρεις συστάδες.



Διάγραμμα: U-matrix



Διάγραμμα: Δείκτης Silhouette για 3 συστάδες



Διάγραμμα: Εφαρμογή k-means στον U-matrix

		CARAPELLI	HEDIARD	JARRE d'OR	KOLYMVARI	LERIDA	PUGET
1	31,22%	0,616754678	0,724938186	0,474308739	0,569964657	0,64834395	0,64965232
2	38,10%	0,509069952	0,465393866	0,414287979	0,602876637	0,334377835	0,709154795
3	30,69%	0,785365566	0,688648729	0,734747382	0,810092161	0,526971595	0,86878714

Πίνακας: 1^η στήλη :Ποσοστά του συνόλου των καταναλωτών

Επόμενες στήλες: Μέσες ολικές χρησιμότητες για κάθε προϊόν για κάθε συστάδα

CARAPELLI	HEDIARD	JARRE d'OR	KOLYMVARI	LERIDA	PUGET
12	18	3	9	5	12
20,34%	30,51%	5,08%	15,25%	8,47%	20,34%

Πίνακας: 1^η συστάδα,

1^η γραμμή: αριθμός καταναλωτών που τοποθέτησαν το προϊόν ως 1^η επιλογή

2^η γραμμή: ποσοστό καταναλωτών που τοποθέτησαν το προϊόν ως 1^η επιλογή

CARAPELLI	HEDIARD	JARRE d'OR	KOLYMVARI	LERIDA	PUGET
7	6	0	13	0	46
10%	8%	0%	18%	0%	64%

Πίνακας: 2^η συστάδα

1^η γραμμή: αριθμός καταναλωτών που τοποθέτησαν το προϊόν ως 1^η επιλογή

2^η γραμμή: ποσοστό καταναλωτών που τοποθέτησαν το προϊόν ως 1^η επιλογή

CARAPELLI	HEDIARD	JARRE d'OR	KOLYMVARI	LERIDA	PUGET
10	2	5	10	0	31
17%	3%	9%	17%	0%	53%

Πίνακας: 3^η συστάδα

1^η γραμμή: αριθμός καταναλωτών που τοποθέτησαν το προϊόν ως 1^η επιλογή
 2^η γραμμή: ποσοστό καταναλωτών που τοποθέτησαν το προϊόν ως 1^η επιλογή

Product	Country	Image	Colour	Odour	Taste	Packaging	Price
CARAPELLI	Italian	Unaware	Natural	Natural	Natural	Good	31
LERIDA	Spanish	Unaware-Good	Natural	Natural	Natural-Delicious	Fair-Good	65
KOLYMVARI	Cretan	Unaware-Good	Natural	Natural	Natural	Fair	20
Hediart	French 1	Good	Natural	Natural	Natural-Delicious	Good	48
JARRE d'OR	French 2	Unaware	Unnatural	Natural	Natural	Fair	37
Puget	French 3	Good	Natural	Natural	Natural	Fair-Good	18

Πίνακας: Χαρακτηριστικά των προϊόντων λαδιού που αξιολογούν οι καταναλωτές από την έρευνα

Τα συμπεράσματα που προκύπτουν από τις συστάδες που προέκυψαν είναι τα παρακάτω:

Η πρώτη συστάδα αποτελείται από το 31,22% των συνολικών καταναλωτών της έρευνας. Για την 1^η συστάδα το 30,51% των καταναλωτών έχουν ως πρώτη τους επιλογή το HEDIARD, το 20,34% των καταναλωτών έχουν ως πρώτη επιλογή το PUGET και το 20,34% των καταναλωτών έχουν ως πρώτη επιλογή το CARAPELLI.

Επίσης από το μέσο όρο των ολικών χρησιμότητων φαίνεται ότι προτιμούν HEDIARD 0,7249 και στην συνέχεια ακολουθεί το PUGET 0,6497 και το LERIDA 0,6483.

Συμπεραίνεται ότι τα τέσσερα προϊόντα που αναφέρονται είναι ανταγωνιστικά στη συγκεκριμένη συστάδα και στο συγκεκριμένο προφίλ καταναλωτών. Το HEDIARD έχει υψηλή τιμή, καλή συσκευασία, φυσικό χρώμα καθώς και φυσική και απολαυστική γεύση. Το PUGET ξεχωρίζει με την χαμηλότερη τιμή, έχει μέτρια προς καλή συσκευασία, φυσικό χρώμα και φυσική γεύση. Το CARAPELLI κυμαίνεται στις χαμηλές τιμές στις επιλογές των καταναλωτών, με καλή συσκευασία, φυσικό χρώμα και φυσική γεύση. Τέλος το LERIDA έχει την ακριβότερη τιμή, μέτρια προς καλή συσκευασία, φυσικό χρώμα, φυσική και απολαυστική γεύση. Σύμφωνα με τα χαρακτηριστικά των ανταγωνιστικών προϊόντων της συστάδας συμπεραίνεται ότι οι καταναλωτές της συγκεκριμένης συστάδας δεν αποφασίζουν με βάση την τιμή αφού παρατηρείτε μεγάλο εύρος τιμών στις επιλογές τους, αλλά δίνουν μεγάλη σημαντικότητα στη γεύση και στην συσκευασία.

Η δεύτερη συστάδα αποτελείται από το 38,10% των συνολικών καταναλωτών της έρευνας. Για την 2^η συστάδα το 64% των καταναλωτών κατατάσσουν το PUGET ως πρώτη επιλογή και το 18% των καταναλωτών της συστάδας επιλέγουν ως πρώτη επιλογή το KOLYMVARI.

Το ίδιο παρατηρείται και από τον πίνακα με τις μέσες ολικές χρησιμότητες με τιμές PUGET 0,7092 και KOLYMVARI 0,6029.

Τα προϊόντα που ανταγωνίζονται στην δεύτερη συστάδα και στο συγκεκριμένο προφίλ καταναλωτών είναι το PUGET και το KOLYMVARI. Τα δύο προϊόντα κατέχουν τις χαμηλότερες τιμές της αγοράς και οι καταναλωτές τα προτιμούν για την τιμή τους. Η διαφορά τους είναι ότι το PUGET είναι Γαλλικής προέλευσης ενώ το KOLYMVARI Ελληνικής προέλευσης, το PUGET έχει καλύτερη συσκευασία και είναι κατά ελάχιστα πιο φτηνό.

Η τρίτη συστάδα αποτελείται από το 30,7% των συνολικών καταναλωτών της έρευνας. Για την τρίτη συστάδα το 53% των καταναλωτών κατατάσσουν ως πρώτη επιλογή το PUGET, το 17% το KOLYMVARI και το 17% το CARAPELLI.

Από τον μέσο όρο των ολικών χρησιμότητων προκύπτουν οι τιμές για το PUGET 0,8687, για το KOLYMVARI 0,81 και για το CARAPELLI 0,7854.

Τα προϊόντα που ανταγωνίζονται στην τρίτη συστάδα και στο συγκεκριμένο προφίλ καταναλωτών είναι το PUGET, το KOLYMVARI και το CARAPELLI. Παρατηρείται μια επιθυμία προς τα προϊόντα με την χαμηλότερη τιμή καθώς, το PUGET είναι η φτηνότερη εναλλακτική, στην συνέχεια είναι το KOLYMVARI και επόμενο με πιο ανεβασμένη τιμή ακολουθεί το CARAPELLI. Ξεκάθαρα συμπεραίνουμε ότι οι καταναλωτές της τρίτης συστάδας αποφασίζουν με βάση την τιμή του προϊόντος.

Από τα παραπάνω συμπεραίνεται ότι το μεγαλύτερο μερίδιο της αγοράς το κατέχει το PUGET και έχει καταφέρει να απευθύνεται σε όλα τα προφίλ των καταναλωτών. Σημαντικό μερίδιο της αγοράς κατέχει και το KOLYMVARI και το CARAPELLI.

5.7 Περιπτώσεις Ακραίων Τιμών

Οι περιπτώσεις ακραίων τιμών στα βάρη αφορούν περιπτώσεις που οι καταναλωτές δεν λαμβάνουν σχεδόν καθόλου υπόψη τους ένα κριτήριο. Θέτοντας ένα κατώφλι σημαντικότητας των κριτηρίων 0,02 προκύπτουν τα παρακάτω αποτελέσματα.

Color	Influence	Odour	Packaging	Price	Taste
6	1	9	4	8	2
2,94%	0,49%	4,41%	1,96%	3,92%	0,98%

Συγκεντρωτικά φαίνεται οι καταναλωτές κατά ένα ποσοστό 4,41% φαίνεται ότι αποφασίζουν για την αγορά του λαδιού τους χωρίς να παίρνουν υπόψη τους το άρωμα του λαδιού. Επίσης κατά ένα 3,92% αποφασίζουν χωρίς να υπολογίζουν την τιμή και κατά ένα 2,94% το χρώμα του λαδιού. Τέλος με ποσοστό 1,96% αποφασίζουν χωρίς να παίρνουν υπόψη τους την συσκευασία του προϊόντος.

5.8 Τμηματοποίηση Βάσει των Βαρών των Χαρακτηριστικών

Στο παρών υποκεφάλαιο πραγματοποιείται συσταδοποίηση των καταναλωτών της αγοράς με γνώμονα τα βάρη των κριτηρίων. Τα βάρη των κριτηρίων προέρχονται από την εφαρμογή της Utastar σε περιβάλλον MARKEX (Matsatsinis and Siskos, 1999) στα δεδομένα της έρευνας που πραγματοποιήθηκε στην Γαλλία και πιο συγκεκριμένα στο Παρίσι στον τομέα του γεωργικού μάρκετινγκ, για τους καταναλωτές της αγοράς λαδιού. (Siskos et al. 2001) (Παράγραφος 5.2)

Τα βάρη των κριτηρίων έχουν μορφή πίνακα έξι(6) x διακόσια τέσσερα (204), καθώς αποτελείται από τα έξι(6) χαρακτηριστικά, χρώμα, επιρροή, άρωμα, συσκευασία, τιμή, γεύση και τους διακόσιους τέσσερεις(204) καταναλωτές.

Αντλείται μια βιβλιοθήκη, ανοιχτή προς όλους τους χρήστες, από ιστότοπο με αλγορίθμους (github.com), ονόματος sompy. Η συγκεκριμένη βιβλιοθήκη περιέχει τον αλγόριθμο Self Organizing Maps. Η εφαρμογή και τροποποίησή της γίνεται σε περιβάλλον Python και πιο συγκεκριμένα στο Pycharm.

Πραγματοποιείται τμηματοποίηση της αγοράς με γνώμονα τα βάρη των χαρακτηριστικών με το νευρωνικό δίκτυο Self Organizing Maps. Με την μέθοδο δοκιμής και σφάλματος καθορίστηκαν:

- Το μέγεθος του πλέγματος του SOM ίσο με 20x20 νευρώνες
- Ο συντελεστής `train_len_factor` ίσος με 5
- Τυχαία αρχικοποίηση των βαρών

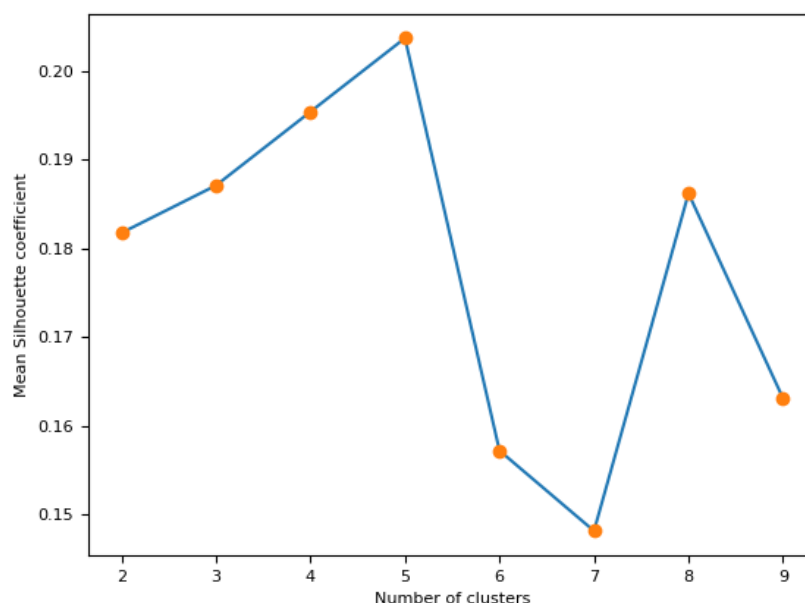
Οι υπόλοιπες παράμετροι είναι στην προκαθορισμένη επιλογή της βιβλιοθήκης.

Τα βάρη των χαρακτηριστικών συνδέονται με τους νευρώνες του πλέγματος SOM διαμέσου των βαρών του νευρωνικού δικτύου. Τα βάρη του νευρωνικού δικτύου είναι διανύσματα που έχουν τις ίδιες διαστάσεις με το διάνυσμα των βαρών των χαρακτηριστικών. Η συνάρτηση που συνδέει τα δύο μεταξύ τους, βάρη των χαρακτηριστικών και νευρώνες στο πλέγμα του SOM, είναι η Ευκλείδεια απόσταση του διανύσματος εισόδου με το διάνυσμα του βάρους του νευρωνικού δικτύου που αντιστοιχεί στον κάθε νευρώνα.

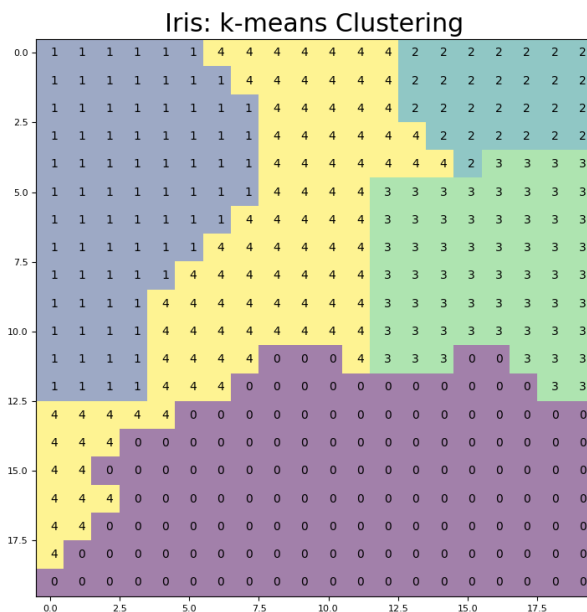
Οι νευρώνες πλέον στο πλέγμα του SOM(U-Matrix) τοποθετούνται σε κοντινές αποστάσεις ανάλογα με την ομοιότητα τους. Σε αυτό το σημείο μέσα στον αλγόριθμο εφαρμόζεται ο αλγόριθμος K-means ώστε να τοποθετηθούν οι αντίστοιχες ταμπέλες στους ήδη τμηματοποιημένους καταναλωτές.

Η αγορά τμηματοποιείται σε πέντε συστάδες έπειτα από την εφαρμογή του αλγορίθμου. Με γνώμονα τον δείκτη silhouette για την επιλογή του αριθμού των συστάδων, οι συστάδες που χωρίστηκαν οι καταναλωτές είναι πέντε σε αριθμό. Αυτό συνέβη διότι η μεγαλύτερη τιμή του δείκτη silhouette ήταν 0.204 κατά την επανάληψη του αλγορίθμου που τμηματοποιεί την αγορά σε πέντε συστάδες.

Attributes: Mean Silhouette for various cluster number



Διάγραμμα: Δείκτης Silhouette για τις διάφορες συστάδες



Διάγραμμα: Εφαρμογή k-means στον U-matrix

Η αγορά τμηματοποιείται σε πέντε συστάδες και παρουσιάζονται στους παρακάτω πίνακες τα αποτελέσματα της ανάλυσης.

	Color	Influence	Odour	Packaging	Price	Taste
1	0,10839115	0,12707417	0,10451845	0,14066639	0,38058958	0,13876025
2	0,13299483	0,16193681	0,15227443	0,14723312	0,12866078	0,27690003
3	0,12615939	0,31655108	0,09646509	0,18621338	0,15046949	0,12414156
4	0,13228537	0,15353377	0,1073324	0,31840068	0,16474704	0,12370074
5	0,18032082	0,15264724	0,17753597	0,15307545	0,19929368	0,13712685

Πίνακας: Μέσα βάρη για την κάθε συστάδα

1	0%	0%	0%	0%	100%	0%
2	4%	21%	13%	32%	0%	29%
3	22%	26%	18%	25%	0%	24%

Πίνακας: 1^η Συστάδα

1^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 1^η θέση

2^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 2^η θέση

3^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 3^η θέση

1	2%	12%	7%	0%	2%	76%
2	12%	17%	12%	22%	22%	15%
3	12%	27%	17%	27%	17%	7%

Πίνακας: 2^η Συστάδα

1^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 1^η θέση

2^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 2^η θέση

3^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 3^η θέση

1	0%	88%	0%	6%	6%	0%
2	6%	13%	0%	44%	31%	6%
3	25%	0%	6%	31%	19%	25%

Πίνακας: 3^η Συστάδα

1^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 1^η θέση

2^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 2^η θέση

3^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 3^η θέση

1	0%	0%	0%	100%	0%	0%
2	13%	22%	9%	0%	48%	9%
3	22%	22%	9%	0%	17%	30%

Πίνακας: 4^η Συστάδα

1^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 1^η θέση

2^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 2^η θέση

3^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 3^η θέση

1	25%	6%	15%	8%	42%	4%
2	25%	15%	25%	21%	17%	6%
3	10%	21%	15%	13%	17%	23%

Πίνακας: 5^η Συστάδα

1^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 1^η θέση

2^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 2^η θέση

3^η γραμμή: Ποσοστό καταναλωτών που η σημαντικότητα του κριτηρίου είναι σε 3^η θέση

1	34,52%
2	20,81%
3	8,12%
4	11,68%
5	24,37%

Πίνακας: Ποσοστά κάθε συστάδας

Τα συμπεράσματα που προκύπτουν από τα παραπάνω στοιχεία των συστάδων είναι τα παρακάτω:

Η πρώτη συστάδα αποτελείται από το 34,52% των συνολικών καταναλωτών της έρευνας. Από τα μέσα βάρη της συστάδας προκύπτει ότι το συγκεκριμένο προφίλ καταναλωτών θεωρούν πιο σημαντικό την τιμή του λαδιού που αγοράζουν με σημαντικότητα 0,3806. Όλο το σύνολο των καταναλωτών έχει ως πιο σημαντικό κριτήριο την τιμή καθώς για το 100% της συστάδας είναι το μεγαλύτερη σημαντικότητα έναντι των υπολοίπων.

Τα χαρακτηριστικά που δεν θεωρούν καθόλου σημαντικά οι καταναλωτές της πρώτης συστάδας είναι το άρωμα και το χρώμα του λαδιού με τιμές μέσου βάρους 0,1045 και 0,1084 αντίστοιχα.

Το προφίλ των καταναλωτών της πρώτης συστάδας θυσιάζει τα υπόλοιπα χαρακτηριστικά για την χαμηλή τιμή του προϊόντος και αποτελεί το μεγαλύτερο ποσοστό των ερωτηθέντων της έρευνας.

Η δεύτερη συστάδα αποτελείται από το 20,81% των συνολικών καταναλωτών της έρευνας. Από τα μέσα βάρη συμπεραίνεται ότι το σημαντικότερο χαρακτηριστικό για τους καταναλωτές της δεύτερης συστάδας είναι η γεύση του λαδιού με σημαντικότητα 0,2769. Για το 76% της συστάδας η σημαντικότητα της γεύσης ξεπερνάει κάθε άλλη σημαντικότητα.

Τα λιγότερο σημαντικά χαρακτηριστικά για την δεύτερη συστάδα είναι η τιμή και στην συνέχεια το χρώμα με τιμές μέσου βάρους 0,1287 και 0,1330 αντίστοιχα.

Το προφίλ των καταναλωτών της δεύτερης συστάδας έχει ανάγκη για ένα γευστικό λάδι ανεξαρτήτως τιμής.

Η τρίτη συστάδα αποτελείται από το 8,12% των συνολικών καταναλωτών και είναι η μικρότερη σε αριθμό καταναλωτών συστάδα. Από τα μέσα βάρη της συστάδας συμπεραίνεται ότι το σημαντικότερο χαρακτηριστικό για την συστάδα είναι η επιρροή που έχει το προϊόν με σημαντικότητα 0,3166. Το 88% των καταναλωτών της συστάδας έχει ως μεγαλύτερη σημαντικότητα έναντι των υπολοίπων, την επιρροή που έχει το προϊόν.

Το λιγότερο σημαντικό χαρακτηριστικό για την συγκεκριμένη συστάδα είναι η άρωμα του λαδιού με συνολική σημαντικότητα 0,0964. Στην συνέχεια έρχονται η γεύση και το χρώμα με συνολικές σημαντικότητες 0,1241 και 0,1262 αντίστοιχα.

Το προφίλ των καταναλωτών της τρίτης συστάδας δίνει σημασία στην επιρροή που έχει το λάδι που επιλέγει να αγοράσει χωρίς να δίνει σημασία στην γεύση και το χρώμα του λαδιού. Το συγκεκριμένο προφίλ καταναλωτών είναι ένα αρκετά μικρό ποσοστό του δείγματος.

Η τέταρτη συστάδα αποτελεί πάλι μικρό κομμάτι του συνολικού δείγματος των καταναλωτών με μόλις 11,68% των συνολικών καταναλωτών. Από τα μέσα βάρη εξάγεται ότι οι καταναλωτές της συστάδας θεωρούν πιο σημαντικό χαρακτηριστικό την συσκευασία του προϊόντος με μέση σημαντικότητα 0,3184. Το σύνολο της συστάδας, δηλαδή το 100% των καταναλωτών της τέταρτης συστάδας, έχει ως πιο σημαντικό κριτήριο την συσκευασία του προϊόντος έναντι των άλλων κριτηρίων.

Ως λιγότερα σημαντικά κριτήρια για τους καταναλωτές της τέταρτης συστάδας είναι το άρωμα του λαδιού και η γεύση με μέσα βάρη 0,1073 και 0,1237 αντίστοιχα.

Το προφίλ των καταναλωτών της τέταρτης συστάδας πάλι αποτελεί μικρό ποσοστό του συνολικού αριθμού των καταναλωτών της έρευνας αλλά αποφασίζει το λάδι που θα αγοράσει από την συσκευασία του. Το συγκεκριμένο προφίλ καταναλωτών επηρεάζεται αρκετά από την εικόνα του προϊόντος παρά από τις αισθήσεις της όσφρησης και της γεύσης.

Η πέμπτη συστάδα αποτελείται από το 24,37% των συνολικών καταναλωτών του δείγματος καταναλωτών. Φαίνεται ότι από τα μέσα βάρη θεωρεί πιο σημαντικό το χαρακτηριστικό της τιμής με μέση σημαντικότητα 0,1993, αλλά πολύ κοντά είναι και η τιμή του μέσου βάρους για το χρώμα και το άρωμα με τιμές 0,1803 και 0,1775 αντίστοιχα. Από τα μέσα βάρη καθώς και από την κατανομή των καταναλωτών σε σχέση με το ποια σημαντικότητα είναι μεγαλύτερο έναντι των υπολοίπων παρατηρείται μια αντίστοιχη κατανομή. Δηλαδή το 42% των καταναλωτών της πέμπτης συστάδας έχει μεγαλύτερη τιμή βάρους για την τιμή του λαδιού, το 25% των καταναλωτών της συστάδας έχει μεγαλύτερη τιμή βάρους για το χρώμα του λαδιού και το 15% έχει μεγαλύτερη τιμή βάρους για το άρωμα του λαδιού, έναντι των υπολοίπων.

Το ίδιο φαινόμενο παρατηρείται και στα λιγότερο σημαντικά κριτήρια της πέμπτης συστάδας. Δηλαδή ως λιγότερο σημαντικό κριτήριο είναι αυτό της γεύσης με τιμή μέσου βάρους 0,1371 και στην συνέχεια με πολύ κοντινές τιμές είναι τα χαρακτηριστικά της επιρροής του προϊόντος και της συσκευασίας του λαδιού με τιμές 0,1526 και 0,1531 αντίστοιχα.

Συμπεραίνεται ότι η πέμπτη συστάδα δεν έχει ένα ξεκάθαρο προφίλ καταναλωτών αλλά ένα συνονθύλευμα από καταναλωτές που δεν καταναλωτών τόση σκέψη στην αγορά του λαδιού.

Color	Influence	Odour	Packaging	Price	Taste
13	22	10	29	90	33
7%	11%	5%	15%	46%	17%






Πίνακας: Το χαρακτηριστικό που θεωρούν πιο σημαντικό στην απόφαση τους συνολικά οι καταναλωτές

Συνολικά οι καταναλωτές στην απόφασή τους θεωρούν πιο σημαντική την τιμή κατά 46% και στην συνέχεια είναι η γεύση και η συσκευασία με 17% και 15% των καταναλωτών αντίστοιχα. Οι καταναλωτές που θεωρούν πιο σημαντικό κριτήριο για την απόφαση τους την επιρροή που έχει το προϊόν είναι σε ποσοστό 11% και οι καταναλωτές που αποφασίζουν με βάση το χρώμα ή το άρωμα ως πιο σημαντικό κριτήριο είναι 7% και 5% αντίστοιχα.

5.9 Συνδυασμός Συμπερασμάτων με τα Αποτελέσματα από το MARKEK

5.9.1 Βάρη Των Κριτηρίων

Πραγματοποιήθηκε δημιουργία συστάδων που έχουν τα ίδια χαρακτηριστικά με τις συστάδες οι οποίες δημιουργήθηκαν από τα βάρη των κριτηρίων από την ανάλυση με το SOM.

Cluster Name	Criteria	Analysis Method	Consumers Number	Actions
Cluster 1	Color (0.134, Non Significant)	Average Weights	45 (22.06%)	
	Odour (0.135, Non Significant)			
	Price (0.237, Significant)			
Cluster 2	Color (0.134, Non Significant)	Average Weights	38 (18.63%)	
	Taste (0.166, Significant)			
	Price (0.237, Non Significant)			
Cluster 3	Influence (0.161, Significant)	Average Weights	19 (9.31%)	
	Color (0.134, Non Significant)			
	Odour (0.135, Non Significant)			
	Taste (0.166, Non Significant)			
Cluster 4	Odour (0.135, Non Significant)	Average Weights	40 (19.61%)	
	Taste (0.166, Non Significant)			
	Packaging (0.168, Significant)			
Cluster 5	Taste (0.166, Non Significant)	Average Weights	66 (32.35%)	
	Price (0.237, Significant)			

Πίνακας: Συσταδοποίηση με το σύστημα MARKEK, με βάση τα κριτήρια

1	34,52%
2	20,81%
3	8,12%
4	11,68%
5	24,37%

Πίνακας: Ποσοστό καταναλωτών από την τμηματοποίηση των βαρών των κριτηρίων με SOM.

Παρατηρείται ότι με την ίδια σημαντικότητα στα χαρακτηριστικά, υπάρχει ομοιότητα στον αριθμό των καταναλωτών που ανήκουν σε κάθε συστάδα. Δηλαδή, για την πρώτη συστάδα με το χαρακτηριστικό της τιμής να έχει την μεγαλύτερη σημαντικότητα και τα χαρακτηριστικά του χρώματος και του αρώματος να μην είναι καθόλου σημαντικά, δημιουργείτε μια συστάδα που αποτελεί το 22,06% του συνολικού δείγματος των καταναλωτών. Υπάρχει μια διαφορά 12,46% στην πρώτη συστάδα που δημιουργήθηκε από την τμηματοποίηση με το SOM με ποσοστό 34,52%.

Για την δεύτερη συστάδα δεν παρατηρείται σχεδόν καμία διαφορά στον αριθμό των καταναλωτών που την αποτελούν, καθώς από την ανάλυση του MARKEX η δεύτερη συστάδα αποτελεί το 18,63% ενώ στην τμηματοποίηση από τον SOM είναι 20,81%. Τα κριτήρια που την χαρακτηρίζουν είναι η μεγάλη σημαντικότητα που δίνουν οι καταναλωτές στην γεύση και στην μικρή σημαντικότητα που δίνουν οι καταναλωτές στο χρώμα και στην τιμή.



Για την τρίτη συστάδα υπάρχει μια μικρή διαφορά στα ποσοστά των δύο αναλύσεων. Στην τμηματοποίηση που γίνεται από τον SOM η Τρίτη συστάδα αποτελεί το 8,12% των καταναλωτών ενώ από την ανάλυση του MARKEX το 9,31% των καταναλωτών. Τα κριτήρια που είναι δίνουν μεγάλη σημαντικότητα οι καταναλωτές είναι η επιρροή που έχει το προϊόν ενώ αυτά που δεν θεωρούν σημαντικά είναι το άρωμα, το χρώμα και η γεύση.

Για την τέταρτη συστάδα υπάρχει απόκλιση στον αριθμό των καταναλωτών που την απαρτίζουν. Η συσταδοποίηση που πραγματοποιήθηκε με το SOM δημιούργησε μια συστάδα με ποσοστό 11,68% των συνολικών καταναλωτών, ενώ με το MARKEX δημιουργήθηκε μια συστάδα με 19,61%. Τα κριτήρια που την χαρακτηρίζουν είναι ότι οι καταναλωτές θεωρούν μεγάλης σημαντικότητας την συσκευασία του προϊόντος και με καθόλου σημαντικότητα την γεύση και το άρωμα.

Για την πέμπτη συστάδα εμφανίζεται πάλι μια διαφορά στα ποσοστά των δύο συστάδων που τα κριτήρια που τις χαρακτηρίζουν είναι όμοια. Η τμηματοποίηση από το MARKEX δημιουργεί την πέμπτη συστάδα που αποτελείται από το 32,35% των συνολικών καταναλωτών σε αντίθεση με την ανάλυση από το SOM που αποτελείται από το 24,37%. Τα κριτήρια που την χαρακτηρίζουν είναι η μεγάλη σημαντικότητα στην τιμή και η καθόλου σημαντικότητα στην γεύση.

5.9.2 Ολικές Χρησιμότητες Καταναλωτών

Πραγματοποιήθηκε δημιουργία συστάδων που έχουν τα ίδια χαρακτηριστικά με τις συστάδες οι οποίες δημιουργήθηκαν από τις ολικές χρησιμότητες των κριτηρίων από την ανάλυση με το SOM.

Cluster 1 -Ολικές Χρησιμότητες	Taste (0.166, Non Significant)	Average Weights	62 (30.39%)	
	Packaging (0.168, Non Significant)			
Cluster 2 -Ολικές Χρησιμότητες-Μono τιμή	Price (0.237, Significant)	Average Weights	91 (44.61%)	

Πίνακας: Πρώτη Συστάδα: Καθόλου σημαντικά: Γεύση & συσκευασία, Δεύτερη Συστάδα σημαντικό κριτήριο: Τιμή

Έγινε εντοπισμός της σημαντικότητας των κριτηρίων σε κάθε συστάδα που δημιουργήθηκε από τις ολικές χρησιμότητες με τον SOM. Πραγματοποιήθηκε έπειτα προσπάθεια δημιουργίας τριών συστάδων, που περιλαμβάνουν την ίδια σημαντικότητα στα κριτήρια που περιέχουν και οι συστάδες από την ανάλυση με το SOM.

Πίνακας: Διαφορετικοί συνδυασμοί συστάδων με σημαντικό κριτήριο την τιμή του λαδιού.

Influence (0.161, Non Significant)	Average Weights	69 (33.82%)
Price (0.237, Significant)		
Color (0.134, Non Significant)	Average Weights	65 (31.86%)
Price (0.237, Significant)		
Odour (0.135, Non Significant)	Average Weights	60 (29.41%)
Price (0.237, Significant)		
Taste (0.166, Non Significant)	Average Weights	66 (32.35%)
Price (0.237, Significant)		
Packaging (0.168, Non Significant)	Average Weights	62 (30.39%)
Price (0.237, Significant)		

Η πρώτη συστάδα από την τμηματοποίηση με το SOM αποτελείται από το 31,22% των συνολικών καταναλωτών και από την τμηματοποίηση με την βοήθεια του MARKEX από το 30,39%. Τα κριτήρια που θεωρεί σημαντικά η συγκεκριμένη συστάδα δεν υπήρχε τρόπος να εντοπιστούν καθώς δεν υπήρχε κάτι ξεκάθαρο που χαρακτήριζε όλες τις επιλογές των προϊόντων των καταναλωτών που χαρακτηρίζονται από μεγάλο αριθμό ολικής χρησιμότητας. Για αυτό τον λόγο η εντοπίστηκαν τα κριτήρια που δεν δίνει καθόλου σημασία η συστάδα, που είναι η γεύση και η συσκευασία.

Η δεύτερη συστάδα αποτελείται από το 38,10% και η τρίτη από το 30,7% των συνολικών καταναλωτών της έρευνας. Η δεύτερη συστάδα καθώς και η τρίτη χαρακτηρίζονται από την μεγάλη σημαντικότητα που δίνουν στην τιμή του προϊόντος. Δεν είναι ξεκάθαρο όμως ούτε εφικτό να εντοπιστεί το κριτήριο που είναι μη σημαντικό και χαρακτηρίζει την μια από τις δύο. Οπότε παρουσιάζονται πέντε διαφορετικές συστάδες που ο συνδυασμός δύο εκ αυτών αντιπροσωπεύει τις συστάδες που δημιουργήθηκαν με την τμηματοποίηση του SOM.

Σε όλες τις παραπάνω εναλλακτικές το σημαντικό κριτήριο είναι η τιμή και διαφοροποιείται το κριτήριο που δεν θεωρούν σημαντικό. Όπως φαίνεται στο παραπάνω πίνακα δημιουργήθηκαν πέντε συστάδες όλες αποτελούμενες κοντά στο 30% των συνολικών καταναλωτών.

- Η πρώτη συστάδα με ποσοστό 33,82% δεν θεωρεί σημαντικό κριτήριο την επιρροή που έχει το προϊόν.
- Η δεύτερη συστάδα με ποσοστό 31,86% δεν θεωρεί σημαντικό κριτήριο το χρώμα του λαδιού.
- Η τρίτη συστάδα με ποσοστό 29,41% δεν θεωρεί σημαντικό κριτήριο το άρωμα του λαδιού.
- Η τέταρτη συστάδα με ποσοστό 32,35% δεν θεωρεί σημαντικό κριτήριο τη γεύση του λαδιού.
- Η πέμπτη συστάδα με ποσοστό 30,39% δεν θεωρεί σημαντικό κριτήριο την συσκευασία του λαδιού.

Όλες οι παραπάνω συστάδες είναι πιθανές να αντιπροσωπεύουν τις συστάδες που παρουσιάστηκαν κατά την ανάλυση του SOM, καθώς έχουν κριτήριο μεγάλης σημαντικότητας την τιμή και τα ποσοστά είναι παρόμοια, δηλαδή κοντά στο 30%.

Κεφάλαιο 6ο: Συμπεράσματα

Από την τμηματοποίηση της αγοράς από τα βάρη των κριτηρίων και από τις ολικές χρησιμότητες των προϊόντων των καταναλωτών προκύπτουν διάφορα συμπεράσματα.

Με την τμηματοποίηση της αγοράς με τις ολικές χρησιμότητες των καταναλωτών εξάγονται τα ανταγωνιστικά προϊόντα σε κάθε συστάδα.

Με την τμηματοποίηση της αγοράς με τα βάρη των κριτηρίων εξάγονται τα κριτήρια που είναι πιο σημαντικά για την κάθε συστάδα. Επίσης εξάγονται τα κριτήρια που είναι λιγότερο σημαντικά για κάθε συστάδα.

Με την εύρεση των κριτηρίων που είναι λιγότερο σημαντικά για κάθε συστάδα μπορεί να εφαρμοστεί μια μεθοδολογία που ονομάζεται «feature selection», επιλογή κριτηρίων.

Σκοπός της μεθόδου είναι η βελτίωση της απόδοσης του μοντέλου συσταδοποίησης χρησιμοποιώντας την μείωση των διαστάσεων του προβλήματος, την μείωση της πιθανότητας υπερεκπαίδευσης του νευρωνικού δικτύου και την καλύτερη ερμηνευσιμότητα του νευρωνικού δικτύου.

Θεωρώντας επίσης ότι οι συστάδες του δείγματος που χρησιμοποιούνται ανταποκρίνονται σε αντιστοιχία στην αγορά. Μια εταιρία θα πρέπει να εντοπίσει το προφίλ των πελατών που απευθύνεται και να στοχεύσει με την ανάλογη στρατηγική μάρκετινγκ ή την ανάλογη δημιουργία νέου προϊόντος.

Το νέο προϊόν κατέχει τα ίδια χαρακτηριστικά στα χαρακτηριστικά που δίνουν μεγαλύτερη σημαντικότητα οι καταναλωτές στα προϊόντα που διαλέγουν στις διαφορές συστάδες.

Η σύγκριση των συστάδων που δημιουργήθηκαν από το Markex και από τον SOM δείχνουν κοντινά αποτελέσματα. Το συμπέρασμα είναι ότι η τμηματοποίηση της αγοράς με το Νευρωνικό Δίκτυο Self Organizing Maps αποδίδει ορθά αποτελέσματα και μπορεί να χρησιμοποιηθεί κατά αυτό τον τρόπο.

Βιβλιογραφία

- Aggarwal, C.C., C. K. Reddy (2013), *Data Clustering: Algorithms and Applications*, CRC Press.
- Azadnia, A. H., Mat Saman, M. Z., Wong, K. Y., Ghadimi, P., & Zakuanc, N. (2012). *Sustainable supplier selection based on self-organizing map neural network and multi-criteria decision-making approaches*. *Procedia - Social and Behavioral Sciences*, 65, 879–884.
- Barone, A. (2023, October 19). *Marketing strategy: What it is, how it works, how to create one*. Investopedia. Retrieved from <https://www.investopedia.com/terms/m/marketing-strategy.asp>
- Bell, J. (2020), *Machine Learning: Hands-On for Developers and Technical Professionals* (2nd Ed.), John Wiley & Sons.
- Blei, D.M., A.Y. Ng, M.I. Jordan (2003), *Latent dirichlet allocation*, *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022.
- Bloom, J. Z. (2005). *Market segmentation: A neural network application*. *Annals of Tourism Research*, 32(1), 93–111. <https://doi.org/10.1016/j.annals.2004.05.001>
- Borden, N., & Marshall, M. (1959). *Advertising Management: Text and Cases*. Homewood, Ill, Richard D. Irwin.
- Chattopadhyay, M., Dan, P. K., Majumdar, S., & Chakraborty, P. S. *Application of Artificial Neural Network in Market Segmentation: A Review on Recent Trends*.
- Chen, J. (2022). *What is a neural network?* Investopedia. Retrieved from <https://www.investopedia.com/terms/n/neuralnetwork.asp#:~:text=A%20neural%20network%20is%20a,organic%20or%20artificial%20in%20nature>
- Chondromatidou, L. (2019). *Ταξινόμηση βίντεο με Αναδρομικά Νευρωνικά Δίκτυα* (dissertation).
- Chong, K. W. (2003). *The Role of Pricing in Relationship Marketing - A Study of the Singapore Heavy Equipment Spare*
- Christopher, V. V. (2020). *How does self-organizing algorithm work?* Medium. Retrieved from <https://medium.com/analytics-vidhya/how-does-self-organizing-algorithm-works-f0664af9bf04>

- Dodds, W. B., Monroe, K. B., & Grewal, D. (1991). Effects of price, brand, and store information on buyers' product evaluations. *Journal of marketing research*, 28(3), 307- 319.
- Duarte, V., Zuniga-Jara, S., & Contreras, S. (2022). *Machine Learning and Marketing: A Systematic Literature Review*. *IEEE Access*. Advance online publication.
- Grewal, D., Hulland, J., Kopalle, P. K., & Karahanna, E. (2020). The future of technology and marketing: A multidisciplinary perspective. *Journal of the Academy of Marketing Science*, 48, 1-8.
- Jacquet-Lagrèze, E., & Siskos (1982), Assessing a set of additive utility functions for multicriteria decision making: The UTA method, *European Journal of Operational Research*, no.10 (151-164)
- Kiang, M. Y., Hu, M. Y., & Fisher, D. M. (2004). An extended self-organizing map network for market segmentation—a telecommunication example. *Expert Systems with Applications*, 27(2), 261-270.
- Kohonen, T., & Honkela (2007). Kohonen Network. *Scholarpedia*. 2 (1): 1568. Bibcode: 2007SchpJ...2.1568K. doi:10.4249/scholarpedia.1568.
- Kohonen, T.: Automatic formation of topological maps of patterns in a self-organizing system. In: *Proc. 2nd Scand. Conf. on Image Analysis*, pp. 214–220, Oja, E., Simula, O. (eds.).
- Kotler, P., Keller, L. K., Brady, M., Goodman, M., & Hansen, T. (2000). *Marketing management and consumer behavior*.
- Li, S. (n.d.). Illustration of neuron structure of the neural network. From https://researchgate.net/figure/Illustration-of-neuron-structure-of-the-neural-network_fig2_329990010
- Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. *Physics Procedia*, 25, 1104–1109. doi:10.1016/j.phpro.2012.03.206
- Little, J.D.C.(1970), Models and managers : The concept of decision calculus, *Management Science*, vol.16, no. 8, pp.466-489.
- Liu, B. (2015), *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, Cambridge University Press.
- Liu, J., Liao, X., Huang, W., & Liao, X. (2019). Market segmentation: A multiple criteria approach combining preference analysis and segmentation decision. *Omega*, 83, 1–13.

- Matsatsinis, N., & Siskos, Y. (1999). "MARKEX: An intelligent decision support system for product development decisions". *European Journal of Operational Research*, 113(2), pp. 336-354.
- Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.
- Morgan, N., Whitler, K., Feng, H., & Chari, S. (2019). Research in marketing strategy. *Journal of the Academy of Marketing Science*, σσ. 4-29
- Nilashi, M., Mardani, A., Liao, H., Ahmadi, H., Abdul Manaf, A., & Almkadi, W. (2019). A Hybrid Method with TOPSIS and Machine Learning Techniques for Sustainable Development of Green Hotels Considering Online Reviews. *Sustainability*, 11(21), 5908.
- Ning, X., & Lovell, M.R. (2002). On the Sliding Friction Characteristics of Unidirectional Continuous FRP Composites. *ASME J. Tribol.*, v124(1), pp. 5-13. doi: 10.1115/1.1398295
- Palmer, A. (2004), *Introduction to Marketing: Theory and Practice*, Oxford University Press.
- Perceptrons and multi-layer perceptrons. IndianTechWarrior. (2022, June 5). Retrieved from <https://indiantechwarrior.com/perceptrons-and-multi-layer-perceptrons/>
- Peter, J. P., & Olson, J. C. (2010). *Consumer Behavior & Marketing Strategy*.
- Ralhan, A. (2018). Self organizing maps. Medium. Retrieved from <https://medium.com/@abhinavr8/self-organizing-maps-ff5853a118d4>
- Robert L. Thorndike (1953). "Who Belongs in the Family?". *Psychometrika*. 18 (4): 267–276.
- Rosenblatt, F. (1958) The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65, 386.
- Roy, B. (1976), From optimization to multicriteria decision aid: Three main operational attitudes, in: Thiriez, H. and S. Zionts (eds.), *Multiple Criteria Decision Making*, Springer-Verlag, 130, pp. 1-32.
- Roy, B. (1990), The outranking approach and the foundations of ELECTRE methods, in C.A. Bana e Costa (ed.), *Readings in Multiple Criteria Decision Aid*, Springer-Verlag, pp.155-183.
- Saeed, M. (2022). Self-organizing maps: Theory and implementation in python with NumPy. Stack Abuse. Retrieved November 2, 2022, from

<https://stackabuse.com/self-organizing-maps-theory-and-implementation-inpython-with-numpy/>

- Sagar Sharma, Activation Functions in neural Networks, Towards Data Science (2017)
- Singh, M. (2012). Marketing Mix of 4P'S for Competitive Advantage. *IOSR Journal of Business and Management (IOSRJBM)*, 3(6), 40-45.
- Siskos, J., D. Yannacopoulos (1985), UTAStar: An ordinal regression method for building additive value functions, *Investigacao Operacional*, vol. 5, no. 18, pp.39-53
- Siskos, Y., E. Grigoroudis, N.F. Matsatsinis (2005), Outranking theory and the UTA methods, in: J. Figueira, S. Greco, M. Ehrgott (eds.), *Multiple Criteria Decision Analysis, - State of the Art – Surveys*, International Series in Operations Research and Management Science, pp.297-344, Springer.
- Siskos, Y., E. Grigoroudis, N.F. Matsatsinis (2016), UTA methods, in: S. Greco, M. Ehrgott, J. Figueira (eds.), *Multiple Criteria Decision Analysis, - State of the Art – Surveys (2nd Edition)*, International Series in Operations Research and Management Science, vol. I, pp. 315-362, Springer.
- Siskos, Y., N.F. Matsatsinis, G. Baourakis (2001), Multicriteria analysis in agricultural marketing: The case of French olive oil market, *European Journal of Operational Research*, vol. 130, no. 2, 2001, pp. 315-331, ([https://doi.org/10.1016/S0377-2217\(00\)00043-6](https://doi.org/10.1016/S0377-2217(00)00043-6)).
- study. *International Journal of Social Sciences & Educational Studies*, 4(4).
- Tan, P. N., M. Steinbach, V. Kumar (2014), *Introduction to Data Mining*, Pearson.
- Thabit, T. H., & Younus, S. Q. (2015). The Relationship between Added-Value of Organization and QMA, *International Journal of Engineering Research and Management Technology*, 2(2), 35- 42.
- Thabit, T., & Raewf, M. (2018). The evaluation of marketing mix elements: A case
- Thomson, R. E., & Emery, W. J. (n.d.). Self organizing map. *Self Organizing Map - an overview | ScienceDirect Topics*. Retrieved from <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/self-organizing-map>
- Tung, C.Y. (1982). *Evaporative Heat Transfer in the Contact Line of a Mixture*. Ph.D. Thesis, Rensselaer Polytechnic Institute, Troy, NY.

- Ultsch, Alfred. (2003). U*Matrix: a Tool to visualize Clusters in high dimensional Data.
- Wang, C. (2022). *Efficient customer segmentation in digital marketing using deep learning with swarm intelligence approach*. *Information Processing and Management*, 59, 103085.
- Wedel, M., & Kamakura, W. A. (2000). *Market Segmentation: Conceptual and Methodological Foundations*. Kluwer Academic Publishers.
- Wikimedia Foundation. (2020, July 5). U-matrix. Wikipedia. Retrieved from [https://en.wikipedia.org/wiki/U-matrix#:~:text=The%20U%2Dmatrix%20\(unified%20distance,space%20using%20a%202D%20image](https://en.wikipedia.org/wiki/U-matrix#:~:text=The%20U%2Dmatrix%20(unified%20distance,space%20using%20a%202D%20image)
- Wikimedia Foundation. (2022). Self-organizing map. Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Self-organizing_map
- Ακακιάδου , Γ. (2007). Μελέτη του αλγορίθμου ομαδοποίησης k-means σε δεδομένα του παγκόσμιου ιστού (dissertation).
- Βασίλειου, Κ. (2021). Αναδρομικά νευρωνικά δίκτυα για πρόβλεψη χρηματιστηριακών μεγεθών (dissertation).
- Κυρίτσης Κ. (2014,Μάιος). Νευρωνικά Δίκτυα και Μηχανές Διανυσματικής Υποστήριξης (dissertation).
- Ματσατσίνης, Ν. (2010), Συστήματα Υποστήριξης Αποφάσεων, Εκδόσεις Νέων Τεχνολογιών, Αθήνα.
- Ματσατσίνης, Ν. Φ. (n.d.). Ανάλυση Συμπεριφοράς Καταναλωτών. Χανια; Πολυτεχνειούπολη, Κουνουπιδιανά, 73100.