



Πολυτεχνείο Κρήτης

Σχολή Μηχανικών Παραγωγής και Διοίκησης

**Ανάπτυξη μεθοδολογίας
τμηματοποίησης της αγοράς με χρήση
μεθόδων πολυκριτήριας ανάλυσης,
εξόρυξης δεδομένων και ανάλυσης
κειμένων σε δεδομένα μέσω
κοινωνικής δικτύωσης και
διαδικτύου**

Διπλωματική Εργασία

Ζήσης Γιώργος

Επιβλέπων

Ματσατσίνης Νικόλαος, Καθηγητής

Χανιά, Φεβρουάριος 2023

Ευχαριστίες

Με την ολοκλήρωση της διπλωματικής μου εργασίας, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όσους συνέβαλαν στην εκπόνησή της. Αρχικά, στον επιβλέποντα καθηγητή, Νικόλαο Ματσατσίνη, για τη συνεχή καθοδήγηση και τις πολύτιμες συμβουλές που μου προσέφερε καθ' όλη τη διάρκεια της εκπόνησης της διπλωματικής εργασίας.

Επιπλέον, θα ήθελα να ευχαριστήσω την υποψήφια διδάκτορα Καλαφάτη Φωτεινή, που συνέβαλε στην ολοκλήρωση του έργου με την τακτική και συνεπή της παρουσία και την επίλυση των ποικίλων εμποδίων που εμφανίστηκαν. Καθώς επίσης και όλους τους καθηγητές του Πολυτεχνείου Κρήτης, που προσέφεραν τις γνώσεις τους, τόσο σε θεωρητικό, όσο και σε πρακτικό επίπεδο όλα αυτά τα χρόνια.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου, για την προσπάθεια που έκαναν όλοι τους ξεχωριστά όλα αυτά τα χρόνια μέχρις ότου ολοκληρώσω τις σπουδές μου, καθώς και όλο τον κόσμο που συνέβαλε με το δικό του, μικρό αλλά σημαντικό τρόπο, συμπληρώνοντας το ερωτηματολόγιο για τη συλλογή των δεδομένων.

Σας ευχαριστώ όλους πολύ και τον καθένα ξεχωριστά!

Περίληψη

Από τις αρχές του 21ου αιώνα, ο τομέας της επιστήμης των δεδομένων αναπτύσσεται ολοένα και περισσότερο. Σε αυτή την ανάπτυξη οδήγησε η ύπαρξη μεγάλων όγκων δεδομένων και η ανάγκη για ανάλυσή τους. Η αύξηση των δεδομένων οφείλεται αφενός μεν στην εξέλιξη των Τεχνολογιών Πληροφορικής και Επικοινωνιών (ΤΠΕ) και από την άλλη στη ραγδαία ανάπτυξη της χρήσης του διαδικτύου και ιδιαίτερα των μέσων κοινωνικής δικτύωσης. Οι εξελίξεις αυτές επηρεάζουν τη λειτουργία των οργανισμών και των επιχειρήσεων, με αποτέλεσμα να κρίνεται απαραίτητη η αξιοποίηση των νέων τεχνολογιών και μεθόδων ανάλυσης δεδομένων με στόχο τη βελτίωση της λήψης αποφάσεων.

Ένα από τα θέματα που απασχολούν ιδιαίτερα τους οργανισμούς είναι η διερεύνηση και ανάλυση της συμπεριφοράς των καταναλωτών. Αυτό βέβαια δεν σημαίνει πως η εφαρμογή των προαναφερθέντων μεθόδων βρίσκουν εφαρμογή μόνο στο κομμάτι του marketing. Ειδικότερα, οπουδήποτε υπάρχει μεγάλος όγκος αναξιοποίητων δεδομένων μπορούν να τεθούν σε εφαρμογή αντίστοιχες τεχνικές και μεθοδολογίες. Ο βασικότερος λόγος που τα analytics αποτελούν ένα από τα φλέγοντα ζητήματα συζήτησης είναι λόγω των πολλών πλεονεκτημάτων που μπορούν να προσφέρουν τόσο στις επιχειρήσεις, όσο και στα στελέχη που τα αξιοποιούν για την ορθότερη και αποτελεσματικότερη λήψη αποφάσεων.

Στόχος της εργασίας αυτής λοιπόν, είναι η ανάπτυξη μίας μεθοδολογίας τμηματοποίησης αγορών, η οποία βασίζεται σε αναλύσεις δεδομένων. Οι αναλύσεις αυτές προέρχονται από διαφορετικές κατηγορίες μεθόδων. Στην πρώτη κατηγορία ανήκουν οι μέθοδοι εξόρυξης γνώσης από δεδομένα, όπως για παράδειγμα ο αλγόριθμος K-means, που χρησιμοποιήθηκε και στην προκειμένη περίπτωση, που έχει ως στόχο την συσταδοποίηση των δεδομένων, έτσι ώστε τα κοινά ή παρόμοια δεδομένα να ομαδοποιηθούν μαζί. Η επόμενη κατηγορία είναι αυτή των πολυκριτήριων μεθόδων, στις οποίες ανήκει και η UTASTAR. Μέσω της UTASTAR, προέκυψαν τα βάρη των κριτηρίων, γνωστοποιώντας έτσι τα χαρακτηριστικά που είναι πιο σημαντικά για τους καταναλωτές. Με βάση αυτά, στην συνέχεια αναπτύχθηκε ένα νέο προϊόν που θα μπορούσε να ανταγωνιστεί τα ήδη υπάρχοντα και να εισαχθεί στην αγορά.

Η αγορά στην οποία επικεντρώθηκε η παρούσα διπλωματική εργασία είναι αυτή των συσκευασμένων γιαουρτιών και των επιδόρπιων γιαουρτιού στην Ελλάδα. Εν κατακλείδι, τα προς ανάλυση δεδομένα αποκτήθηκαν μέσω ερωτηματολογίων που διαμοιράστηκαν ηλεκτρονικά εν μέσω της πανδημίας του covid-19 και τα τελικά

αποτελέσματα θα συγκριθούν από αντίστοιχη κλαδική έρευνα που έγινε την προηγούμενη χρονιά.

Πίνακας Περιεχομένων

Ευχαριστίες.....	2
Περίληψη.....	2
Πίνακας Περιεχομένων	4
Κεφάλαιο 1 Εισαγωγή.....	6
1.1. Σκοπός της εργασίας.....	6
1.2. Δομή της εργασίας	7
Κεφάλαιο 2 Υφιστάμενη κατάσταση	8
2.1. Εισαγωγή	8
2.2. Το μοντέλο συμπεριφοράς των καταναλωτών	8
2.3. Τα ράφια των καταστημάτων.....	10
2.4. Παρουσίαση των προϊόντων και της αγοράς τους.....	10
2.4.1. Παρουσίαση προϊόντων και των χαρακτηριστικών τους.....	11
2.4.2. Η αγορά των γιουρτιών στην Ελλάδα	12
Κεφάλαιο 3 Η Μεθοδολογία	13
3.1. Εισαγωγή	13
3.2. Περιγραφική ανάλυση (διάγραμμα ροής).....	14
3.3. Το ερωτηματολόγιο της έρευνας.....	15
3.4. Καθαρισμός και προ - επεξεργασία δεδομένων	15
3.4.1. Ελλιπείς Τιμές.....	15
3.4.2. Ακραίες Τιμές	16
3.4.3. Συντελεστής Συσχέτισης/Kendall's Tau.....	18
3.4.4. Δείκτης Σκιαγράφησης/Silhouette	18
3.5. Στατιστική περιγραφή.....	19
3.6. Οι μέθοδοι που χρησιμοποιήθηκαν.....	19
3.6.1. Πολυκριτήριες μέθοδοι.....	20
3.6.2. Μέθοδοι εξόρυξης δεδομένων	32
Κεφάλαιο 4 Η συλλογή και η προ – επεξεργασία των δεδομένων	46
4.1. Εισαγωγή	46
4.2 Το ερωτηματολόγιο της έρευνας.....	46
4.2.1. Τρόπος σύνταξης ερωτηματολογίου.....	46
4.2.2. Η ανάλυση του ερωτηματολογίου	47
4.3. Τυχαίες και ελλιπείς τιμές	50
4.4. Συντελεστής συσχέτισης τ του Kendall.....	51
4.5. Ακραίες τιμές.....	52

4.6. Δείκτης σκιαγράφησης Silhouette	54
Κεφάλαιο 5 Εφαρμογή μεθόδων	56
5.1. Εισαγωγή	56
5.2. Η στατιστική περιγραφή	56
5.3. Εφαρμογή της UTASTAR	64
5.4. Συσταδοποίηση με χρήση αλγορίθμων του MARKEK	64
5.5. Ο αλγόριθμος K-Means	65
5.6. Ο αλγόριθμος DBScan	70
5.7. Ο αλγόριθμος EM (Expectation – Maximization)	71
5.8. Ο αλγόριθμος Hierarchical Clusterer	73
5.9. Η μέθοδος Canopy	77
5.10. Η μέθοδος UTASTAR	78
5.10.1. Μελέτη πρώτης ομάδας (Συσταδοποίηση με τον K-Means)	81
5.10.2. Μελέτη δεύτερης ομάδας (Συσταδοποίηση με τον K-Means)	85
5.10.3. Μελέτη πρώτης ομάδας (Συσταδοποίηση με τον DBScan)	88
5.10.4. Μελέτη δεύτερης ομάδας (Συσταδοποίηση με τον DBScan)	91
5.10.5. Μελέτη τρίτης ομάδας (Συσταδοποίηση με τον DBScan)	95
5.10.6. Μελέτη τέταρτης ομάδας (Συσταδοποίηση με τον DBScan)	98
5.10.7. Μελέτη πρώτης ομάδας (Συσταδοποίηση με τον EM)	101
5.10.8. Μελέτη δεύτερης ομάδας (Συσταδοποίηση με τον EM)	104
5.10.9. Μελέτη πρώτης ομάδας (Συσταδοποίηση με τον Hierarchical Clusterer)	107
5.10.10. Μελέτη δεύτερης ομάδας (Συσταδοποίηση με τον Hierarchical Clusterer)	110
5.10.11. Μελέτη τρίτης ομάδας (Συσταδοποίηση με τον Hierarchical Clusterer)	113
Κεφάλαιο 6 Σύγκριση μεθόδων συσταδοποίησης	115
Κεφάλαιο 7 Συμπεράσματα	117
Βιβλιογραφία	118
Παράρτημα	121
1. Ερωτηματολόγιο	121
2. Αποτελέσματα UTASTAR για τις ολικές χρησιμότητες των εναλλακτικών	128
3. Αποτελέσματα UTASTAR για τα βάρη των κριτηρίων	132

Κεφάλαιο 1 Εισαγωγή

Σύμφωνα με έρευνα που διενεργήθηκε από το Ίδρυμα Οικονομικών και Βιομηχανικών Ερευνών (“Υιοθέτηση των ΤΠΕ και Ψηφιακή Ανάπτυξη στην Ελλάδα”, 2014), η ραγδαία ανάπτυξη της ψηφιακής τεχνολογίας, και γενικότερα των Τεχνολογιών Πληροφορικής και Επικοινωνιών, τόσο στην Ελλάδα, όσο και παγκοσμίως, έχει προσφέρει αρκετές νέες δυνατότητες τόσο στις επιχειρήσεις και στον δημόσιο τομέα, όσο και στον κόσμο που τις αξιοποιεί. Οι νέες αυτές δυνατότητες και ευκαιρίες που προσφέρονται, αποτελούν έξυπνες τακτικές ανάπτυξης και συμβαδίζουν πλήρως με τις νόρμες της βιωσιμότητας. Σύμφωνα με την Ευρωπαϊκή Επιτροπή (Digital Economy and Society Index (DESI), 2022), προκειμένου να μετρηθεί αυτή η ψηφιακή πρόοδος, αναπτύχθηκε ο δείκτης Ψηφιακής Οικονομίας και Κοινωνίας (DESI). Οι εκθέσεις DESI που αναπτύσσονται βάσει των δεδομένων της προηγούμενης χρονιάς αναφέρουν πως εν μέσω της πανδημίας το 33 - 44% των θέσεων εργασίας επιτρέπουν την τηλεργασία. Επιπλέον, το ηλεκτρονικό εμπόριο και η αυτοματοποίηση σημείωσαν σημαντική αύξηση. Παρόλο αυτά, λόγω έλλειψης ψηφιακών δεξιοτήτων των εργαζομένων και των πολιτών κρίνεται δύσκολη η μελλοντική ανάπτυξη. Για αυτόν ακριβώς τον λόγο, τα τελευταία χρόνια γίνεται ολοένα και μεγαλύτερη συσσώρευση των δεδομένων, χωρίς την ορθότερη και αποτελεσματικότερη αξιοποίησή αυτών.

Η συσσώρευση των δεδομένων αυτών, έφεραν στο προσκήνιο ένα νέο σχετικά φαινόμενο, αυτό των Μεγάλων Δεδομένων (Big Data), όπως ονομάζεται. Σύμφωνα με τους Τσακανίκας κα. (2014), κάθε ένα λεπτό παράγονται περίπου δύο τρισεκατομμύρια bytes δεδομένων, ενώ σημαντικό είναι να αναφερθεί ότι τα τελευταία χρόνια τα ψηφιακά δεδομένα που έχουν παραχθεί είναι περισσότερα από αυτά που έχουν παραχθεί στην υπόλοιπη ανθρώπινη ιστορία. Βάσει υπολογισμών, ο ρυθμός ανάπτυξης των δεδομένων είναι ίσος με 40%. Παρόλο αυτά όμως, στην Ελλάδα τα επίπεδα εκμετάλλευσης των δεδομένων είναι ακόμα χαμηλά σε σύγκριση με τις υπόλοιπες ευρωπαϊκές χώρες. Η σύγκριση αυτή γίνεται χάρη σε συγκεκριμένους δείκτες ανάπτυξης, που διενεργήθηκαν από το Ίδρυμα Οικονομικών και Βιομηχανικών Ερευνών (“Υιοθέτηση των ΤΠΕ και Ψηφιακή Ανάπτυξη στην Ελλάδα”, 2014), εκ των οποίων μόλις στους είκοσι δύο από αυτούς, η Ελλάδα βρίσκεται πάνω από τον ευρωπαϊκό μέσο όρο. Το συνολικό πλήθος των δεικτών αυτών είναι ίσο με ογδόντα τέσσερις και σχετίζονται κυρίως με την ευρυζωνική διείσδυση, τις συναλλαγές μέσω διαδικτύου και τη συχνότητα χρήσης του.

1.1. Σκοπός της εργασίας

Σκοπός της διπλωματικής αυτής είναι η ανάπτυξη μίας μεθοδολογίας τμηματοποίησης αγοράς, η οποία θα παρέχει ορθή και έγκυρη πληροφόρηση σχετικά με τη συμπεριφορά και τις προτιμήσεις των καταναλωτών. Στην προκειμένη περίπτωση, ως παράδειγμα χρησιμοποιήθηκε η αγορά του γιαουρτιού και των επιδόρπων γιαουρτιού.

1.2. Δομή της εργασίας

Έχοντας ως στόχο τη διευκόλυνση των αναγνωστών, στη συνέχεια παρουσιάζεται η δομή ολόκληρης της εργασίας. Η παρούσα διπλωματική εργασία αποτελείται συνολικά από επτά κεφάλαια, μέσα στα οποία συνυπολογίζονται τόσο το κεφάλαιο της “Εισαγωγής”, όσο και αυτό των συμπερασμάτων, ενώ στη συνέχεια παρουσιάζονται και τα αντίστοιχα παραρτήματα μαζί με τη βιβλιογραφία. Ακολούθως, αναλύεται περιεκτικά το καθένα κεφάλαιο ξεχωριστά, δίνοντας μία σύντομη περιγραφή αυτού. Στο πρώτο κεφάλαιο, ο αναγνώστης ουσιαστικά πληροφορείται για το θέμα της διπλωματικής εργασίας, αποκτώντας μία πολύ γενική εικόνα για το τι ακριβώς πρόκειται να ακολουθήσει, ενώ παράλληλα ενημερώνεται για τις γενεσιουργές αιτίες αυτού. Επόμενο στάδιο, είναι η παρουσίαση της αγοράς που μελετάται, τόσο ως προς την ποικιλομορφία των προϊόντων που εμφανίζει, όσο και ως προς τις τάσεις και τις συνήθειες της εποχής και των καταναλωτών κατ’ επέκταση. Στο δεύτερο κεφάλαιο επιπλέον, παρουσιάζονται και κάποια βασικά κριτήρια τα οποία λαμβάνονται υπόψη από τους καταναλωτές πριν την αγορά κάποιου προϊόντος. Εν συνεχεία, έχοντας πλέον αποκτήσει μία καλή εικόνα του προς μελέτη θέματος, στο τρίτο κεφάλαιο παρουσιάζεται η μεθοδολογία της διπλωματικής εργασίας. Ειδικότερα, γίνεται αναφορά τόσο στην ανάπτυξη του ερωτηματολογίου και τη συγκέντρωση των δεδομένων, όσο και στο θεωρητικό υπόβαθρο των μεθόδων που χρησιμοποιήθηκαν κατά την προ – επεξεργασία αλλά και αργότερα. Στο τέταρτο κεφάλαιο, γίνεται αναλυτική παρουσίαση των βημάτων της προ – επεξεργασίας στην πράξη, ενώ στο πέμπτο κεφάλαιο παρουσιάζονται οι μέθοδοι πάνω στα δεδομένα που συγκεντρώθηκαν. Στο τελευταίο κεφάλαιο παρουσιάζονται τα συμπεράσματα και τα τελικά αποτελέσματα της έρευνας σε συγκεντρωτική μορφή. Τέλος, ακολουθεί η σχετική βιβλιογραφία, πάνω στην οποία βασίστηκε η έρευνα, μαζί με τα σχετικά παραρτήματα.

Κεφάλαιο 2 Υφιστάμενη κατάσταση

2.1. Εισαγωγή

Από τα παλαιά χρόνια, φαίνεται πως τα γαλακτοκομικά προϊόντα αποτελούν αναπόσπαστο κομμάτι της διατροφής του ανθρώπου, ανεξαρτήτως ηλικίας, φύλου και καταγωγής. Ειδικότερα, τα γιαούρτια, όπως και τα επιδόρπια γιαουρτιού, φαίνεται να έχουν αυξητική τάση τα τελευταία χρόνια ως προς τη ζήτηση τους (Google Trends). Λόγω της αυξημένης αυτής ζήτησης και σύμφωνα με την ενδιαμέση εκθέση της κλαδικής έρευνας (2020) που υλοποίησε η επιτροπή ανταγωνισμού, πολλές αλυσίδες supermarket καταφεύγουν στην τοποθέτηση δικών τους προϊόντων στα ράφια, δίνοντας ακόμα περισσότερες επιλογές στους καταναλωτές, και αυξάνοντας αυτόματα τον ανταγωνισμό στην αγορά. Ο έντονος αυτός ανταγωνισμός επηρεάζει ως έναν βαθμό τις ήδη υπάρχουσες ισορροπίες στα μερίδια αγοράς, καθώς τα συγκεκριμένα προϊόντα δεν είναι ιδιαιτέρως ακριβά και ο καταναλωτής μπορεί να αλλάξει εύκολα εταιρείες. Για αυτόν ακριβώς τον λόγο, η συγκεκριμένη διπλωματική εργασία στοχεύει στην εύρεση όλων των παραγόντων που επηρεάζουν τους καταναλωτές, καθώς και να σκιαγραφήσει όσο το δυνατόν καλύτερα τη συμπεριφορά αυτών, δημιουργώντας μία επιτυχημένη στρατηγική για την προώθηση των εκάστοτε προϊόντων ή και την εισαγωγή ενός νέου προϊόντος στην αγορά.

2.2. Το μοντέλο συμπεριφοράς των καταναλωτών

Προσπαθώντας να αναλυθεί η συμπεριφορά του καταναλωτή, θα πρέπει πρώτα να αναφερθούν τα χαρακτηριστικά των προϊόντων που μελετάμε, καθώς και πως αυτά μεταβάλλονται με το πέρασμα του χρόνου και τις όποιες κοινωνικο – οικονομικές αλλαγές. Από την προηγούμενη δεκαετία, ίσως και λίγο πιο νωρίς, η ελληνική κοινωνία δοκιμάστηκε αρκετά και κυρίως στον οικονομικό της τομέα, λόγω της οικονομικής κρίσης που μάστιζε τη χώρα. Οι συνθήκες αυτές ανάγκασαν τους καταναλωτές να καταφύγουν στην αγορά φθηνότερων προϊόντων και ειδών πρώτης ανάγκης, αδιαφορώντας για την ποιότητά τους. Το φαινόμενο αυτό αποτέλεσε την αρχή για την ανάπτυξη και την εισαγωγή πολλών νέων προϊόντων ιδιωτικής ετικέτας στα ράφια των supermarket. Με το πέρασμα του χρόνου, οι καταναλωτές άρχισαν να αγοράζουν και να εμπιστεύονται τα προϊόντα αυτά, αλλάζοντας τα μερίδια αγοράς στο χώρο της γαλακτοβιομηχανίας και των γιαουρτιών κατ' επέκταση. Πέραν από την οικονομική κρίση, ο πληθωρισμός και ο μειωμένος οικογενειακός προϋπολογισμός συνέβαλαν στην αγορά προϊόντων ιδιωτικής ετικέτας. Εν συνεχεία, αφήνοντας πίσω τα τελευταία χρόνια, και παίρνοντας ως σημείο αναφοράς το τέλος του 2021 και τις αρχές του 2022, φαίνεται πως η τιμή των ειδών πρώτης ανάγκης (και των γαλακτοκομικών προϊόντων) στα ράφια των supermarket αυξήθηκε, λόγω της αύξησης της τιμής του πετρελαίου. Όπως γίνεται αντιληπτό, το κριτήριο της τιμής παίζει σίγουρα καταλυτικό ρόλο σε αυτή τη μελέτη, καθώς είναι άρρηκτα συνδεδεμένο με την προτίμηση των καταναλωτών και τις κοινωνικο – οικονομικές συνθήκες που επικρατούν στην χώρα.

Επιπρόσθετα, τα τέσσερα επόμενα χαρακτηριστικά που θα αναλυθούν στην παρούσα έρευνα σχετίζονται περισσότερο με τις προσωπικές προτιμήσεις του καταναλωτή.

- Το κριτήριο της γεύσης. Δηλαδή, η προτίμηση του καταναλωτή μπορεί να διαφέρει ανάλογα με το είδος του γάλακτος που χρησιμοποιείται στο εκάστοτε γιαούρτι. Επί παραδείγματι, το αγελαδινό γιαούρτι διαφέρει από το πρόβειο, όπως και από αυτό που προέρχεται από σκόνη γάλακτος.
- Το κριτήριο του αρώματος. Το άρωμα του εκάστοτε γιαουρτιού διαφέρει ανάλογα την πρώτη ύλη που χρησιμοποιείται. Στο πρόβειο για παράδειγμα, η οσμή του γάλακτος είναι πιο έντονη από ένα προϊόν γιαουρτιού που προέρχεται από σκόνη γάλακτος.
- Το κριτήριο της υφής. Στα πρόβεια παραδοσιακά γιαούρτια σχηματίζεται υμένιο (πέτσα) στο πάνω μέρος της γιαούρτης, ενώ στα αγελαδινά στραγγιστά όχι.
- Το κριτήριο της προέλευσης του γάλακτος. Δηλαδή, ίσως είναι πιο σημαντικό για τον ίδιο να αγοράσει ένα καθαρά ελληνικό προϊόν, στο οποίο οι πρώτες ύλες είναι εγχώριες, παρά ένα προϊόν που παράγεται ή επεξεργάζεται σε ξένη χώρα.

Συνεπώς, γίνεται αντιληπτό πως η δύναμη της συνήθειας και οι προσωπικές επιλογές είναι αυτές που παίζουν σημαντικότερο ρόλο στην αξιολόγηση αυτών των τεσσάρων κριτηρίων. Έτσι, καθίσταται δύστοκη η ερμηνεία των καταναλωτικών προτιμήσεων.

Τα τελευταία δύο κριτήρια που λαμβάνονται υπόψη στη συγκεκριμένη έρευνα, σχετίζονται περισσότερο με τις αντιλήψεις και τις πεποιθήσεις των καταναλωτών. Με το πέρασμα του χρόνου, την ολοένα πιο ένθερμη, και απαραίτητη, ευαισθητοποίηση των ανθρώπων για τα περιβαλλοντικά ζητήματα του πλανήτη, καθώς και τη μόλυνση του περιβάλλοντος, τόσο από αέριους ρύπους, όσο και από τα πλαστικά, πολλοί είναι εκείνοι οι καταναλωτές που αναζητούν και ενστερνίζονται πιο φιλικές ως προς το περιβάλλον συνήθειες. Χαρακτηριστικό παράδειγμα αποτελεί, η αντικατάσταση του πλαστικού της συσκευασίας γιαουρτιού (κεσεδάκι), σε πήλινο. Με τον τρόπο αυτό, η συσκευασία μπορεί να χρησιμοποιηθεί εκ νέου, ή και για μία διαφορετική χρήση από τον καταναλωτή. Εναλλακτικό παράδειγμα είναι, τόσο η χρήση ανακυκλώσιμου πλαστικού στην συσκευασία, όσο και η μείωση του πλαστικού σε αυτή. Τελευταίο κριτήριο της μελέτης είναι η επωνυμία (brand name). Πολύ συχνό είναι το φαινόμενο οι καταναλωτές να αγοράζουν ένα συγκεκριμένο προϊόν από μία εταιρεία, χωρίς να λαμβάνουν υπόψη άλλα κριτήρια. Το φαινόμενο αυτό μπορεί να συμβαίνει ανεξαρτήτως αν το συγκεκριμένο προϊόν να είναι το καλύτερο στην αγορά και αυτομάτως γίνεται πιο δύσκολο το έργο των ειδικών για την ερμηνεία των καταναλωτικών συνηθειών.

Συνοψίζοντας, η κατανόηση της συμπεριφοράς των καταναλωτών είναι μία αρκετά δύσκολη υπόθεση και επηρεάζεται από αρκετούς παράγοντες. Οι παράγοντες αυτοί, μπορεί να είναι τόσο υποκειμενικοί, και να βασίζονται στις προσωπικές αντιλήψεις και συνήθειες, όσο και πιο αντικειμενικοί ώστε να μπορούν να προκύψουν μέσα από τεκμήρια και υπαρκτά δεδομένα.

2.3. Τα ράφια των καταστημάτων

Όσον αφορά τώρα τις αγορές των προϊόντων γιαουρτιού και επιδόρππων γιαουρτιού, φαίνεται πως υπάρχει τεράστιος ανταγωνισμός. Ο ανταγωνισμός αυτός επιβεβαιώνεται και από τα ράφια των supermarket, τόσο από τη μεγάλη ποικιλία προϊόντων, όσο και από τον χώρο που καταλαμβάνουν στο σύνολό τους. Πιο συγκεκριμένα, εκτός από τα προϊόντα που έρχονται πρώτα σε προτίμηση (κατέχουν τα μεγαλύτερα μερίδια αγοράς), παρατηρείται η παράλληλη ύπαρξη και άλλων λιγότερο ανταγωνιστικών προϊόντων. Ωστόσο, οι θέσεις των προϊόντων αυτών δεν εξυπηρετούν το ίδιο καλά τον καταναλωτή. Αυτό το φαινόμενο σχετίζεται κυρίως με την επιρροή και τη δύναμη που έχει μία ηγέτιδα εταιρεία του κλάδου, καθώς και με τη φερεγγυότητα του προμηθευτή. Εκτός αυτού όμως, η είσοδος στα υψηλά ράφια συνεπάγεται και υψηλότερο κόστος, πέραν των καλών συνεργασιών και του ονόματος της εταιρείας. Συμπερασματικά, η διαδικασία καθιέρωσης ενός προϊόντος σε μία καλή θέση στα ράφια των καταστημάτων, εξαρτάται από πολλούς παράγοντες αλλά και από την διαπραγματευτική ισχύ της εταιρείας.

Ερχόμενοι στις συνθήκες του σήμερα, εν μέσω πανδημίας, φαίνεται πως η κατάσταση στα ράφια του supermarket δεν έχει επηρεαστεί. Παρόλο αυτά, έχει έρθει στο προσκήνιο ένα καινούριο φαινόμενο αυτό των ηλεκτρονικών αγορών. Παρότι στις υπόλοιπες ευρωπαϊκές χώρες οι πολίτες είναι πιο εξοικειωμένοι με τις ηλεκτρονικές αγορές, σύμφωνα με την επιτροπή ανταγωνισμού, στην Ελλάδα μέχρι πρότινος, μόλις το 0.33% των αγορών για είδη πρώτης ανάγκης γίνονταν ψηφιακά. Σύμφωνα με εκτιμήσεις, υπολογίζεται ότι το ποσοστό αυτό θα παραμείνει σχετικά σταθερό μόλις επιστρέψουμε στους κανονικούς ρυθμούς της καθημερινότητας, παρότι εν μέσω της πανδημίας covid είχε αυξηθεί. Αν και υπάρχουν κάποιες αισιόδοξες ενδείξεις για μία μικρή αύξηση του ποσοστού αυτού, ακόμα δεν υπάρχουν απτά στοιχεία. Για αυτό ακριβώς τον λόγο, κρίνεται σκόπιμη μία περαιτέρω μελέτη για τον τρόπο παρουσίασης των προϊόντων στις ψηφιακές αγορές.

2.4. Παρουσίαση των προϊόντων και της αγοράς τους

Στην υποενότητα αυτή θα παρουσιαστούν, τόσο τα προϊόντα που χρησιμοποιήθηκαν για την ανάπτυξη του ερωτηματολογίου, όσο και η κατάταξη των εταιρειών, μαζί με τα μερίδια αγοράς, στην ενιαία αγορά του γιαουρτιού. Παράλληλα, θα δοθούν κάποιοι βασικοί ορισμοί για την περαιτέρω επεξήγηση των γιαουρτιών και των χαρακτηριστικών τους.

2.4.1. Παρουσίαση προϊόντων και των χαρακτηριστικών τους

Τις τελευταίες δεκαετίες στην Ελλάδα, ο κλάδος της παραγωγής γιαουρτιού φαίνεται να ακμάζει. Αποτέλεσμα αυτού, είναι η έντονη διαφοροποίηση μεταξύ των τύπων γιαουρτιού. Ωστόσο, εύλογο είναι το ερώτημα πώς ορίζεται η γιαούρτη, αλλά και πώς διαφοροποιείται απ' όλα τα υπόλοιπα προϊόντα. Σύμφωνα με τους Αγγελίδης κα. (2015), ως “γιαούρτη” μπορεί να χαρακτηριστεί μόνο το προϊόν που παρασκευάζεται αποκλειστικά και μόνο από γάλα και οξυγαλακτική καλλιέργεια. Όλοι οι υπόλοιποι τύποι γιαουρτιού που υπάρχουν αποτελούν επιδόρπια γιαουρτιού, και θα πρέπει να ταξινομούνται βάσει του γάλακτος.

Μερικά από τα είδη γιαουρτιού που παράγονται και είναι πιο δημοφιλή στην Ελλάδα παρουσιάζονται στην συνέχεια.

- Πρώτος τύπος γιαουρτιού είναι το παραδοσιακό. Ο τρόπος παρασκευής αναπτύχθηκε σε βάθος χρόνου, μέσα από την συστηματική παρατήρηση και την εμπειρία. Ωστόσο, επειδή σαν τρόπος παρασκευής ενέχει κινδύνους για τους καταναλωτές, οι γαλακτοβιομηχανίες έχουν κρατήσει την βασική τεχνολογία ίδια, με αλλαγές ωστόσο στα υπόλοιπα στάδια της παραγωγής. Χαρακτηριστικό γνώρισμα αυτού του είδους του γιαουρτιού, είναι ο σχηματισμός υμένιου (πέτσας) στο πάνω μέρος.
- Δεύτερος τύπος γιαουρτιού είναι το στραγγιστό γιαούρτι. Είναι ο πιο διαδεδομένος τύπος γιαουρτιού και προκύπτει είτε μέσα από τη στράγγιση της πρώτης ύλης σε υφασμάτινους σάκους, είτε μέσω της φυγοκέντρωσης. Σύμφωνα με την ελληνική νομοθεσία, το προϊόν αυτό προκύπτει από πλήρες γιαούρτι από το οποίο έχει αφαιρεθεί μέρος του νερού. Με την αφαίρεση του νερού, αφαιρούνται και διαλυμένα συστατικά, όπως γαλακτοζάχαρο και άλατα. Το ποσοστό σε λίπος πρέπει να είναι τουλάχιστον 8% και παραπάνω από 5% για το στραγγιστό αγελαδινό γιαούρτι.
- Τρίτος τύπος γιαουρτιού είναι η συνεκτική (set type). Στην περίπτωση αυτή, το γάλα ομογενοποιείται. Ως απόρροια αυτού, δεν σχηματίζεται υμένιο. Επιπλέον, η επώαση γίνεται σε περιέκτες, και μετά την πήξη το πήγμα δε διαταράσσεται. Θα πρέπει να σημειωθεί πως ο τύπος αυτός, φαίνεται να είναι δημοφιλής στην χώρα μας και γι' αυτό αποκαλείται και ως φυσική γιαούρτη. Ένας άλλος ορισμός είναι και αυτός του επιδόρπιου γιαουρτιού.
- Τελευταίος τύπος γιαούρτης είναι ο αναμεμιγμένος. Στην κατηγορία αυτή, το γάλα επωάζεται σε δεξαμενές και στη συνέχεια το πήγμα θραύεται, ψύχεται, αναμιγνύεται ή όχι με φρούτα και συσκευάζεται σε ερμητικά κλειστούς περιέκτες.

2.4.2. Η αγορά των γιαουρτιών

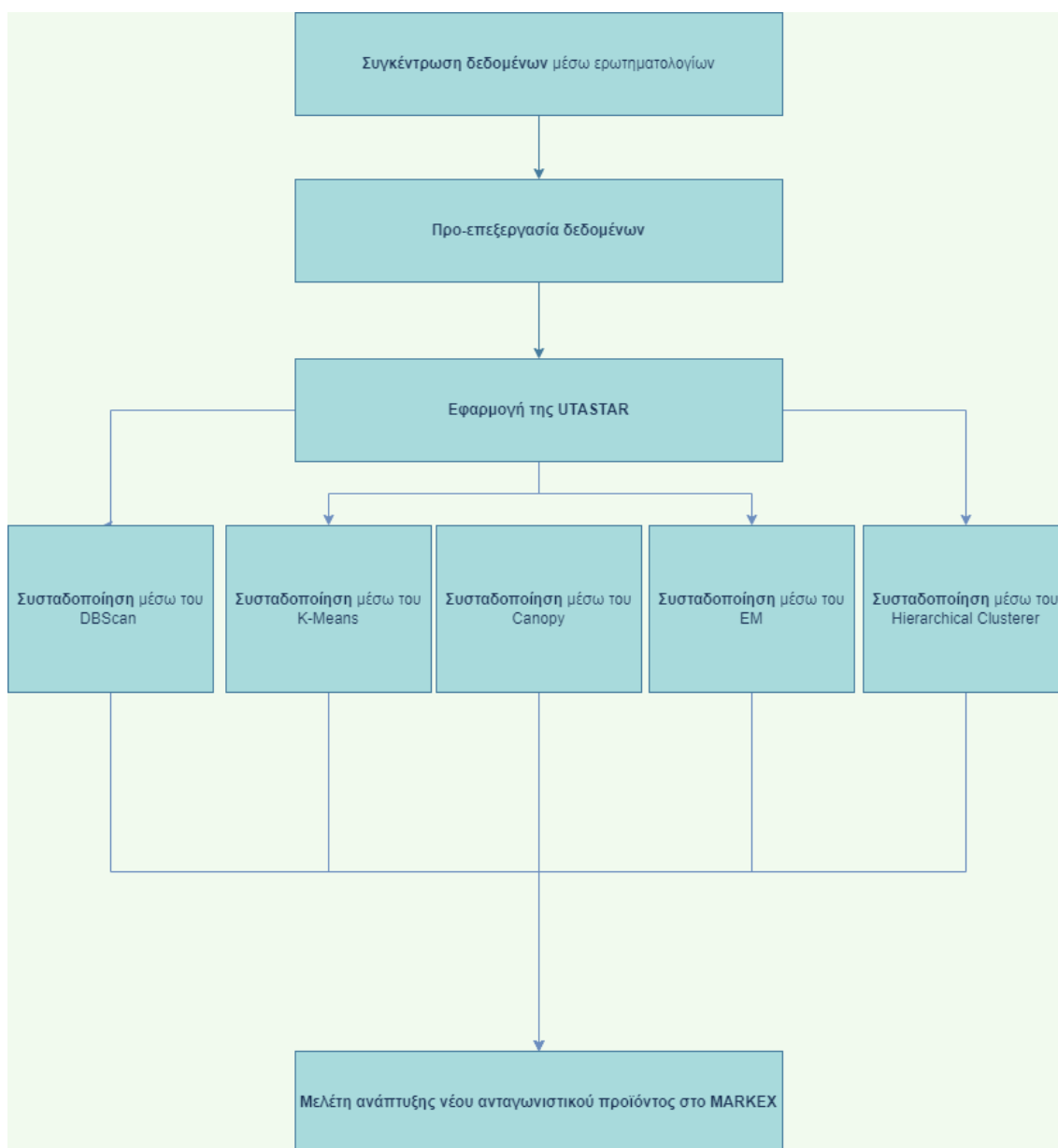
Σύμφωνα με την Επιτροπή Ανταγωνισμού (“Κλαδική έρευνα στον κλάδο παραγωγής, διανομής και εμπορίας βασικών καταναλωτικών ειδών και ιδίως ειδών διατροφής, καθώς και ειδών καθαριότητας και προσωπικής υγιεινής”, 2020) με την είσοδο των προϊόντων ιδιωτικής ετικέτας, σε συνδυασμό με τις κοινωνικο – οικονομικές εξελίξεις, τα μερίδια αγοράς στο κλάδο της γαλακτοβιομηχανίας επηρεάστηκαν αισθητά. Σύμφωνα με κλαδική έρευνα που υλοποίησε η επιτροπή ανταγωνισμού, οι γαλακτοβιομηχανίες φαίνεται να στρέφονται περισσότερο στην αγορά του γιαουρτιού, και όχι σε αυτή των παγωτών ή του γάλακτος, καθώς τόσο τα περιθώρια κέρδους μπορούν να είναι μεγαλύτερα, όσο και πιο εύκολο να εισαχθούν τα προϊόντα στην αγορά με μικρότερο κόστος και χωρίς διαφημιστικές δαπάνες. Έτσι, πρωταγωνιστικό ρόλο στον κλάδο παίζει το στραγγιστό γιαούρτι, το οποίο αφήνει και το υψηλότερο περιθώριο κέρδους. Η ηγέτιδα εταιρεία του κλάδου συνεχίζει να είναι η “ΦΑΓΕ”, όπως και τα προηγούμενα χρόνια. Ωστόσο, τα τελευταία χρόνια το μερίδιο αγοράς της έχει μειωθεί σε 21.2%. Αμέσως μετά, ακολουθεί με ποσοστό 14.3% η “ΚΡΙ-ΚΡΙ”, ενώ μετά είναι προϊόντα ιδιωτικής ετικέτας με ποσοστό 12.9%. Στην τέταρτη θέση βρίσκεται η εταιρεία “VIVARTIA (ΔΕΛΤΑ & ΜΕΒΓΑΛ)” με ποσοστό 11.9%. Αμέσως μετά, έρχεται η θεσσαλική εταιρεία “ΟΛΥΜΠΟΣ” με μερίδιο 9.3% και με μικρή διαφορά η “ΔΩΔΩΝΗ” (8%). Στην όγδοη θέση είναι τα γιαούρτια “ΝΟΥΝΟΥ (FRIESLAND)” 3.5% και την δεκάδα κλείνουν η γαλλική “DANONE” με 3.2% και η “ΦΑΡΜΑ ΚΟΥΚΑΚΗ” με 2.4%.

Τόσο σε ευρωπαϊκό, όσο και σε παγκόσμιο επίπεδο η κατανάλωση γιαουρτιού φαίνεται ότι έχει αυξητικές τάσεις. Ειδικότερα, οι πολίτες πλέον είναι περισσότερο ενημερωμένοι για την υψηλή διατροφική αξία του συγκεκριμένου προϊόντος, καθώς και τα θετικά που προσφέρει στην υγεία του ανθρώπου με την μακροχρόνια κατανάλωσή του. Οι ρυθμοί κατανάλωσης του γιαουρτιού μπορούν να διατυπωθούν με την βοήθεια του σύνθετου ετήσιου ρυθμού ανάπτυξης (CAGR). Σε ευρωπαϊκό επίπεδο ο CAGR ισούται με 3.2%, ενώ σε παγκόσμιο επίπεδο η αύξηση αυτή είναι μεγαλύτερη και ισούται με 6.25%. Αξίζει να σημειωθεί πως οι τιμές αυτές αναφέρονται σε βάθος πενταετίας από το 2022 μέχρι και το 2027. Τέλος, η Γερμανία και η Γαλλία φαίνεται να έχουν τις μεγαλύτερες αγορές γιαουρτιών βάσει πωλήσεων στην Ευρώπη, ενώ παράλληλα η Γερμανία φαίνεται να υποστηρίζει ολοένα και περισσότερο την κατανάλωση ελληνικού γιαουρτιού (<https://www.mordorintelligence.com/industry-reports/europe-yogurt-market>, n.d.).

Κεφάλαιο 3 Η Μεθοδολογία

3.1. Εισαγωγή

Σε αυτό το κεφάλαιο πρόκειται να παρουσιαστεί όλη η διαδικασία που ακολουθήθηκε για την υλοποίηση της συγκεκριμένης έρευνας. Πέραν από την περιγραφή της συνολικής μεθοδολογίας, θα γίνει και μία σύντομη γραφική παρουσίαση αυτής, καθώς και πως συνδέονται τα επιμέρους τμήματά της. Αναπόσπαστο κομμάτι του κεφαλαίου, είναι το θεωρητικό υπόβαθρο, τόσο των μεθόδων που χρησιμοποιήθηκαν, μαζί με τα αντίστοιχα εργαλεία, όσο και η προ – επεξεργασία των δεδομένων.



Σχήμα 3.1 Διάγραμμα ροής της μεθοδολογία

Τέλος, θα πρέπει να σημειωθεί πως η δομή της υπόλοιπης εργασίας και τα κεφάλαια, ακολουθούν την δομή του διαγράμματος ροής (σχήμα 3.1) που αναπτύχθηκε, και που εμφανίζεται στην ενότητα 3.2.

3.2. Περιγραφική ανάλυση (διάγραμμα ροής)

Στόχος της παρούσας διπλωματικής είναι η ανάπτυξη μίας μεθοδολογίας τμηματοποίησης αγορών, η οποία βασίζεται σε αναλύσεις δεδομένων. Οι αναλύσεις αυτές προέρχονται από διαφορετικές κατηγορίες μεθόδων. Για να επιτευχθεί ο στόχος της εργασίας αυτής, ακολουθήθηκαν τα βήματα που παρουσιάστηκαν στο διάγραμμα ροής της μεθοδολογίας (Σχήμα 3.1) τα οποία επεξηγούνται ακολούθως:

- Πρώτο βήμα είναι η συγκέντρωση των δεδομένων. Για την υλοποίηση της έρευνας κρίνεται απαραίτητη η συγκέντρωση των απαραίτητων δεδομένων, μέσω του ερωτηματολογίου που αναπτύχθηκε.
- Δεύτερο βήμα είναι η προ - επεξεργασία των δεδομένων. Η προ - επεξεργασία αυτή είναι απαραίτητη, καθώς έτσι τα τελικά αποτελέσματα θα είναι αρκετά πιο αξιόπιστα.
- Τρίτο βήμα είναι η εφαρμογή της μεθόδου UTASTAR τον πολυκριτήριο πίνακα κάθε καταναλωτή χωριστά. Από την εφαρμογή της UTASTAR υπολογίζονται για κάθε καταναλωτή οι μερικές χρησιμότητες, τα βάρη των κριτηρίων, καθώς και οι ολικές χρησιμότητες των εναλλακτικών επιλογών.
- Τέταρτο βήμα είναι η αναζήτηση τμημάτων της αγοράς μέσω της εφαρμογής αλγορίθμων συσταδοποίησης. Πιο αναλυτικά, τόσο το αρχείο με τα βάρη των κριτηρίων, όσο και το αρχείο με τις ολικές χρησιμότητες, αναλύονται με τη βοήθεια των πακέτων λογισμικού MATLAB (<https://www.mathworks.com/products/matlab.html>) και WEKA (<https://www.cs.waikato.ac.nz/ml/weka/>), για να εφαρμοστούν αντίστοιχα ο αλγόριθμος συσταδοποίησης K-Means, καθώς και οι αλγόριθμοι DBScan, Hierarchical Clusterer, Canopy και ο αλγόριθμος EM. Οι συγκεκριμένοι αλγόριθμοι επιλέχθηκαν, τόσο για να εξεταστούν διαφορετικά σενάρια κατά την συσταδοποίηση, και να επιλεγεί το πλέον προτιμότερο σενάριο συσταδοποίησης, όσο και για την απλότητα που εμφανίζουν στην εφαρμογή τους. Με τον τρόπο αυτό, γίνεται η ομαδοποίηση των καταναλωτών όλου του δείγματος, σύμφωνα με τις προτιμήσεις τους και της σημαντικότητας που εξέφρασαν για το κάθε κριτήριο. Τμηματοποιώντας το συνολικό δείγμα, στον επιθυμητό αριθμό ομάδων (από την εκάστοτε συσταδοποίηση), δημιουργούνται αντίστοιχα και τα αρχεία καταναλωτών που χρησιμοποιούνται ως είσοδοι στο MARKEX.
- Τελευταίο βήμα είναι η προσομοίωση της αγοράς για καθεμία από αυτές τις ομάδες/συστάδες/τμήματα και η ανάπτυξη ανταγωνιστικών

προϊόντων για όλες τις ομάδες αυτές. Για να επιτευχθεί αυτό, χρησιμοποιήθηκαν προτιμησιακά μοντέλα.

3.3. Το ερωτηματολόγιο της έρευνας

Στην σημερινή εποχή, η συλλογή των δεδομένων μπορεί να επιτευχθεί με ποικίλους τρόπους. Οι πιο δημοφιλείς μέθοδοι είναι είτε μέσω της συμπλήρωσης ερωτηματολογίων, είτε μέσω προσωπικών συνεντεύξεων, είτε μέσω απλών παρατηρήσεων. Στην προκειμένη περίπτωση θα εφαρμοστεί η πρώτη μέθοδος. Στο σημείο αυτό θα πρέπει να σημειωθεί πως η συλλογή των ερωτηματολογίων έγινε κυρίως ηλεκτρονικά, αποστέλλοντας το ερωτηματολόγιο σε καταναλωτές. Ακόμα, οι ερωτώμενοι καλούνται να απαντήσουν στις ερωτήσεις με μία συγκεκριμένα σειρά. Οι ερωτήσεις αυτές σχετίζονται με τα χαρακτηριστικά των προϊόντων, τις προτιμήσεις των καταναλωτών και τις συνήθειες αυτών. Τέλος, η δομή του ερωτηματολογίου αποτελεί κρίσιμη εργασία και χρήζει ιδιαίτερης προσοχής για την επιτυχία της έρευνας.

3.4. Καθαρισμός και προ - επεξεργασία δεδομένων

Για την ορθότερη και αποτελεσματικότερη διαχείριση και, κατ' επέκταση, λήψη αποφάσεων, κρίνεται απαραίτητος ο καθαρισμός και η προ - επεξεργασία των δεδομένων, ο οποίος έπεται της συλλογής των δεδομένων, όπως ακριβώς παρουσιάζεται και στο διάγραμμα ροής στο σχήμα 3.1. Ως καθαρισμό και προ - επεξεργασία δεδομένων, νοείται κάθε ενέργεια που συντελείται και έχει ως στόχο την αφαίρεση λανθασμένων ή μη χρήσιμων, ως προς την έρευνα, δεδομένων και κωδικοποίηση αυτών, ή η μείωση θορύβου έτσι ώστε να εξαχθεί ένα ασφαλέστερο, βέβαιο και λογικό αποτέλεσμα. Στην προκειμένη περίπτωση, οι έλεγχοι που έγιναν σχετίζονταν με τις ακραίες και τις ελλιπείς τιμές του δείγματος, τη συσχέτιση αυτών και την αξιοπιστία της ομαδοποίησης των δεδομένων. Στις τέσσερις υποενότητες που ακολουθούν, θα παρουσιαστούν οι έννοιες αυτές, καθώς και οι τρόποι εντοπισμού και αντιμετώπισης αυτών (Hyun Kang, 2013).

3.4.1. Ελλιπείς Τιμές

Η έλλειψη δεδομένων είναι ένα αρκετά συνηθισμένο πρόβλημα στις στατιστικές έρευνες που πραγματοποιούνται, και που το οποίο μπορεί να επηρεάσει τόσο τα αποτελέσματα της έρευνας, σε συνθήκες λάθους χειρισμού, όσο και να συμβάλλει στην αδυναμία εφαρμογής συνηθισμένων στατιστικών μεθόδων. Ωστόσο, προκειμένου να γίνει πιο κατανοητό το συγκεκριμένο πρόβλημα, δίνεται το ακόλουθο παράδειγμα. Έστω ότι αναλύεται το αρχείο excel που εξήχθη από το Google Forms. Αν υπάρχουν περιπτώσεις - πλειάδες στις οποίες δεν εμφανίζονται καταγεγραμμένες τιμές για ποικίλα χαρακτηριστικά, όπως το πλήθος των μελών της οικογένειας που καταναλώνουν γιαούρτι ή η ηλικία, τότε έχουμε ελλιπείς τιμές (missing values). Δηλαδή, υπάρχουν τιμές κάποιων χαρακτηριστικών, οι οποίες δεν καταγράφηκαν ή καταγράφηκαν λανθασμένα για κάποια μεταβλητή. Εναλλακτικά, τα στοιχεία αυτά

μπορεί να μην είναι πλέον διαθέσιμα, καθώς χαρακτηρίστηκαν λανθασμένα ασήμαντα, και διαγράφηκαν.

Για να αποφευχθεί το πρόβλημα αυτό, έχουν αναπτυχθεί ποικίλοι τρόποι αντιμετώπισης. Προτού περιγράψουν οι τρόποι αυτοί, θα πρέπει να σημειωθεί πως οι προς συμπλήρωση τιμές δεν αποτελούν πραγματικές τιμές, αλλά εκτιμήσεις αυτών. Πρώτος τρόπος αντιμετώπισης λοιπόν, είναι η παράληψη ολόκληρης της περίπτωσης – πλειάδας. Η μέθοδος αυτή, δεν είναι αρκετά αποτελεσματική, καθώς αυτόματα απορρίπτονται και τα υπόλοιπα συμπληρωμένα δεδομένα, και υπάρχει απώλεια πληροφόρησης. Η δεύτερη τεχνική που έχει αναπτυχθεί είναι η χειροκίνητη συμπλήρωση. Δηλαδή, ο ερευνητής μελετά την κάθε περίπτωση ξεχωριστά και συμπληρώνει τις τιμές που λείπουν με το χέρι. Όπως γίνεται αντιληπτό, η συγκεκριμένη τεχνική είναι αδύνατη στην περίπτωση ενός πολύ μεγάλου σετ δεδομένων, καθώς σπαταλάτε πολύς χρόνος.

Οι άλλες τέσσερις τεχνικές που χρησιμοποιούνται, εισάγουν έναν βαθμό ανακρίβειας (bias), καθώς η τιμή που εισάγεται αποτελεί πρόβλεψη. Ένας από αυτούς τους τρόπους αντιμετώπισης είναι η συμπλήρωση του κενού με ένα σύμβολο. Το σύμβολο αυτό, μπορεί να είναι ένας μεγάλος αριθμός, ή ένας χαρακτήρας. Η μέθοδος αυτή ωστόσο, δεν είναι και η πιο ασφαλής, καθώς ένας αλγόριθμος μπορεί να αναγνωρίσει τον συγκεκριμένο χαρακτήρα και να τον μελετήσει ως ξεχωριστή τιμή του προβλήματος. Επόμενη εναλλακτική είναι η χρήση του μέτρου της κεντρικής τάσης του χαρακτήρα. Στην περίπτωση αυτή ουσιαστικά χρησιμοποιείται η μέση τιμή ή η διάμεση μέση τιμή, ανάλογα την κάθε περίπτωση συμμετρίας, και συμπληρώνεται η κατάλληλη τιμή. Εξειδικευμένη περίπτωση της προηγούμενης τεχνικής είναι η χρήση της μέσης τιμής ή της διάμεσης τιμής για μία συγκεκριμένη κλάση από το σετ δεδομένων. Τελευταίος τρόπος αντιμετώπισης είναι η συμπλήρωση με την πιο πιθανή τιμή (Ματσατσίνης, 2021).

Στην παρούσα διπλωματική εργασία, οι μέθοδοι που χρησιμοποιήθηκαν ήταν οι δύο πρώτες, η παράλειψη περιπτώσεων και η χειροκίνητη συμπλήρωση. Επιπρόσθετα, αξίζει να σημειωθεί πως εντοπίστηκαν και άλλα κενά πεδία εντός του σετ δεδομένων. Παρόλο αυτά, δεν αποτελούν όλα αυτά τα κενά ελλείπουσες τιμές, καθώς στις περιπτώσεις αυτές ο χρήστης είχε την δυνατότητα να μην απαντήσει στο ερωτηματολόγιο. Συνεπώς, και δεδομένου ότι χρησιμοποιήθηκαν οι δύο πρώτοι τρόποι αντιμετώπισης, δεν χρειάστηκαν να ληφθούν ειδικοί κανόνες για τη διαχείριση των ελλειπών τιμών.

3.4.2. Ακραίες Τιμές

Ένα άλλο συχνό πρόβλημα που εμφανίζεται στις στατιστικές έρευνες είναι αυτό των ακραίων τιμών (outlier). Ως ακραία τιμή λοιπόν ορίζεται ένα σημείο δεδομένων που διαφοροποιείται αισθητά από το υπόλοιπο σύνολο δεδομένων. Είναι δηλαδή ένα στοιχείο που αποκλίνει από τα υπόλοιπα δεδομένα, σε τέτοιο βαθμό, ώστε να

δημιουργούνται αμφιβολίες για τον τρόπο που έχει προκύψει (Ματατσίνης, 2021).

Ωστόσο, θα πρέπει να δοθεί ιδιαίτερη προσοχή και στα δεδομένα με θόρυβο, τα οποία πολλές φορές συγχέονται μεταξύ τους. Ειδικότερα, θόρυβος (noise) καλείται ένα τυχαίο σφάλμα και δεν έχει κάποια ιδιαίτερη σημασία για περαιτέρω ανάλυση. Αντίθετα, η ακραία τιμή μπορεί να αποτελείται από λάθη ή ασυνέπεια, και προφανώς χρήζει ανάλυσης. Συνεπώς, βάσει αυτών, ο θόρυβος θα πρέπει να περιορίζεται και να μηδενίζεται εντελώς, αν αυτό είναι εφικτό, ενώ η ακραία τιμή να ερμηνεύεται, καθώς μπορεί να προσφέρει χρήσιμη πληροφορία για το εκάστοτε πρόβλημα που μελετάται κάθε φορά.

Όσον αφορά τώρα την ταξινόμηση των ακραίων τιμών, αυτές μπορούν να ταξινομηθούν σε τρεις κατηγορίες. Πρώτη περίπτωση είναι αυτή των ολικών ακραίων τιμών (global), όπου ουσιαστικά ένα αντικείμενο διαφέρει αισθητά απ' όλο το σετ δεδομένων. Η κατηγορία αυτή συνήθως είναι και η πιο εύκολη στον εντοπισμό και είναι και η πιο απλή. Στην συνέχεια, επόμενη κατηγορία είναι αυτή των συμφραζόμενων, οι οποίες ορίζονται ως τα δεδομένα που αποκλίνουν από ένα συγκεκριμένο πλαίσιο. Γι' αυτό ακριβώς τον λόγο, αποκαλούνται και υπό όρους ακραίες τιμές, αφού εξαρτάται από αυτούς. Εν κατακλείδι, τελευταία κατηγορία είναι αυτή των συναφών ακραίων τιμών. Ουσιαστικά, κάποιο δεδομένο του αντικειμένου που μελετάται είναι ακραία τιμή. Γι' αυτόν τον λόγο, ονομάζονται και υπό συνθήκη ακραίες τιμές.

Γνωρίζοντας πλέον ποιες είναι, και πώς ορίζονται οι ακραίες τιμές, θα πρέπει να σημειωθούν και όλοι οι τρόποι με τους οποίους εντοπίζονται. Ο πρώτος τρόπος κατηγοριοποίησης βασίζεται στη μοντελοποίηση των δεδομένων, ελέγχοντας την ύπαρξη κανονικότητας. Δηλαδή, υπάρχει ένας ταξινομητής ο οποίος εκπαιδεύεται με ένα σετ δεδομένων και ο οποίος στην συνέχεια προσπαθεί να αναγνωρίσει τις ακραίες τιμές και να τις διαχωρίσει. Επειδή λοιπόν, οι μέθοδοι αυτές λειτουργούν με αυτόν το συγκεκριμένο τρόπο, ονομάστηκαν εποπτευόμενες (supervised). Δεύτερη κατηγορία είναι αυτή των ημι – εποπτευόμενων μεθόδων (semi – supervised), όπου ουσιαστικά τα περισσότερα δεδομένα δεν έχουν κάποια ετικέτα, όπως στην προηγούμενη περίπτωση και η εκπαίδευση του μοντέλου γίνεται βάσει των ήδη υπάρχουσών ετικετών, και των δεδομένων που είναι αρκετά κοντά σε αυτές. Έτσι το μοντέλο, μπορεί να διαχωρίσει τις ακραίες τιμές από τα κανονικά αντικείμενα. Τελευταία κατηγορία είναι οι μη εποπτευόμενες μέθοδοι (unsupervised), οι οποίες ουσιαστικά λειτουργούν έχοντας ως αρχή τους, ότι τα κανονικά αντικείμενα είναι μερικώς ομαδοποιημένα. Με τον τρόπο αυτό, δημιουργούνται αρκετές ομάδες με διαφορετικά χαρακτηριστικά ή μία από τη άλλη και οι ακραίες τιμές βρίσκονται πιο μακριά από κάποια ομάδα (Matsatsinis, 2021).

Μία εναλλακτική κατηγοριοποίηση θα μπορούσε να είναι η ακόλουθη. Η πρώτη κατηγορία να βασίζεται σε συσταδοποιήσεις, δηλαδή τα κανονικά αντικείμενα να είναι αρκετά κοντά το ένα στο ένα για την κάθε συστάδα, ενώ οι ακραίες τιμές να είναι σε αραιές συστάδες ή να μην ανήκουν σε κάποια συστάδα. Δεύτερη κατηγορία είναι αυτή που βασίζεται σε μοντέλα. Δηλαδή, ότι τα αντικείμενα διαχωρίζονται χάρη σε ένα

στοχαστικό μοντέλο, και πως τα στοιχεία εκείνα που δε συμβαδίζουν με το μοντέλο αυτό είναι οι ακραίες τιμές. Οι μέθοδοι αυτές, υποθέτουν πως οι ακραίες τιμές βρίσκονται σε κάποια μακρινή απόσταση από τα υπόλοιπα στοιχεία και τους πλησιέστερους γείτονες (Matsatsinis, 2021).

Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκε μη εποπτευόμενη μέθοδος για τον εντοπισμό των ακραίων τιμών. Η μέθοδος αυτή εφαρμόστηκε στο περιβάλλον της WEKA. Παρόλο αυτά δεν βρέθηκαν ακραίες τιμές στο δείγμα όποτε δε χρειάστηκε να εφαρμοστεί κάποια άλλη μέθοδος για να αφαιρεθούν τα αντικείμενα αυτά.

3.4.3. Συντελεστής Συσχέτισης/Kendall's Tau

Για την ύπαρξη αξιόλογων δεδομένων, τα οποία με τη σειρά τους θα δώσουν και ορθότερα αποτελέσματα, ένα ακόμα μέτρο που λαμβάνεται υπόψη, είναι ο συντελεστής συσχέτισης τ του Kendall. Το μέτρο αυτό, χρησιμοποιείται για την κατανόηση της ισχύος μεταξύ δύο μεταβλητών. Δηλαδή, είναι ένα μέτρο, όπου ελέγχεται η ομοιότητα των δεδομένων όταν ταξινομούνται (Maurice Kendall, 1938). Βασικό πλεονέκτημα του συντελεστή αυτού είναι η εύκολη ερμηνεία του, η οποία βασίζεται στις πιθανότητες και την παρατήρηση εναρμονισμένων και μη ζευγών τιμών. Η μαθηματική του έκφραση είναι η εξής:

$$\tau = 2 * \frac{(N_c - N_d)}{n(n-1)} \quad (3.1)$$

Όπου,

- N_c : σύμφωνες τιμές του δείγματος
- N_d : μη σύμφωνες τιμές του δείγματος

Ως εναρμονισμένα, ορίζονται τα ζεύγη (X_j, Y_j) και (X_k, Y_k) αν οι δύο τιμές της μίας παρατήρησης είναι μεγαλύτερες ή μικρότερες από τις αντίστοιχες τιμές της άλλης παρατήρησης. Επιπλέον, ως μη εναρμονισμένα ζεύγη ορίζονται τα ίδια ζεύγη τιμών, αν η διάταξη των τιμών για την πρώτη μεταβλητή είναι αντίθετη από τη διάταξη για τη δεύτερη μεταβλητή. Αν όλα τα ζεύγη τιμών είναι εναρμονισμένα, τότε $\tau=1$, ενώ αν είναι μη εναρμονισμένα παίρνει τιμή $\tau=-1$ ("Kendall rank correlation coefficient", 2021).

3.4.4. Δείκτης Σκιαγράφησης/Silhouette

Τελευταίο βήμα της προ – επεξεργασίας του συνολικού σετ των δεδομένων, είναι η εφαρμογή του δείκτη σκιαγράφησης (silhouette). Ωστόσο, πως ορίζεται αυτός ο δείκτης; Σύμφωνα με τη μαθηματική έκφραση (Rousseeuw, 1986) ο μέσος $s(i)$ του συνόλου των σημείων ορίζεται ως:

$$S(i) = \begin{cases} \frac{1-a(i)}{b(i)}, & \text{αν } a(i) < b(i) \\ 0 & \text{αν } a(i) = b(i) \\ \frac{b(i)}{1-a(i)}, & \text{αν } a(i) > b(i) \end{cases} \quad (3.2)$$

Όπου,

- $a(i)$: μέση απόσταση i σημείου από τα σημεία της συστάδας
- $b(i)$: μέση απόσταση i από όλα τα σημεία κάθε άλλης συστάδας

Επιπλέον, ο δείκτης σκιαγράφησης κυμαίνεται $-1 \leq s(i) \leq 1$ και ουσιαστικά αποτελεί ένα μέτρο απόδοσης της ομοιότητας των υπό μελέτη δεδομένων. Δείχνει δηλαδή πόσο καλά έγινε η συσταδοποίηση του αντικειμένου, άρα και πόσο κοντά βρίσκονται τα στοιχεία της κάθε ομάδας μεταξύ τους, ενώ παράλληλα πόσο πιο απόμακρα είναι από τα στοιχεία των άλλων ομάδων. Όσο πιο κοντά στη μονάδα είναι η τιμή του δείκτη, τόσο καλύτερα έχει ομαδοποιηθεί το αντικείμενο. Αντίθετα, όσο πιο κάτω από το μηδέν είναι ο δείκτης, τόσο πιο ακατάλληλη κρίνεται η ομαδοποίηση. Επιπλέον, θα πρέπει να σημειωθεί πως δίνεται η δυνατότητα γραφικής αναπαράστασης του δείκτη για κάθε αντικείμενο ξεχωριστά (silhouette clustering, 2021). Εκτός αυτού όμως, σημασία έχει και η συνολική τιμή των δεικτών για την αξιολόγηση του εκάστοτε data set στο σύνολό του. Έτσι, και σε αυτή την έρευνα, υπολογίστηκε η μέση τιμή για όλους τους δείκτες. Προφανώς, η μέση τιμή φράζεται και αυτή εντός των ίδιων ορίων, και για την ερμηνεία των αποτελεσμάτων ισχύει ακριβώς και ότι στον δείκτη σκιαγράφησης.

3.5. Στατιστική περιγραφή

Πριν από την εφαρμογή των αλγορίθμων, ένα πρόσθετο βήμα είναι η περιγραφική στατιστική. Το βήμα αυτό, είναι αρκετά βασικό, καθώς έτσι κατανοούνται καλύτερα τα προς εξέταση δεδομένα. Ειδικότερα, έχοντας πλέον προ – επεξεργαστεί τα δεδομένα, είναι ευκολότερο να αναλυθούν και να οργανωθούν σε πίνακες, καθώς και να χρησιμοποιηθούν στατιστικά μεγέθη. Στην προκειμένη περίπτωση, η οργάνωση των στοιχείων αυτών θα γίνει με χρήση συγκεντρωτικών πινάκων. Ως απόρροια αυτού, τα αποτελέσματα που θα εξαχθούν θα είναι πλέον μετρήσιμα και θα διαφαίνεται και πιο εύκολα η όποια πιθανή τάση των στοιχείων. Τέλος, πέραν από τους πίνακες και τα στατιστικά μεγέθη, θα χρησιμοποιηθούν και διαγράμματα για την οπτικοποίηση των αποτελεσμάτων.

3.6. Οι μέθοδοι που χρησιμοποιήθηκαν

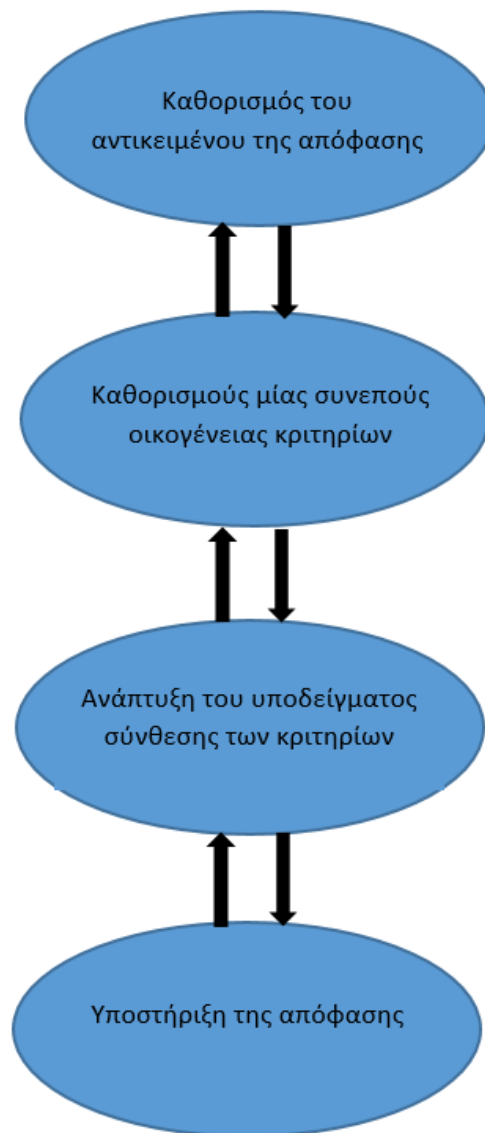
Στην ενότητα αυτή, πρόκειται να εξηγηθούν οι μέθοδοι και οι τεχνικές που χρησιμοποιήθηκαν καθ' όλη τη διάρκεια της έρευνας, καθώς και τα περιβάλλοντα που χρησιμοποιήθηκαν για την εφαρμογή των μεθόδων. Πιο αναλυτικά, στην αρχή θα εξηγηθούν οι πολυκριτήριες μέθοδοι, όπως και η UTASTAR που χρησιμοποιήθηκε στην προκειμένη περίπτωση και έπειτα οι αλγόριθμοι συσταδοποίησης, που ήταν

απαραίτητοι για την ομαδοποίηση των δεδομένων. Οι αλγόριθμοι παρουσιάζονται με την σειρά που παρουσιάζονται και στο διάγραμμα ροής που παρουσιάστηκε στην ενότητα 3.2. Επιπλέον, θα πρέπει να σημειωθεί πως θα δοθούν παραδείγματα για την καλύτερη κατανόηση των μεθόδων, καθώς και τα μαθηματικά μοντέλα, όπου αυτό είναι δυνατόν.

3.6.1. Πολυκριτήριες μέθοδοι

Προτού γίνει αναφορά στις πολυκριτήριες μεθόδους, είναι σημαντική η αναφορά στους στόχους μίας πολυκριτήριας ανάλυσης, καθώς και μία πολύ σύντομη ιστορική αναδρομή. Ειδικότερα, δεδομένου ότι τα περισσότερα προβλήματα που αντιμετωπίζει στην καθημερινότητά του ο άνθρωπος δεν είναι απλά και δεν αντιμετωπίζονται με μία μονοδιάστατη ανάλυση, τις τελευταίες δεκαετίες άρχισε να αναπτύσσεται ένας καινούριος κλάδος για την αντιμετώπιση αυτών των προβλημάτων. Ο κλάδος αυτός, είναι ο κλάδος της πολυκριτήριας ανάλυσης αποφάσεων, ο οποίος ευδοκίμησε τόσο σε πιο θεωρητικό επίπεδο, όσο και στο πιο πρακτικό (Ματσατσίνης, 2022). Για να γίνει πιο αντιληπτό αυτό, σε θεωρητικό επίπεδο οι άνθρωποι κατανόησαν και ενστερνίστηκαν πλήρως την δήλωση ότι: "Τα περισσότερα προβλήματα δεν αντιμετωπίζονται με μία μονόπλευρη ανάλυση". Επιπλέον, η εύρεση μίας λύσης βασίζεται σε διαφορετικούς παράγοντες για το εκάστοτε πρόβλημα που εξετάζεται, ενώ πολύ σημαντικό ρόλο παίζει και η εμπειρία του αποφασίζοντα και οι συνθήκες που επικρατούν (Σίσκος, 2008). Σε πρακτικό επίπεδο τώρα, η μορφή του προβλήματος, η σημαντικότητα του εκάστοτε χαρακτηριστικού, οι προϋποθέσεις και οι περιορισμοί που υπάρχουν, άρχισαν να αποκτούν μία μαθηματική μορφή και σταδιακά να μοντελοποιούνται ολοένα και περισσότερο, απλουστεύοντας έτσι τη διαδικασία της πολυκριτήριας λήψης αποφάσεων (Dourmos and Zorounidis, 2001).

Πρώτη επιτυχημένη προσπάθεια αντιμετώπισης αυτού του είδους των προβλημάτων έγινε από τον Pareto (1896), ο οποίος και εισήγαγε την έννοια της αποτελεσματικότητας. Στην συνέχεια, ακολούθησαν αρκετοί μεταξύ αυτών οι Von Neumann και Morgenstern (1944), που εισήγαγαν τη θεωρία της χρησιμότητας και ο Roy (1968), ο οποίος ανέφερε τη θεωρία των σχέσεων υπεροχής. Από το 1970 έως το 1990, ο συγκεκριμένος κλάδος φαίνεται να αναπτύσσεται με πολύ γρήγορους ρυθμούς. Συνεπώς, ένας από τους πολλούς ορισμούς που προέκυψαν για την πολυκριτήρια ανάλυση, είναι αυτός του Ζοπουνίδη (1999) σύμφωνα με τον οποίο, ως πολυκριτήρια ανάλυση ορίζεται ένα σύνολο μεθόδων που επιτρέπουν τη σύνθεση πολλών κριτηρίων εκτίμησης, έτσι ώστε να γίνει δυνατή η επιλογή, η κατάταξη, η περιγραφή και η ταξινόμηση ενός συνόλου εναλλακτικών ενεργειών.



Σχήμα 3.2 Στάδια λήψης αποφάσεων

Η μεθοδολογία μοντελοποίησης Πολυκριτήριας Ανάλυσης Αποφάσεων σύμφωνα με τον Roy (1976; 1990), αποτελείται από τα εξής στάδια (Σχήμα 3.2):

- Πρώτο στάδιο είναι αυτό του καθορισμού του εκάστοτε αντικειμένου της απόφασης. Δηλαδή, εντοπίζονται όλες οι πιθανές εφικτές λύσεις του προβλήματος και επιπλέον αποσαφηνίζεται πλήρως το εκάστοτε πρόβλημα. Όλες οι πιθανές, εφικτές λύσεις του προβλήματος αποτελούν το σύνολο A, και στο οποίο εφαρμόζονται μία εκ των τεσσάρων προβληματικών. Οι προβληματικές που μπορούν να εφαρμοστούν είναι η κατάταξη των εναλλακτικών, η ταξινόμηση, η περιγραφή ή η επιλογή.
- Δεύτερο στάδιο της διαδικασίας είναι αυτό του καθορισμού μίας συνεπούς οικογένειας κριτηρίων. Ως συνεπή οικογένεια κριτηρίων, ορίζεται το σύνολο των κριτηρίων που εντοπίζονται στο εξεταζόμενο πρόβλημα και για το οποίο

ισχύουν οι ιδιότητες της μονοτονίας, της επάρκειας και του μη πλεονασμού. Η μονοτονία του κριτηρίου μπορεί να χαρακτηριστεί ως φθίνουσα ή αύξουσα, ενώ αν με την αφαίρεση κάποιου κριτηρίου οι άλλες δύο ιδιότητες ανατραπούν, τότε το σύνολο δεν είναι πλεονασματικό. Όσον αφορά την επάρκεια τώρα, όταν δύο εναλλακτικές δραστηριότητες είναι ισοδύναμες μεταξύ τους, για κάθε κριτήριο, τότε υπάρχει επάρκεια.

- Τρίτο στάδιο της διαδικασίας είναι η ανάπτυξη του υποδείγματος σύνθεσης των κριτηρίων, για το εξεταζόμενο πρόβλημα.
- Τέταρτο στάδιο είναι η υποστήριξη της απόφασης, περιλαμβάνονται όλα εκείνα τα στοιχεία που μπορούν να βοηθήσουν τον αποφασίζοντα να επιλέξει και να κατανοήσει την καλύτερη εναλλακτική, επιχειρηματολογώντας πλήρως με απτά δεδομένα τα αποτελέσματα που εξήχθησαν.

3.6.1.1. Η μέθοδος UTASTAR

Για την αντιμετώπιση των πολυκριτήριων προβλημάτων, δημιουργήθηκαν πολλές διαφορετικές μεθοδολογικές προσεγγίσεις, οι οποίες φαίνεται να διαφέρουν όχι μόνο στη μορφή τους, αλλά και στη διαδικασία ανάπτυξης των υποδειγμάτων. Μία από τις πολλές, διαφορετικές ομαδοποιήσεις που προέκυψαν ήταν αυτή των Pardalos et al. (1995). Η πρώτη κατηγορία είναι αυτή του πολυκριτηρίου ή πολυστοχικού μαθηματικού προγραμματισμού (Multiobjective Mathematical Programming). Η δεύτερη κατηγορία είναι αυτή της πολυκριτήριας θεωρίας χρησιμότητας (Multiattribute Utility Theory), ενώ τρίτη στην σειρά έρχεται η θεωρία σχέσεων υπεροχής (Outranking relations). Τελευταία κατηγορία είναι αυτή της αναλυτικής - συνθετικής προσέγγισης (Preference disaggregation Approach), στην οποία ανήκει και η UTASTAR.

Μελετώντας περισσότερο το συγκεκριμένο θεωρητικό ρεύμα και αναλύοντας την συγκεκριμένη μέθοδο, γίνεται αντιληπτό ότι πρόκειται για χρήση μοντέλων ανάλυσης παλινδρόμησης, που έχουν ως στόχο την βέλτιστη αναπαράσταση των προτιμήσεων των αποφασιζόντων. Τις δύο εκείνες δεκαετίες όπου ο παρών κλάδος άνθιζε περισσότερο (1970-1990), οι Jacquet – Lagreze και ο Siskos (1982), πρωτοπαρουσίασαν την μέθοδο UTA. Σύμφωνα με τους Siskos et al. (2016), μία βελτιωμένη εκδοχή της UTA, είναι η UTASTAR την οποία και παρουσίασαν τρία χρόνια αργότερα. Η συγκεκριμένη μέθοδος εισήγαγε μία νέα τεχνική, σε αυτού του είδους τα μοντέλα, επαναληπτικού χαρακτήρα. Αναλυτικότερα, ως πρώτο βήμα είναι η λήψη μίας απόφασης. Η απόφαση αυτή λαμβάνεται τόσο βάσει των κριτηρίων, όσο και βάσει προβληματικής και στην συνέχεια εφαρμόζεται. Επόμενο βήμα, είναι η ανάλυση και η παρατήρηση των αποτελεσμάτων, ενώ τα όποια συμπεράσματα και αποτελέσματα προκύπτουν, συμβάλουν στην επανεκτίμηση των κριτηρίων και στην εκ νέου σύνθεσή τους. Για να προκύψει δηλαδή ένα αποδεκτό μοντέλο, θα πρέπει να ελεγχθεί ο βαθμός συσχέτισης αν είναι ικανοποιητικός από την πλευρά του αποφασίζοντα. Αν δεν είναι αρκετά ικανοποιητικός ο βαθμός αυτός, θα πρέπει να γίνει επανεκτίμηση είτε στον

τρόπο που λήφθηκε η απόφαση, είτε στην απλοποίηση της σύνθεσης του μοντέλου, είτε στην συνεπή οικογένεια κριτηρίων, καθώς πολύ πιθανόν είναι να μην λήφθηκαν οι κατάλληλοι δείκτες μέτρησης.

Στις επόμενες γραμμές ακολουθούν κάποιοι βασικοί συμβολισμοί και έννοιες έτσι ώστε να επεξηγηθεί καλύτερα η μέθοδος. Έστω ότι το σύνολο των εναλλακτικών συμβολίζεται με A_R , και το οποίο αποτελείται από a_m εναλλακτικές, δηλαδή είναι της μορφής $A_R = \{a_1, a_2, a_3, \dots, a_m\}$. Επίσης, γνωρίζουμε ότι για κάθε εναλλακτική a_i , υπάρχει και ένα μοναδικό σφάλμα $\sigma(a)$, ενώ ως σύνολο $J = \{1, 2, 3, \dots, p\}$, ορίζεται το σύνολο των καταναλωτών. Τέλος, τα κριτήρια βάσει των οποίων γίνεται η βαθμολόγηση από τους καταναλωτές, συμβολίζονται με $g_1, g_2, g_3, \dots, g_n$. Τέλος, για κάθε προϊόν a_i , υπάρχει και ένα διάνυσμα $g(a)$, που αναπαριστά την πολυκριτήρια αποτίμηση του προϊόντος για τον κάθε καταναλωτή. Στη συγκεκριμένη έρευνα, το πλήθος των καταναλωτών είναι ίσο με το πλήθος των ερωτηματολογίων μετά την προ – επεξεργασία, οι εναλλακτικές είναι τα επτά διαφορετικά προϊόντα γιαουρτιού, όπως επτά είναι και τα κριτήρια που χρησιμοποιήθηκαν για την βαθμολόγηση των προϊόντων. Σύμφωνα με τους Siskos & Yannacopoulos (1985), υπάρχει ένας κανόνας για τα n κριτήρια, αυτός της προσθετικής συνάρτησης αξιών, ο οποίος έχει την μορφή:

$$\bar{u}(g) = u_1(g_1) + u_2(g_2) + u_3(g_3) \dots u_n(g_n) \quad (3.3)$$

Ουσιαστικά, η εξίσωση 3.3 υλοποιείται μέσω της UTASTAR για τον κάθε καταναλωτή. Ακόμα, θα πρέπει να σημειωθεί πως ο κάθε καταναλωτής έχει διαφορετικές δομές δεδομένων, μέσα στις οποίες αποθηκεύονται τα διανύσματα των μερικών χρησιμοτήτων ανά κριτήριο. Δεδομένου ότι το άθροισμα όλων των βαρών ισούται με την μονάδα, οι συναρτήσεις των μερικών χρησιμοτήτων ορίζονται στο διάστημα $[0,1]$. Έτσι, για να προκύψουν τα βάρη των κριτηρίων κανονικοποιημένα, και κατ' επέκταση σε μορφή ποσοστού, διαιρούνται οι μερικές χρησιμότητες $u_i(g_i)$ με την μέγιστη τιμή του βάρους ανά κριτήριο p_i . Όσον αφορά τώρα την ολική αξία για το καθένα προϊόν, δίνεται από τον τύπο:

$$u[g(a)] = p_1 * u_1[g_1(a)] + p_2 * u_2[g_2(a)] + \dots + p_n * u_n[g_n(a)] \quad (3.4)$$

Η βέλτιστη λύση του προβλήματος αυτού, έρχεται μέσω της συνάρτησης χρησιμότητας, με την βοήθεια γραμμικού προγραμματισμού. Το επόμενο βήμα της διαδικασίας είναι ο έλεγχος ευστάθειας, έτσι ώστε να συνεχιστεί η διαδικασία και να βρεθούν και οι υπόλοιπες πολλαπλές λύσεις. Οι λύσεις αυτές δηλαδή, προκύπτουν από την μεταβελτιστοποίηση και την επίλυση άλλου ένα γραμμικού προβλήματος. Τέλος, η βελτιστοποίηση και η μεταβελτιστοποίηση της λύσης γίνεται με την μεγιστοποίηση των βαρών.

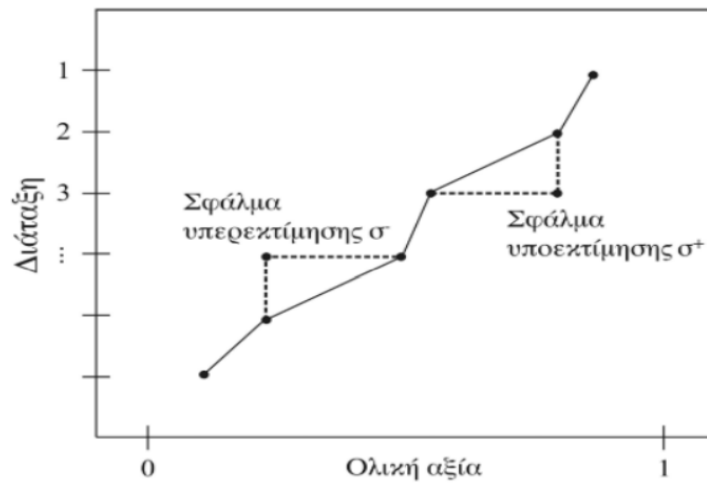
Βάσει μαθηματικών συμβολισμών και εννοιών, καθώς και των βημάτων που περιεγράφηκαν, η UTASTAR μπορεί να επεξηγηθεί στα ακόλουθα τέσσερα στάδια.

Πρώτο βασικό στάδιο του μαθηματικού μοντέλου είναι η μετατροπή των μεταβλητών. Δηλαδή, ελέγχεται η μονοτονία των κριτηρίων στους περιορισμούς του προβλήματος και μετατρέπεται ως εξής:

$$w_{ij} = u_i(g_i^{j+1}) - u_i(g_i^j) \geq 0 \quad \forall i, j \quad (3.5)$$

Στο δεύτερο στάδιο εισάγεται μία συνάρτηση διπλού σφάλματος. Τα σφάλματα που εισάγονται είναι αυτό της υπερεκτίμησης (σ_a^+) και της υποεκτίμησης (σ_a^-) για κάθε ζεύγος διαδοχικών δράσεων.

$$\Delta(\alpha, b) = u(g(\alpha)) - u(g(b)) + \sigma^+(\alpha) - \sigma^-(\alpha) - \sigma^+(b) + \sigma^-(b) \quad (3.6)$$



Σχήμα 3.3 Επεξήγηση σφαλμάτων

Για την καλύτερη κατανόηση των σφαλμάτων στο σχήμα 3.3 δίνεται γραφικά η επεξήγηση αυτών, σε σχέση με την καμπύλη ποιοτικής παλινδρόμησης.

Το τρίτο στάδιο της μεθόδου, είναι η επίλυση του γραμμικού προβλήματος. Η αντικειμενική συνάρτηση του προβλήματος έχει την μορφή ελαχιστοποίησης και ως στόχο έχει την μείωση των σφαλμάτων, τόσο αυτών των υποεκτιμήσεων ($\sigma^+(\alpha)$), όσο και των υπερεκτιμήσεων ($\sigma^-(\alpha)$). Η αντικειμενική συνάρτηση είναι η εξής:

$$[min] \quad F = \sum_{a \in A} \{\sigma^+(\alpha) + \sigma^-(\alpha)\} \quad (3.7)$$

Όσον αφορά τους περιορισμούς του προβλήματος τώρα, αυτοί χωρίζονται σε δύο ομάδες. Η πρώτη ομάδα περιορισμών φαίνεται στην συνέχεια:

$$\Delta(\alpha, b) \geq \delta, \text{ όταν } \alpha > b \quad (3.8)$$

Ο περιορισμός αυτός ουσιαστικά καταφέρει τα σφάλματα για το κάθε ζεύγος ενεργειών να είναι οριακά θετικά, όταν το a υπερσχύει του b , καθώς ως δ ορίζεται μία μικρή θετική τιμή που ισούται με την ελάχιστη απόσταση των τιμών μεταξύ των ολικών χρησιμότητων των εναλλακτικών επιλογών a, b .

$$\Delta(a,b)=0, \text{ όταν } a \sim b \quad (3.9)$$

Με την ίδια ακριβώς λογική, όπως και στην σχέση 3.8, όταν τα σφάλματα για το κάθε ζεύγος ενεργειών είναι ίσα με 0, όταν το a δεν υπερσχύει του b και αντίστροφα, είναι δηλαδή αδιάφορα ως προς την προτίμηση.

$$\sum_i \sum_j w_{ij} = 1 \quad (3.10)$$

Σύμφωνα με την σχέση 3.10, το άθροισμα των μεταβλητών που μετατράπηκαν, βάσει των περιορισμών μονοτονικότητας είναι ίσο με την μονάδα.

Δεύτερη ομάδα περιορισμών είναι αυτή της μη αρνητικότητας, όπου ουσιαστικά τα μεγέθη θα πρέπει να είναι θετικά, ή ίσα με το μηδέν. Τα μεγέθη αυτά είναι τα σφάλματα υποεκτίμησης, υπερεκτίμησης και τα οι μεταβλητές που μετατράπηκαν αρχικά στο πρόβλημα.

$$w_{ij} \geq 0 \quad (3.11)$$

$$\sigma^+(a) \geq 0 \quad (3.12)$$

$$\sigma^-(a) \geq 0 \quad (3.13)$$

Το τελευταίο βήμα της συγκεκριμένης μεθοδολογίας, πραγματοποιείται σε περίπτωση που η πολυκριτήρια εκτίμηση ενός αποφασίζοντα δεν είναι ακέραια δυνατή τιμή του κριτηρίου, και ουσιαστικά η μερική χρησιμότητα του υπολογίζεται με γραμμική παρεμβολή. Η διαδικασία αυτή αποτελεί έναν έλεγχο ύπαρξης πολλαπλών βέλτιστων ή ημιβέλτιστων λύσεων. Εν κατακλείδι, τα βάρη αποκτούν την εξής μορφή:

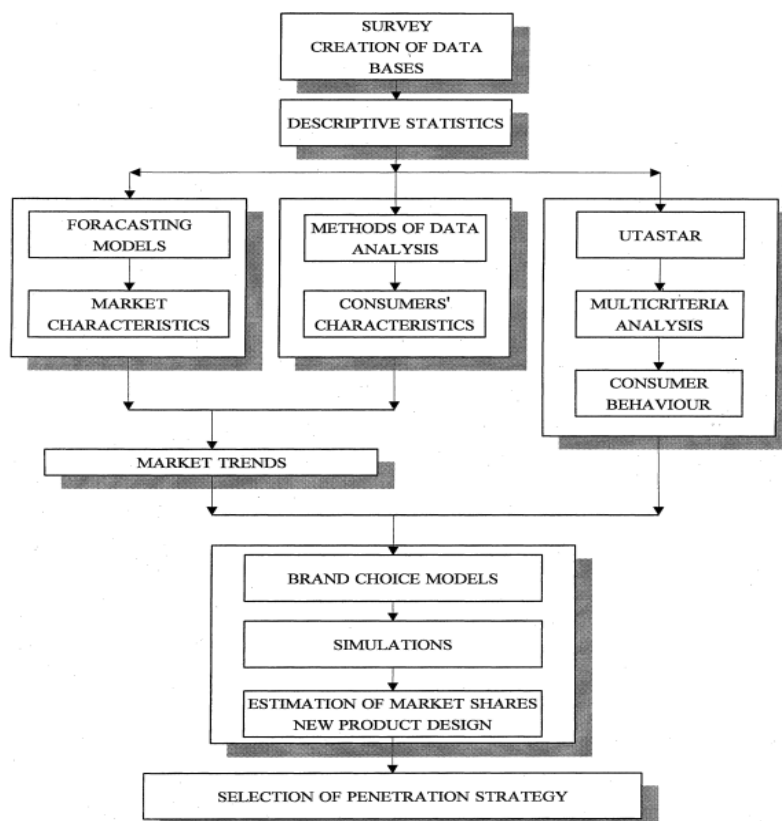
$$u_i(g_i^*) = \sum_{k=1}^{a_i-1} w_{ik} \quad (3.14)$$

3.6.1.2. To Markex

Για την βιωσιμότητα των επιχειρήσεων, είναι πολύ σημαντική η εισαγωγή νέων προϊόντων στην αγορά. Η άποψη αυτή επιβεβαιώνεται από αρκετούς ειδικούς. Σύμφωνα με τους Nylen (1990) και Urban & Hauser (1993), η εισαγωγή των νέων προϊόντων στην αγορά δεν είναι εύκολη διαδικασία. Παράλληλα, εκτός από την

δυσκολία ένταξης, υπάρχει και η δυσκολία από οικονομικής απόψεως. Σε κάθε περίπτωση όμως, θα πρέπει να σημειωθεί πως οι οικονομικές επιπτώσεις είναι εντονότερες όταν το νέον προϊόν δεν εισάγεται εν τέλει στην αγορά επιτυχώς, και όχι κατά την διαδικασία εισαγωγής του.

Για να αποφευχθούν τέτοια προβλήματα, αναπτύχθηκαν διάφορα συστήματα λήψης αποφάσεων. Τα αρχικά αυτά συστήματα λήψης αποφάσεων είχαν αρκετά μειονεκτήματα, όπως για παράδειγμα την περιορισμένη τους εφαρμογή σε συγκεκριμένες φάσεις της παραγωγικής διαδικασίας και την ανικανότητα διαχείρισης ποιοτικών δεδομένων. Έτσι, οι Siskos και Matsatsinis (1993), Ματσατσίνης (1995) και Matsatsinis and Siskos (1999; 2003) ανέπτυξαν μία νέα μεθοδολογία ανάπτυξης της αγοράς, βασιζόμενη στην συμπεριφορά των καταναλωτών, το MARKEK (MARKEt EXpert). Επιπλέον, το MARKEK παρουσιάζει αρκετά πλεονεκτήματα, καθώς είναι αρκετά φιλικό προς τους χρήστες, δίνοντας διαγράμματα και γραφήματα, και η εγκατάστασή του είναι αρκετά εύκολη.



Σχήμα 3.4 Το διάγραμμα ροής του MARKEK

Ο τρόπος λειτουργίας του MARKEK περιγράφεται συνοπτικά στο σχήμα 3.4 που ακολουθεί. Αναλυτικότερα, πρώτο βήμα είναι η δημιουργία μίας σωστής βάσης δεδομένων για την εκάστοτε έρευνα που διεξάγεται. Στην συνέχεια, δίνεται η δυνατότητα για εμφάνιση περιγραφικών στατιστικών. Από εκεί και πέρα, με την σωστή

χρήση του συστήματος, μπορούν να εφαρμοστούν μέθοδοι ανάλυσης δεδομένων, καθώς και μοντέλα πρόβλεψης. Με τον τρόπο αυτό, εντοπίζονται οι τάσεις της αγοράς που μελετάται σε κάθε περίπτωση, όπως είναι για παράδειγμα τα χαρακτηριστικά της αγοράς ή των καταναλωτών. Με εφαρμογή της UTASTAR, και σε συνδυασμό με τα εναλλακτικά σενάρια εισαγωγής ενός νέου προϊόντος στην αγορά, και τα ποικίλα μοντέλα συμπεριφοράς, εντοπίζεται το καταλληλότερο στρατηγικό πλάνο για το προϊόν.

3.6.1.3. Μοντέλα προσωπικής επιλογής καταναλωτή

Όπως αναφέρθηκε και στην ενότητα 2.2, η αγοραστική συμπεριφορά διαφέρει από καταναλωτή σε καταναλωτή. Σε μία προσπάθεια εκτίμησης και μοντελοποίησης των καταναλωτικών συμπεριφορών, αναπτύχθηκαν τα μοντέλα προσωπικής επιλογής. Τα μοντέλα αυτά βασίζονται σε μαθηματικές συναρτήσεις και διαφέρουν, τόσο ως προς τον τρόπο αντιμετώπισης των εξωτερικών παραγόντων/ερεθισμάτων, όσο και ως προς τον τρόπο αντιμετώπισης των προβλημάτων αυτών εσωτερικά, κατά την εκτέλεση διεργασιών. Για να γίνει πιο κατανοητό, γίνεται η υπόθεση ότι ο καταναλωτής αξιολογεί όλες τις εναλλακτικές, βάσει των κριτηρίων, θέτοντας ως στόχο το μέγιστο προσωπικό όφελος. Η συνάρτηση του μοντέλου τώρα, λαμβάνει υπόψη την ατομική προτίμηση του καταναλωτή και τα χαρακτηριστικά της εκάστοτε εναλλακτικής. Η συνάρτηση χρησιμότητας μπορεί να συμβολιστεί ως εξής:

$$U_{ij} = F_i(X_{ijk} * B_{i,k} * C_i) + e_{ij} \quad (3.18)$$

Όπου,

- i : το πλήθος των καταναλωτών
- j : το πλήθος των προϊόντων
- k : το πλήθος των χαρακτηριστικών του προϊόντος
- U_{ij} : η χρησιμότητα που προσδίδει ο καταναλωτής i στο προϊόν j
- F_i : η συνάρτηση που προσεγγίζει καλύτερα τον καταναλωτή i
- X_{ijk} : η αξία του χαρακτηριστικού k του προϊόντος j για τον καταναλωτή i
- B_i : η σημαντικότητα του k για τον καταναλωτή i
- C_i : συντελεστής προσαρμογής του καταναλωτή i
- e_{ij} : σφάλμα μέτρησης της χρησιμότητας, το οποίο ακολουθεί μία συγκεκριμένη κατανομή

Δεδομένου ότι υπάρχουν αρκετές και διαφορετικές, καταναλωτικές συμπεριφορές, έχουν αναπτυχθεί πληθώρα μοντέλων, με στόχο την καλύτερη, δυνατή εκτίμηση και επεξήγηση των παραγόντων που ώθησαν τον καταναλωτή σε αυτή του την επιλογή. Στην παρούσα διπλωματική εργασία, θα αναπτυχθούν τα οκτώ από αυτά, τα οποία απαντώνται στο MARKEX και εμφανίζονται και στον πίνακα 3.3.

Πίνακας 3.1 Μοντέλα συμπεριφοράς καταναλωτή

Μοντέλα προσωπικής επιλογής καταναλωτή	Μαθηματικός τύπος
Luce	$P_{ij}(C) = \frac{U_{ij}}{\sum_{k \in C} U_{ik}}$
Lesourne	$P_{ij}(C) = \frac{U_{ij}^2}{\sum_{k \in C} U_{ij}^2}$
Multinomial McFadden-1	$P_{ij}(C) = \frac{e^{U_{ij}}}{\sum_{k \in C} e^{U_{ik}}}$
Slightly Reinforced McFadden-2	$P_{ij}(C) = \frac{e^{2U_{ij}}}{\sum_{k \in C} e^{2U_{ik}}}$
Width of utilities 1	$P_{ij}(C) = \frac{U_{ij}^{U_{max}-U_{min}}}{\sum_{k \in C} U_{ik}^{U_{max}-U_{min}}}$
Width of utilities 2	$P_{ij}(C) = \frac{e^{2(U_{i,max}-U_{i,min})}}{\sum_{k \in C} e^{2(U_{i,max}-U_{i,min})}}$
Maximum of utilities	$P_{ij}(j C) = \begin{cases} \frac{1}{m}, & \text{όταν } U_{i,max} \geq U_{i,j} \geq U_{i,min} - \varepsilon_i \\ 0, & \text{αλλιού} \end{cases}$
Equal Possibilities	$P_j = \frac{1}{m}, \text{όταν } U_{i,max} - U_{i,min} \leq 0,1$

3.6.1.3.1. Το μοντέλο Luce

Η προσέγγιση που αναπτύσσεται στο συγκεκριμένο μοντέλο αναφέρει πως οι εκτιμήσεις του καταναλωτή παραμένουν σταθερές. Συνεπώς, τα προϊόντα που έχουν ίδιες εκτιμήσεις και ίδια σειρά προτίμησης, θα πρέπει να έχουν και ίδιες χρησιμότητες. Εναλλακτικά, το αξίωμα του Luce, όπως αλλιώς ονομάζεται αναφέρει πως εάν ένα χ προϊόν ανήκει σε ένα σύνολο εναλλακτικών C , και S ένα υποσύνολο του C , η πιθανότητα που έχει το χ να επιλεγεί απ' όλο το σύνολο των εναλλακτικών, ισούται με το γινόμενο της πιθανότητας που έχει το χ να επιλεγεί από το S και της πιθανότητας που έχει ένα στοιχείο του S να επιλεγεί από το σύνολο C (Tsafarakis, 2007). Ο μαθηματικός ορισμός διατυπώνεται ως εξής:

$$\Pr(\chi|C) = \Pr(\chi|S) * \Pr(S|C) \quad (3.19)$$

Τώρα για την εναλλακτική επιλογή χ , η σταθερή χρησιμότητα u_χ έναντι μίας άλλης εναλλακτικής z δίνεται από την σχέση:

$$u_\chi = \frac{\Pr(\chi|C)}{\Pr(z|C)} \quad (3.20)$$

Με την ίδια ακριβώς λογική για δύο εναλλακτικές επιλογές χ και γ , οι σταθερές χρησιμότητες είναι ίσες με:

$$\frac{U_x}{U_y} = \frac{\Pr(x|C)}{\Pr(y|C)} \quad (3.21)$$

Για όλες τις εναλλακτικές του συνόλου C, ισχύει ότι:

$$\frac{u_x}{\sum_{y \in C} u_y} = \frac{\Pr(x|C)}{\sum_{y \in C} \Pr(y|C)} \quad (3.22)$$

Ωστόσο, δεδομένου ότι ο παρονομαστής της σχέσης 3.22 ισούται με την μονάδα και γενικεύοντας παράλληλα όσα προαναφέρθηκαν, προκύπτει η σχέση 3.23.

$$P_{ij}(C_i) = \frac{V_{ij}}{\{\sum_{k \in C} V_{ij}\}} \quad (3.23)$$

Όπου,

- $P_{ij}(C_i)$: η πιθανότητα ο πελάτης i θα επιλέξει την μάρκα j από το σύνολο C.
- $V_{ij}(C_i)$: η χρησιμότητα που αναμένει το άτομο i να αποκομίσει από την επιλογή της μάρκας j.
- C_i : το σύνολο των εναλλακτικών που αξιολογεί ο καταναλωτής i.

3.6.1.3.2. Το μοντέλο Lesourne

Το μοντέλο αυτό αποτελεί μία παραλλαγή του προηγούμενου μοντέλου που αναπτύχθηκε στην ενότητα 3.6.1.3.2., καθώς και αυτό είναι μοντέλο σταθερής χρησιμότητας. Η διαφοροποίηση του έρχεται στον υπολογισμό της πιθανότητας $P_{ij}(C_i)$, καθώς ισούται με τον λόγο του τετραγώνου της χρησιμότητας U_{ij} που ο καταναλωτής i προσδοκά να αποκομίσει από την επιλογή του προϊόντος j, προς το άθροισμα των τετραγώνων των χρησιμοτήτων που αποδίδει στο σύνολο των προϊόντων του C. Σε πρακτικό επίπεδο, οι πιθανότητες με τις μεγαλύτερες χρησιμότητες υπερτερούν έναντι των άλλων εναλλακτικών, λόγω της ύπαρξης του τετραγώνου που υπάρχει στον εκθέτη του τύπου 3.24.

$$P_{ij}(C) = \frac{U_{ij}^2}{\sum_{k \in C} U_{ij}^2} \quad (3.24)$$

3.6.1.3.3. Το μοντέλο Multinomial McFadden-1

Αντίθετα από τα δύο προηγούμενα μοντέλα σταθερής χρησιμότητας, αυτό το μοντέλο βασίζεται σε τυχαίες διακυμάνσεις. Δηλαδή, οι χρησιμότητες που αποδίδει ο καταναλωτής στο εκάστοτε προϊόν, μπορεί να διαφέρουν ανάλογα με την κάθε περίπτωση. Με τον τρόπο αυτό, το μοντέλο εκφράζει την τάση του καταναλωτή να ελαχιστοποιήσει τις διαφορές στις προτιμήσεις του, όσον αφορά τα προϊόντα της αγοράς. Τέλος, η πιθανότητα ο πελάτης i να επιλέξει την μάρκα j, ισούται με:

$$P_{ij}(C) = \frac{e^{u_{ij}}}{\sum_{k \in C} e^{u_k}} \quad (3.25)$$

3.6.1.3.4. Το μοντέλο Slightly Reinforced McFadden-2

Το μοντέλο αυτό, αποτελεί μία παραλλαγή του μοντέλου που αναπτύχθηκε στην ενότητα 3.6.1.3.3.. Πιο αναλυτικά, εκφράζει την διάθεση του καταναλωτή να διαχωρίσει τα προϊόντα μεταξύ τους, χωρίς ωστόσο να δίνει μία μεγάλη προτίμηση στο πρώτο προϊόν έναντι όλων των υπολοίπων. Από μαθηματικής απόψεως, το μοντέλο εκφράζεται ως:

$$P_{ij}(C) = \frac{e^{2u_{ij}}}{\sum_{k \in C} e^{2u_{ik}}} \quad (3.26)$$

3.6.1.3.5. Το μοντέλο Width of utilities 1 & 2

Μία πιο ευρύτερη κατηγορία μοντέλων, από αυτά των Luce και Lesourne, αναπτύχθηκε από τους Pessemier et al. (1971). Σύμφωνα με αυτούς, για να εφαρμοστεί επιτυχώς το μοντέλο, θα πρέπει πρώτον τα προς μελέτη προϊόντα να είναι ανταγωνιστικά μεταξύ τους και δεύτερον, να έχει επέλθει σταθεροποίηση στην αγορά. Έχοντας σε ισχύ αυτές τις δύο προϋποθέσεις, φαίνεται να ισχύει η εξής σχέση:

$$P_{ij}(C) = \frac{u_{ij}^a}{\sum_{k \in C} u_{ik}^a} \quad (3.27)$$

Δηλαδή, η πιθανότητα ο καταναλωτής i να επιλέξει το προϊόν j , είναι ίση με τον λόγο της χρησιμότητας U_{ij} , υψωμένη σε μία δύναμη a , προς το άθροισμα των χρησιμότητων που αποδίδει ο ίδιος καταναλωτής στο σύνολο των προϊόντων, υψωμένο στην ίδια δύναμη. Όταν η παράμετρος a παίρνει την τιμή 1, τότε εφαρμόζεται το μοντέλο Luce, ενώ όταν παίρνει την τιμή 2 εφαρμόζεται το μοντέλο Lesourne. Αξίζει να σημειωθεί, πως η τιμή του a , παραμένει σταθερή και μετά την είσοδο του προϊόντος στην αγορά. Επιπλέον, όσο μεγαλύτερες τιμές παίρνει η παράμετρος a , τόσο ενισχύονται οι πιθανότητες επιλογής προϊόντων με την μέγιστη χρησιμότητα. Βασιζόμενος σε αυτό το μοντέλο, ο Matsatsinis προχώρησε στην αλλαγή της παραμέτρου a σε απόσταση μεταξύ της μέγιστης και της μικρότερης χρησιμότητας, και εκφράζει ουσιαστικά την δυσκολία που αντιμετωπίζει ο καταναλωτής στη επιλογή μίας εναλλακτικής. Όσο μεγαλύτερη τιμή παίρνει ο δείκτης a , τόσο πιο εύκολο είναι για τον καταναλωτή να επιλέξει ένα προϊόν, ενώ όσο μικρότερη είναι η τιμή του a , τόσο πιο δύσκολα επιλέγει ο καταναλωτής το πιο προτιμώμενο προϊόν. Η διαφορά αυτή μεταξύ της μέγιστης και της ελάχιστης χρησιμότητας ονομάζεται εύρος και συμβολίζεται με δ . Τέλος, όσον αφορά την μαθηματική έκφραση για το πρώτο και το δεύτερο μοντέλο ισχύει αντίστοιχα ότι:

$$P_{ij}(C) = \frac{U_{ij}^{U_{max}-U_{min}}}{\sum_{k \in C} U_{ik}^{U_{max}-U_{min}}} \quad (3.28)$$

$$P_{ij}(C) = \frac{e^{2(U_{i,max}-U_{i,min})}}{\sum_{k \in C} e^{2(U_{i,max}-U_{i,min})}} \quad (3.29)$$

3.6.1.3.6. Το μοντέλο Maximum of utilities

Το μοντέλο μέγιστων χρησιμότητων βασίστηκε στο γεγονός ότι ο καταναλωτής επιλέγει τα προϊόντα από τα οποία μπορεί κάθε φορά να αποκομίσει την μέγιστη χρησιμότητα (Ματσατσίνης, 1995). Ως απόρροια αυτού, πιθανότητα αγοράς έχουν τα προϊόντα που ανήκουν στην ομάδα μέγιστων χρησιμότητων. Τώρα, ως ομάδα μέγιστων χρησιμότητων ορίζεται ως το σύνολο των προϊόντων, που έχουν επιλεγεί από τον καταναλωτή, και ανήκουν στην περιοχή των μέγιστων χρησιμότητων. Με την σειρά, ως περιοχή μέγιστων χρησιμότητων ορίζεται μία περιοχή μεταξύ του $U_{i,max}$ και του $U_{i,min}-\varepsilon_i$, όπου ε_i είναι το εύρος της περιοχής των μέγιστων χρησιμότητων και δίνεται από τον τύπο:

$$\varepsilon = \frac{\delta}{v-1} \quad (3.30)$$

- $\delta_i = U_{i,max} - U_{i,min}$
- v : το πλήθος των προϊόντων/ εναλλακτικών

Βάσει μαθηματικής έκφρασης, η πιθανότητα ένας καταναλωτής i να επιλέξει το προϊόν j , σε περιοχή μέγιστων χρησιμότητων, όπου βρίσκονται m προϊόντα ισούται με:

$$P_{ij}(j|C) = \begin{cases} \frac{1}{m}, & \text{όταν } U_{i,max} \geq U_{i,j} \geq U_{i,min} - \varepsilon_i \\ 0, & \text{αλλιού} \end{cases} \quad (3.31)$$

3.6.1.3.7. Το μοντέλο Equal Possibilities

Όγδοο και τελευταίο μοντέλο, με το οποίο θα ασχοληθούμε, είναι αυτό των ίσων χρησιμότητων. Το μοντέλο αυτό έχει πρακτική σημασία στην περίπτωση όπου ο καταναλωτής i αδυνατεί ή δυσκολεύεται να επιλέξει ένα προϊόν σαν το πλέον προτιμότερο. Με άλλα λόγια, ισχύει όταν το εύρος των χρησιμότητων είναι μικρότερο ή ίσο του 0.1. Κάτω από αυτές τις συνθήκες, φαίνεται πως η πιθανότητα ο καταναλωτής i να επιλέξει το προϊόν j ισουαί με:

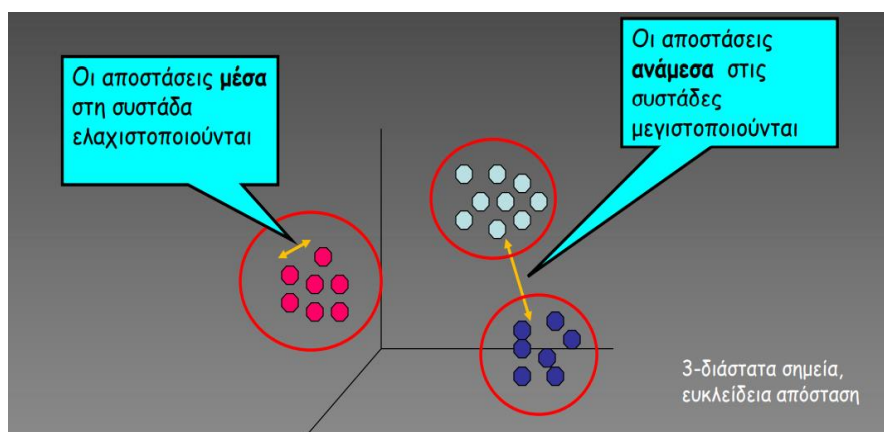
$$P_j = \frac{1}{m}, \text{όταν } U_{i,max} - U_{i,min} \leq 0,1 \quad (3.32)$$

3.6.2. Μέθοδοι εξόρυξης δεδομένων

Οι τεράστιοι όγκοι δεδομένων που έχουν αναπτυχθεί τα τελευταία χρόνια σε συνδυασμό με την ανάγκη των ανθρώπων για διαχείριση αυτών των όγκων και ανάκτηση όλης της απαραίτητης πληροφορίας, οδήγησε στην ανάπτυξη του κλάδου της εξόρυξης δεδομένων (data mining). Ο κλάδος αυτός επηρεάστηκε αισθητά και από άλλους κλάδους, για αυτό ακριβώς τον λόγο υπάρχουν ποικίλοι ορισμοί ανάλογα με τον επιστήμονα, που τους διατύπωσε, και τον κλάδο στον οποίο απασχολείται. Επί παραδείγματι, οι Witten και Frank (2000), ανέφεραν ότι οι μέθοδοι αυτοί ορίζουν πρότυπα για την εκάστοτε διαδικασία βάσει των δεδομένων που χρησιμοποιούν. Από την άλλη, οι Han και Kamber (2001) πρότειναν τις μεθόδους αυτές ως μία ευκαιρία για ανακάλυψη γνώσης μέσα από τα δεδομένα. Τώρα, χαρακτηριστικοί κλάδοι που εφαρμόζουν τέτοιου είδους τεχνικές είναι ο κλάδος της μηχανικής μάθησης, η οπτικοποίηση, η στατιστική, οι βάσεις δεδομένων κ.α.

Μελετώντας τον συγκεκριμένο κλάδο και προσπαθώντας να κατηγοριοποιήσουμε τις μεθόδους και τις τεχνικές, διακρίνονται τέσσερις διαφορετικές ενότητες:

- Πρώτη ενότητα είναι αυτή της ταξινόμησης, όπου τα δεδομένα τοποθετούνται σε προκαθορισμένες ομάδες.
- Δεύτερη ενότητα είναι οι κανόνες συσχέτισης που βρίσκουν κοινά σημεία μεταξύ των δεδομένων (κανόνες).
- Τρίτη ενότητα είναι αυτή των συνοψίσεων, στην οποία βρίσκεται μία συνοπτική περιγραφή για ένα σύνολο ή υποσύνολο δεδομένων.
- Τέταρτη ενότητα, είναι και αυτή που θα μελετηθεί, είναι η συσταδοποίηση, όπου τα δεδομένα ομαδοποιούνται σε μη προκαθορισμένες ομάδες και φέρουν κοινά χαρακτηριστικά μεταξύ τους. Στην κατηγορία αυτή ανήκει και ο αλγόριθμος K-Means που εφαρμόστηκε στην παρούσα έρευνα.



Σχήμα 3.5 Δημιουργία συστάδων

Γενικότερα, οι αλγόριθμοι συσταδοποίησης έχουν ως στόχο την δημιουργία ομάδων και ανήκουν στην κατηγορία των περιγραφικών μοντέλων. Τα στοιχεία της κάθε ομάδας έχουν όσο το δυνατόν πιο όμοια χαρακτηριστικά μεταξύ τους, είναι

δηλαδή αρκετά κοντά το ένα στο άλλο, ενώ παράλληλα η κάθε ομάδα βρίσκεται όσο το δυνατόν πιο μακριά η μία από την άλλη. Για να γίνει πιο κατανοητό, στην συνέχεια παρουσιάζεται το σχήμα 3.5, στο οποίο παρουσιάζεται μία απεικόνιση δεδομένων στον τρισδιάστατο χώρο. Όπως φαίνεται και στο σχήμα, στο εσωτερικό μέρος της ομάδας οι αποστάσεις θα πρέπει αν ελαχιστοποιούνται, ενώ η απόσταση μεταξύ των ομάδων αυξάνεται.

3.6.2.1. Ο αλγόριθμος K-Means

Ο όρος “K-Means” διατυπώθηκε για πρώτη φορά από τον James MacQueen (1967), παρότι σαν ιδέα υπήρχε μία δεκαετία νωρίτερα. Παρόλο αυτά, επίσημα σε κάποιο επιστημονικό περιοδικό δημοσιεύτηκε από τον Lloyd (1957). Κάποια βασικά χαρακτηριστικά του συγκεκριμένου αλγορίθμου ακολουθούν στις επόμενες γραμμές. Πρώτα απ’ όλα, ο K-Means αν και είναι μία αποτελεσματική μέθοδος, βρίσκει εφαρμογή μόνο σε αντικείμενα που είναι σε έναν συνεχή χώρο, σε αντίθεση με άλλες μεθόδους. Επιπλέον, δεδομένου ότι θα πρέπει να ορίζεται το πλήθος των συστάδων εκ των προτέρων, ως είσοδο στον αλγόριθμο, σίγουρα αποτελεί μειονέκτημα της μεθόδου, καθώς απαιτείται παραπάνω χρόνος είτε για την επιλογή αυτής της τιμής, είτε για την δοκιμή διαφόρων πιθανών τιμών. Τέλος, η μέθοδος αυτή παρουσιάζει αδυναμίες τόσο στα μη κυρτά σχήματα, όσο και στα σετ δεδομένων που περιέχουν ακραίες τιμές και θόρυβο.

Γνωρίζοντας πλέον τα βασικά χαρακτηριστικά του αλγορίθμου, θα γίνει επεξήγηση της μεθόδου περιγραφικά, καθώς και πως λειτουργεί ο K-Means σε βήματα. Αρχικά, δίνονται τα δεδομένα εισόδου. Τα δεδομένα αυτά είναι το πλήθος των k συστάδων που θα προκύψουν και οι n πλειάδες που χρησιμοποιούνται για ομαδοποίηση. Έχοντας γίνει πλέον η ανάγνωση των δεδομένων, επιλέγονται τυχαία k αρχικά στοιχεία για να γίνουν τα κεντροειδή/κέντρα των συστάδων. Η διαδικασία αυτή υλοποιείται με την βοήθεια του τύπου 3.33 που ακολουθεί, όπου k το πλήθος των συστάδων και n οι διαστάσεις του προβλήματος.

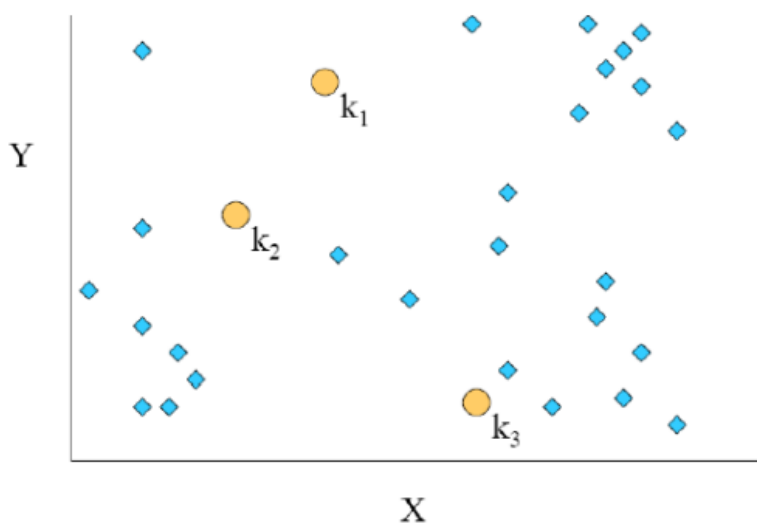
$$CP(x_1, x_2, \dots, x_k) = \left(\frac{\sum_{i=1}^k x_{1st_i}}{k}, \frac{\sum_{i=1}^k x_{2nd_i}}{k}, \dots, \frac{\sum_{i=1}^k x_{nth_i}}{k} \right) \quad (3.33)$$

Έχοντας πλέον βρεθεί τα κέντρα των συστάδων (centroids), ξεκινάει η ανάθεση του κάθε αντικειμένου στο πλησιέστερο κεντρικό σημείο, βάσει μιας μετρικής απόστασης, όπως για παράδειγμα αυτής της ευκλείδειας απόστασης. Η απόσταση αυτή ουσιαστικά ποσοτικοποιεί τις παρατηρήσεις, κατά πόσον είναι κοντά ή όχι η μία από την άλλη. Η σχέση 3.34 που ακολουθεί, παρουσιάζει τον μαθηματικό υπολογισμό της ευκλείδειας απόστασης.

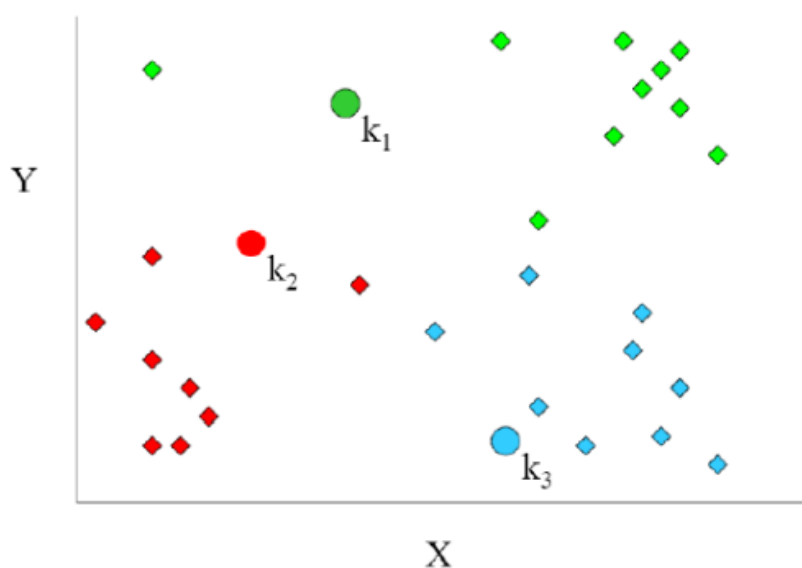
$$d(x, y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \dots + (y_n - x_n)^2} = \sqrt{\sum_{i=1}^n (y_i - x_i)^2} \quad (3.34)$$

Αξίζει να σημειωθεί, πως υπάρχουν και άλλες εναλλακτικές που χρησιμοποιούνται αντί της ευκλείδειας απόστασης. Για παράδειγμα, η απόσταση Manhattan, που

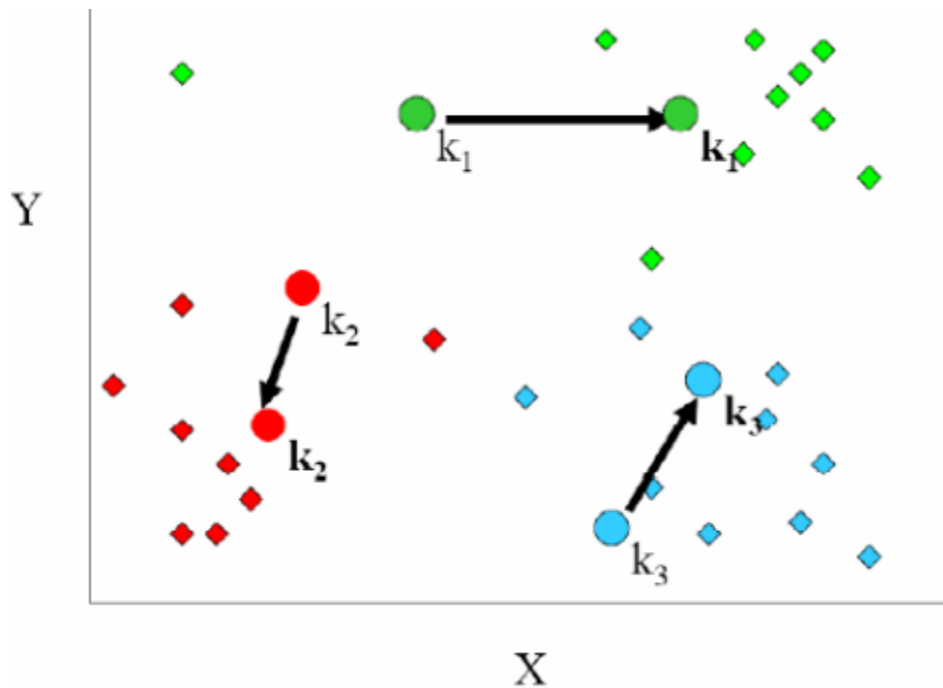
χρησιμοποιεί απόλυτες αποκλίσεις αντί για τετραγωνικές, ή η απόσταση Mahalanobis, η οποία λαμβάνει υπόψη τις συνδυασκόμενες μεταξύ των στοιχείων. Επόμενο βήμα του αλγορίθμου είναι ο επαναπροσδιορισμός του κεντροειδούς για την κάθε συστάδα, και επανάληψη των προηγούμενων βημάτων. Η διαδικασία αυτή ολοκληρώνεται όταν πλέον, τα κεντροειδή της κάθε συστάδας δεν αλλάζουν.



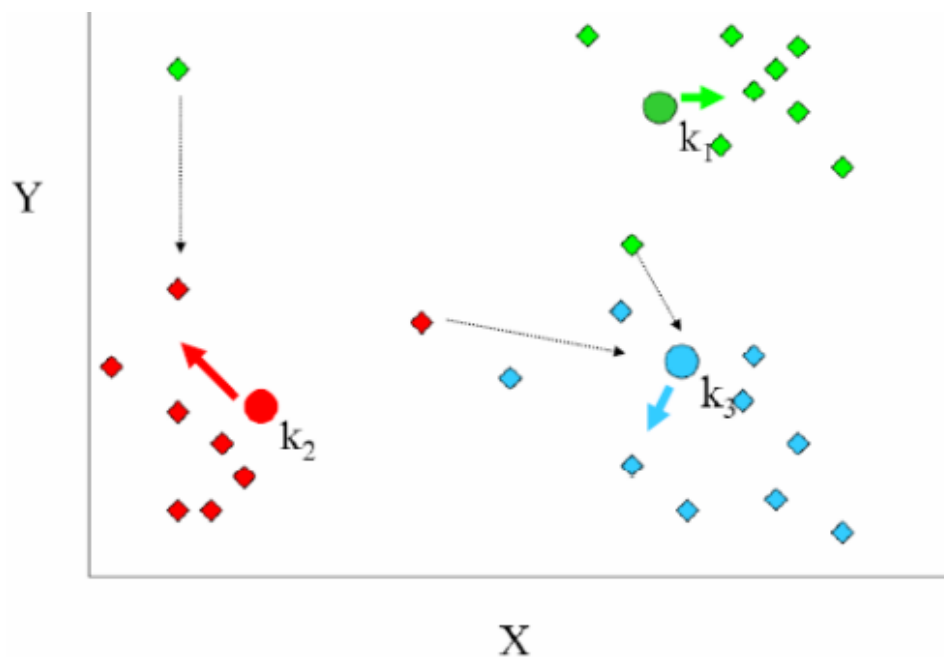
Σχήμα 3.6 Πρώτο βήμα του αλγορίθμου



Σχήμα 3.7 Δεύτερο βήμα αλγορίθμου



Σχήμα 3.8 Τρίτο βήμα αλγορίθμου



Σχήμα 3.9 Τέταρτο βήμα του αλγορίθμου

Όσον αφορά την γραφική αναπαράσταση της μεθόδου, αυτή περιγράφεται στα σχήματα 3.6-3.9 (Tan et al., 2014). Αναλυτικότερα, στο σχήμα 3.6 δίνονται τυχαία τα αρχικά σημεία και υπολογίζονται τα τρία κεντροειδή. Στην συνέχεια στο σχήμα 3.7, έχει γίνει η αρχική ανάθεση όλων των αντικειμένων του δείγματος στις συστάδες. Όπως φαίνεται, τα στοιχεία της πρώτης συστάδας ανήκουν στο πάνω μέρος του

διαγράμματος και είναι με πράσινο χρώμα, ενώ με κόκκινο και γαλάζιο είναι τα αντικείμενα της δεύτερης και τρίτης συστάδας αντίστοιχα. Έπειτα, στο σχήμα 3.8, παρατηρείται ο εκ νέου υπολογισμός των κεντροειδών, και η αντικατάστασή των παλιών από τα νέα κέντρα των συστάδων, ενώ στο σχήμα 3.9 γίνεται η αντικατάσταση των ήδη αρχικά τοποθετημένων αντικειμένων στις νέες συστάδες.

Συνοψίζοντας, ο αλγόριθμος K-Means περιγράφεται στα τέσσερα βήμα που ακολουθούν.

Βήμα 1^ο : Επιλογή k σημείων ως κέντρα των συστάδων.

Βήμα 2^ο : Ανάθεση όλων των αντικειμένων στις συστάδες.

Βήμα 3^ο : Εκ νέου υπολογισμός των κεντροειδών.

Βήμα 4^ο : Επιστροφή στο βήμα 2, μέχρι τα κέντρα των συστάδων να μην αλλάζουν.

3.6.2.1.1. Matlab

Το MATLAB (MATrix LABoratory) θα μπορούσε να χαρακτηριστεί ως μία ιδιαίτερη γλώσσα προγραμματισμού πολλαπλών παραδειγμάτων και με περιβάλλον αριθμητικού υπολογισμού, το οποίο αναπτύχθηκε την δεκαετία του 1960 στα πλαίσια της διδακτορικής διατριβής του Clever Moler (MATLAB, 2022). Εκτός αυτού όμως, πρόκειται και για ένα εργαλείο γραφικής αναπαράστασης δεδομένων, που δίνει πληθώρα δυνατοτήτων στους χρήστες του. Ένα ακόμα σημαντικό χαρακτηριστικό είναι τα toolboxes που περιέχει, τα οποία διευκολύνουν κατά πολύ την εμπειρία του χρήστη. Ως απόρροια αυτού, το MATLAB φαίνεται να βρίσκει εφαρμογή σε ποικίλους κλάδους, όπως για παράδειγμα στον μαθηματικό, τον επιστημονικό και τον τεχνικό.

Παρόλο αυτά, θα πρέπει να σημειωθεί πως αρχικά δεν είχε την σημερινή μορφή, αλλά ούτε αποτελούσε γλώσσα προγραμματισμού, καθώς δεν είχε ούτε προγράμματα, ούτε εργαλείοθήκες, ούτε γραφικά. Ήταν μόνο μία διαδραστική αριθμομηχανή μήτρας. Ωστόσο, με το πέρασμα του χρόνου το περιβάλλον αυτό επαναπρογραμματίστηκε σε C και το 1984 έφτασε η εμπορική του ανάπτυξη. Έκτοτε, γίνονται αρκετές αλλαγές και βελτιώσεις (Hanselman & Littlefield, 2005).

Στην προκειμένη περίπτωση, το MATLAB χρησίμευσε τόσο στην δημιουργία ενός σύντομου κώδικα, όσο και στην οπτικοποίηση των δεδομένων. Όσον αφορά την περιγραφή του κώδικα που αναπτύχθηκε, αυτή εντοπίζεται στο κεφάλαιο “Εφαρμογή μεθόδων”, όπως και τα οπτικοποιημένα αποτελέσματα.

3.6.2.2. Ο Αλγόριθμος DBScan

Ο αλγόριθμος DBScan (Density – Based Spatial Clustering of Applications with Noise) είναι ένας αλγόριθμος ομαδοποίησης δεδομένων, ο οποίος προτάθηκε από τους Ester et al. (1996). Επιπλέον, αξίζει να σημειωθεί πως ο DBScan είναι ένας από τους συνηθέστερους ως προς την εφαρμογή αλγορίθμους και έχει βραβευτεί με το test of time στο συνέδριο εξόρυξης δεδομένων ACM SIGKDD (DBSCAN, 2022). Από θεωρητικής απόψεως, ο αλγόριθμος αυτός χρησιμοποιεί την έννοια της πυκνότητας. Η έννοια της

πυκνότητας είναι γνωστή ήδη από άλλες θετικές επιστήμες, όπως είναι η φυσική, όπου μελετάται το πλήθος των σημείων ανά μονάδα όγκου. Με την ίδια ακριβώς λογική εφαρμόζεται και η πυκνότητα σε αυτόν τον αλγόριθμο. Δηλαδή, για έναν δεδομένο χώρο μελέτης εντοπίζονται τα στοιχεία που ανήκουν σε αυτόν. Όσα περισσότερα στοιχεία ανήκουν σε αυτόν τον χώρο, τόσο μεγαλύτερη είναι και η συγκέντρωση. Ο χώρος αυτός τώρα, ορίζεται βάσει μέτρησης κάποιας απόστασης, όπως είναι για παράδειγμα η ευκλείδεια απόσταση. Στην συνέχεια, από τυχαία σημεία στον χώρο, που είναι το σύνολο των δεδομένων, συγκρίνονται ανά δύο για το πόσο πυκνά είναι μεταξύ τους και σχηματίζονται ομάδες. Οι ομάδες αυτές με την σειρά τους ελέγχονται για το πόσο κοντά είναι με άλλες ομάδες και επαναλαμβάνεται η διαδικασία. Με την ολοκλήρωση αυτής της επαναληπτικής διαδικασίας, όλα εκείνα τα σημεία που δεν έχουν τοποθετηθεί σε κάποια ομάδα, ορίζονται ως θόρυβος (Noise Points).

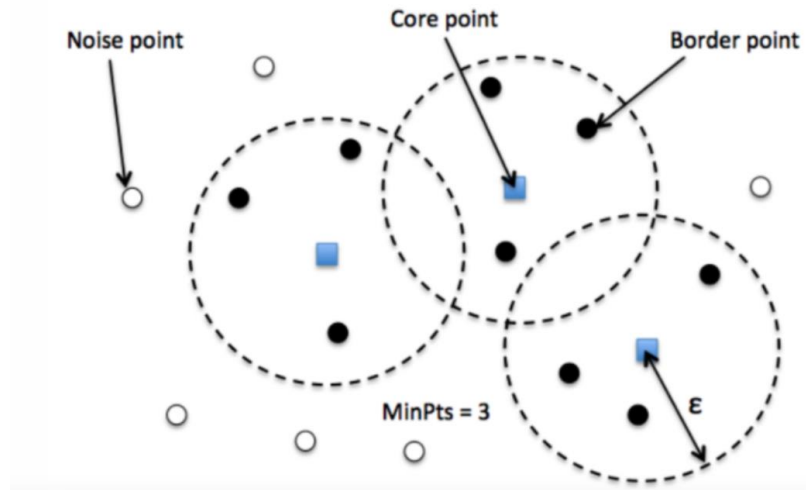
Πέραν αυτών όμως, λαμβάνονται υπόψη και δύο ακόμα παράμετροι κατά την εφαρμογή του αλγορίθμου. Η πρώτη παράμετρος (*eps*) καθορίζει κατά πόσο τα στοιχεία που ελέγχονται κάθε φορά είναι γειτονικά μεταξύ τους (ανήκουν στο ίδιο σύμπλεγμα). Αν τα στοιχεία αυτά βρίσκονται σε απόσταση μικρότερη ή ίση με την τιμή *eps*, τότε τα στοιχεία αυτά μπορούν να θεωρηθούν ότι είναι αρκετά κοντά, άρα και ότι ανήκουν στο ίδιο σύμπλεγμα. Η δεύτερη παράμετρος τώρα, ονομάζεται *MinPoints* και ουσιαστικά ορίζει το ελάχιστο πλήθος των σημείων που είναι απαραίτητο έτσι ώστε να δημιουργηθεί μία πυκνή περιοχή. Για παράδειγμα, αν η παράμετρος οριστεί με την τιμή 4, τότε θα πρέπει να χρησιμοποιηθούν τουλάχιστον τέσσερα σημεία για την δημιουργία της πυκνής περιοχής. Από μαθηματικής απόψεως, ένα σημείο *p*, προσεγγίζεται εύκολα από ένα σημείο *q*, αν και μόνο αν ισχύουν οι σχέσεις 3.35 και 3.36.

$$dis(p, q) \leq Eps \quad (3.35)$$

$$|\{dis(p, q) \leq Eps\}| \geq MinPoints \quad (3.36)$$

Εναλλακτικά, προκειμένου να γίνει πλήρως κατανοητό, στο σχήμα που ακολουθεί δίνονται όλες οι απαραίτητες έννοιες σε μορφή παραδείγματος. Αναλυτικότερα, με μπλε χρώμα σημειώνονται όλα εκείνα τα σημεία που αποτελούν πυρήνες για τον αλγόριθμο. Οι πυρήνες αφορούν το δεύτερο μέρος του περιορισμού (*Eps*) και δείχνουν ότι διατίθενται τα σημεία αρκετά κοντά μεταξύ τους, με αποτέλεσμα να σχηματίζουν μία συστάδα. Ως οριακό σημείο (Border Point) ορίζεται εκείνο το σημείο, το οποίο είναι άμεσα προσεγγίσιμο, και συμβαδίζει με τους περιορισμούς 3.35 και 3.36. Το σημείο αυτό θα πρέπει να ανήκει σε μια συστάδα, καθώς είναι κοντά σε ένα σημείο, το οποίο καλείται πυρήνας αλλά δεν είναι απαραίτητο το ίδιο να είναι πυρήνας. Επιπλέον, ως θόρυβος (Noise Point), νοείται κάθε σημείο που δεν ανήκει σε κάποια συστάδα. Τέλος, έχοντας την τιμή 3 στην παράμετρο *MinPoints*, παρατηρούμε ότι κάθε ομάδα

αποτελείται από 3 τουλάχιστον στοιχεία, ενώ με ϵ (Eps) ορίζεται η απόσταση του πυρήνα από τα υπόλοιπα στοιχεία.



Σχήμα 3.10 Γραφική απεικόνιση βασικών εννοιών του DBScan (DBSCAN Clustering Algorithm in Machine Learning, 2022).

Συνοψίζοντας, η εκτίμηση της τιμής των παραμέτρων δεν είναι και τόσο εύκολη διαδικασία. Συνεπώς, θα πρέπει πρώτον να είναι ξεκάθαρη η χρήση της κάθε παραμέτρου, κατανοώντας την σε βάθος, και δεύτερον να υπάρχει μία γνώση ως προς το σύνολο δεδομένων που θα χρησιμοποιηθεί.



Σχήμα 3.11 Εφαρμογή του K – Means και του DBScan για διαφορετικές κατανομές (DBSCAN Clustering Algorithm in Machine Learning, 2022).

Σε πρακτικό επίπεδο, σημαντικό είναι να αναφερθούν και όλες οι δυνατότητες που προσφέρει ο αλγόριθμος αυτός, καθώς και τα πιθανά μειονεκτήματα που παρουσιάζει. Η βασικότερη διαφορά σε σχέση με άλλους αλγορίθμους, όπως ο K - Means, είναι ο υπολογισμός του πλήθους των συστάδων. Ειδικότερα, ενώ στον K – Means δίνεται το πλήθος των συστάδων ως δεδομένο στον αλγόριθμο, στον DBScan προκύπτει ως αποτέλεσμα μετά την εκτέλεση του αλγορίθμου. Επιπλέον, θα πρέπει να σημειωθεί, πως ο συγκεκριμένος αλγόριθμος φέρει πολύ καλά αποτελέσματα ακόμα

και όταν στα δεδομένα υπάρχει θόρυβος. Για να γίνει αυτό αντιληπτό, δίνεται το παράδειγμα που ακολουθεί. Έστω ένα σετ δεδομένων, στο οποίο εφαρμόζεται τόσο ο αλγόριθμος K – Means, όσο και ο DBScan. Εφαρμόζοντας τον K – Means τα δεδομένα θα χωριστούν σε ομάδες ανεξάρτητα από το αν τα στοιχεία βρίσκονται κοντά το ένα στο άλλο. Αντίθετα, με την χρήση του DBScan τα στοιχεία που βρίσκονται πιο απομακρυσμένα θα οριστούν ως θόρυβος και δεν θα ενταχθούν σε κάποια ομάδα. Τέλος, δεδομένου ότι λειτουργεί βάσει πυκνότητας μπορεί να φέρει καλύτερα αποτελέσματα από τον K – Means σε μία ποικιλία άλλων κατανομών. Η διαπίστωση αυτή απεικονίζεται στο σχήμα 3.11.

Από την άλλη πλευρά, βασικό μειονέκτημα της μεθόδου είναι η πολυπλοκότητα του αλγορίθμου, που έχει σαν αποτέλεσμα να αυξάνεται ο χρόνος εκτέλεσης του. Επιπρόσθετα, ένα ακόμα αρνητικό στοιχείο είναι η εκτίμηση της τιμής των παραμέτρων. Πολλές φορές προκειμένου να εξαχθούν ορθά αποτελέσματα, θα πρέπει να υπάρχει τόσο γνώση πάνω στο σετ δεδομένων που χρησιμοποιείται κάθε φορά, όσο και να δημιουργηθεί κάποιο πιθανό διάγραμμα για την καλύτερη κατανόηση των δεδομένων στον χώρο.

Έχοντας πλέον ολοκληρώσει με την επεξήγηση του αλγορίθμου και τα βασικά χαρακτηριστικά αυτού, στην συνέχεια δίνεται ένα πρακτικό παράδειγμα, όπου βρίσκεται εφαρμογή ο συγκεκριμένος αλγόριθμος. Έστω ότι ο κλάδος ο οποίος θα μελετηθεί είναι αυτός του ηλεκτρονικού εμπορίου και στόχος της εκάστοτε επιχείρησης είναι η εύρεση των αντικειμένων που αγοράζει ο κάθε πελάτης. Δεδομένου ότι δεν γνωρίζουμε τι χρειάζεται ο κάθε πελάτης, ανατρέχοντας σε ένα σύνολο δεδομένων, στο οποίο έχουν αποθηκευτεί παλιότερες αγορές πελατών, είναι πιο εύκολο να προβλεφθούν τα προϊόντα που θα αγοράσει ο κάθε πελάτης. Έτσι, έστω για παράδειγμα ότι ένας πελάτης έχει αγοράσει στο παρελθόν ένα στυλό, ένα μολύβι και ένα τετράδιο, και ένας νέος πελάτης αγοράζει ένα στυλό και το τετράδιο. Με την βοήθεια του αλγορίθμου, δίνεται η δυνατότητα να προτείνουμε στον νέο πελάτη να αγοράσει και το μολύβι μαζί με τα άλλα δύο προϊόντα. Ακόμα, θα πρέπει να τονιστεί πως ο αλγόριθμος μπορεί να βρει εφαρμογή σε πολλούς διαφορετικούς τομείς. Συμπερασματικά, ο αλγόριθμος βρίσκει ομοιότητες μεταξύ των εκάστοτε συμπλεγμάτων, εντοπίζοντας έτσι συσχετίσεις και δομές στα προς μελέτη δεδομένα. Εκτός αυτού όμως, είναι χρήσιμος για τη εύρεση μοτίβων και την πρόβλεψη τάσεων.

3.6.2.3. Ο αλγόριθμος EM

Ο αλγόριθμος EM (Expectation Maximization) εξηγήθηκε για πρώτη φορά επίσημα από τους Dempster et al. (1977). Παρόλο αυτά, ήδη από το 1958 είχε χρησιμοποιηθεί από το HO Hartley (Expectation-Maximization Algorithm, 2022). Δεδομένου ότι υπολογίζονται οι πιθανότητες κάθε εναλλακτικής και παράλληλα γίνεται εκτίμηση των κατανομών από τις πιθανότητες των ομάδων, ο αλγόριθμος EM αποτελεί μία γενίκευση του K – Means σε πιθανοκρατικό πλαίσιο. Για αυτόν ακριβώς τον λόγο, ο συγκεκριμένος αλγόριθμος βρίσκει εφαρμογή ευρέως στην εξόρυξη δεδομένων, ακόμα και όταν υπάρχουν ελλιπή δεδομένα. Όπως και οι περισσότεροι επαναληπτικοί

αλγόριθμοι, έτσι και αυτός αποτελείται από δύο βασικά βήματα τα οποία εναλλάσσονται μεταξύ τους. Το πρώτο βασικό βήμα, ονομάζεται βήμα ευρέσεως προσδοκώμενης τιμής και έχει ως στόχο τον υπολογισμό της αναμενόμενης τιμής μίας πιθανότητας που εκτιμάται, χρησιμοποιώντας τις εκτιμήσεις των παραμέτρων ενός στατιστικού μοντέλου. Επόμενο βήμα του αλγορίθμου είναι το βήμα μεγιστοποίησης, το οποίο υπολογίζει τις παραμέτρους του στατιστικού μοντέλου, μεγιστοποιώντας την αναμενόμενη τιμή της πιθανότητας που βρέθηκε στο προηγούμενο βήμα. Ο αλγόριθμος τερματίζει όταν δεν υπάρχει πλέον βελτίωση και μεταβολές στις τιμές των παραμέτρων των μοντέλων από βήμα σε βήμα. Για την επιβεβαίωση αυτού χρησιμοποιείται ο δείκτης λογαριθμικής πιθανοφάνειας (log likelihood). Πιο αναλυτικά, χρησιμοποιώντας το περιβάλλον της WEKA, από τις παραμέτρους εισόδου, χρησιμοποιήθηκε μόνο το numClusters για την δημιουργία συγκεκριμένου αριθμού συστάδων. Στη συνέχεια, τα αποτελέσματα που εξήχθησαν, εκτός από τον αριθμό των συστάδων όπου ήταν γνωστός, ήταν ο αριθμός των επαναλήψεων του αλγορίθμου, ο χρόνος ολοκλήρωσης της συσταδοποίησης, το σύνολο των ποσοστών των εγγράφων για καθεμία συστάδα και ο δείκτης λογαριθμικής πιθανοφάνειας. Στις εικόνες που ακολουθούν παρουσιάζονται, στο πλαίσιο επεξήγησης του αλγορίθμου, τα αποτελέσματα για την δημιουργία δύο συστάδων.

```
EM
==

Number of clusters: 2
Number of iterations performed: 0

      Cluster
Attribute    0    1
              (0.73) (0.27)
```

Σχήμα 3.12 Αποτελέσματα του αλγορίθμου EM Vol.1

Πρώτα απ' όλα, εμφανίζεται ο αριθμός των συστάδων που δημιουργήθηκαν και οι απαραίτητες επαναλήψεις που χρειάστηκαν για να σχηματιστεί το μοντέλο. Στην συνέχεια, εμφανίζεται το ποσοστό του δείγματος που ανήκει σε μία από τις δύο συστάδες. Δηλαδή, 73% του δείγματος ανήκει στην πρώτη συστάδα και 27% ανήκει στο δεύτερο cluster.

```

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      108 ( 73%)
1       39 ( 27%)

Log likelihood: -31.52001

```

Σχήμα 3.13 Αποτελέσματα του αλγορίθμου EM Vol.2

Ακόμα, στο σχήμα 3.13, εμφανίζεται ο χρόνος εκτέλεσης του αλγορίθμου, ο οποίος ισούται με 0.01 sec και ο αριθμός των στοιχείων που ανήκουν σε κάθε ομάδα. Τέλος, με τιμή -31.52001 εμφανίζεται και ο δείκτης πιθανοφάνειας. Όσο μεγαλύτερη τιμή φέρει ο δείκτης αυτός τόσο καλύτερη εφαρμογή έχει το μοντέλο.

```

0,800044      2      1
0,892877      2      1
0,635793      2      1
[total]      243  174
DODONI
0,864273      2      1
0,842901      2      1
0,800072      3      1

```

Σχήμα 3.14 Αποτελέσματα του αλγορίθμου EM Vol.3

Τέλος, δεδομένου ότι είναι αλγόριθμος που λειτουργεί με λανθάνουσες μεταβλητές (Dahua Lin, n.d.), θα πρέπει να δοθεί ιδιαίτερη έμφαση στον τρόπο επεξήγησης και τον τρόπο σύγκρισης των αποτελεσμάτων. Στη συγκεκριμένη περίπτωση, τα ποσοτικά χαρακτηριστικά που χρησιμοποιήθηκαν ήταν τα βάρη και οι ολικές χρησιμότητες των ερωτηθέντων. Τα χαρακτηριστικά αυτά αποτυπώνονται στην πρώτη στήλη της εικόνας 3.14. Προκειμένου να υπολογιστεί η πιθανότητα υλοποίησης του κάθε χαρακτηριστικού, θα πρέπει να χρησιμοποιηθούν οι τιμές των άλλων δύο στηλών, προς το σύνολο των τιμών ([total]), έχοντας ως στόχο την κανονικοποίηση. Προκειμένου να γίνει πιο κατανοητό στην συνέχεια δίνεται ένα παράδειγμα για την πρώτη εναλλακτική που φαίνεται στην εικόνα 3.14 και που ανήκει στην πρώτη συστάδα. Έτσι, η πιθανότητα υλοποίησης του χαρακτηριστικού αυτού, ισούται με $2/243 = 0.00823$.

Στο σημείο αυτό, θα πρέπει να αναφερθούν τόσο τα πλεονεκτήματα που φέρει ο συγκεκριμένος αλγόριθμος, όσο και τα πιθανά μειονεκτήματα αυτού. Ειδικότερα, τα

αποτελέσματα τα οποία προκύπτουν έχουν ιδιαίτερη στατιστική σημασία και μπορούν να οδηγήσουν σε περαιτέρω στατιστικά συμπεράσματα. Επιπλέον, από προγραμματιστικής απόψεως είναι αρκετά εύκολος και δίνει ικανοποιητικές εκτιμήσεις, όταν οι αρχικές τιμές είναι σε αποδεκτό εύρος. Επιπλέον, είναι ιδανικός στην χρήση όταν εντοπίζονται αρκετές ελλειπείς τιμές. Από την άλλη πλευρά, βασικό μειονέκτημα της μεθόδου είναι ότι φέρει αργή σύγκλιση, καθώς και ότι εξαρτάται από τις εκάστοτε αρχικές τιμές που δίνονται. Η αργή αυτή σύγκλιση μπορεί να οφείλεται στο γεγονός ότι είτε λείπει αρκετή πληροφορία, είτε η αρχική τιμή που δόθηκε είναι πολύ κοντά στην επιθυμητή λύση. Τέλος, απαιτεί την χρήση πολλών αρχικών τιμών, καθώς εντοπίζει τοπικά μέγιστα και όχι ολικά.

3.6.2.3. Ο αλγόριθμος *Hierarchical Clusterer*

Στην εξόρυξη δεδομένων και στην στατιστική, η ιεραρχική ομαδοποίηση, ή αλλιώς ιεραρχική ανάλυση συστάδων (*Hierarchical Cluster Analysis*), είναι μία μέθοδος ανάλυσης συστάδων που έχει ως στόχο την ιεραρχία. Ως ιεραρχία ορίζεται η δυνατότητα διάταξης των στοιχείων «πάνω», «κάτω» ή «στο ίδιο επίπεδο» με άλλα στοιχεία (*Hierarchical Clustering*, 2022). Αυτού του είδους η συσταδοποίηση μπορεί να είναι είτε διαιρετική (*divisive*), είτε συσσωρευτική (*agglomerative*). Η διαιρετική συσταδοποίηση ξεκινάει από μία συστάδα, η οποία αποτελείται από ολόκληρο το προς μελέτη σύνολο στοιχείων. Στην συνέχεια, αυτή η μοναδική συστάδα διαχωρίζεται σε μικρότερες συστάδες, μέχρις ότου ο αριθμός των συστάδων να ισούται με τον αριθμό των πλειάδων. Για να υλοποιηθεί αυτή η διαδικασία, θα πρέπει σε κάθε βήμα να αποφασίζεται τόσο ποια συστάδα θα διαχωριστεί, όσο και ο τρόπος με τον οποίον θα γίνει ο διαχωρισμός. Από την άλλη πλευρά, οι συσσωρευτικοί αλγόριθμοι ξεκινούν ακριβώς αντίθετα. Δηλαδή, αρχικά όλα τα αντικείμενα αποτελούν ξεχωριστές συστάδες, οι οποίες στην συνέχεια, βάσει απόστασης, συγχωνεύονται μέχρις ότου να υπάρξει μία συστάδα. Τέλος, τα αποτελέσματα της ιεραρχικής ομαδοποίησης παρουσιάζονται συνήθως σε μορφή «δέντρου» (δενδροδιάγραμμα).

Στην προκειμένη περίπτωση, ο προς μελέτη αλγόριθμος ανήκει στην δεύτερη κατηγορία, αυτή των συσσωρευτικών αλγορίθμων. Έχοντας ως στόχο την καλύτερη κατανόηση του, στην συνέχεια παρουσιάζονται συνοπτικά τα βήματα όλων των συσσωρευτικών αλγορίθμων (Ματσατσίνης, 2021).

- Πρώτο βήμα είναι η δημιουργία του πίνακα εγγύτητας για όλες τις πλειάδες. Δηλαδή, χρησιμοποιώντας κάθε φορά το εκάστοτε μέτρο για τον υπολογισμό της απόστασης (π.χ. ευκλείδεια απόσταση), βρίσκεται η απόσταση του κάθε στοιχείου με τα υπόλοιπα. Επειδή, η απόσταση του σημείου A από το σημείο B είναι ίδια με την απόσταση του σημείου B από το A, ο πίνακας που προκύπτει είναι συμμετρικός.
- Δεύτερο βήμα είναι η συγχώνευση των κοντινότερων δύο συστάδων. Έχοντας πλέον βρεθεί οι αποστάσεις μεταξύ των σημείων, και λαμβάνοντας υπόψη τα κριτήρια απόστασης, γίνεται η συγχώνευση των συστάδων.

- Τρίτο βήμα είναι η ενημέρωση των δεδομένων του πίνακα εγγύτητας. Μετά την συγχώνευση των ομάδων, είναι αναμενόμενο οι αποστάσεις να έχουν διαφοροποιηθούν. Γι αυτόν ακριβώς τον λόγο ελέγχονται οι αποστάσεις μεταξύ των σημείων των συστάδων με τα υπόλοιπα στοιχεία/ομάδες, και ο πίνακας ενημερώνεται με την μικρότερη δυνατή τιμή που έχει υπολογιστεί ως απόσταση.
- Τέταρτο και τελευταίο βήμα είναι η συνθήκη τερματισμού. Δηλαδή, μία επαναληπτική διαδικασία μέχρις ότου να υπάρξει μία και μόνο συστάδα.

Όσον αφορά τώρα κάποια βασικά χαρακτηριστικά γνωρίσματα του αλγορίθμου τόσο σε σύγκριση με άλλους αλγορίθμους, όσο και ξεχωριστά μόνος του εμφανίζονται στην συνέχεια. Πρώτα απ' όλα, δεδομένου ότι τα αποτελέσματα του Hierarchical Clusterer μπορούν να αποτυπωθούν σε δένδροδιάγραμμα, αυτόματα τον καθιστούν πιο κατανοητό και πιο φιλικό ως προς τους αναγνώστες. Παράλληλα, αφού είναι πιο κατανοητός, είναι και πιο εύκολο να ερμηνευτεί και να εντοπιστεί το πλήθος των συστάδων. Στο σημείο αυτό αξίζει να σημειωθεί, πως ο χρήστης επιλεγεί κάθε φορά το πλήθος των συστάδων που τον εξυπηρετεί ερμηνεύοντας το δένδροδιάγραμμα. Ακόμα, ο αλγόριθμος αυτός βρίσκει εφαρμογή σε οποιουδήποτε τύπου χαρακτηριστικά, ανεξαρτήτως απόστασης και ομοιότητας. Ωστόσο, λόγω της χρονικής πολυπλοκότητας και της μνήμης που χρειάζεται να δεσμεύσει για την δημιουργία του πίνακα εγγύτητας, ο αλγόριθμος αυτός μπορεί να γίνει αρκετά αργός και ακριβός για μεγάλα σύνολα δεδομένων, το οποίο είναι και το βασικό του μειονέκτημα. Για αυτόν ακριβώς τον λόγο, σε τέτοιες περιπτώσεις συνήθως επιλέγεται η εφαρμογή του αλγορίθμου K – Means.

Στο πρακτικό επίπεδο τώρα, και προκειμένου να υπάρξει μία πιο ολοκληρωμένη εικόνα του αλγορίθμου, θα γίνει μία αρκετά σύντομη αναφορά στις παραμέτρους που θα πρέπει να ρυθμιστούν στο λογισμικό της WEKA πριν την εφαρμογή του αλγορίθμου. Εκτός από το πλήθος των συστάδων, θα χρειαστεί να επιλεγεί και η συνάρτηση για τον υπολογισμό της απόστασης των σημείων. Μερικές εναλλακτικές που προτείνονται είναι η ευκλείδεια απόσταση, η Chebyshev, η Manhattan κ.α. Στην συγκεκριμένη περίπτωση επιλέχθηκε η ευκλείδεια απόσταση, η οποία υπολογίζεται από τον τύπο 3.37.

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^d (x_i - y_i)^2} \quad (3.37)$$

Επόμενη παράμετρο που θα πρέπει να ρυθμιστεί είναι αυτή της μεθόδου του υπολογισμού της απόστασης μεταξύ των συστάδων. Μερικές από τις εναλλακτικές που προτείνονται είναι αυτοί του απλού συνδέσμου (single link), του πλήρους συνδέσμου (complete link), της μέσης τιμής ομάδας (average link), του κεντροειδούς συστάδων (Distance between Centroids), της μεθόδου Ward κ.α. Από αυτές επιλέχθηκε η πρώτη. Στην περίπτωση του απλού συνδέσμου, η εγγύτητα ορίζεται από την απόσταση μεταξύ των δύο πλησιέστερων σημείων δύο γειτονικών ομάδων. Ο τύπος που χρησιμοποιείται για τον υπολογισμό αυτό είναι ο ακόλουθος.

$$d(C_i, C_j) = \min\{d(x, y) \mid x \in C_i, y \in C_j; i = 1, 2, \dots, n; j = 1, 2, \dots, m\} \quad (3.38)$$

Για την καλύτερη κατανόηση του αλγορίθμου, στις επόμενες γραμμές δίνεται το εξής παράδειγμα. Έστω το ακόλουθο σετ δεδομένων με τις συγκεκριμένες συντεταγμένες, όπως αυτές φαίνονται στον πίνακα που ακολουθεί.

Πίνακας 3.2 Συντεταγμένες παραδείγματος για τον Hierarchical Clusterer

Σημεία	X	Y
A	4,0	5,3
B	2,2	3,8
Γ	3,5	3,2
Δ	2,6	1,9
E	0,8	4,1
Z	4,5	3,0

Πριν ξεκινήσει η συσταδοποίηση, θα πρέπει πρώτα να υπολογιστούν οι αποστάσεις μεταξύ όλων των σημείων για να δημιουργηθεί ο πίνακας εγγύτητας. Χρησιμοποιώντας τον τύπο 3.37, υπολογίζεται η ευκλείδεια απόσταση. Στην συνέχεια δίνεται ως παράδειγμα ο υπολογισμός της απόστασης μεταξύ των σημείων A και B.

$$d(A, B) = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} = \sqrt{(4 - 2,2)^2 + (5,3 - 3,8)^2} = 2,34$$

Με όμοιο τρόπο υπολογίζονται και οι υπόλοιπες αποστάσεις για όλα τα σημεία. Ο πίνακας 3.9 που παρουσιάζεται στην συνέχεια είναι ο πίνακας ευκλείδειας απόστασης. Ο λόγος για τον οποίο είναι συμπληρωμένος ο μισός από αυτόν, οφείλεται στην συμμετρία, δηλαδή ότι η απόσταση από το σημείο A στο B είναι ίδια με την απόσταση από το B στο A.

Πίνακας 3.3 Πίνακας ευκλείδειας απόστασης.

Σημεία	A	B	Γ	Δ	E	Z
A	0					
B	2,34	0				
Γ	2,16	1,43	0			
Δ	3,68	1,94	1,6	0		
E	3,42	1,43	2,85	2,6	0	
Z	2,35	2,44	1,02	2,2	3,86	0

Δεδομένου ότι στην συγκεκριμένη περίπτωση εφαρμόστηκε η μέθοδος του απλού συνδέσμου, χρησιμοποιείται ο τύπος 3.38 για τον υπολογισμό της απόστασης μεταξύ των συστάδων. Έστω ότι θέλουμε να υπολογίσουμε την απόσταση της συστάδας (Γ,Z)

και της (B,E) που έχουν δημιουργηθεί από την συνένωση των αρχικών συστάδων. Ισχύει ότι:

$$d((\Gamma, Z), (B, E)) = \min(d(\Gamma, B); d(Z, B); d(\Gamma, E); d(Z, E)) \\ = \min(1, 43; 2, 44; 2, 85; 3, 86) = 1, 43$$

Με όμοιο τρόπο υπολογίζονται και οι αποστάσεις των άλλων συστάδων.

3.6.2.4. Ο αλγόριθμος Canopy

Ο αλγόριθμος Canopy αποτελεί ουσιαστικά έναν αλγόριθμο προ – ομαδοποίησης χωρίς επίβλεψη, ο οποίος χρησιμοποιείται ως βήμα προ – επεξεργασίας τόσο για τον αλγόριθμο K – Means, όσο και για τον αλγόριθμο ιεραρχικής ομαδοποίησης. Διατυπώθηκε για πρώτη φορά από τους Andrew McCallum, Kamal Nigam και Lyle Ungar και βασικότερος στόχος του συγκεκριμένου αλγορίθμου είναι η γρηγορότερη ομαδοποίηση μεγάλων συνόλων δεδομένων (Canopy clustering algorithm, September 2022). Όπως και στον αλγόριθμο K – Means, έτσι και στον αλγόριθμο Canopy χρησιμοποιούνται μέτρα απόστασης για την εκχώρηση των δεδομένων σε ομάδες, ενώ παράλληλα ως είσοδο στον αλγόριθμο δίνονται και οι τιμές T_1 και T_2 που ορίζονται ως κατώφλια. Πιο αναλυτικά, για τα δύο κατώφλια T_1 και T_2 ισχύει η σχέση $T_1 > T_2$. Για να γίνει πιο κατανοητή η γρηγορότερη ομαδοποίηση των δεδομένων, κρίνεται απαραίτητη η ανάλυση της μεθόδου και η επεξήγησή της σε βήματα.

- Πρώτο βήμα του αλγορίθμου είναι η δημιουργία μίας πρώιμης ομάδας. Η ομάδα αυτή, προκύπτει από μία τυχαία επιλογή μίας εγγραφής από το σύνολο δεδομένων.
- Δεύτερο βήμα είναι ο υπολογισμός των αποστάσεων των υπόλοιπων σημείων από αυτό της πρώιμης ομάδας. Αν η απόσταση των σημείων αυτών είναι μικρότερη από την τιμή T_1 τότε τα στοιχεία αυτά εκχωρούνται στην ομάδα.
- Στο τρίτο βήμα γίνεται η πιθανή αφαίρεση δεδομένων. Όσα στοιχεία έχουν απόσταση μικρότερη από T_2 , τότε τα δεδομένα αυτά αφαιρούνται εντελώς ώστε να αποφευχθεί η περαιτέρω επεξεργασία τους.
- Το τέταρτο βήμα είναι η επανάληψη των προηγούμενων βημάτων του αλγορίθμου. Πρόκειται ουσιαστικά για μία επαναληπτική διαδικασία, μέσω της οποίας προκύπτει ένα σύνολο από ομάδες (canopies), όπου η καθεμία αποτελείται από ένα ή περισσότερα στοιχεία.
- Πέμπτο και τελευταίο βήμα είναι η συνθήκη τερματισμού. Ο αλγόριθμος ολοκληρώνεται όταν αφαιρεθούν όλα τα στοιχεία από το αρχικό σύνολο.

Κεφάλαιο 4 Η συλλογή και η προ – επεξεργασία των δεδομένων

4.1. Εισαγωγή

Στο κεφάλαιο αυτό, γίνεται αναλυτική περιγραφή όλων των βημάτων που ακολουθήθηκαν στην συγκεκριμένη έρευνα, όσον αφορά το κομμάτι της συλλογής και της προ – επεξεργασίας των δεδομένων. Ειδικότερα, θα παρουσιαστεί ο τρόπος σύνταξης του ερωτηματολογίου, ο τρόπος ανάπτυξής του, καθώς και ερωτήσεις αυτού. Επιπλέον, θα γίνει αναφορά στον καθαρισμό των δεδομένων από τυχαίες και ελλιπείς τιμές που υπήρχαν στις απαντήσεις των ερωτηματολογίων. Ως τυχαίες τιμές ορίζονται οι τιμές που συμπληρώθηκαν από τους χρήστες και ακολουθούσαν ένα συγκεκριμένο μοτίβο (π.χ. συμπλήρωση τρίτης επιλογής σε όλες τις ερωτήσεις), ή οι απαντήσεις που αναιρούσαν προηγούμενες απαντήσεις. Η αφαίρεση ή η επεξεργασία αυτών των τιμών έγινε απευθείας στο excel. Επιπρόσθετος καθαρισμός έγινε και με την βοήθεια του συντελεστή συσχέτισης τ του Kendall, με την βοήθεια του MARKEX. Τελευταίο βήμα ήταν ο έλεγχος για την ύπαρξη ακραίων τιμών, μέσω του κατάλληλου φίλτρου στη WEKA, καθώς και ο έλεγχος για την αξιοπιστία της συσταδοποίησης με τον δείκτη silhouette, που έγινε στο MATLAB. Τέλος, σε καθένα από τα βήματα αυτά, θα γίνεται ταυτόχρονη παρουσίαση των αποτελεσμάτων ή των αρχείων που παράγονται, με στόχο την καλύτερη κατανόηση.

4.2 Το ερωτηματολόγιο της έρευνας

Στη δεύτερη υποενότητα, πρόκειται να παρουσιαστεί ο τρόπος μέσω του οποίου συγκεντρώθηκαν τα δεδομένα. Δηλαδή, τόσο το πώς κατασκευάστηκε το ερωτηματολόγιο και το πώς διαμοιράστηκαν αυτά, όσο και ο λόγος που δομήθηκε με αυτόν τον τρόπο, καθώς και τι εξυπηρετεί η καθεμία ερώτηση στην έρευνα.

4.2.1. Τρόπος σύνταξης ερωτηματολογίου

Προκειμένου να αναπτυχθεί το συγκεκριμένο ερωτηματολόγιο, χρειάστηκε πρώτα να γίνει μία έρευνα στο διαδίκτυο και να βρεθούν τα προϊόντα που είναι ανταγωνιστικά μεταξύ τους. Δηλαδή, το είδος, ο τύπος του γιαουρτιού καθώς και τις εταιρείες που προτιμούν οι καταναλωτές. Βάσει αυτών των στοιχείων επιλέχθηκαν τα επτά πιο ανταγωνιστικά προϊόντα. Έπειτα, επιλέχθηκαν επτά διαφορετικά κριτήρια, όπως η γεύση, το άρωμα και η τιμή. Τα κριτήρια αυτά ουσιαστικά είναι κοινά γνωρίσματα, που φέρουν τα συγκεκριμένα προϊόντα μεταξύ τους, και τα οποία λαμβάνονται υπόψη από τους χρήστες προτού αγοράσουν το εκάστοτε γιαούρτι. Στην συνέχεια, το ερωτηματολόγιο αναπτύχθηκε στο “Google Forms”, με το μεγαλύτερο μέρος αυτών να αποστέλλονται ηλεκτρονικά, ενώ ένα πολύ μικρό ποσοστό απαντήθηκε μέσω τηλεφωνικής επικοινωνίας. Η χρονική διάρκεια που απαιτήθηκε μέχρις ότου να συγκεντρωθεί ένα ικανοποιητικό δείγμα, διακόσια είκοσι ένα στον αριθμό, ήταν ίση με τέσσερις μήνες. Πιο ειδικά, από τον Απρίλιο του 2021 έως τον Ιούλιο του 2021. Τέλος,

τα ερωτηματολόγια μοιράστηκαν κυρίως με αυτούς τους δύο τρόπους, λόγω της πανδημίας covid-19.

4.2.2. Η ανάλυση του ερωτηματολογίου

Όπως έχει ήδη αναφερθεί σε προηγούμενες ενότητες (κεφ. 4.2.1. και στην περίληψη), τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα διπλωματική εργασία, προέκυψαν μέσα από ένα ερωτηματολόγιο που αναπτύχθηκε. Το ερωτηματολόγιο αυτό, αποτελείται συνολικά από είκοσι δύο ερωτήσεις, δύο εκ των οποίων είναι προαιρετικές. Πιο αναλυτικά, χωρίζεται σε τρία διαφορετικά επίπεδα ανάγνωσης. Το πρώτο επίπεδο είναι αυτό των ερωτήσεων πολλαπλής επιλογής, όπου ο χρήστης δίνει πληροφορίες γενικού χαρακτήρα ή εκφράζει τις προτιμήσεις του. Το δεύτερο επίπεδο είναι οι ερωτήσεις που χρησιμοποιήθηκαν κατά την πολυκριτήρια ανάλυση. Δηλαδή, ερωτήσεις όπου οι χρήστες δηλώνουν τη σημαντικότητα των εκάστοτε χαρακτηριστικών που διερωτώνται. Τέλος, υπάρχουν ερωτήσεις σύντομης ανάπτυξης, όπου ο κάθε χρήστης μπορεί να γράψει την άποψή του για οτιδήποτε του ζητείται.

Στο σημείο αυτό θα πρέπει να σημειωθεί πως όλες οι απαντήσεις είναι ανώνυμες και τα στοιχεία που συλλέχθηκαν χρησιμοποιήθηκαν αποκλειστικά και μόνο για τους σκοπούς της διπλωματικής εργασίας. Οι χρήστες είναι ενήμεροι προτού συμμετάσχουν στην έρευνα και συμφωνούν πλήρως με την προηγούμενη δήλωση.

4.2.2.1. Γενικές ερωτήσεις πολλαπλής επιλογής

Το πρώτο είδος των ερωτήσεων που καλείται να απαντήσει ο καταναλωτής αποτελείται συνολικά από 9 (ή 10 ερωτήσεις αν υπολογιστεί και η γενική πολυκριτήρια) ερωτήσεις. Το είδος των ερωτήσεων αυτών, σχετίζονται κυρίως με γενικά στοιχεία δημογραφικού χαρακτήρα, ερωτήσεις προτίμησης και ερωτήσεις καταναλωτικής συμπεριφοράς. Ειδικότερα, στην αρχή του ερωτηματολογίου ο χρήστης καλείται να απαντήσει στις προτάσεις που παρουσιάζονται στην συνέχεια.

1. Ποιο είναι το φύλο σας;

Η συγκεκριμένη ερώτηση αποτελεί κλασικό παράδειγμα ερώτησης δημογραφικού χαρακτήρα, η οποία είναι απαραίτητη για την περαιτέρω ανάλυση των δεδομένων. Χρησιμοποιείται προκειμένου να ελεγχθεί αν υπάρχει κάποιο χαρακτηριστικό ή κριτήριο το οποίο εμφανίζεται συχνότερα σε ένα από τα δύο φύλα. Δηλαδή, αν οι άνδρες ή οι γυναίκες έχουν κάποια συγκεκριμένη προτίμηση ή συμπεριφορά.

2. Ποια είναι η ηλικία σας;

Στην συνέχεια, ο ερωτώμενος καλείται να δηλώσει την ηλικία του. Η ερώτηση αυτή, όπως και η προηγούμενη είναι υποχρεωτική. Δεδομένου ότι τα ερωτηματολόγια που συλλέγονται είναι ανώνυμα, ο χρήστης μπορεί να

συμπληρώσει πολύ πιο εύκολα το ευαίσθητο αυτό προσωπικό στοιχείο. Η ερώτηση αυτή είναι δημογραφικού χαρακτήρα, όπως και η προηγούμενη, και σαν απώτερο σκοπό έχει τη δημιουργία πιθανών ηλικιακών ομάδων με κοινά χαρακτηριστικά.

3. Από πόσα μέλη αποτελείται συνολικά η οικογένειά σας για την οποία κάνετε την αγορά του γιαουρτιού;

Στην ερώτηση αυτή ο χρήστης δηλώνει απλά έναν αριθμό. Ως στόχο έχει την εύρεση του πλήθους των ατόμων που καταναλώνουν το εκάστοτε προϊόν γιαουρτιού. Αυτό συμβαίνει προκειμένου να δούμε αν υπάρχουν μέλη στην οικογένεια που δεν καταναλώνουν το συγκεκριμένο γαλακτοκομικό προϊόν, και κατ' επέκταση να αποκτηθεί μία καλύτερη εικόνα γύρω από την κατανάλωση του γιαουρτιού σε οικογενειακό επίπεδο.

4. Όσον αφορά το μορφωτικό επίπεδο, σε ποια κατηγορία ανήκετε;

Πρόσθετη σημαντική πληροφορία για την έρευνα αποτελεί το μορφωτικό επίπεδο του κάθε ερωτώμενου. Η ερώτηση αυτή είναι επίσης δημογραφικού χαρακτήρα και υποχρεωτική, ως προς τη συμπλήρωση, και συμβάλλει στην καλύτερη κατανόηση του προφίλ του κάθε καταναλωτή.

5. Ποιο είναι το επάγγελμά σας;

Με την ίδια ακριβώς λογική, όπως και στην προηγούμενη ερώτηση, ο ερωτώμενος θα απαντήσει στην ερώτηση για το ποιο είναι το επάγγελμά του, δίνοντας παραπάνω πληροφορίες για τον ίδιο.

6. Πόσο συχνά αγοράζετε γιαούρτι;

Στη συγκεκριμένη ερώτηση ο καταναλωτής θα πρέπει να επιλέξει την απάντηση που ανταποκρίνεται περισσότερο στη δική του περίπτωση. Δηλαδή, το πόσο συχνά αγοράζει γιαούρτι ή επιδόρπια γιαουρτιού. Η ερώτηση αυτή σχετίζεται με την καταναλωτική του συμπεριφορά, και αποτελεί βασική πληροφορία για την παρούσα έρευνα.

7. Ποιο είναι το είδος του γιαουρτιού που προτιμάτε κυρίως να καταναλώνετε;

Στην ερώτηση αυτή τώρα, ο χρήστης θα πρέπει να δηλώσει το είδος του γιαουρτιού που προτιμάει να καταναλώνει περισσότερο. Με τον τρόπο αυτό, είναι πιο εύκολο να μελετηθούν ξεχωριστά οι ερωτώμενοι σε μικρότερες ομάδες (εδώ βάσει του είδους τους γιαουρτιού που καταναλώνουν), παρατηρώντας όχι μόνο την απήχηση που έχει το κάθε είδος γιαουρτιού, αλλά και πιθανές άλλες κοινές συμπεριφορές μεταξύ των καταναλωτών.

8. Ποιον τύπο γιαουρτιού προτιμάτε να καταναλώνετε συνήθως;

Στο σημείο αυτό και προκειμένου να αποφευχθεί κάποια σύγχυση του ερωτώμενου, δίνονται επεξηγήσεις για τους περισσότερους τύπους γιαουρτιού που υπάρχουν. Η ερώτηση αυτή είναι ίδιας λογικής με την προηγούμενη.

9. Από πού αγοράζετε συνήθως το συγκεκριμένο προϊόν; (πολλαπλές απαντήσεις)

Στην ένατη ερώτηση πλέον, ο ερωτώμενος έχει την δυνατότητα να επιλέξει παραπάνω από μία απαντήσεις. Στόχος εδώ είναι να βρεθεί το μέρος όπου οι καταναλωτές συνηθίζουν να αγοράζουν το συγκεκριμένο προϊόν όπου μελετάται, καθώς και να βρεθεί σε ποιο σημείο γίνονται οι περισσότερες πωλήσεις αυτού.

10. Σημειώστε, στο αντίστοιχο τετράγωνο, το βαθμό σημαντικότητας που αποδίδετε σε κάθε ένα από τα παρακάτω κριτήρια-χαρακτηριστικά, κατά τη διαδικασία επιλογής του γιαουρτιού σας, χρησιμοποιώντας την κλίμακα 0-10 (Όπου με 0 σημειώνετε όταν δεν το λαμβάνετε καθόλου υπόψη ενώ με 10 αυτό με την μέγιστη σημαντικότητα)

	1	2	3	4	5	6	7	8	9	10
Επωνυμία Εταιρείας (Brand Name)										
Γεύση										
Προέλευση Γάλακτος(τόπος παραγωγής)										
Υφή										
Άρωμα										
Συσκευασία										
Τιμή										

Η συγκεκριμένη ερώτηση λειτουργεί επικουρικά και ως στόχο έχει την επιβεβαίωση των τελικών αποτελεσμάτων. Να γίνει γνωστό δηλαδή, ποιο κριτήριο είναι σημαντικότερο για την πλειοψηφία του δείγματος.

4.2.2.2. Πολυκριτήριες ερωτήσεις

Το δεύτερο μέρος του ερωτηματολογίου αποτελείται από ερωτήσεις προτίμησης, όπου ο κάθε καταναλωτής δηλώνει ποιο ανταγωνιστικό προϊόν του αρέσει περισσότερο βάσει κάποιων συγκεκριμένων κριτηρίων. Τα κριτήρια αυτά είναι επτά στο σύνολό τους και πιο συγκεκριμένα είναι: η γεύση, η τιμή, η προέλευση του γάλακτος που χρησιμοποιείται στην παρασκευή του γιαουρτιού, το άρωμα, η

συσκευασία, η υφή και η επωνυμία. Στο είδος αυτό των ερωτήσεων, ο καταναλωτής θα πρέπει να βαθμολογήσει σε μια πενταβάθμια κλίμακα το καθένα χαρακτηριστικό για καθένα από τα επτά διαφορετικά ανταγωνιστικά προϊόντα. Στη συνέχεια ακολουθεί ένα ενδεικτικό παράδειγμα των ερωτήσεων αυτών. Επιπλέον, αξίζει να σημειωθεί πως με την συμπλήρωση αυτών των ερωτήσεων κατασκευάζεται και ο πολυκριτήριος πίνακας που θα χρησιμοποιηθεί για την είσοδο των δεδομένων στο MARKEX.

1. Ποια είναι η γνώμη σας για το κριτήριο 'Γεύση' των παρακάτω προϊόντων γιαουρτιού;

	Άσχημη	Περίεργη	Αδιάφορη	Ωραία	Πολύ ωραία
Κρι-Κρι αγελαδινό στραγγιστό 'αγαπημένο'					
Δωδώνη Ηπειρωτικό αγελαδινό στραγγιστό					
ΔΕΛΤΑ διπλοστραγγιστό 'Μικρές οικογενειακές Φάρμες'					
ΦΑΓΕ Total					
ΝΟΥΝΟΥ αγελαδινό στραγγιστό γιαούρτι					
Όλυμπος κατσικίσιο γιαούρτι					
Ροδόπη κατσικίσιο γιαούρτι					

4.3. Τυχαίες και ελλιπείς τιμές

Πρώτο βήμα για την υλοποίηση της εργασίας είναι η συγκέντρωση των απαραίτητων δεδομένων. Αφού ολοκληρώθηκε η συγκέντρωση των ηλεκτρονικών ερωτηματολογίων, μέσω του Google Forms, δημιουργήθηκε ένα αρχείο excel που αποτελούνταν από 221 ερωτηματολόγια. Ο πρώτος έλεγχος που έγινε ήταν πολύ γρήγορος, καθώς αφαιρέθηκαν ερωτηματολόγια τα οποία είχαν συμπληρωθεί στην τύχη από τους ερωτώμενους. Δηλαδή, παρουσίαζαν συγκεκριμένα επαναλαμβανόμενα

μοτίβα στις απαντήσεις τους ή υπήρχαν ερωτήσεις που αναιρούσαν τις απαντήσεις που είχαν δώσει οι χρήστες προηγουμένως. Από την διαδικασία αυτή αφαιρέθηκαν 19 ερωτηματολόγια.

Επόμενο βήμα για την προ - επεξεργασία των δεδομένων ήταν η συμπλήρωση των ελλιπών τιμών. Στο σημείο αυτό, αξίζει να σημειωθεί πως όπου εμφανίζονταν ελλιπείς τιμές, το κάθε ερωτηματολόγιο επεξεργάζονταν ξεχωριστά. Ειδικότερα, στις απλές περιπτώσεις όπου ήταν εύκολο να εξαχθούν συμπεράσματα για τις αγοραστικές και προσωπικές συνήθειες του κάθε ατόμου, οι ελλιπείς τιμές συμπληρώνονταν. Για να γίνει πιο κατανοητό αυτό, παρακάτω ακολουθεί ένα απλό παράδειγμα. Έστω ότι ο ερωτώμενος, ο οποίος συμπληρώνει το ερωτηματολόγιο, αγοράζει κάθε εβδομάδα μία οικογενειακή συσκευασία, του κιλού, γιαούρτι, ενώ παράλληλα είναι ο μόνος που καταναλώνει γιαούρτι στην οικογένεια. Γίνεται ευκόλως αντιληπτό πως θα πρέπει να καταναλώνει γιαούρτι σχεδόν κάθε μέρα έτσι ώστε την επόμενη εβδομάδα να αγοράσει εκ νέου μία νέα, αντίστοιχη συσκευασία. Με την λογική αυτή συμπληρώθηκαν κάποιες από τις ελλιπείς τιμές. Οι ελλιπείς τιμές που εμφανίστηκαν συνολικά για όλο το δείγμα των 202 ερωτηματολογίων δεν ξεπερνούσαν τις δέκα, ενώ μόλις σε δύο από αυτές δεν μπορούσε να εξαχθεί κάποιο ασφαλές συμπέρασμα, και διαγράφηκαν πλήρως τα ερωτηματολόγια. Με την ολοκλήρωση της διαδικασίας αυτής, το συνολικό δείγμα αποτελούνταν από 200 ερωτηματολόγια.

4.4. Συντελεστής συσχέτισης τ του Kendall

Εν συνεχεία, έχοντας πλέον αφαιρέσει τις τυχαίες και τις ελλιπείς τιμές στο σύνολο του δείγματος, διαμορφώθηκε κατάλληλα ένα αρχείο excel για την είσοδο των δεδομένων στο MARKEK. Το αρχείο αυτό αποτελούνταν από τέσσερα διαφορετικά φύλλα. Ειδικότερα, το πρώτο φύλλο περιείχε γενικά στοιχεία για την μέθοδο UTASTAR, όπως το πλήθος των ερωτηματολογίων, των εναλλακτικών-ανταγωνιστικών προϊόντων και των κριτηρίων που χρησιμοποιήθηκαν, καθώς και η κωδικοποίηση για την απαραίτητη μεταβελτιστοποίηση του προβλήματος. Το δεύτερο φύλλο, με όνομα "Criteria" φέρει όλα τα χαρακτηριστικά εκείνα τα οποία αξιολογεί ο χρήστης πριν την αγορά του προϊόντος. Όπως έχει αναφερθεί και προηγουμένως, τα κριτήρια αυτά ήταν επτά στο σύνολο και πιο συγκεκριμένα ήταν η γεύση, το brand name, το άρωμα, η υφή, η συσκευασία, η προέλευση και η τιμή. Επιπρόσθετα στοιχεία του φύλλου αυτού, είναι η μονοτονία του κάθε κριτηρίου (αύξον ή φθίνον), ο τύπος του κριτηρίου (ποιοτικό ή ποσοτικό), η μέγιστη και η ελάχιστη τιμή που μπορούν να φέρουν, καθώς και το πλήθος των τιμών που μπορεί να πάρει το κάθε κριτήριο. Στο τρίτο φύλλο παρουσιάζονται μόνο τα ονόματα των επτά εναλλακτικών και τέλος στο τέταρτο φύλλο περιέχονται οι κωδικοποιημένες απαντήσεις που είχαν δώσει οι ερωτώμενοι. Πιο αναλυτικά, οι επτά πρώτες στήλες έφεραν την βαθμολογία για τα επτά κριτήρια αντίστοιχα του πρώτου προϊόντος, οι επόμενες επτά στήλες την βαθμολογία για τα επτά κριτήρια για το δεύτερον προϊόν και ούτω καθεξής.

Έχοντας πλέον έτοιμο το αρχείο εισόδου για το σύνολο του δείγματος, και αφού εισήχθη στο MARKEK, για ανάγνωση των δεδομένων, εφαρμόστηκε η UTASTAR. Από

την καρτέλα “Consumers Analysis”, και επιλέγοντας το post optimization, εμφανίζεται το πλήθος των στοιχείων που ο δείκτης συσχέτισης είναι ίσος με ένα, το πλήθος των στοιχείων που έχουν από 0.9-1 και των στοιχείων εκείνων που είναι από 0.9 και κάτω. Αξίζει να σημειωθεί ότι βάσει του δείκτη συσχέτισης (Kendall’s Tau) έγινε εκ νέου καθαρισμός των δεδομένων, αφαιρώντας όλες εκείνες τις τιμές που οι τιμές τους ήταν κάτω από 0.7. Τα αποτελέσματα του συντελεστή συσχέτισης παρουσιάζονται αναλυτικά στον παρακάτω πίνακα 4.1.

Πίνακας 4.1 Ο δείκτης τ του Kendall για το συνολικό δείγμα

Kendall’s Tau	<0.7	0.7-0.8	0.8-0.9	0.9-1
Πλήθος	9	25	40	126

Μετά την μείωση του δείγματος, σύμφωνα πάντα με τον συντελεστή συσχέτισης επαναλήφθηκε η παραπάνω διαδικασία. Η μόνη διαφορά ήταν ότι το νέο αρχείο εισόδου πλέον περιείχε μικρότερο πλήθος ερωτηματολογίων (191), καθώς είχαν αφαιρεθεί όλα εκείνα τα ερωτηματολόγια με Kendall’s Tau μικρότερο του 0.7. Εφαρμόζοντας εκ νέου την UTASTAR για τα νέα δεδομένα, από την καρτέλα “Consumers Analysis”, με την επιλογή post optimization πάντα, και παράλληλα με την επιλογή τύπος αποτελεσμάτων “Global Utilities”, βρέθηκε όλο το πλήθος των ολικών χρησιμοτήτων για όλο το δείγμα της έρευνας. Το δείγμα αυτό αποθηκεύτηκε σε ένα αρχείο excel για περαιτέρω χρήση, στην WEKA και στο MATLAB, που παρουσιάζεται στην υποενότητα 4.4 και 4.5 αντίστοιχα. Με την ίδια ακριβώς λογική, και επιλέγοντας ως τύπο αποτελεσμάτων το “Criteria Weights”, αποθηκεύτηκαν σε ένα άλλο αρχείο excel τα βάρη των κριτηρίων για καθένα ερωτώμενο ξεχωριστά.

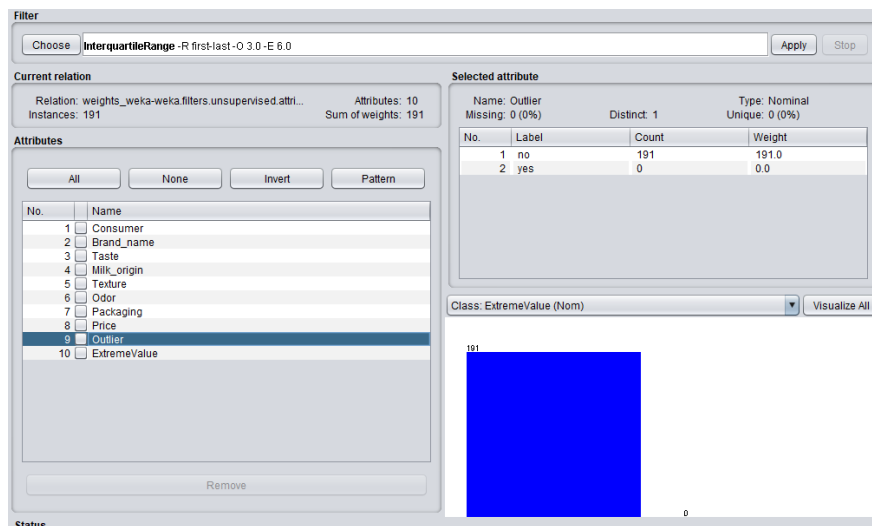
4.5. Ακραίες τιμές

Έχοντας πλέον τα βάρη και τις ολικές χρησιμότητες, από την διαδικασία που ακολουθήθηκε στην υποενότητα 4.4 για όλο το πλήθος του δείγματος, επόμενο βήμα είναι η συσταδοποίηση. Δηλαδή, ο διαχωρισμός των ερωτώμενων σε ομάδες που φέρουν κοινά χαρακτηριστικά μεταξύ τους. Πριν από την αρχική εφαρμογή του αλγορίθμου K-Means στο περιβάλλον της MATLAB, κρίνεται απαραίτητος ο έλεγχος ακραίων τιμών στο δείγμα. Για τον λόγο αυτό, χρησιμοποιήθηκε το περιβάλλον της WEKA. Στην συνέχεια παρουσιάζονται μερικές βασικές πληροφορίες για το περιβάλλον αυτό. Πρώτα απ’ όλα φαίνεται πως πήρε το λογισμικό αυτό το συγκεκριμένο όνομα (Waikato Environment for Knowledge Learning). Από το πανεπιστήμιο στο οποίο αναπτύχθηκε, το Waikato στην Νέα Ζηλανδία. Ως στόχο έχει την εξόρυξη γνώσης από δεδομένα, με χρήση εργαλείων οπτικοποίησης και μία μεγάλη ποικιλία αλγορίθμων. Μερικά χαρακτηριστικά παραδείγματα κατηγοριών/οικογένειες μεθόδων που εντοπίζονται στο λογισμικό είναι η συσταδοποίηση, η παλινδρόμηση και η επεξεργασία δεδομένων γενικότερα. Όσον αφορά τώρα την πρόσβαση σε αυτό (<https://www.cs.waikato.ac.nz/ml/weka/>, n.d.), ο καθένας μπορεί να το χρησιμοποιήσει και να αποκτήσει πρόσβαση δωρεάν. Στον σύνδεσμό αυτό παρέχονται

επιπλέον πληροφορίες για το λογισμικό, καθώς και αρχεία για εφαρμογή των αλγορίθμων (Πεπενιδης, κα., 2015).

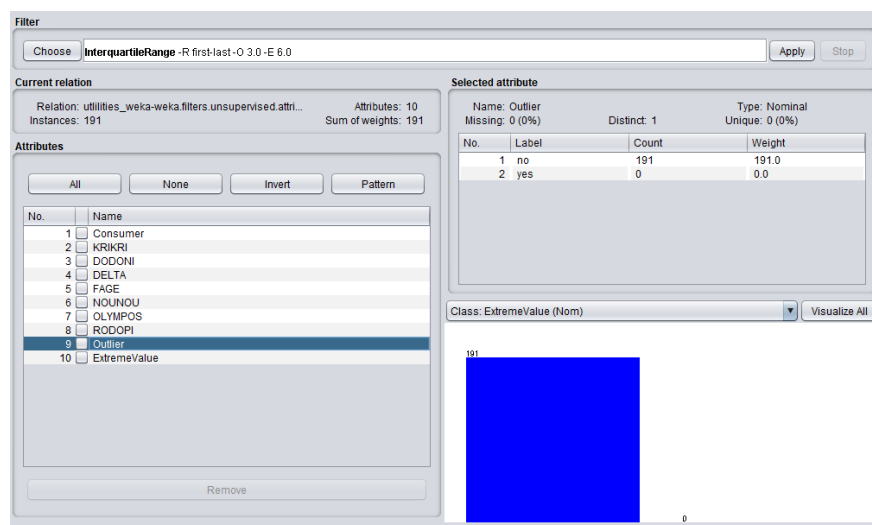
Αφού ανέβηκε και διαβάστηκε το αρχείο των βαρών, έχοντας την απαραίτητη μορφή, επιλέχθηκε το φίλτρο “Interquartile Range”. Η διαδικασία επιλογής ήταν πολύ απλή. Το μονοπάτι για την εύρεση του συγκεκριμένου φίλτρου φαίνεται στην συνέχεια.

Choose→Filters→Unsupervised→Attribute



Σχήμα 4.1 Ακραίες τιμές στο αρχείο βαρών

Πατώντας “Apply”, στο δεξιό μέρος της οθόνης, εμφανίζεται στο αρχείο μία επιπλέον στήλη, όπου φαίνεται αν υπάρχουν ακραίες τιμές ή όχι. Επιλέγοντας το χαρακτηριστικό των ακραίων τιμών και το κατάλληλο class, επιλέγεται το “Visualize All” και οπτικοποιείται το ζητούμενο, όπως αυτό φαίνεται στο σχήμα 4.1.

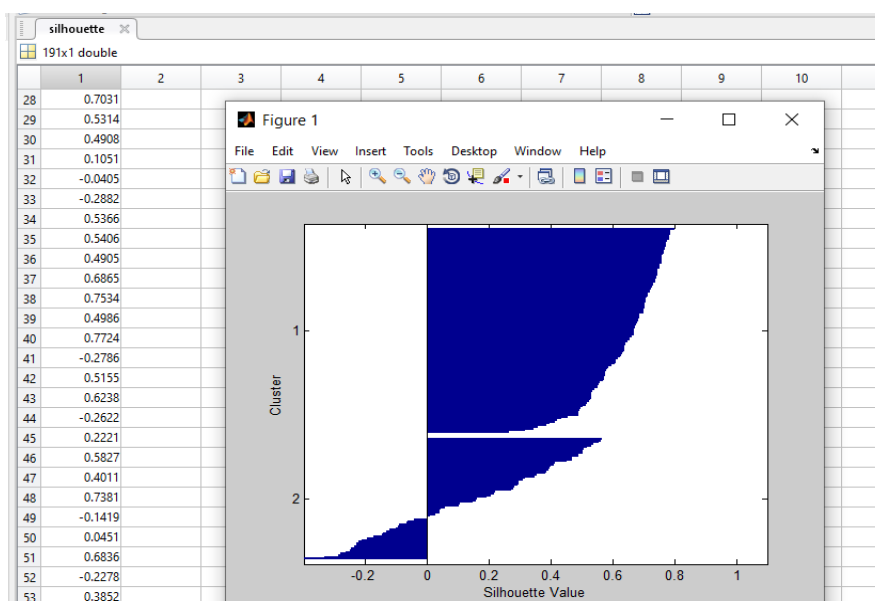


Σχήμα 4.2 Ακραίες τιμές στο αρχείο ολικών χρησιμότητων

Η ίδια ακριβώς διαδικασία ακολουθήθηκε και για το αρχείο των ολικών χρησιμότητων, και το αποτέλεσμα εμφανίζονται στο σχήμα 4.2. Αξίζει να σημειωθεί πως και τα δύο αρχεία δεν περιείχαν ακραίες τιμές, συνεπώς το δείγμα της έρευνας δεν μειώθηκε περαιτέρω και παρέμεινε στα 191 ερωτηματολόγια.

4.6. Δείκτης σκιαγράφησης Silhouette

Με την εφαρμογή του αλγορίθμου K-Means στο MATLAB, ήταν απαραίτητος και ο έλεγχος του δείκτη σκιαγράφησης (silhouette), για τον έλεγχο αξιοπιστίας της συσταδοποίησης (Rousseeuw, 1987). Πιο αναλυτικά, παίρνοντας τα αποτελέσματα από την εφαρμογή του αλγορίθμου, τόσο για το αρχείο με τα βάρη των κριτηρίων, όσο και για το αρχείο με τις ολικές χρησιμότητες, στόχος ήταν η ύπαρξη όσο το δυνατόν μεγαλύτερης τιμής για τον μέσο δείκτη silhouette, ενώ παράλληλα να μην υπάρχουν αρνητικές τιμές σε κανέναν δείκτη silhouette μεμονωμένα. Το εύρος των τιμών του δείκτη σκιαγράφησης ορίζεται μεταξύ του -1 και του +1. Όσο πιο αρνητική είναι η τιμή του τόσο περισσότερα δεδομένα έχουν συσταδοποιηθεί λάθος, έχουν εκχωρηθεί δηλαδή σε λάθος ομάδα τα στοιχεία. Αντίθετα, όσο πιο θετικός και κοντά στην μονάδα είναι ο δείκτης, τόσο πιο επιτυχημένη είναι η συσταδοποίηση. Αξίζει να σημειωθεί πως οι τιμές κοντά στο μηδέν υποδηλώνουν επικαλυπτόμενες συστάδες. Δηλαδή, ο δείκτης σκιαγράφησης ενός ερωτώμενου, καλό θα ήταν να είναι θετικός, έτσι ώστε να προκύψει μία μεγαλύτερη τιμή στον μέσο δείκτη silhouette για όλο το δείγμα. Όσες απαντήσεις είχαν αρνητικές τιμές ημειώνονταν και στην συνέχεια αφαιρούνταν και για τα δύο αρχεία. Η διαδικασία αυτή ολοκληρώθηκε όταν πλέον δεν υπήρχαν άλλες αρνητικές τιμές. Στο σημείο αυτό το τελικό δείγμα είχε μειωθεί κατά 44 ερωτηματολόγια (147 εναπομείναντα).



Σχήμα 4.3 Γραφική απεικόνιση δείκτη silhouette

Η επόμενη εικόνα που ακολουθεί (σχήμα 4.3), είναι ενδεικτική και χρησιμοποιείται για την καλύτερη κατανόηση της διαδικασίας. Τα τελικά αποτελέσματα παρουσιάζονται στο κεφάλαιο 5.2 μαζί με την εφαρμογή του αλγορίθμου K-Means. Ειδικότερα, στην αριστερή πλευρά του σχήματος, στην στήλη 1, εμφανίζονται οι τιμές του κάθε δείκτη silhouette για τον αντίστοιχο ερωτώμενο. Για παράδειγμα, στις γραμμές 32, 33, 41, 44, 49 και 52 παρατηρούμε αρνητικές τιμές. Οι τιμές αυτές σημειώθηκαν και στην συνέχεια αφαιρέθηκαν και από το αρχείο με τα βάρη των κριτηρίων, και από αυτό με τις ολικές χρησιμότητες. Η αφαίρεση αυτή έγινε και στα δύο αρχεία, προκειμένου να εξεταστούν οι ίδιοι ερωτώμενοι. Αντίστοιχα, στο γράφημα παρουσιάζεται ο δείκτης silhouette για όλο το σετ δεδομένων. Στο γράφημα, τα στοιχεία που βρίσκονται στο κάτω μέρος είναι αυτά με την αρνητική τιμή και τα οποία αφαιρέθηκαν.

Κεφάλαιο 5 Εφαρμογή μεθόδων

5.1. Εισαγωγή

Στο κεφάλαιο αυτό, γίνεται πλήρης παρουσίαση του δείγματος της έρευνας από στατιστικής απόψεως, καθώς και η αναλυτική περιγραφή όλων των βημάτων που ακολουθήθηκαν στην συγκεκριμένη έρευνα, όσον αφορά τις μεθόδους που χρησιμοποιήθηκαν. Πιο αναλυτικά, αναφέρονται ποια αρχεία χρησιμοποιήθηκαν ως είσοδο στην εκάστοτε περίπτωση, καθώς και τι περιείχαν τα αρχεία αυτά. Ακόμα, η περιγραφική αυτή διαδικασία θα γίνει τόσο για την UTASTAR, που υλοποιήθηκε στο λογισμικό MARKECH, όσο και για τον K-Means και τους υπόλοιπους αλγορίθμους συσταδοποίησης. Ένα επιπλέον κομμάτι της περιγραφικής αυτής διαδικασία είναι και η μελέτη ανάπτυξης και εισαγωγής τριών ανταγωνιστικών προϊόντων στην ενιαία αγορά γιαουρτιού. Τέλος, θα συμπεριληφθούν όλα τα απαραίτητα διαγράμματα και γραφήματα, μαζί με την απαραίτητη επεξήγηση.

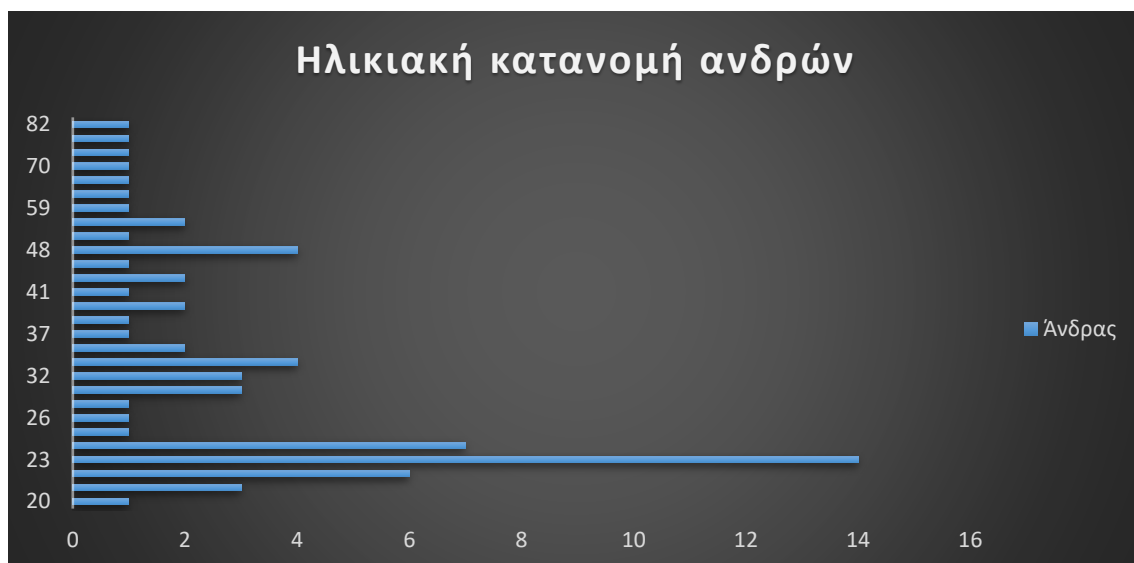
5.2. Η στατιστική περιγραφή

Μελετώντας το δείγμα της έρευνας, στην συνέχεια παρουσιάζεται σε μορφή πινάκων και διαγραμμάτων, τα διάφορα χαρακτηριστικά στοιχεία για την καλύτερη κατανόησή αυτού.



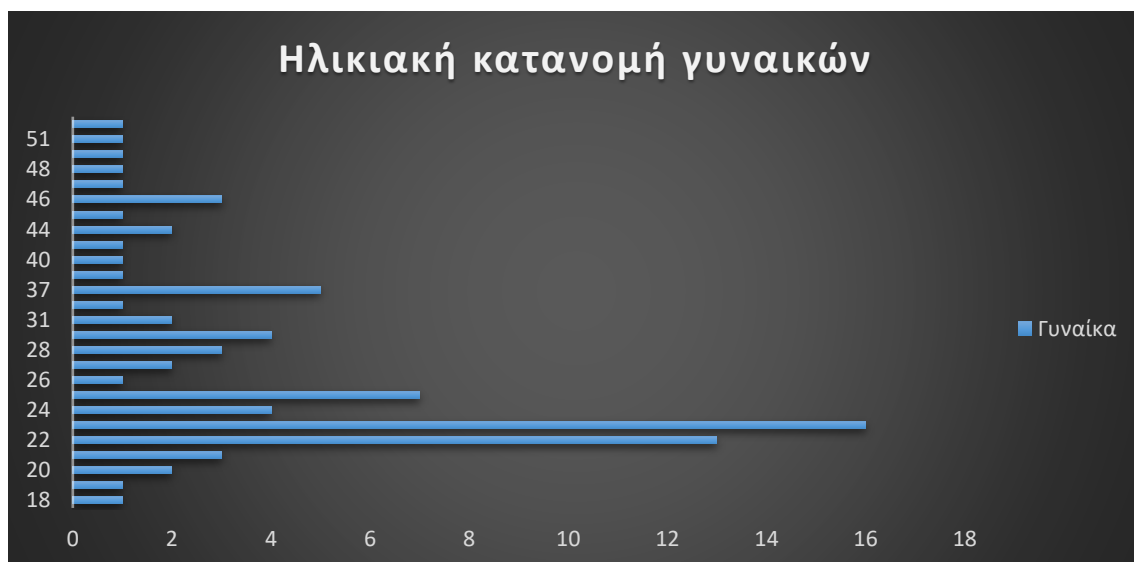
Σχήμα 5.1 Ποσοστιαία κατανομή του φύλου των συμμετεχόντων

Πιο αναλυτικά όπως φαίνεται και στην εικόνα 5.1, το 46% του δείγματος αποτελείται από άνδρες, ενώ το υπόλοιπο 54% από γυναίκες.



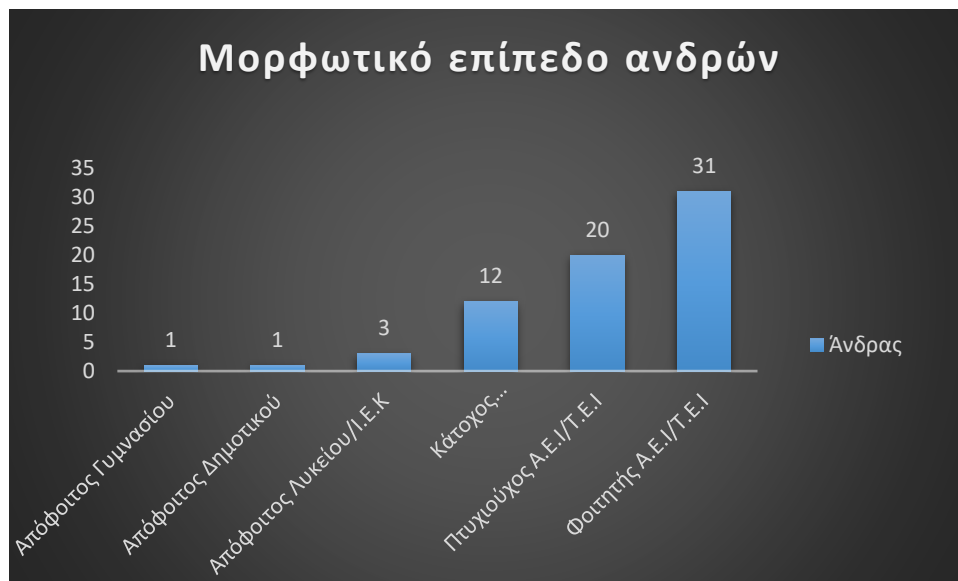
Σχήμα 5.2 Ηλικιακή κατανομή ανδρών

Όσον αφορά την ηλικιακή κατανομή των ανδρών, αυτή παρουσιάζεται στο σχήμα 5.2 που ακολουθεί, ενώ το αντίστοιχο διάγραμμα για την ηλικιακή κατανομή των γυναικών εμφανίζεται στο σχήμα 5.3.



Σχήμα 5.3 Ηλικιακή κατανομή γυναικών

Όπως φαίνεται στο παραπάνω διάγραμμα, οι περισσότεροι άνδρες του δείγματος ανήκουν σε μικρότερες ηλικιακές ομάδες και κυρίως στην δεκαετία των είκοσι έως τριάντα. Ακριβώς τα ίδια ισχύουν και για το γυναικείο δείγμα. Αυτό συμβαίνει καθώς το μεγαλύτερο μέρος του δείγματος αποτελείται από τον κοινωνικό περίγυρο αυτού που ανέπτυξε το ερωτηματολόγιο.



Σχήμα 5.4 Μορφωτικό επίπεδο ανδρών

Το μορφωτικό επίπεδο των ανδρών (σχήμα 5.4) αποτελείται κυρίως από φοιτητές και απόφοιτους πανεπιστημίων. Επιπλέον, δώδεκα από τους συνολικούς εξήντα οκτώ είναι κάτοχοι κάποιου μεταπτυχιακού, ενώ υπάρχουν και συνολικά δύο απόφοιτοι του δημοτικού και γυμνασίου.



Σχήμα 5.5 Μορφωτικό επίπεδο γυναικών

Από την άλλη πλευρά, το δείγμα των γυναικών φαίνεται να μην αποτελείται καθόλου από απόφοιτους γυμνασίου ή δημοτικού, αλλά η πλειοψηφία αυτού αποτελείται από φοιτήτριες και πτυχιούχους. Τέλος, αξίζει να σημειωθεί πως δώδεκα γυναίκες είναι

επίσης κάτοχοι μεταπτυχιακού. Όσον αφορά τώρα την κατανομή των επαγγελμάτων, τόσο στις γυναίκες όσο και στους άνδρες, εμφανίζονται τα αντίστοιχα διαγράμματα στην συνέχεια. Τα δύο φύλα ως κοινό χαρακτηριστικό έχουν την ύπαρξη των φοιτητών ως την πλέον επικρατέστερη ομάδα. Στους άνδρες η δεύτερη πιο πολυπληθής ομάδα είναι αυτή των ιδιωτικών υπαλλήλων, ενώ η ομάδα αυτή στις γυναίκες έρχεται τρίτη, αμέσως μετά των δημοσίων υπαλλήλων. Όσοι συνταξιούχοι εντοπίστηκαν στο δείγμα της έρευνας είναι όλοι άνδρες.



Σχήμα 5.6 Επαγγέλματα ανδρών



Σχήμα 5.7 Επαγγέλματα γυναικών



Σχήμα 5.8 Συχνότητα αγοράς γιαουρτιού (άνδρες)

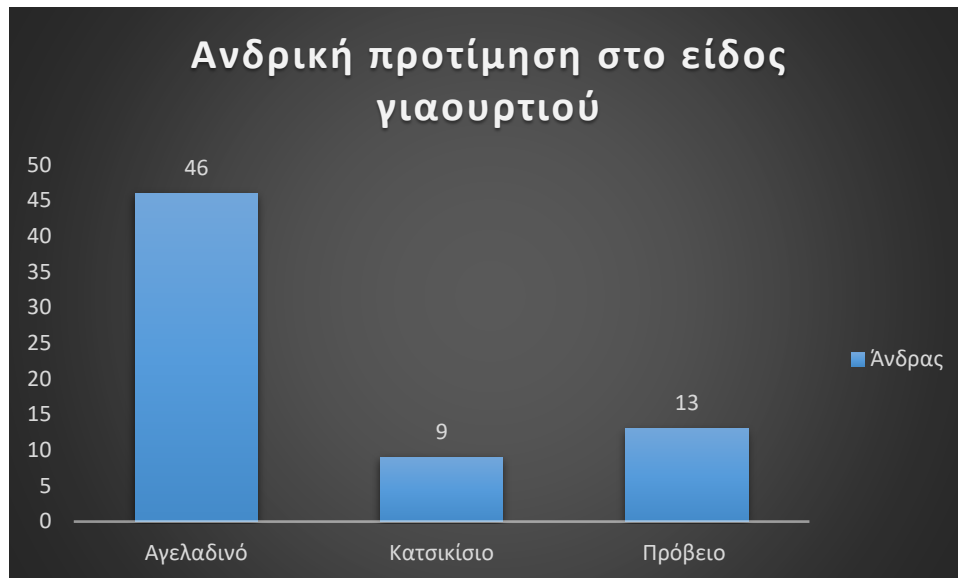
Όπως φαίνεται και στο σχήμα 5.8, η πλειοψηφία των ανδρών φαίνεται να αγοράζει γιαούρτι κάθε εβδομάδα, ενώ η αμέσως επόμενη επιλογή είναι 2 με 3 φορές τον μήνα. Ως τρίτη επιλογή έρχεται η αγορά κάθε δύο ή τρεις φορές την εβδομάδα με διαφορά 2 βαθμών από την δεύτερη. Όμοια συμπεριφορά και αποτελέσματα στην κατάταξη, έχει και το γυναικείο δείγμα.



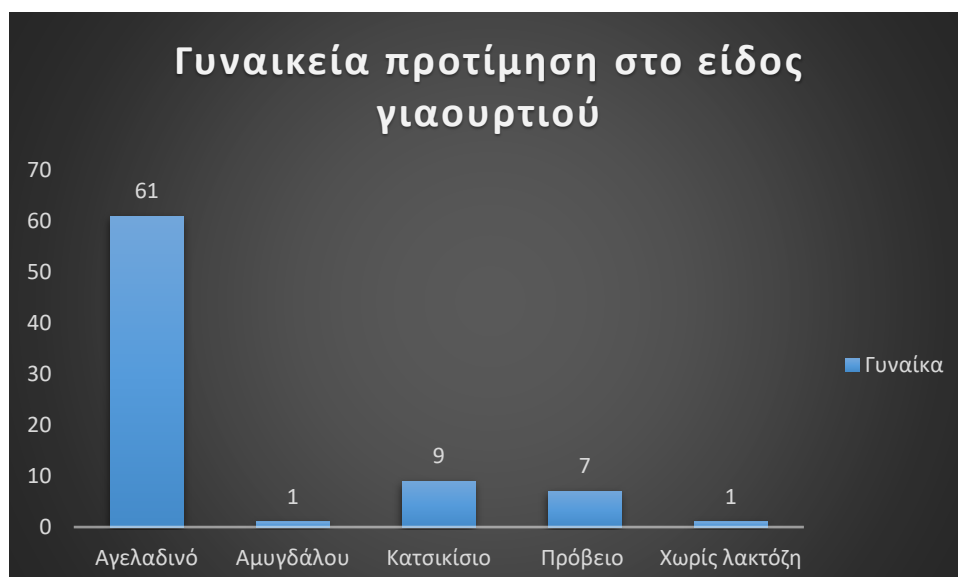
Σχήμα 5.9 Συχνότητα αγοράς γιαουρτιού (γυναίκες)

Η προτίμηση στα είδη γάλακτος φαίνεται να διαφέρουν ανάλογα το φύλο του ανθρώπου. Αναλυτικότερα, οι γυναίκες φαίνεται να επιλέγουν περισσότερα είδη

γάλακτος προς κατανάλωση, σε αντίθεση με τους άνδρες που καταφεύγουν στα τρία πιο γνωστά και συνηθισμένα είδη (αγελαδινό, κατσικίσιο και πρόβειο). Επιπλέον, τόσο οι άνδρες, όσο και οι γυναίκες ως δημοφιλέστερο είδος γιαουρτιού επιλέγουν το αγελαδινό. Εκτός αυτού όμως, η επιλογή αυτή έρχεται πρώτη σε προτίμηση με μεγάλη διαφορά από το δεύτερο είδος. Το δεύτερο είδος σε προτίμηση για τους άνδρες είναι το πρόβειο, ενώ για τις γυναίκες το κατσικίσιο.

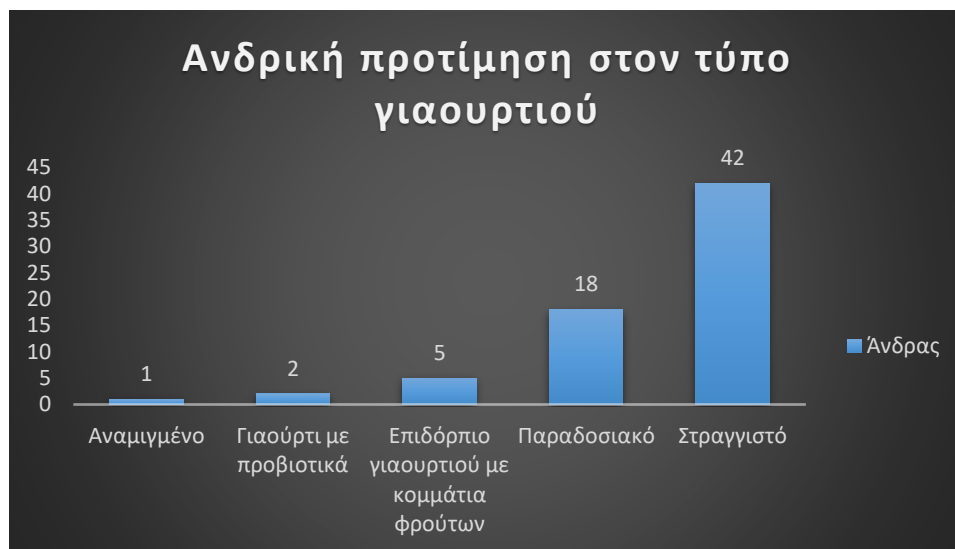


Σχήμα 5.10 Κατανομή ανδρικών προτιμήσεων για το είδος του γιαουρτιού

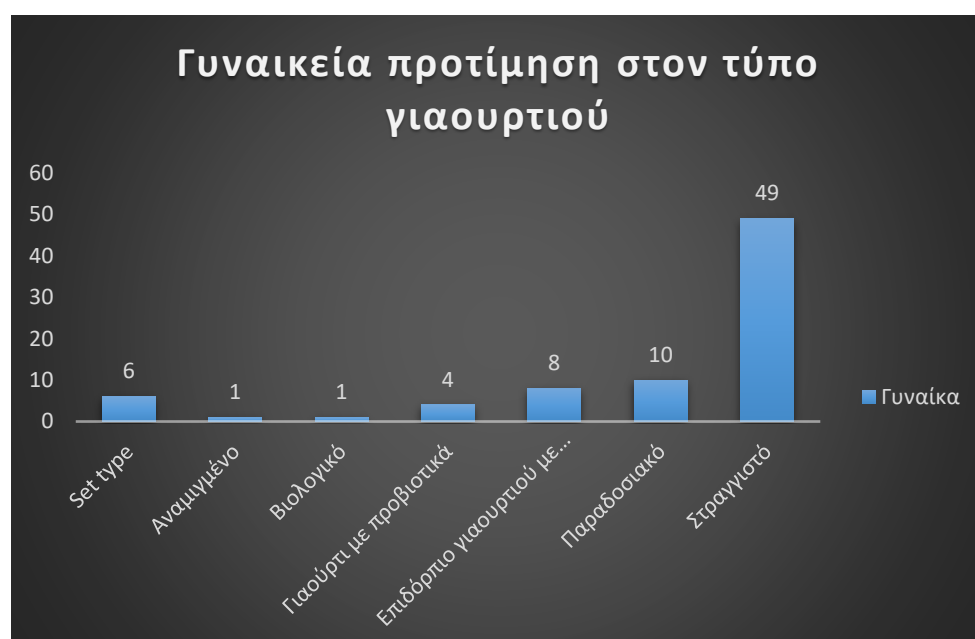


Σχήμα 5.11 Κατανομή γυναικείων προτιμήσεων για το είδος του γιαουρτιού

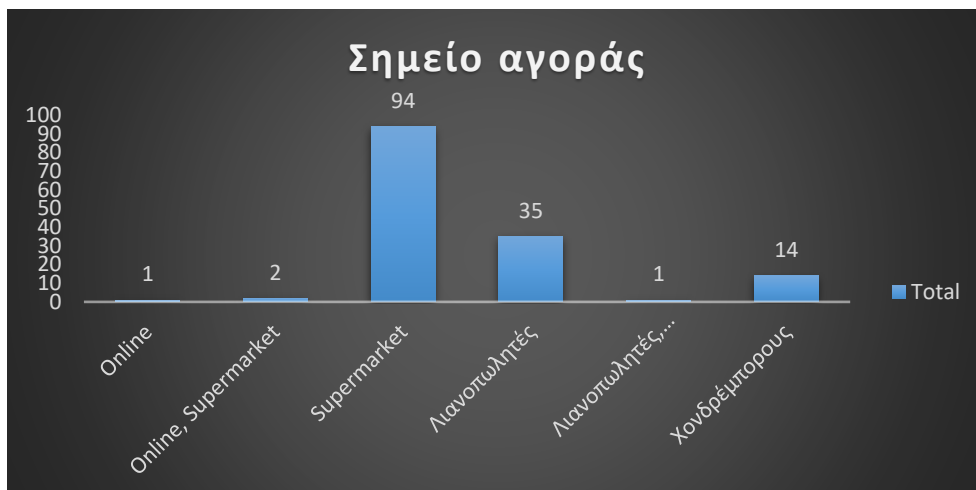
Οι προτιμήσεις στον τύπο του γιαουρτιού και για τα δύο φύλα είναι σχεδόν ίδιες. Πρώτη επιλογή έρχεται το στραγγιστό γιαούρτι με μεγάλη διαφορά έναντι του παραδοσιακού που βρίσκεται αρκετά πιο πίσω. Στην τρίτη επιλογή εμφανίζονται τα επιδόρπια γιαουρτιού. Μικρότερα μερίδια στην αγορά κατέχουν τα γιαούρτια με τα προβιοτικά, το set type και τα βιολογικά.



Σχήμα 5.12 Κατανομή ανδρικών προτιμήσεων για τον τύπο του γιαουρτιού

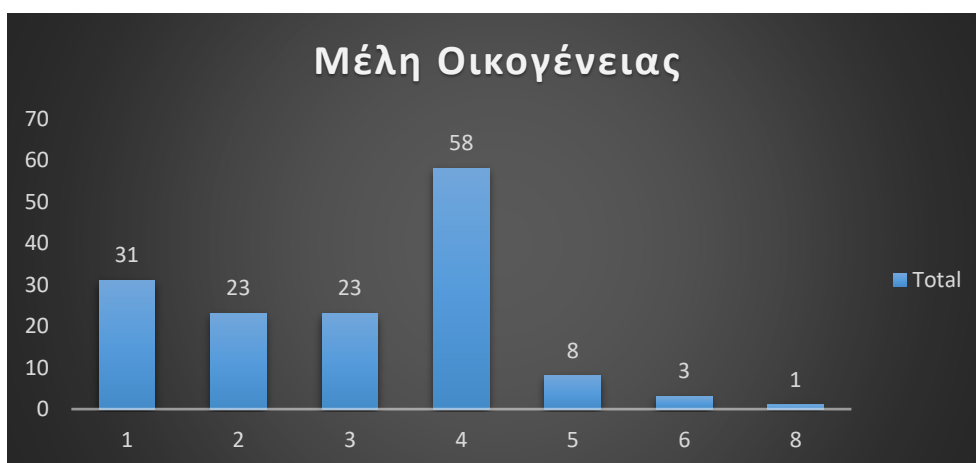


Σχήμα 5.13 Κατανομή γυναικείων προτιμήσεων για τον τύπο του γιαουρτιού



Σχήμα 5.14 Προτίμηση στην τοποθεσία αγοράς

Όπως φαίνεται και στο σχήμα 5.14, η πλειοψηφία του δείγματος φαίνεται να προτιμά το supermarket ως σημείο αγοράς των συγκεκριμένων προϊόντων, ενώ αμέσως επόμενη επιλογή, με αρκετά μεγάλη διαφορά, είναι η αγορά από τους λιανοπωλητές. Τρίτη στην σειρά είναι η αγορά του γιαουρτιού από τους χονδρέμπορους, ενώ υπάρχουν και κάποιοι χρήστες που αγοράζουν τόσο από τους λιανοπωλητές όσο και από το supermarket. Τέλος, μόλις 2 ερωτώμενοι αγοράζουν ηλεκτρονικά.



Σχήμα 5.15 Πλήθος μελών οικογένειας που καταναλώνουν γιαούρτι

Στο σχήμα 5.15 της περιγραφικής στατιστικής, παρουσιάζεται το πλήθος των μελών της κάθε οικογένειας που καταναλώνουν γιαούρτι για το σύνολο του δείγματος. Πιο αναλυτικά, οι περισσότερες οικογένειες φαίνεται να αποτελούνται από τέσσερα άτομα, ενώ στην τρίτη θέση υπάρχει ισοψηφία μεταξύ των οικογενειών που αποτελούνται από δύο και τρία άτομα. Όσον αφορά την δεύτερη θέση είναι αυτή του ενός ατόμου.

5.3. Εφαρμογή της UTASTAR

Όπως αναφέρθηκε και στο τέταρτο κεφάλαιο, έχοντας πλέον δημιουργήσει τον πολυκριτήριο πίνακα, χρησιμοποιώντας το ερωτηματολόγιο, εφαρμόστηκε ο αλγόριθμος UTASTAR στο MARKEX. Εν συνεχεία επιλέγεται το Post optimization, δηλαδή η μεταβελτιστοποίηση των αποτελεσμάτων, με σκοπό την άντληση των επιθυμητών αποτελεσμάτων. Μέσω αυτής της διαδικασίας προκύπτουν οι ολικές χρησιμότητες για τον κάθε αποφασίζοντα και τα βάρη των κριτηρίων τα οποία θα χρησιμοποιηθούν μετέπειτα. Έπειτα, γίνεται η συσταδοποίηση των καταναλωτών, όπως φαίνεται στην ενότητα 5.4, και η προσομοίωση των προϊόντων με χρήση μοντέλων προσωπικής επιλογής, όπως στην ενότητα 5.10.

Average Criteria Weights - Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.178204	0.172274	0.132876	0.136884	0.138258	0.121624	0.119880

Σχήμα 5.16 Μέσα βάρη κριτηρίων όλου του δείγματος.

Average Alternatives Utilities - Post Optimization						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPUS	RODOPI
0.806627	0.836116	0.821929	0.878040	0.788556	0.800352	0.764135

Σχήμα 5.17 Μέσες ολικές χρησιμότητες




Από την καρτέλα «Average Values», εντοπίζονται, τόσο τα μέσα βάρη των κριτηρίων, όσο και οι ολικές χρησιμότητες, τα οποία απαρουσιάζονται στην εικόνα 5.17 και 5.18 αντίστοιχα. Η πλειοψηφία του δείγματος φαίνεται πως θεωρεί πιο σημαντικό το κριτήριο της επωνυμίας με τιμή ίση με 0.178204, ενώ το δεύτερο σημαντικότερο κριτήριο, με μικρή διαφορά, είναι αυτό της γεύσης. Στην τρίτη θέση έρχεται το κριτήριο της οσμής και της υφής. Βάσει των βαρών που δόθηκαν το πλέον προτιμότερο προϊόν είναι αυτό της εταιρείας «ΦΑΓΕ», ενώ αμέσως μετά ακολουθεί αυτό της εταιρείας «ΔΩΔΩΝΗ» και της «ΔΕΛΤΑ».

5.4. Συσταδοποίηση με χρήση αλγορίθμων του MARKEX

Για την συσταδοποίηση των καταναλωτών χρησιμοποιήθηκαν τα κριτήρια με το μεγαλύτερο βάρος, καθώς και πιθανοί συνδυασμοί αυτών. Μεγαλύτερες τιμές επιλέχθηκαν για τη σημαντικότητα, ενώ μικρότερες για τη μη σημαντικότητα και την αδιαφορία. Θεωρείται λοιπόν ότι για τους καταναλωτές που παρουσιάζουν μεγαλύτερες τιμές από το εκάστοτε κατώφλι κάθε κριτηρίου είναι σημαντικό το εν λόγω κριτήριο, ενώ για τα κριτήρια στα οποία παρουσιάζονται χαμηλότερες τιμές από

το κατώφλι τους δεν είναι σημαντικά για αυτούς. Ως απόρροια αυτού, όταν το κατώφλι ενός σημαντικού κριτηρίου αυξάνεται, τόσο μειώνεται το πλήθος των καταναλωτών μίας ομάδας, ενώ για ένα κριτήριο το οποίο δεν είναι σημαντικό ή αδιάφορο παρατηρείται ότι όσο αυξάνεται η τιμή του κατωφλίου, τόσο αυξάνονται και οι καταναλωτές σε κάθε ομάδα. Στον πίνακα που ακολουθεί παρουσιάζονται 3 συστάδες για όλους τους καταναλωτές του δείγματος. Η πρώτη ομάδα που δημιουργήθηκε, προέκυψε με βάση τα δύο πιο σημαντικά κριτήρια του δείγματος. Η δεύτερη ομάδα σχηματίστηκε βάσει του πιο σημαντικού κριτηρίου και δύο λιγότερο σημαντικών κριτηρίων, αυτά της προέλευσης του γάλακτος και της συσκευασίας, ενώ η τρίτη και τελευταία ομάδα αποτελείται από το δεύτερο, το τρίτο και το τελευταίο πιο σημαντικό κριτήριο.

Πίνακας 5.1 Ομαδοποίηση καταναλωτών στο MARKEK

Cluster Name	Criteria	Analysis Method	Consumers Number	Actions
Cluster 1	Brand_name (0.178, Significant)	Average Weights	15 (10.20%)	
	Taste (0.172, Significant)			
Cluster 2	Brand_name (0.178, Significant)	Average Weights	42 (28.57%)	
	Milk_origin (0.133, Non Significant)			
	Packaging (0.122, Non Significant)			
Cluster 3	Taste (0.172, Significant)	Average Weights	28 (19.05%)	
	Odor (0.138, Non Significant)			
	Price (0.12, Non Significant)			

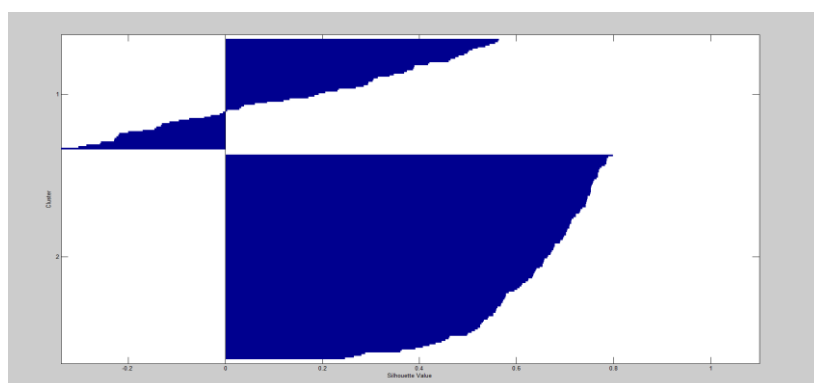
5.5. Ο αλγόριθμος K-Means

Με την ολοκλήρωση της προ - επεξεργασίας, και δεδομένου ότι δεν υπήρχαν ακραίες τιμές από τα αρχεία, εφαρμόστηκε ο αλγόριθμος K-Means τόσο στο αρχείο των ολικών χρησιμότητων, όσο και σε αυτό με τα βάρη των κριτηρίων. Τα αρχεία αυτά, καθώς και ο τρόπος που προέκυψαν, περιγράφηκαν στην ενότητα 4.4. Εκτός από την εντολή που υπάρχει στο MATLAB για την εφαρμογή του αλγορίθμου, θα έπρεπε να υπολογιστεί ο δείκτης σκιαγράφησης silhouette για όλο το πλήθος των συστάδων. Για τον λόγο αυτό, αναπτύχθηκε ένα μικρό script επτά γραμμών. Οι πρώτες γραμμές αυτού, χρησίμευαν για τον καθαρισμό του command window και την ανάγνωση των αρχείων από το excel. Στις επόμενες γραμμές του κώδικα, εφαρμόζεται ο αλγόριθμος K-Means και στην συνέχεια υπολογίζεται τόσο ο δείκτης silhouette για καθένα από τα ερωτηματολόγια, όσο και ο μέσος δείκτης silhouette για όλο το δείγμα. Ακόμα, θα πρέπει να σημειωθεί πως η διαδικασία αυτή επαναλήφθηκε για δύο, τρεις και για τέσσερις συστάδες. Σε κάθε περίπτωση, εξετάζονταν αν υπήρχαν αρνητικά silhouette και έπειτα αφαιρούνταν από το αρχικό δείγμα και επαναλαμβάνονταν η διαδικασία, μέχρις ότου πρώτον, να είναι όλοι οι δείκτες silhouette μεγαλύτεροι του μηδενός και δεύτερον, ο μέσος δείκτης silhouette να είναι όσο το δυνατόν πλησιέστερα στην μονάδα. Σε κάθε επανάληψη τα ερωτηματολόγια που αφαιρούνταν ήταν τόσο αυτά που είχαν αρνητικό silhouette στο αρχείο με τα βάρη, όσο και τα αντίστοιχα που είχαν αρνητικά silhouette στο αρχείο με τις ολικές χρησιμότητες, έτσι ώστε να μελετηθούν οι ίδιοι ερωτώμενοι.

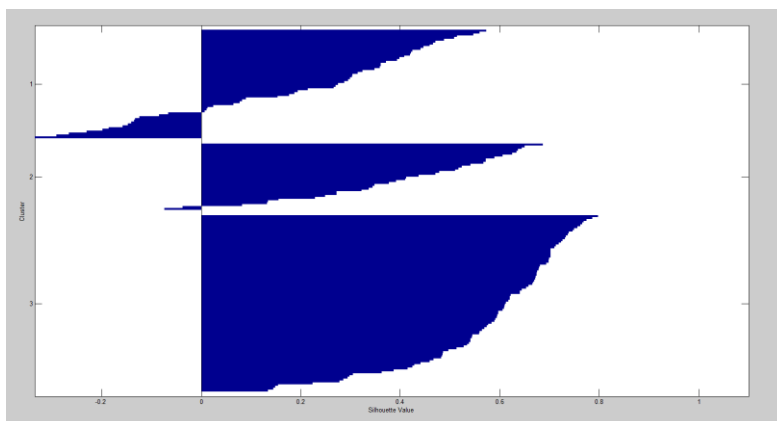
Πίνακας 5.2 Αποτελέσματα K-Means για την πρώτη επανάληψη

CLUSTERS	2	3	4
Ερωτηματολόγια που αφαιρέθηκαν	3,7,32,33,40,41,44,49,52,60,81,88,96,113,120,124,127,134,141,148,154,156,157,161,163,167,172,173,179,181,182,184,186,191	1,4,28,31,38,40,44,49,50,51,52,56,58,61,63,66,67,85,87,96,97,100,105,107,109,114,116,118,120,123,124,128,130,131,136,140,141,142,143,145,151,153,156,161,176,179,181,184,185,190,191	12,13,23,35,43,56,58,64,68,74,84,85,102,103,104,109,116,128,131,133,139,142,143,144,150,159,160,169,183,187,190
Εναπομείναντα ερωτηματολόγια	191-34 = 157	191-51 = 140	191-32 = 159
Μέσος δείκτης σκιαγράφησης/ silhouette (utilities)	0.5842	0.4441	0.4359
Μέσος δείκτης σκιαγράφησης/ silhouette (weights)	0.4507	0.4793	0.3968

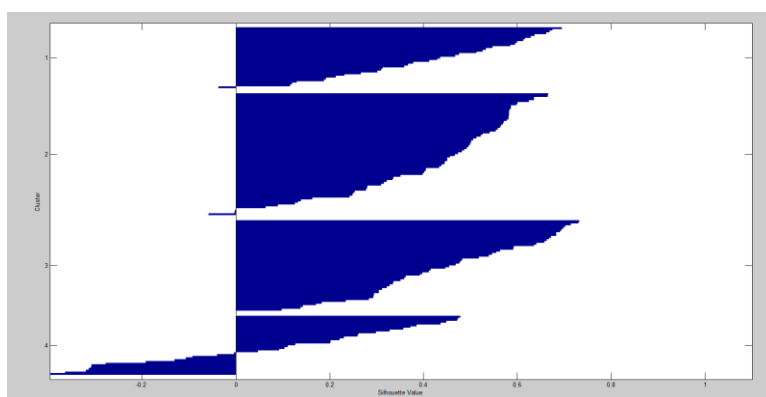
Με εφαρμογή του script για το αρχείο των ολικών χρησιμοτήτων, που περιγράφηκε στην προηγούμενη παράγραφο, εντοπίζονται τα ερωτηματολόγια που θα πρέπει να αφαιρεθούν. Με όμοιο ακριβώς τρόπο, εφαρμόζεται και για το δεύτερο αρχείο με τα βάρη των κριτηρίων και εντοπίζονται και τα υπόλοιπα προς αφαίρεση ερωτηματολόγια. Η διαδικασία αυτή εφαρμόζεται για δύο, τρεις και τέσσερις συστάδες. Τα αποτελέσματα για την πρώτη επανάληψη εμφανίζονται στον πίνακα 5.2.



Σχήμα 5.18 Δείκτης silhouette για 2 clusters (utilities)

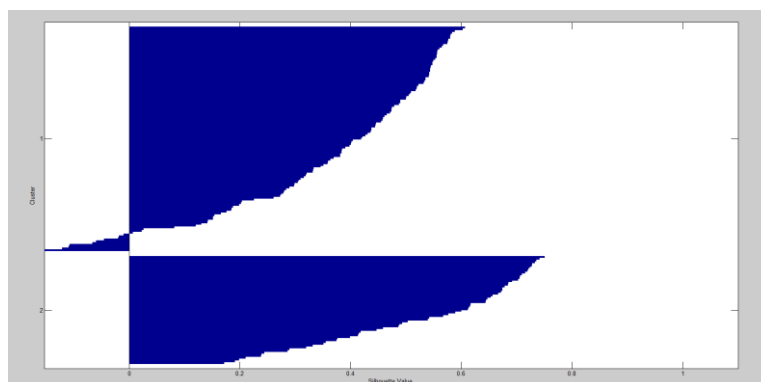


Σχήμα 5.19 Δείκτης silhouette για 3 clusters (utilities)

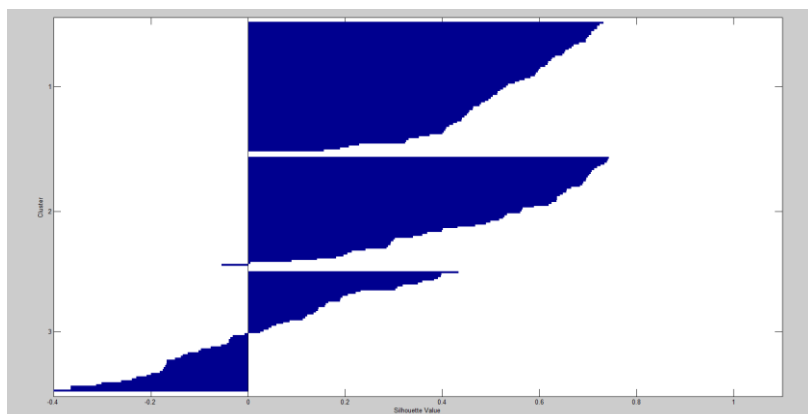


Σχήμα 5.20 Δείκτης silhouette για 4 clusters (utilities)

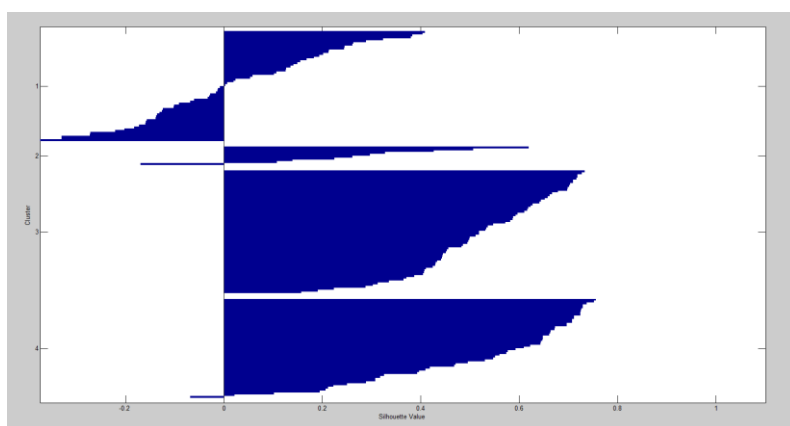
Τα τρία διαγράμματα (5.19 - 5.21) αναφέρονται στο αρχείο των ολικών χρησιμότητων και τα διαφορετικά είδη συσταδοποίησης. Η σειρά παρουσίασης τους γίνεται βάσει αριθμών συστάδων κατά αύξοντα τρόπο. Επιπλέον, οι τιμές που προεκτείνονται αριστερά του του μηδενός είναι οι τιμές με αρνητικό silhouette, και οι οποίες αφαιρέθηκαν.



Σχήμα 5.21 Δείκτης silhouette για 2 clusters (weights)



Σχήμα 5.22 Δείκτης silhouette για 3 clusters (weights)



Σχήμα 5.23 Δείκτης silhouette για 4 clusters (weights)

Τα τρία διαγράμματα (5.22-5.24) είναι τα αντίστοιχα με τα προηγούμενα, αλλά για το αρχείο των βαρών.

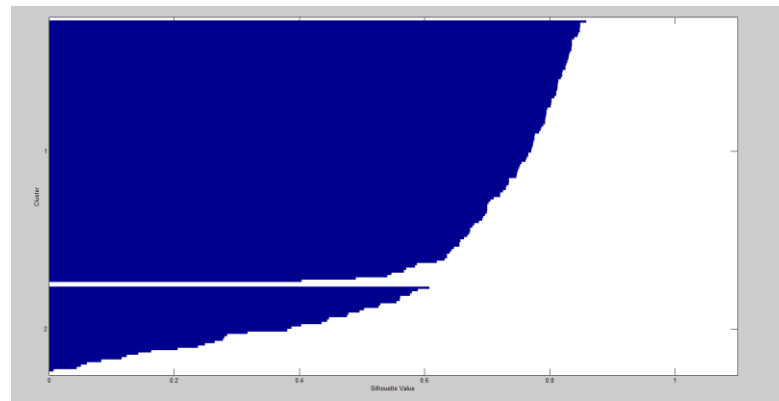
Με την ολοκλήρωση της παραπάνω διαδικασίας, ως καλύτερη επιλογή επιλέχθηκε η πρώτη, δηλαδή η δημιουργία δύο συστάδων, δεδομένου ότι ο μέσος δείκτης σκιαγράφησης ήταν ο καλύτερος δυνατός και με την μεγαλύτερη θετική τιμή, σε σχέση με τις υπόλοιπες εναλλακτικές.

Όπως ήδη αναφέρθηκε και στο κεφάλαιο 4.5, με αφαίρεση όλων των απαραίτητων ερωτηματολογίων το τελικό δείγμα φαίνεται να αποτελείται από 147 ερωτηματολόγια. Η πρώτη ομάδα που προέκυψε, για το αρχείο των ολικών χρησιμοτήτων, αποτελείται από 36 εγγραφές, ενώ οι υπόλοιπες 111 ανήκουν στο δεύτερο cluster. Από την άλλη πλευρά, όσον αφορά το αρχείο των βαρών, φαίνεται ότι η πιο πολυπληθής ομάδα είναι η πρώτη με 104 εγγραφές. Η δεύτερη ομάδα από την άλλη, περιέχει 43 ερωτηματολόγια.

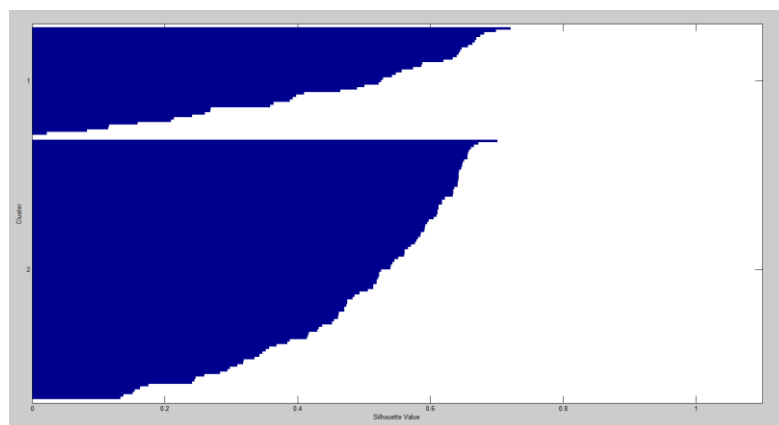
Πίνακας 5.3 Αποτελέσματα K-Means για την τελευταία επανάληψη

CLUSTERS	2	3	4
Μέσος δείκτης σκιαγράφηση - silhouette (utilities)	0.6476	0.4786	0.4581
Μέσος δείκτης σκιαγράφηση - silhouette (weights)	0.4884	0.4992	0.4754

Στην συνέχεια, στον πίνακα 5.3. εμφανίζονται τα αποτελέσματα της τελευταίας επανάληψης του αλγορίθμου K-Means. Συγκεκριμένα, με πράσινο χρώμα σημειώθηκαν τα καλύτερα αποτελέσματα. Ο μέσος δείκτης σκιαγράφησης στο αρχείο με τις ολικές χρησιμότητες ισούται με 0.6476, ενώ για το αρχείο με τα βάρη είναι 0.4884.



Σχήμα 5.24 Δείκτης silhouette για 2 clusters στο τελικό δείγμα (utilities)



Σχήμα 5.25 Δείκτης silhouette για 2 clusters στο τελικό δείγμα (weights)

Εν κατακλείδι, τα αντίστοιχα διαγράμματα με όλα τα θετικά silhouette εμφανίζεται παρακάτω (σχήμα 5.25-5.26).

5.6. Ο αλγόριθμος DBScan

Όσον αφορά τον αλγόριθμο DBScan, αυτός εφαρμόστηκε με τη βοήθεια του λογισμικού της WEKA. Πιο αναλυτικά, αφού πρώτα μετατράπηκαν σε csv μορφή το αρχείο των βαρών και των ολικών χρησιμοτήτων, μέσω του excel, έγινε η ανάγνωση των δεδομένων και στην συνέχεια από την καρτέλα “cluster”, επιλέχθηκε ο αλγόριθμος και με τις κατάλληλες αλλαγές στις παραμέτρους του αλγορίθμου, βρέθηκε η καλύτερη εναλλακτική. Οι παράμετροι που τροποποιήθηκαν ήταν τόσο αυτός του ελάχιστου αριθμού στοιχείων που ανήκουν στην κάθε συστάδα (MinPts), όσο και η απόσταση των στοιχείων από τον πυρήνα της κάθε συστάδας (Eps).

Έχοντας ως στόχο την μείωση του θορύβου, αρχικά επιλέγεται κατάλληλα ο παράγοντας Eps, και στην συνέχεια με τις απαραίτητες αυξομειώσεις του MinPts, βρίσκεται η καλύτερη συσταδοποίηση. Στην καλύτερη συσταδοποίηση, για το αρχείο των ολικών χρησιμοτήτων, η τιμή του Eps ισούται με 2.6, ενώ η τιμή του MinPts παίρνει την τιμή 2. Με τον τρόπο αυτό δημιουργούνται τέσσερις συστάδες και 50 από τα 147 στοιχεία δεν συσταδοποιούνται. Από τις τέσσερις αυτές συστάδες, η πλειοψηφία του δείγματος ανήκουν στην πρώτη συστάδα με 90 εγγραφές, ενώ τα υπόλοιπα 6 στοιχεία μοιράζονται στις υπόλοιπες συστάδες. Επιπλέον, θα πρέπει να σημειωθεί πως σχεδόν ίδια αποτελέσματα προκύπτουν και για την δημιουργία δύο συστάδων. Στο σενάριο αυτό, η βασική διαφορά είναι ότι υπάρχουν τέσσερα παραπάνω στοιχεία που δεν έχουν ομαδοποιηθεί. Τα στοιχεία αυτά αφαιρέθηκαν από τις τρεις συστάδες με τα λιγότερα στοιχεία της προηγούμενης περίπτωσης. Ακόμα, θα πρέπει να σημειωθεί πως δεν προέκυψε κάποια εναλλακτική για την δημιουργία τριών συστάδων. Τέλος, στον πίνακα που ακολουθεί, απεικονίζονται συνοπτικά όλες οι εναλλακτικές που προαναφέρθηκαν.

Πίνακας 5.4 Συνοπτική απεικόνιση του DBScan στο αρχείο των ολικών χρησιμοτήτων

Σενάριο	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Unclustered
Δημιουργία 2 συστάδων	90 στοιχεία	3 στοιχεία	-	-	54 στοιχεία
Δημιουργία 3 συστάδων	-	-	-	-	-
Δημιουργία 4 συστάδων	90 στοιχεία	2 στοιχεία	3 στοιχεία	2 στοιχεία	50 στοιχεία

Ακολουθώντας την ίδια διαδικασία για το αρχείο των βαρών, παρατηρείται πως η καλύτερη συσταδοποίηση εμφανίζεται με την δημιουργία τριών συστάδων. Στην περίπτωση αυτή, η παράμετρος Eps ισούται με 2.4, ενώ η παράμετρος MinPts παίρνει την τιμή 3. Όσον αφορά τώρα το πλήθος των στοιχείων ανά συστάδα, η πρώτη ομάδα αποτελείται από 109 εγγραφές, ενώ η δεύτερη και η τρίτη από 3 και 4 αντίστοιχα. Τέλος, τα στοιχεία που δεν συσταδοποιήθηκαν είναι 31 στον αριθμό. Ενδιαφέρον έχει η περίπτωση όπου δημιουργούνται 2 συστάδες, με το Eps να παίρνει την τιμή 2.5 και το MinPts να είναι ίσο με 70. Τα αποτελέσματα που εξάγονται είναι παρόμοια με αυτά

που προέκυψαν στο αρχείο των ολικών χρησιμότητων. Ειδικότερα, εκτός από την παράμετρο MinPts που πήρε την τιμή 70, το Eps ισούται με 2.5 και τα μη συσταδοποιημένα στοιχεία είναι 52 στον αριθμό. Η πρώτη ομάδα αποτελείται από 25 στοιχεία και η δεύτερη από 70. Στο σημείο αυτό θα πρέπει να σημειωθεί πως σε καμία περίπτωση δεν προέκυψε σενάριο με την δημιουργία τεσσάρων ομάδων. Όλα όσα προαναφέρθηκαν, συνοψίζονται στον πίνακα 5.5 που ακολουθεί.

Πίνακας 5.5 Συνοπτική απεικόνιση του DBScan στο αρχείο των βαρών

Σενάριο	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Unclustered
Δημιουργία 2 συστάδων	25 στοιχεία	70 στοιχεία	-	-	52 στοιχεία
Δημιουργία 3 συστάδων	109 στοιχεία	3 στοιχεία	4 στοιχεία	-	31 στοιχεία
Δημιουργία 4 συστάδων	-	-	-	-	-

5.7. Ο αλγόριθμος EM (Expectation – Maximization)

Έχοντας ολοκληρώσει με τον αλγόριθμο DBScan, επαναλαμβάνεται η διαδικασία για τα δύο αρχεία, επιλέγοντας ωστόσο διαφορετικό αλγόριθμο από την καρτέλα “Cluster”, τον EM. Από τις παραμέτρους του αλγορίθμου επιλέγεται το πλήθος των συστάδων που θα προκύψουν.

Η πρώτη εφαρμογή του αλγορίθμου θα γίνει για το αρχείο των ολικών χρησιμότητων και για δύο συστάδες. Στην περίπτωση αυτή, τα 108 από τα 147 στοιχεία ανήκουν στην πρώτη ομάδα, ενώ τα υπόλοιπα 39 στοιχεία στην δεύτερη ομάδα. Η πιθανότητα μία εναλλακτική να ανήκει στην πρώτη ομάδα ισούται με 73%, και 27% για να ανήκει στην δεύτερη. Επιπλέον, θα πρέπει να τονισθεί πως το καλύτερο δυνατό αποτέλεσμα προκύπτει με την δημιουργία δύο συστάδων, καθώς και να σημειωθεί ο δείκτης λογαριθμικής πιθανοφάνειας.

$$\log - \text{likelihood} = -31.52001 \quad (5.1)$$

Επιλέγοντας την δημιουργία τριών συστάδων, η πλειοψηφία του δείγματος (132 εγγραφές) ανήκουν στην πρώτη ομάδα, μία μόλις ανήκει στη δεύτερη ομάδα και 14 στην τρίτη ομάδα. Η πιθανότητα ένα στοιχείο να ανήκει στην πρώτη ομάδα ισούται με 90%, ενώ για την δεύτερη και τρίτη ομάδα 1% και 10% αντίστοιχα. Τέλος, ο δείκτης λογαριθμικής πιθανοφάνειας φαίνεται παρουσιάζεται παρακάτω:

$$\log - \text{likelihood} = -32.05107 \quad (5.2)$$

Στο σενάριο όπου δημιουργούνται τέσσερις συστάδες, η πλειοψηφία του δείγματος (131 εγγραφές) ανήκει στην πρώτη ομάδα, η δεύτερη και η τέταρτη ομάδα

αποτελούνται μόλις από μία εγγραφή η καθεμία, ενώ οι υπόλοιπες 14 εγγραφές ανήκουν στην τρίτη ομάδα. Η πιθανότητα ένα στοιχείο να ανήκει στην πρώτη ομάδα ισούται με 89%, για την δεύτερη και την τέταρτη ομάδα η πιθανότητα ισούται με 1%, ενώ για την τρίτη ομάδα 10%. Όμοια με προηγουμένως, ο δείκτης λογαριθμικής πιθανοφάνειας παρουσιάζεται ακριβώς από κάτω:

$$\log - likelihood = -32.0315 \quad (5.3)$$

Τα αποτελέσματα των συσταδοποιήσεων που προαναφέρθηκαν, παρουσιάζονται συνοπτικά στον πίνακα 5.6.

Πίνακας 5.6 Συνοπτική απεικόνιση του EM στο αρχείο των ολικών χρησιμοτήτων

Σενάριο	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Log - Likelihood
Δημιουργία 2 συστάδων	108 στοιχεία	39 στοιχεία	-	-	-31.52001
Δημιουργία 3 συστάδων	132 στοιχεία	1 στοιχείο	14 στοιχεία	-	-32.05107
Δημιουργία 4 συστάδων	131 στοιχεία	1 στοιχείο	14 στοιχεία	1 στοιχείο	-32.03015

Κάνοντας ανάγνωση το αρχείο των βαρών, αυτή την φορά, επαναλαμβάνεται η διαδικασία για δύο, τρεις και τέσσερις συστάδες αντίστοιχα. Στην περίπτωση που δημιουργούνται δύο συστάδες οι 65 από τις 147 εγγραφές τοποθετούνται στην πρώτη ομάδα, ενώ οι υπόλοιπες 82 στην δεύτερη ομάδα. Έτσι, η πιθανότητα εμφάνισης μίας εναλλακτικής στην πρώτη ομάδα ισούται με 43% και η αντίστοιχη πιθανότητα για την εμφάνιση στην δεύτερη ομάδα ισούται με 57%. Ο δείκτης αξιοπιστίας ισούται με -24.81428. Με την δημιουργία τριών ομάδων ο δείκτης αξιοπιστίας του μοντέλου φαίνεται πως χειροτερεύει με τιμή που ισούται με -25.11695. Η πρώτη ομάδα αποτελείται από 92 στοιχεία, η δεύτερη ομάδα από 32 στοιχεία και η τελευταία ομάδα από 23 στοιχεία. Με άλλα λόγια η πιθανότητα εμφάνισης μίας εναλλακτικής στην πρώτη ομάδα ισούται με 63%, για την εύρεση στην δεύτερη ομάδα η πιθανότητα είναι ίση με 22%, ενώ η εύρεσή της στην τρίτη ομάδα ισούται με 16%. Στην περίπτωση όπου δημιουργούνται τέσσερις συστάδες για το δείγμα της έρευνας, τα αποτελέσματα φαίνονται να είναι τα καλύτερα δυνατά, λαμβάνοντας υπόψιν τον δείκτη λογαριθμικής πιθανοφάνειας.

$$\log - likelihood = -24.55641 \quad (5.4)$$

Η πρώτη ομάδα αποτελείται από 53 εγγραφές, ενώ η δεύτερη ομάδα από 50 εγγραφές. Η τρίτη ομάδα αποτελείται από 17 στοιχεία και η τελευταία συστάδα από

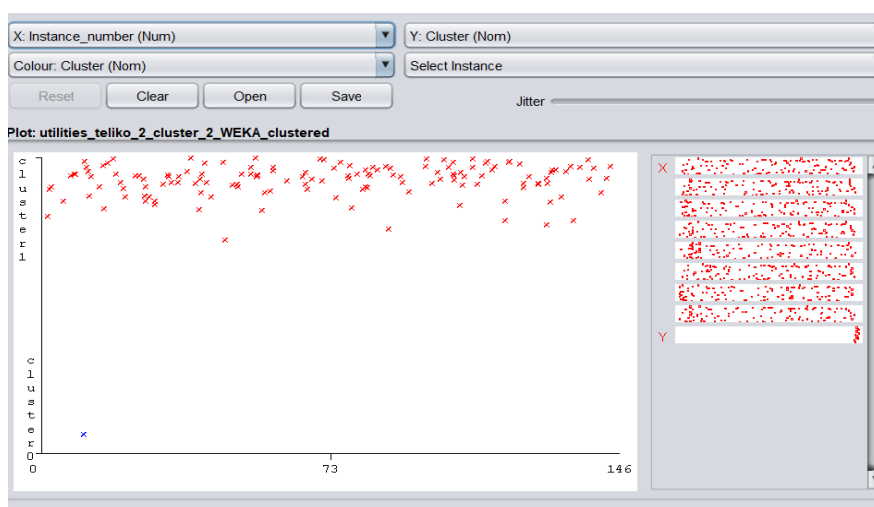
27. Συνεπώς, η πιθανότητα μία εναλλακτική να ανήκει στην σε καθεμία από τις ομάδες αντίστοιχα είναι ίσες με 36%, 34%, 12% και 18%. Στον πίνακα 5.7 που ακολουθεί συνοψίζονται όλα τα παραπάνω σενάρια.

Πίνακας 5.7 Συνοπτική απεικόνιση του EM στο αρχείο των βαρών

Σενάριο	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Log - Likelihood
Δημιουργία 2 συστάδων	65 στοιχεία	82 στοιχεία	-	-	-24.81428
Δημιουργία 3 συστάδων	92 στοιχεία	32 στοιχείο	23 στοιχεία	-	-25.11695
Δημιουργία 4 συστάδων	53 στοιχεία	50 στοιχείο	17 στοιχεία	27 στοιχείο	-24.55641

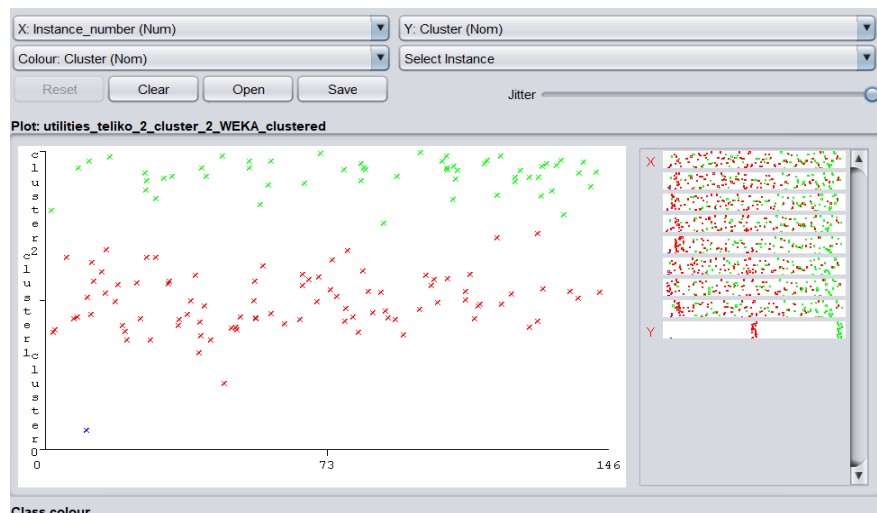
5.8. Ο αλγόριθμος Hierarchical Clusterer

Έχοντας ολοκληρώσει με τους άλλους δύο αλγορίθμους συσταδοποίησης, εφαρμόζεται και ο Hierarchical Clusterer πρώτα για το αρχείο των ολικών χρησιμότητων και στην συνέχεια για το αρχείο των βαρών. Από τις παραμέτρους του αλγορίθμου επιλέγεται το πλήθος των συστάδων που θα δημιουργηθούν, και ως αποτελέσματα προκύπτουν τόσο τα στοιχεία της κάθε ομάδας, όσο και η γραφική αναπαράσταση της κάθε συστάδας. Επιπλέον, επιλέγεται η ευκλείδεια απόσταση για τον υπολογισμό της απόστασης των σημείων, καθώς και η συσταδοποίηση βάσει του κοντινότερου γείτονα ή απλού συνδέσμου (single link). Πιο αναλυτικά, στην περίπτωση όπου δημιουργούνται δύο συστάδες όλες οι εγγραφές εκτός από μία ανήκουν στην πρώτη ομάδα. Συνεπώς, η πιθανότητα ένα στοιχείο να ανήκει στην πρώτη ομάδα ισούται με 99%.



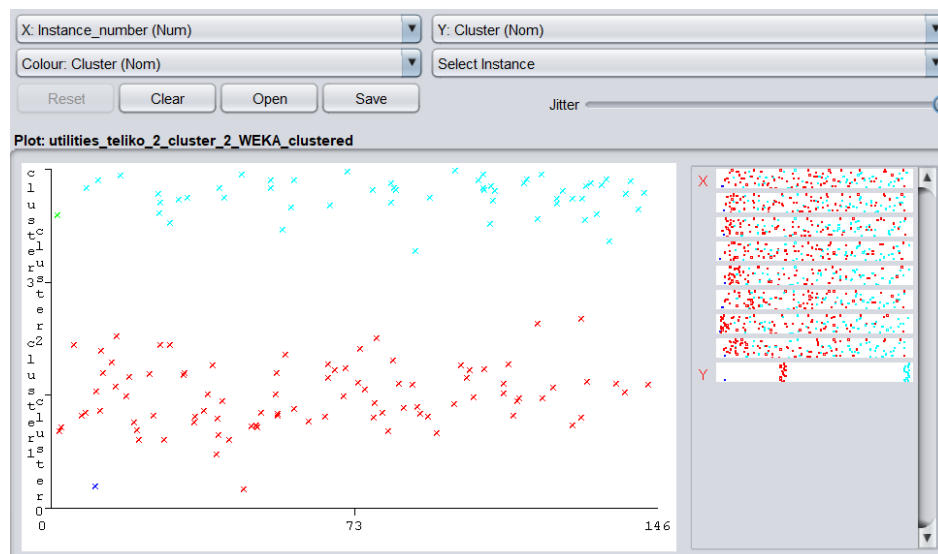
Σχήμα 5.26 Γραφική απεικόνιση των δύο συστάδων στο αρχείο των ολικών χρησιμότητων για τον αλγόριθμο Hierarchical Clusterer

Η γραφική απεικόνιση παρουσιάζεται στο σχήμα 5.27, στο οποίο με κόκκινο χρώμα απεικονίζονται τα στοιχεία της πρώτης ομάδας, ενώ με μπλε χρώμα το μοναδικό στοιχείο της δεύτερης ομάδας. Στην περίπτωση όπου δημιουργούνται τρεις ομάδες για το συνολικό δείγμα, η πρώτη ομάδα φαίνεται να απαρτίζεται μόλις από ένα στοιχείο, ενώ η δεύτερη και η τρίτη από 90 και 56 αντίστοιχα.



Σχήμα 5.27 Γραφική απεικόνιση των τριών συστάδων στο αρχείο των ολικών χρησιμοτήτων για τον αλγόριθμο Hierarchical Clusterer

Η γραφική απεικόνιση των τριών συστάδων παρουσιάζεται στο σχήμα 5.28. Με πράσινο χρώμα σημειώνονται τα στοιχεία της τρίτης ομάδας, με κόκκινο χρώμα τα στοιχεία της δεύτερης ομάδας και με μπλε χρώμα το στοιχείο της πρώτης ομάδας.



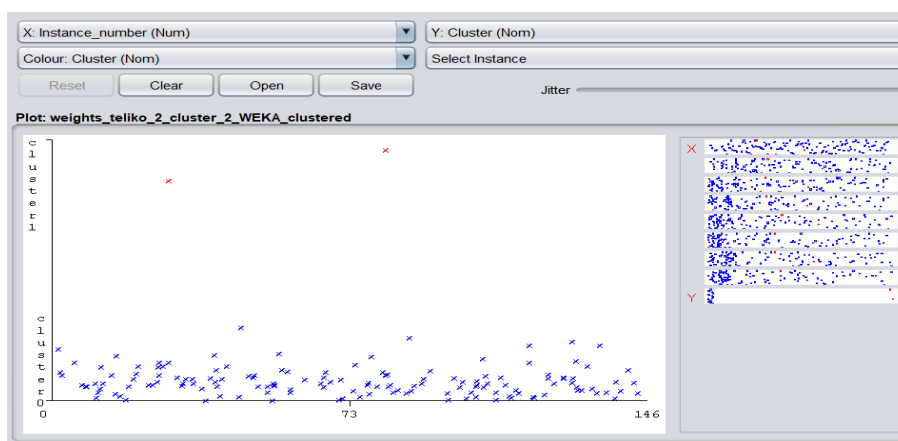
Σχήμα 5.28 Γραφική απεικόνιση των τεσσάρων συστάδων στο αρχείο των ολικών χρησιμοτήτων για τον αλγόριθμο Hierarchical Clusterer

Στην περίπτωση όπου δημιουργούνται τέσσερις συστάδες, τόσο η πρώτη ομάδα, όσο και η τρίτη αποτελούνται μόλις από ένα στοιχείο. Η δεύτερη ομάδα αποτελείται από 90 εγγραφές και η τελευταία ομάδα από 55 εγγραφές. Η γραφική απεικόνιση αυτών διατυπώνεται στο σχήμα 5.29. Με μπλε και πράσινο χρώμα αποτυπώνονται τα στοιχεία της πρώτης και της τρίτης ομάδας. Με κόκκινο χρώμα σημειώνονται τα στοιχεία της δεύτερης, ενώ με γαλάζιο τα στοιχεία της τέταρτης ομάδας. Όλα όσα προαναφέρθηκαν αποτυπώνονται συνοπτικά στον πίνακα 5.8 που ακολουθεί.

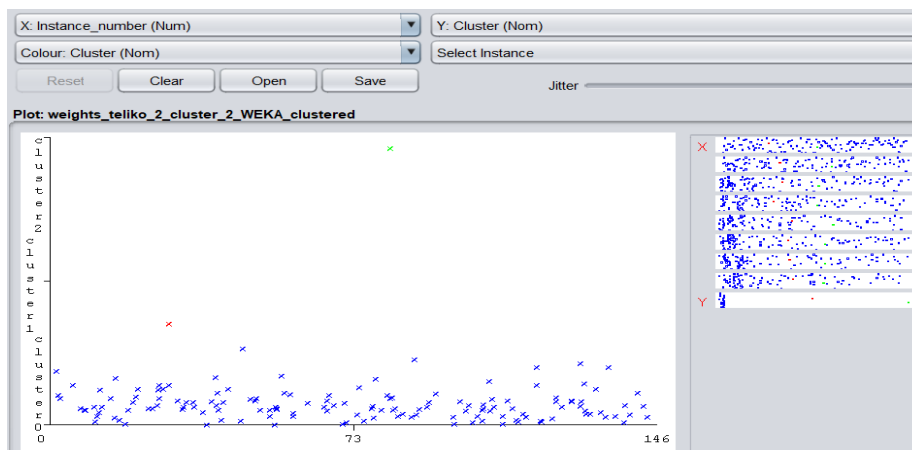
Πίνακας 5.8 Συνοπτική απεικόνιση του Hierarchical Clusterer στο αρχείο των ολικών χρησιμότητων

Σενάριο	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Δημιουργία 2 συστάδων	146 στοιχεία	1 στοιχείο	-	-
Δημιουργία 3 συστάδων	1 στοιχείο	90 στοιχεία	56 στοιχεία	-
Δημιουργία 4 συστάδων	1 στοιχείο	90 στοιχεία	1 στοιχείο	55 στοιχεία

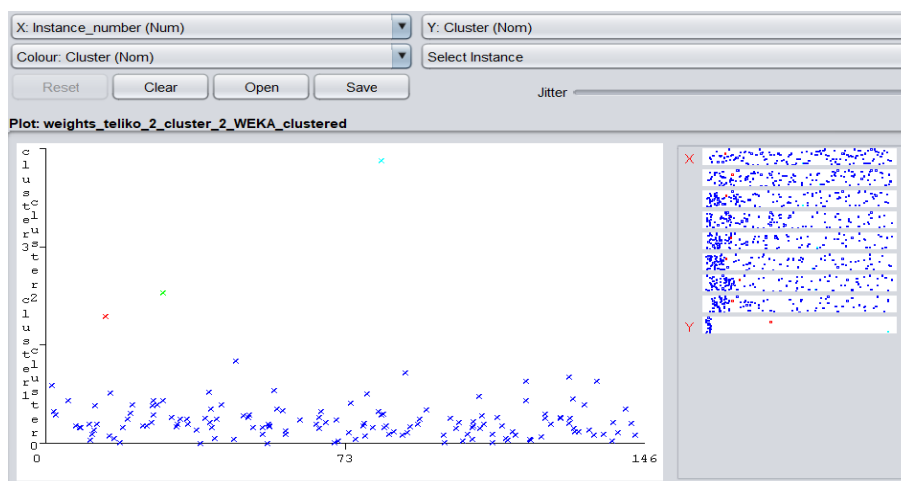
Μελετώντας το αρχείο των βαρών, για την δημιουργία δύο συστάδων παρατηρούμε ότι όλα σχεδόν τα στοιχεία εντοπίζονται στην πρώτη συστάδα, ενώ μόλις δύο στοιχεία από τα 147, που είναι στο σύνολο τους, κατατάσσονται στην δεύτερη συστάδα. Αξίζει να σημειωθεί πως όσο αυξάνεται το πλήθος των ομάδων, παρατηρείται ότι όλες οι υπόλοιπες ομάδες, εκτός της πρώτης, αποτελούνται μόλις από ένα στοιχείο. Η συμπεριφορά αυτή εντοπίζεται τόσο στις τρεις όσο και στις τέσσερις συστάδες.



Σχήμα 5.29 Γραφική απεικόνιση των δύο συστάδων στο αρχείο των βαρών για τον αλγόριθμο Hierarchical Clusterer



Σχήμα 5.30 Γραφική απεικόνιση των τριών συστάδων στο αρχείο των βαρών για τον αλγόριθμο Hierarchical Clusterer



Σχήμα 5.31 Γραφική απεικόνιση των τεσσάρων συστάδων στο αρχείο των βαρών για τον αλγόριθμο Hierarchical Clusterer

Στα σχήματα 5.30-5.32 παρουσιάζονται οι γραφικές απεικονίσεις των συστάδων. Με μπλε χρώμα απεικονίζονται τα στοιχεία της πρώτης ομάδας, ενώ με κόκκινο τα στοιχεία της δεύτερης. Τα στοιχεία της τρίτης ομάδας σημειώνονται με πράσινο χρώμα και τα στοιχεία της τελευταίας ομάδας με γαλάζιο. Στον πίνακα 5.9 που ακολουθεί, συνοψίζονται τα αποτελέσματα της μεθόδου, όπως και στις προηγούμενες περιπτώσεις.

Πίνακας 5.9 Συνοπτική απεικόνιση του Hierarchical Clusterer στο αρχείο των βαρών

Σενάριο	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Δημιουργία 2 συστάδων	145 στοιχεία	2 στοιχεία	-	-
Δημιουργία 3 συστάδων	145 στοιχεία	1 στοιχείο	1 στοιχείο	-
Δημιουργία 4 συστάδων	144 στοιχεία	1 στοιχείο	1 στοιχείο	1 στοιχείο

5.9. Η μέθοδος Canopy

Έχοντας ολοκληρώσει τους προηγούμενους αλγορίθμους συσταδοποίησης, εφαρμόζεται και ο αλγόριθμος Canopy πρώτα για το αρχείο των ολικών χρησιμότητων και στην συνέχεια για το αρχείο των βαρών. Από τις παραμέτρους του αλγορίθμου επιλέγεται το πλήθος των συστάδων που θα δημιουργηθούν σε κάθε περίπτωση, όπως επίσης και οι τιμές των κατωφλίων. Επιπλέον, επιλέγεται η ελάχιστη τιμή της πυκνότητας η οποία καθορίζεται από τον παράγοντα T_2 . Για τιμές των παραμέτρων $T_1 = 1.25$, $T_2 = 1.0$, minimum canopy density=1.0 και την δημιουργία 2 συστάδων, παρατηρείται ότι όλα τα δεδομένα τοποθετούνται στην πρώτη συστάδα εκτός ενός στοιχείου που τοποθετείται στην δεύτερη. Όμοια, για την δημιουργία 3 συστάδων η πρώτη ομάδα αποτελείται από 145 εγγραφές, ενώ η δεύτερη και τρίτη συστάδα μόλις με μία εγγραφή η καθεμία. Σε αντίθεση με τις προηγούμενες δύο περιπτώσεις, όταν δημιουργούνται τέσσερις συστάδες φαίνεται πως η νέα ομάδα που προκύπτει αποτελείται από 27 εγγραφές, ενώ η πρώτη ομάδα με 118. Όσον αφορά τις άλλες δύο ομάδες, εξακολουθούν να παρουσιάζουν ακριβώς την ίδια συμπεριφορά με αρχικά. Οι παραπάνω περιπτώσεις συνοψίζονται στον πίνακα 5.10 που ακολουθεί.

Πίνακας 5.10 Συνοπτική απεικόνιση των συστάδων στο αρχείο των ολικών χρησιμότητων

Σενάριο	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Δημιουργία 2 συστάδων	146 στοιχεία	1 στοιχείο	-	-
Δημιουργία 3 συστάδων	145 στοιχεία	1 στοιχείο	1 στοιχείο	-
Δημιουργία 4 συστάδων	118 στοιχεία	1 στοιχείο	1 στοιχείο	18 στοιχεία

Επαναλαμβάνοντας ακριβώς την ίδια διαδικασία για το αρχείο των βαρών, παρατηρείται μία πιο ομοιόμορφη κατανομή των στοιχείων στις ομάδες. Ειδικότερα, για τις ίδιες τιμές των παραμέτρων ($T_1 = 1.25$, $T_2 = 1.0$, minimum canopy density=1.0) και την δημιουργία 2 συστάδων, η πρώτη ομάδα που προέκυψε αποτελείται από 120 στοιχεία, ενώ η δεύτερη από 27. Για την δημιουργία 3 συστάδων, η πρώτη ομάδα

αποτελείται με 118 στοιχεία, η δεύτερη με 27 στοιχεία και η τρίτη ομάδα με 2 μόλις στοιχεία. Στην τελευταία περίπτωση όπου δημιουργούνται τέσσερις συστάδες, η πρώτη ομάδα είναι η πολυπληθέστερη με 81 στοιχεία να την απαρτίζουν, αμέσως μετά ακολουθεί η τέταρτη ομάδα με 41 εγγραφές, αμέσως μετά ακολουθεί η δεύτερη ομάδα με 23 στοιχεία και τελευταία η τρίτη ομάδα με μόλις 2 εγγραφές όπως ακριβώς στην προηγούμενη περίπτωση.

Πίνακας 5.11 Συνοπτική απεικόνιση των συστάδων στο αρχείο των βαρών

Σενάριο	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Δημιουργία 2 συστάδων	120 στοιχεία	27 στοιχεία	-	-
Δημιουργία 3 συστάδων	118 στοιχεία	27 στοιχεία	2 στοιχεία	-
Δημιουργία 4 συστάδων	81 στοιχεία	23 στοιχεία	2 στοιχεία	41 στοιχεία

5.10. Η μέθοδος UTASTAR

Επόμενο βήμα της μεθοδολογίας είναι η εφαρμογή της UTASTAR στο MARKEK. Πιο αναλυτικά, έχοντας καταλήξει στην περίπτωση με το καλύτερο μέσο silhouette, αυτόματα έχει γίνει γνωστό το πλήθος των ομάδων, καθώς και ποιοι ερωτώμενοι ανήκουν σε ποια ομάδα. Βάσει αυτού του διαχωρισμού δημιουργήθηκαν δύο νέα αρχεία για την είσοδο στο MARKEK, βάσει των ολικών χρησιμότητων, όπως ακριβώς και στην ενότητα 4.3. Το πλήθος των αρχείων που απαιτούνται σε κάθε περίπτωση δίνεται από το ακόλουθο τύπο:

$$\text{Πλήθος Αρχείων Εισόδου} = X + 1 \quad (5.1)$$

όπου,

- Χ: το πλήθος των ομάδων που προέκυψαν από την συσταδοποίηση
- Το +1 αναφέρεται στο αρχείο όλου του τελικού δείγματος.

Με την δημιουργία των νέων αυτών αρχείων, είναι πλέον εφικτή η υλοποίηση της πολυκριτήριας μελέτης. Στην συνέχεια, παρουσιάζεται αναλυτικά η διαδικασία που θα ακολουθηθεί για το πλήθος όλων των αρχείων που προέκυψαν. Πρώτα απ' όλα, εισάγεται το αρχείο και γίνεται ανάγνωση όλων των δεδομένων. Έπειτα, από την επιλογή "RUN UTASTAR", εκτελείται ο αλγόριθμος και εφαρμόζεται η μέθοδος για το συγκεκριμένο σετ δεδομένων.

Από την καρτέλα "Consumers Analysis", και έχοντας επιλέξει την επίλυση του δεύτερου γραμμικού προβλήματος, επιλέγεται το "Average Values" όπου εμφανίζονται οι τιμές των αποτελεσμάτων της μεθόδου. Τα αποτελέσματα αυτά, είναι οι ολικές χρησιμότητες και τα βάρη των κριτηρίων. Όσο μεγαλύτερη είναι η τιμή τους, τόσο πιο σημαντικά είναι τα χαρακτηριστικά για τους καταναλωτές ή τόσο πιο εύκολα επιλέγεται το

συγκεκριμένο προϊόν. Επιπροσθέτως, τα αναλυτικά αποτελέσματα των βαρών (Consumers Analysis→ Criteria Weights) και των ολικών χρησιμοτήτων (Consumers Analysis→ Global Utilities) για κάθε ερωτώμενο αποθηκεύτηκαν σε ξεχωριστά αρχεία excel.

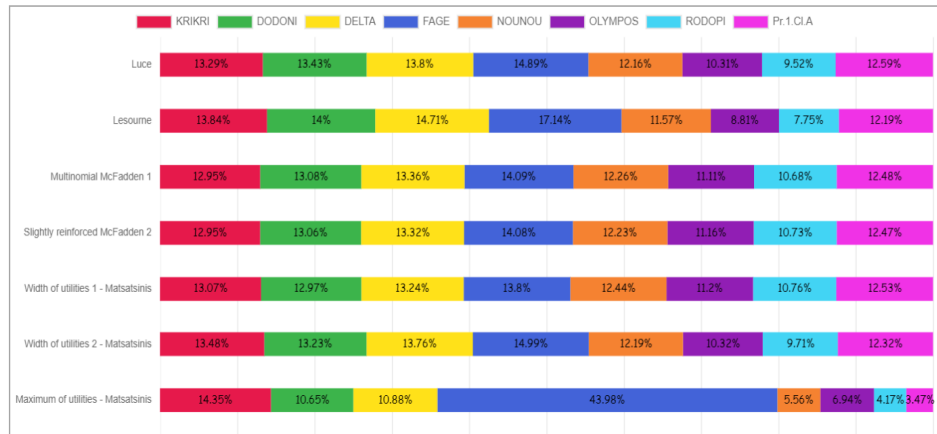
Παράλληλα, σημαντικό μέρος της συγκεκριμένης μεθοδολογίας ανάπτυξης τμηματοποίησης της αγοράς είναι η δημιουργία ενός προϊόντος, το οποίο θα ανταγωνίζεται τα υπόλοιπα παρόμοια προϊόντα της αγοράς. Για την επίτευξη αυτού, θα πρέπει αρχικά να γίνει ανάλυση συμπεριφοράς των καταναλωτών, και στην συνέχεια να αναπτυχθούν τα απαραίτητα τρία σενάρια για την υπάρχουσα αγορά, έτσι ώστε να βρεθούν τα τελικά αποτελέσματα.

Αρχικά, αναπτύσσεται ένα απλό σενάριο για το κάθε νέο ανταγωνιστικό προϊόν.

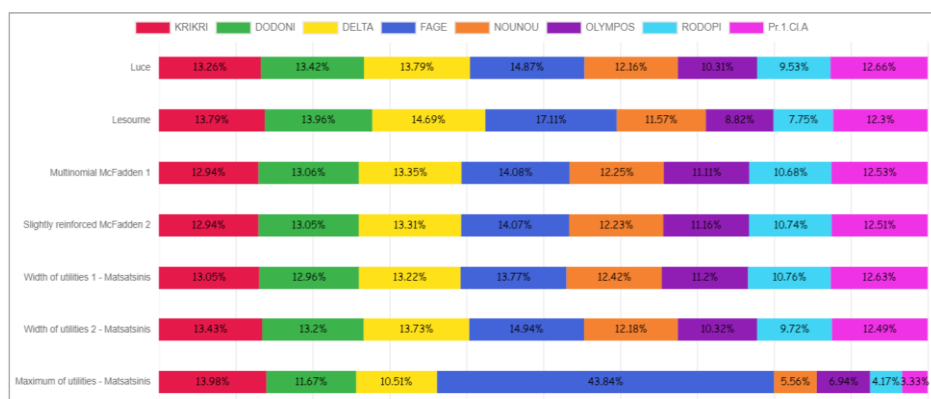
- **Δημιουργία νέας συστάδας:** Όσον αφορά την διαδικασία που ακολουθήθηκε στο MARKEK, αρχικά από την καρτέλα “Criteria Analysis” δημιουργήθηκε ένα νέο cluster, βάσει των κριτηρίων, που έχουν την μεγαλύτερη βαρύτητα για την μελέτη του προϊόντος, και της ομάδας των καταναλωτών που μελετάται.
- **Δημιουργία νέου απλού σεναρίου:** Ύστερα, από την καρτέλα “Market Simulation”, επιλέγεται η δημιουργία ενός απλού σεναρίου και η ταυτόχρονη απεικόνιση όλων των μοντέλων συμπεριφοράς.
- **Δημιουργία νέου προϊόντος:** Από την ενότητα “Actions”, δημιουργείται το νέο προϊόν που θέλουμε να εισάγουμε στην αγορά,
- **Γραφική απεικόνιση των αποτελεσμάτων:** Με την βοήθεια των προτιμησιακών μοντέλων εντοπίζεται, τόσο ποιο προϊόν προτιμάει η συγκεκριμένη ομάδα, όσο και πως επηρεάζονται τα μερίδια αγοράς με την προσθήκη του νέου ανταγωνιστικού προϊόντος.

Έχοντας πλέον ολοκληρώσει την περιγραφική διαδικασία, που αναπτύχθηκε προηγουμένως, στην συνέχεια δίνεται ένα παράδειγμα για την καλύτερη κατανόηση του αναγνώστη. Αρχικά, τα προϊόντα που αναπτύσσονται, θα πρέπει να σημειωθεί πως δεν είναι τυχαία, αλλά βασίζονται στα βάρη των κριτηρίων με τις μέγιστες τιμές. Με τον τρόπο αυτό, επιλέγονται τα κριτήρια που επηρεάζουν περισσότερο τις απόψεις των καταναλωτών και δημιουργούνται πιθανοί συνδυασμοί αυτών. Μέσω του βήματος αυτού, εισάγεται το νέο ανταγωνιστικό προϊόν και μελετάται η διαμόρφωση των μεριδίων αγοράς. Έστω ότι το κριτήριο με το μεγαλύτερο βάρος είναι η τιμή, και δεδομένου ότι η μονοτονία του κριτηρίου είναι φθίνουσα, ξεκινάει η μείωση της τιμής του προϊόντος. Αυτό έχει ως αποτέλεσμα, όσο μειώνεται η χρηματική αξία του προϊόντος, τόσο περισσότερο να αυξάνεται το μερίδιο του νέου προϊόντος στην αγορά. Ωστόσο, κάτι τέτοιο δεν είναι επιθυμητό από την πλευρά των εταιρειών, καθώς δεν θα μπορούσε να αποτελεί βιώσιμη τακτική. Για αυτόν ακριβώς τον λόγο, σχεδιάστηκαν διάφοροι συνδυασμοί μεταξύ των κριτηρίων για να προκύψει ένα επιθυμητό και εφικτό αποτέλεσμα. Τέλος, η διαδικασία αυτή θα υλοποιηθεί τρεις φορές, καθώς θα εξεταστεί η εισαγωγή τριών, διαφορετικών, νέων προϊόντων στην αγορά. Ωστόσο, η

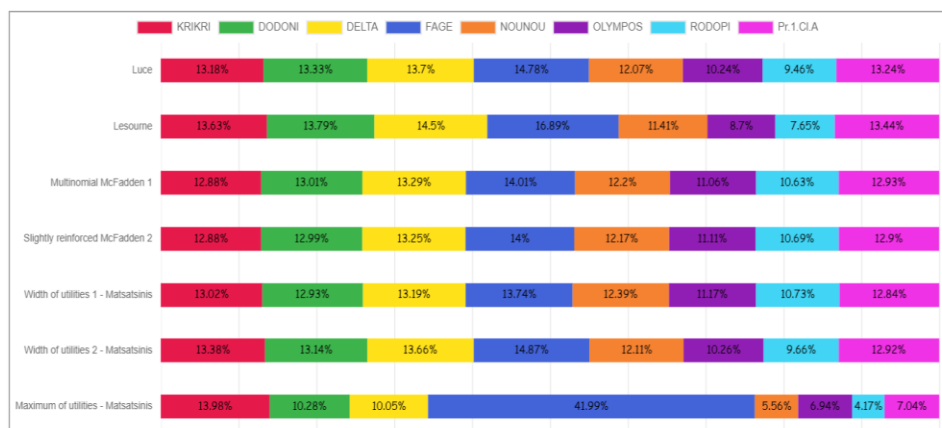
εισαγωγή των τριών αυτών προϊόντων δεν θα είναι ταυτόχρονη. Στις επόμενες τρεις εικόνες που ακολουθούν παρουσιάζεται η εισαγωγή των νέων προϊόντων.



Σχήμα 5.32 Εισαγωγή πρώτου ανταγωνιστικού προϊόντος



Σχήμα 5.33 Εισαγωγή δεύτερου ανταγωνιστικού προϊόντος



Σχήμα 5.34 Εισαγωγή τρίτου ανταγωνιστικού προϊόντος

Εφαρμόζοντας όλα όσα αναφέρθηκαν σε αυτήν την ενότητα για όλα τα διαφορετικά σεντ δεδομένων που προέκυψαν από την συσταδοποίηση στη MATLAB, θα γίνει γνωστό ποια κριτήρια είναι πιο σημαντικά για τους καταναλωτές. Μέσω των κριτηρίων, θα γίνει γνωστό ποιο προϊόν προτιμά να καταναλώνει η κάθε ομάδα. Εκτός αυτού όμως, παρατηρώντας τις αλλαγές στα κριτήρια των νέων προϊόντων, που εισάγονται στην αγορά, θα γίνει πιο κατανοητός ο τρόπος σκέψης των καταναλωτών. Οι αναλύσεις αυτές εντοπίζονται στις αμέσως επόμενες σελίδες (ενότητα 5.10.1 - 5.10.11) και μελετώνται ξεχωριστά. Τέλος, αξίζει να σημειωθεί πως το μοντέλο που ανταποκρίνεται περισσότερο στην πραγματικότητα είναι αυτό του Lesourne. Ειδικότερα, τόσο το προϊόν της εταιρείας «ΦΑΓΕ», όσο και αυτό της «ΚΡΙ – ΚΡΙ», της «ΔΕΛΤΑ» και της «ΟΛΥΜΠΟΣ» έρχονται πολύ κοντά στα πραγματικά μερίδια αγοράς.

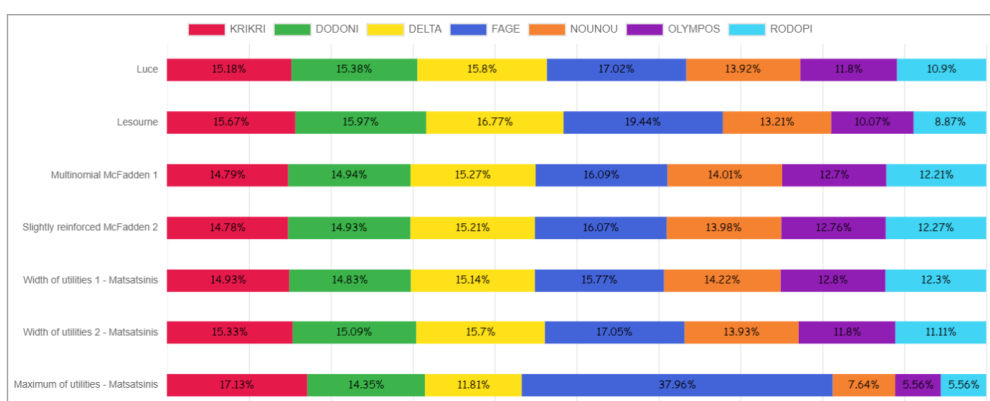
5.10.1. Μελέτη πρώτης ομάδας (Συσταδοποίηση με τον K-Means)

Η πρώτη περίπτωση με την οποία θα ασχοληθούμε, είναι η πρώτη ομάδα που προέκυψε από τον αλγόριθμο K-Means στη MATLAB, η οποία περιέχει τους καταναλωτές, όπως αυτοί προέκυψαν από το αρχείο των ολικών χρησιμοτήτων. Αυτό το σεντ δεδομένων αποτελείται από 36 εγγραφές και μόλις 7 εγγραφές έχουν δείκτη συσχέτισης κάτω από 0.9.

Average Criteria Weights - Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.171615	0.187692	0.127684	0.137898	0.136412	0.113997	0.124702
Average Alternatives Utilities - Post Optimization						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPUS	RODOPI
0.705409	0.719095	0.741513	0.790914	0.654026	0.550846	0.509384

Σχήμα 5.35 Μέσος όρος βαρών και ολικών χρησιμοτήτων πρώτου cluster

Όπως φαίνεται και στο σχήμα 5.36 το πιο σημαντικό κριτήριο για τους καταναλωτές αυτής της ομάδας είναι η γεύση. Αμέσως μετά, ως πιο σημαντικό κριτήριο εμφανίζεται η επωνυμία της εταιρείας και η υφή του γιαουρτιού. Αντίθετα, το λιγότερο σημαντικό κριτήριο είναι η συσκευασία. Επιπλέον, το προϊόν που αγοράζουν πιο εύκολα οι συγκεκριμένοι καταναλωτές είναι από τη εταιρεία «ΦΑΓΕ» και ύστερα από την «ΔΕΛΤΑ». Προσομοιώνοντας την παρούσα κατάσταση για όλες τις εναλλακτικές και με παράλληλη χρήση των μοντέλων προσωπικής επιλογής δίνονται το ακόλουθο διάγραμμα με τα αντίστοιχα μερίδια.



Σχήμα 5.36 Παρουσίαση μεριδίων για όλα τα μοντέλα για το Cluster 1.

Όπως γίνεται αντιληπτό, το προϊόν της εταιρείας «ΦΑΓΕ», κατέχει το μεγαλύτερο μερίδιο σε όλες τις περιπτώσεις. Αμέσως μετά ακολουθεί η εταιρεία «Δέλτα», με εξαίρεση στο μοντέλο Maximum of utilities, όπου εμφανίζεται τέταρτο. Τρίτο έρχεται το προϊόν της εταιρείας «Δωδώνη» για τα τέσσερα πρώτα μοντέλα, ενώ το προϊόν της «Κρι-Κρι» έρχεται αμέσως μετά, με εξαίρεση το τελευταίο μοντέλο (Maximum of utilities), όπου βρίσκεται στην δεύτερη θέση.

Όσον αφορά τώρα το πρώτο απλό σενάριο που δημιουργήθηκε, το νέο προϊόν (Pr.1.Cl.A) που εισάγεται στην αγορά έχει τα εξής χαρακτηριστικά για την αύξουσα, πενταβάθμια κλίμακα των κριτηρίων. Η κλίμακα διαστήματος αυτή αποτυπώνεται στην συνέχεια στον πίνακα 5.12.

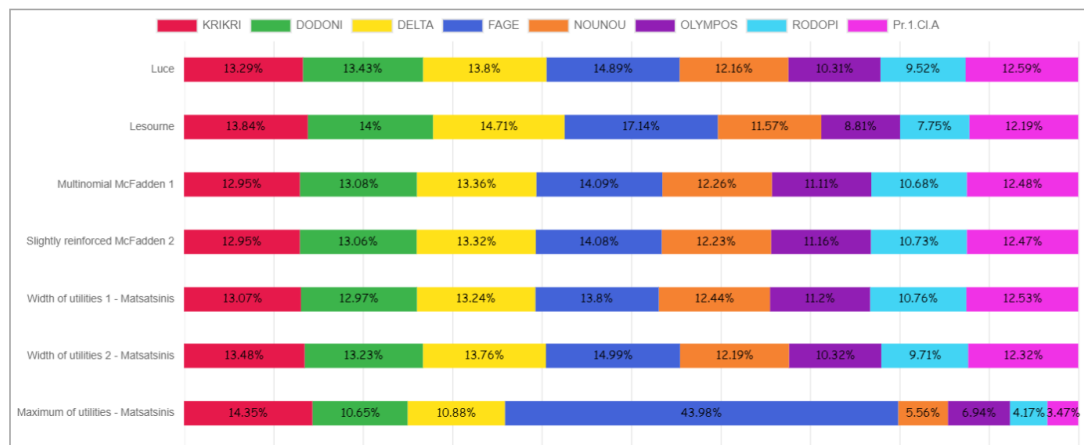
Πίνακας 5.12 Επεξήγηση Κλίμακας Αξιολόγησης Ερωτηματολογίου

Κλίμακα διαστήματος	Πολύ κακή	Κακή	Αδιάφορη ή Λογική	Καλή	Πολύ καλή
Συμβολισμός	1/5	2/5	3/5	4/5	5/5

- a) Επωνυμία: Αδιάφορη (3/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Αναλυτικότερα, έχοντας ως μέγιστη και ταυτόχρονα καλύτερη, δυνατή τιμή, την τιμή 5, το νέο προϊόν (Pr.1.Cl.A) βαθμολογήθηκε όπως φαίνεται παραπάνω. Επιπλέον, θα πρέπει να σημειωθεί πως η τιμή της γεύσης αυξήθηκε κατά μια μονάδα, δεδομένου ότι είναι το κριτήριο με την μεγαλύτερη σημασία για τους καταναλωτές. Όπως φαίνεται και στο παρακάτω διάγραμμα, το νέο προϊόν παρουσιάζεται ως αρκετά ανταγωνιστικό με τα υπόλοιπα. Παρότι δεν βρίσκεται στις πρώτες θέσεις, έχει μικρές αποκλίσεις με τα υπόλοιπα προϊόντα. Έτσι, όλα τα ποσοστά για το συγκεκριμένο προϊόν βρίσκονται

κοντά στο 12% και την πέμπτη θέση, με εξαίρεση το τελευταίο μοντέλο όπου έχει την μικρότερη διείσδυση στην αγορά, μόλις 3.47%.

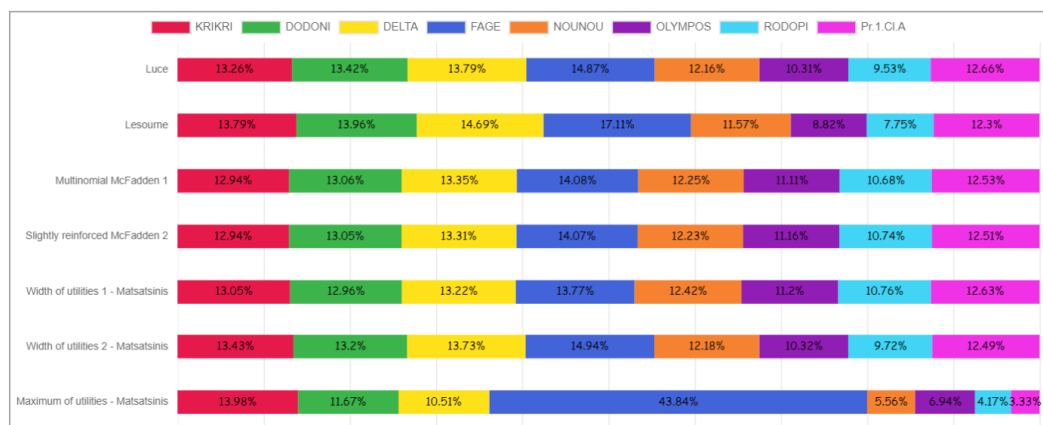


Σχήμα 5.37 Παρουσίαση μεριδίων μετά την εισαγωγή του Pr.1.CI.A για το Cluster A

Η διαδικασία αυτή, θα επαναληφθεί και για την εισαγωγή ενός δεύτερου νέου προϊόντος (Pr.2.CI.A). Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα διαστήματος, όπως αυτή διατυπώθηκε στον πίνακα 5.12, εμφανίζονται στην συνέχεια.

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Αδιάφορη (3/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Όπως στην προηγούμενη περίπτωση, έτσι και σε αυτήν, με μία παραπάνω μονάδα βαθμολογήθηκε, το δεύτερο σημαντικότερο κριτήριο για την συγκεκριμένη ομάδα, που είναι η επωνυμία της εταιρείας. Επίσης, η επωνυμία της εταιρείας αποτελεί και το κριτήριο με την μέγιστη επιρροή για την δεύτερη ομάδα, πιο αναλυτικά ωστόσο θα παρουσιαστεί στην ενότητα 5.10.2. Συγκρίνοντας την δεύτερη αυτή εναλλακτική με την προηγούμενη παρατηρούμε ότι φέρει ελαφρώς μικρότερες τιμές στα ποσοστιαία μερίδια, το οποίο είναι απόλυτα φυσιολογικό αφού έχει επηρεαστεί ένα κριτήριο με μικρότερο βάρος.

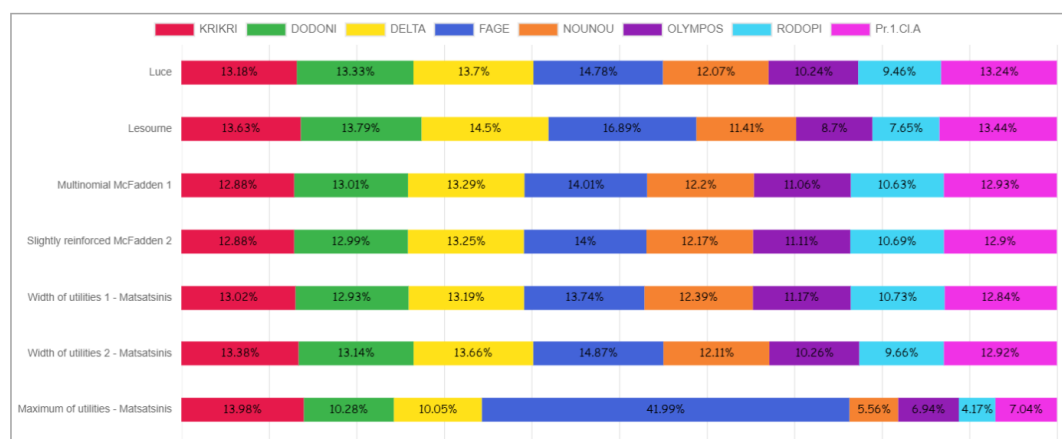


Σχήμα 5.38 Παρουσία μεριδίων μετά την εισαγωγή του Pr.2.CI.A για το Cluster A

Τελευταία περίπτωση για την μελέτη της πρώτης ομάδας, όπως αυτή προέκυψε από την συσταδοποίηση στο MATLAB, είναι η εισαγωγή ενός τρίτου προϊόντος (Pr.3.CI.A), το οποίο έχει τα ακόλουθα χαρακτηριστικά:

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Το τρίτο αυτό σενάριο φέρνει τα καλύτερα δυνατά αποτελέσματα στα ποσοστιαία μερίδια. Το νέο αυτό προϊόν κατέχει στα περισσότερα μοντέλα την τέταρτη θέση και απέχει λιγότερα από 1% από τα προϊόντα που βρίσκονται στην δεύτερη και τρίτη.



Σχήμα 5.39 Παρουσία μεριδίων μετά την εισαγωγή του Pr.3.CI.A για το Cluster A

5.10.2. Μελέτη δεύτερης ομάδας (Συσταδοποίηση με τον K-Means)

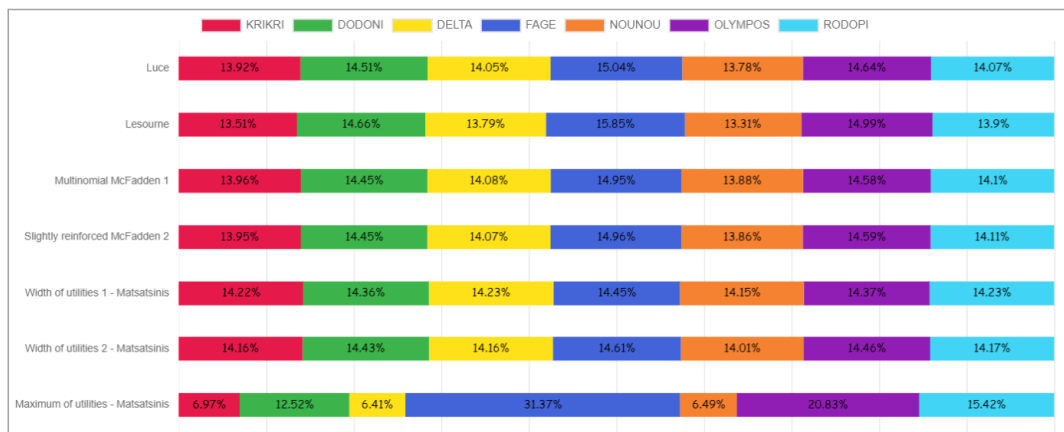
Το δεύτερο σετ δεδομένων αποτελείται από 111 εγγραφές και σαράντα οκτώ εγγραφές έχουν δείκτη συσχέτισης κάτω από 0.9. Όπως φαίνεται και στην εικόνα που ακολουθεί το πιο σημαντικό κριτήριο για τους καταναλωτές αυτής της ομάδας είναι το brand name. Αμέσως μετά, ως πιο σημαντικό κριτήριο εμφανίζεται η γεύση του προϊόντος και μετά το άρωμα του γιαουρτιού. Αντίθετα, το λιγότερο σημαντικό κριτήριο είναι η τιμή. Επιπλέον, το προϊόν που αγοράζουν πιο εύκολα οι συγκεκριμένοι καταναλωτές είναι από τη εταιρεία «ΦΑΓΕ» και ύστερα από την «ΟΛΥΜΠΟΣ» και την «ΔΩΔΩΝΗ».

Average Criteria Weights - Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.180341	0.167274	0.134560	0.136556	0.138857	0.124097	0.118315

Average Alternatives Utilities - Post Optimization						
KRIKRI	DODONI	DELTA	FAΓΕ	NOUNOU	OLYMPUS	RODOPI
0.839455	0.874069	0.848010	0.906297	0.832188	0.881272	0.846757

Σχήμα 5.40 Μέσος όρος βαρών και ολικών χρησιμότητων δεύτερου cluster

Επαναλαμβάνοντας την διαδικασία όπως έγινε στην ενότητα 5.10.1, και προσομοιώνοντας την παρούσα κατάσταση για όλες τις εναλλακτικές, κάνοντας παράλληλη χρήση των μοντέλων προσωπικής επιλογής δίνονται το ακόλουθο διάγραμμα με τα αντίστοιχα μερίδια. Πιο συγκεκριμένα, παρατηρούμε ότι και σε αυτή την περίπτωση το προϊόν της «ΦΑΓΕ» κατέχει την πρώτη θέση σε όλα τα μοντέλα. Στο σημείο αυτό, θα πρέπει να σημειωθεί πως το προϊόν της «ΦΑΓΕ» βρίσκεται πρώτο σε προτίμηση, καθώς οι καταναλωτές έχουν θεωρήσει πως φέρει αρκετά καλά χαρακτηριστικά και εμπιστεύονται το ίδιο το brand name (κριτήριο με την μεγαλύτερη βαρύτητα), σε αντίθεση με τους καταναλωτές της πρώτης συστάδας, οι οποίοι θεωρούσαν πιο σημαντικό το κριτήριο της γεύσης. Στην δεύτερη θέση για όλα τα υπόλοιπα μοντέλα βρίσκεται η εταιρεία «ΟΛΥΜΠΟΣ» με αποκλίσεις από 0.08% έως και κάτι παραπάνω από 10%, το οποίο εμφανίστηκε στο τελευταίο μοντέλο, αυτό του maximum utilities. Έπειτα ακολουθεί η εταιρεία «ΔΩΔΩΝΗ» και μετά η εταιρεία «ΡΟΔΟΠΗ». Ύστερα, έρχεται η «ΔΕΛΤΑ» με σχεδόν μηδενικές αποκλίσεις από την προπορευόμενη εταιρεία. Στην προτελευταία θέση βρίσκεται η «Κρι-Κρι» και στην τελευταία η «Νουνού».

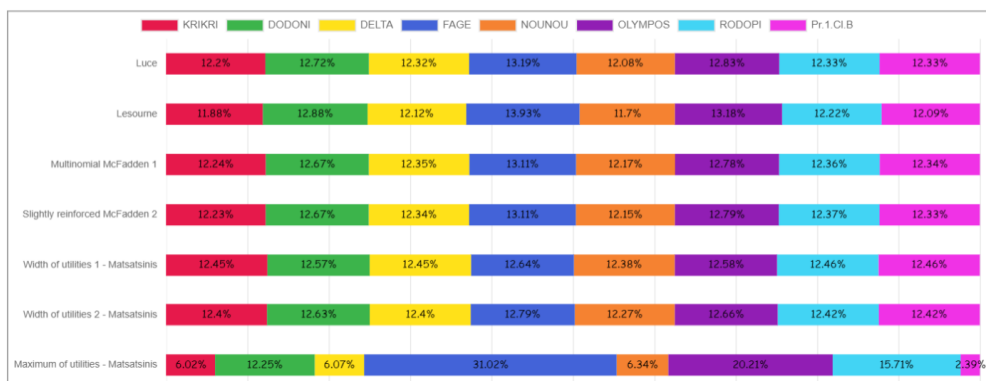


Σχήμα 5.41 Παρουσίαση μεριδίων για όλα τα μοντέλα για το Cluster 2.

Όσον αφορά τώρα το πρώτο απλό σενάριο που δημιουργήθηκε, το νέο προϊόν (Pr.1.Cl.B) που εισάγεται στην αγορά έχει τα εξής χαρακτηριστικά στην αύξουσα, πενταβάθμια κλίμακα των κριτηρίων, όπως αυτή διατυπώθηκε στη συνέχεια:

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Αδιάφορη (3/5)
- c) Προέλευση: Καλή (4/5)
- d) Υφή: Καλή (4/5)
- e) Άρωμα: Αδιάφορο (3/5)
- f) Συσκευασία: Αδιάφορο (3/5)
- g) Τιμή: Λογική (3/5)

Αναλυτικότερα, το νέο προϊόν που εισήχθη στην αγορά φαίνεται αρκετά ανταγωνιστικό παρότι βρίσκεται στην τέταρτη θέση, καθώς απέχει, ποσοστιαία, ελάχιστα από τα προπορευόμενα προϊόντα. Στην μόνη περίπτωση που υστερεί το νέο προϊόν είναι στο μοντέλο maximum of utilities, όπου κατέλαβε την τελευταία θέση.

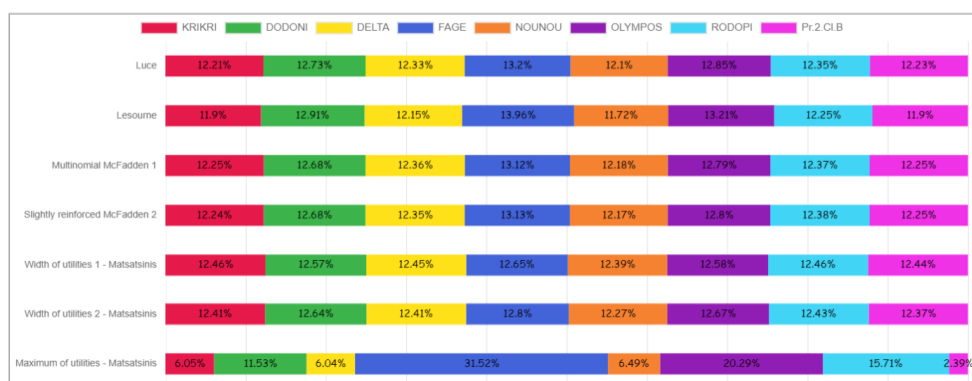


Σχήμα 5.42 Παρουσία μεριδίων μετά την εισαγωγή του Pr.1.Cl.B για το Cluster B

Η διαδικασία αυτή, θα επαναληφθεί και για την εισαγωγή ενός δεύτερου νέου προϊόντος (Pr.2.Cl.A). Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα, είναι:

- a) Επωνυμία: Αδιάφορη (3/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Καλή (4/5)
- d) Υφή: Καλή (4/5)
- e) Άρωμα: Αδιάφορο (3/5)
- f) Συσκευασία: Αδιάφορη (3/5)
- g) Τιμή: Λογική (3/5)

Σε αυτό το νέο προϊόν, αυξήθηκε κατά μία μονάδα το δεύτερο πιο σημαντικό κριτήριο για τους καταναλωτές της δεύτερης ομάδας, το κριτήριο της γεύσης. Τα ποσοστά στα μερίδια είναι σε όλες τις περιπτώσεις ελαφρώς μειωμένα για το νέο προϊόν, το οποίο είναι και λογικό, καθώς επηρεάστηκε ένα κριτήριο με μικρότερο βάρος σε σχέση με το πρώτο προϊόν που εισήχθη στην αγορά.

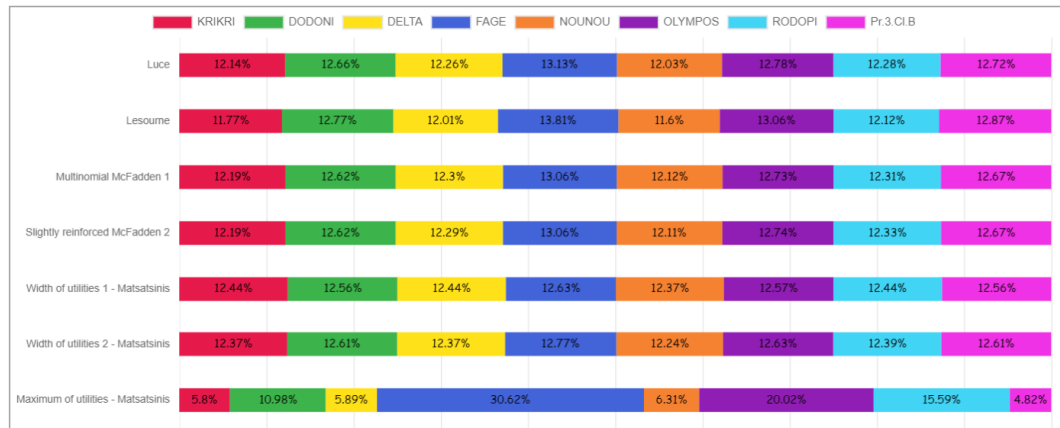


Σχήμα 5.43 Παρουσία μεριδίων μετά την εισαγωγή του Pr.2.Cl.B για το Cluster B

Τελευταία περίπτωση για την μελέτη της πρώτης ομάδας, όπως αυτή προέκυψε από την συσταδοποίηση στο MATLAB, είναι η εισαγωγή ενός τρίτου προϊόντος. Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα διαστήματος, όπως αυτή διατυπώθηκε, εμφανίζονται στην συνέχεια.

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Καλή (4/5)
- d) Υφή: Καλή (4/5)
- e) Άρωμα: Αδιάφορο (3/5)
- f) Συσκευασία: Αδιάφορη (3/5)
- g) Τιμή: Λογική (3/5)

Το τρίτο προϊόν που δοκιμάστηκε στην συγκεκριμένη μελέτη έφερε και τα καλύτερα αποτελέσματα σε σχέση με τα δύο προηγούμενα προϊόντα. Ειδικότερα, φαίνεται πως πλέον ανήκει στην τρίτη καλύτερη θέση συγκρίνοντας τα ποσοστιαία μερίδια.



Σχήμα 5.44 Παρουσία μεριδίων μετά την εισαγωγή του Pr.3.Cl.B για το Cluster B

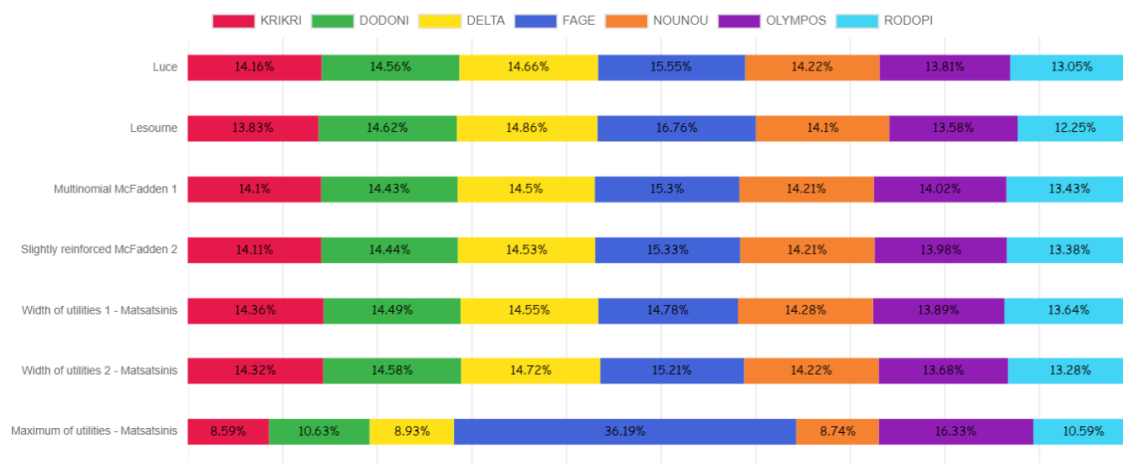
5.10.3. Μελέτη πρώτης ομάδας (Συσταδοποίηση με τον DBScan)

Με όμοια ακριβώς λογική θα επαναληφθεί η παραπάνω διαδικασία και για τον αλγόριθμο DBScan. Σύμφωνα με την ενότητα 5.6, η καλύτερη συσταδοποίηση προέκυψε με την δημιουργία τεσσάρων ομάδων.

Average Criteria Weights - Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.185278	0.166069	0.130205	0.136989	0.138400	0.123606	0.119452
Average Alternatives Utilities - Post Optimization						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPOS	RODOPI
0.842102	0.865117	0.869161	0.920973	0.847818	0.831379	0.787581

Σχήμα 5.45 Μέσος όρος βαρών και ολικών χρησιμοτήτων πρώτου cluster

Το πρώτο σεντ δεδομένων αποτελείται από 90 εγγραφές και οι μισές από αυτές έχουν δείκτη συσχέτισης πάνω από 0.9. Όπως φαίνεται και στο σχήμα 5.46 το πιο σημαντικό κριτήριο για τους καταναλωτές αυτής της ομάδας είναι η επωνυμία. Αμέσως μετά, ως πιο σημαντικό κριτήριο εμφανίζεται η γεύση και το άρωμα του γιαουρτιού. Αντίθετα, το λιγότερο σημαντικό κριτήριο είναι η τιμή. Επιπλέον, το προϊόν που αγοράζουν πιο εύκολα οι συγκεκριμένοι καταναλωτές είναι από τη εταιρεία “ΦΑΓΕ” και ύστερα από την “ΔΕΛΤΑ”, όπως ακριβώς και στην πρώτη ομάδα του που προέκυψε με τον αλγόριθμο K-Means. Προσομοιώνοντας την παρούσα κατάσταση για όλες τις εναλλακτικές και με παράλληλη χρήση των μοντέλων προσωπικής επιλογής δίνονται το ακόλουθο διάγραμμα με τα αντίστοιχα μερίδια.



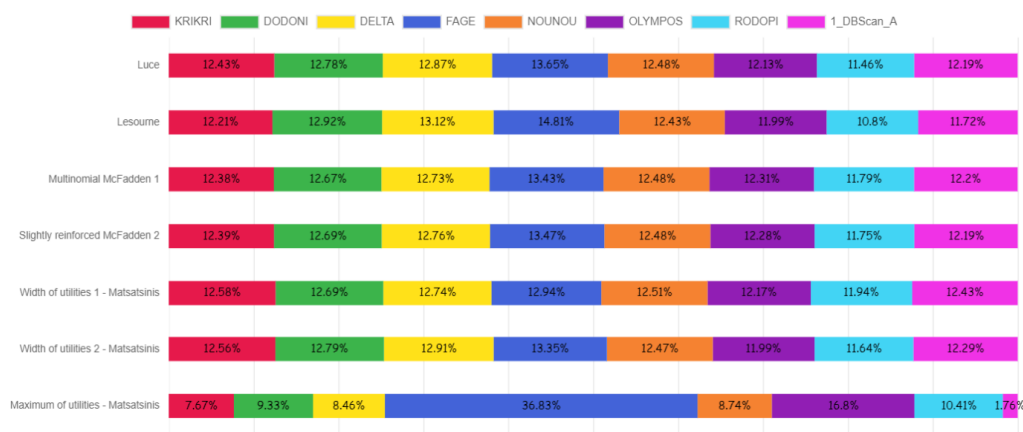
Σχήμα 5.46 Παρουσίαση μεριδίων για όλα τα μοντέλα για το Cluster 1

Όπως γίνεται αντιληπτό, το προϊόν της εταιρείας «ΦΑΓΕ», κατέχει το μεγαλύτερο μερίδιο σε όλες τις περιπτώσεις. Αμέσως μετά ακολουθεί η εταιρεία «Δέλτα», με εξαίρεση στο μοντέλο Maximum of utilities, όπου εμφανίζεται πέμπτο. Αξίζει να σημειωθεί πως η συμπεριφορά και η κατάταξη αυτή είναι παρόμοια σε μεγάλο βαθμό με το δεύτερο σετ δεδομένων, όπως αυτό προέκυψε από την συσταδοποίηση από τον αλγόριθμο K-Means.

Όσον αφορά τώρα το πρώτο απλό σενάριο που δημιουργήθηκε, το νέο προϊόν (1_DBScan_A) που εισάγεται στην αγορά έχει τα εξής χαρακτηριστικά για την αύξουσα, πενταβάθμια κλίμακα των κριτηρίων που χρησιμοποιήθηκε προηγουμένως.

- a) Επωνυμία: Αδιάφορη (3/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Αναλυτικότερα, το νέο προϊόν που εισήχθη στην αγορά φαίνεται να παρουσιάζει αρκετά παρόμοια συμπεριφορά με την δεύτερη ομάδα, όπως αυτή προέκυψε από τον αλγόριθμο K-Means. Η βασικότερη διαφορά είναι ότι το προϊόν αυτό δεν βρίσκεται στην τέταρτη θέση πλέον, αλλά στην πέμπτη.

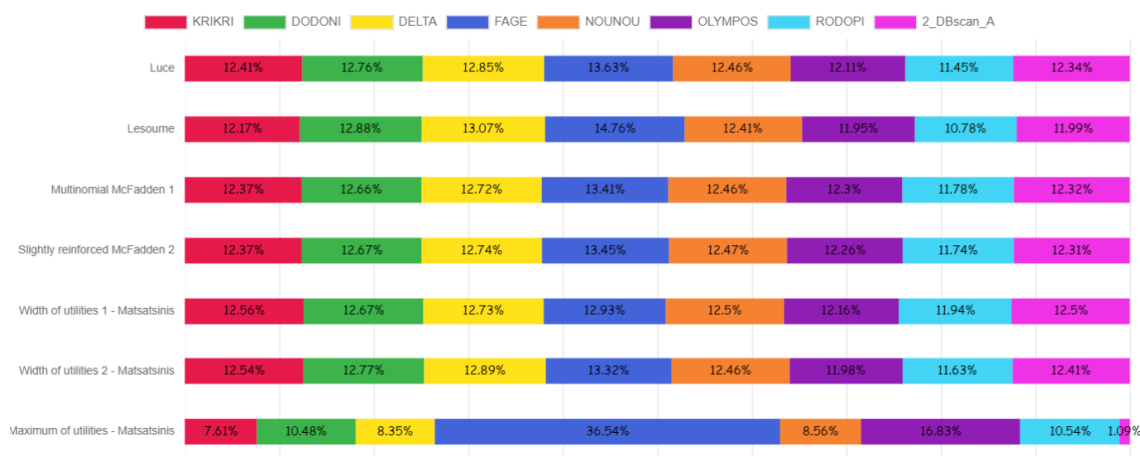


Σχήμα 5.47 Παρουσίαση μεριδίων με την εισαγωγή του 1_DBScan_A για το Cluster A

Η διαδικασία αυτή, θα επαναληφθεί και για την εισαγωγή ενός δεύτερου νέου προϊόντος (2_DBScan_A). Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα διαστήματος, όπως αυτή διατυπώθηκε στον πίνακα 5.12, εμφανίζονται στην συνέχεια.

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Αδιάφορη (3/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Όπως ήταν αναμενόμενο τα μερίδια της αγοράς για το νέο προϊόν βελτιώθηκαν, καθώς αυξήθηκε κατά μία μονάδα το κριτήριο με το μεγαλύτερο βάρος για τους καταναλωτές. Παρόλο αυτά, η αλλαγή αυτή δεν επέφερε μεγάλες αλλαγές σε ότι αφορά την κατάταξη του προϊόντος σε σχέση με τα υπόλοιπα προϊόντα για τα περισσότερα μοντέλα προτίμησης.

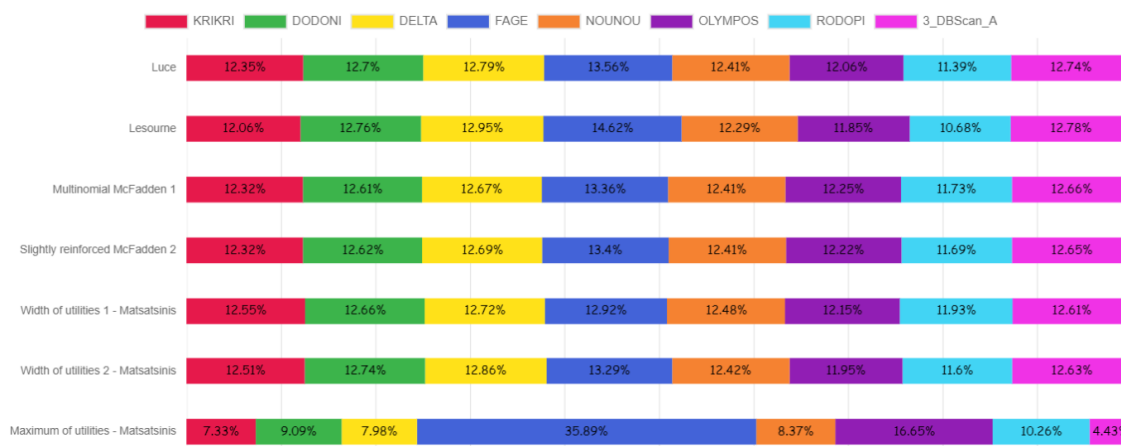


Σχήμα 5.48 Παρουσίαση μεριδίων με την εισαγωγή του 2_DBScan_A για το Cluster A

Τελευταία περίπτωση είναι η εισαγωγή ενός τρίτου προϊόντος (3_DBScan_A), το οποίο έχει τα ακόλουθα χαρακτηριστικά:

- h) Επωνυμία: Καλή (4/5)
- i) Γεύση: Καλή (4/5)
- j) Προέλευση: Αδιάφορη (3/5)
- k) Υφή: Αδιάφορη (3/5)
- l) Άρωμα: Καλό (4/5)
- m) Συσκευασία: Καλή (4/5)
- n) Τιμή: Λογική (3/5)

Το τρίτο αυτό σενάριο φέρνει τα καλύτερα δυνατά αποτελέσματα στα ποσοστιαία μερίδια. Το νέο αυτό προϊόν κατέχει στα περισσότερα μοντέλα την τρίτη θέση και απέχει λιγότερα από 1% από τα προϊόντα που βρίσκονται στην πρώτη και την δεύτερη σε όλα τα μοντέλα εκτός από το Lesourne και το Maximum of utilities.



Σχήμα 5.49 Παρουσίαση μεριδίων με την εισαγωγή του 3_DBScan_A για το Cluster A

5.10.4. Μελέτη δεύτερης ομάδας (Συσταδοποίηση με τον DBScan)

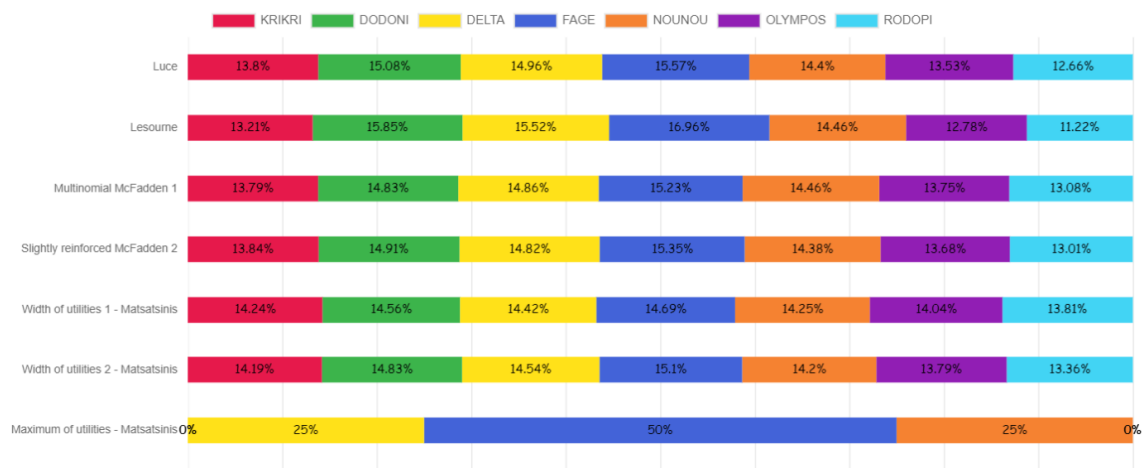
Η συγκεκριμένη ομάδα αποτελείται μόλις από 2 στοιχεία από τα συνολικά 97 που αποτελούντο το αρχικό σετ δεδομένων. Προκειμένου να υπάρξει μία πλήρη εικόνα για όλες τις ομάδες και όλα τα δυνατά αποτελέσματα αυτών, στην συνέχεια θα παρουσιαστούν με ακριβώς την ίδια σειρά η προηγούμενη μεθοδολογία. Πρώτα απ' όλα, κανένα από τα στοιχεία του σετ, δεν έχει δείκτη συσχέτισης κάτω από 0.9. Το κριτήριο με την μεγαλύτερη βαρύτητα για τους καταναλωτές φαίνεται να είναι η γεύση και αμέσως μετά, το άρωμα και η προέλευση του γάλακτος. Σε αντίθεση με την πρώτη ομάδα που μελετήθηκε στην ακριβώς προηγούμενη υποενότητα, η επωνυμία του προϊόντος φαίνεται να παίζει τον λιγότερο σημαντικό ρόλο για τους καταναλωτές.

Average Criteria Weights – Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.100015	0.217847	0.146443	0.132150	0.185661	0.100028	0.117857

Average Alternatives Utilities – Post Optimization						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPUS	RODOPI
0.796465	0.867858	0.871451	0.892858	0.842879	0.792880	0.742866

Σχήμα 5.50 Μέσος όρος βαρών και ολικών χρησιμοτήτων δεύτερου cluster

Προσομοιώνοντας την παρούσα κατάσταση για όλες τις εναλλακτικές και με παράλληλη χρήση των μοντέλων προσωπικής επιλογής δίνονται το ακόλουθο διάγραμμα με τα αντίστοιχα μερίδια.

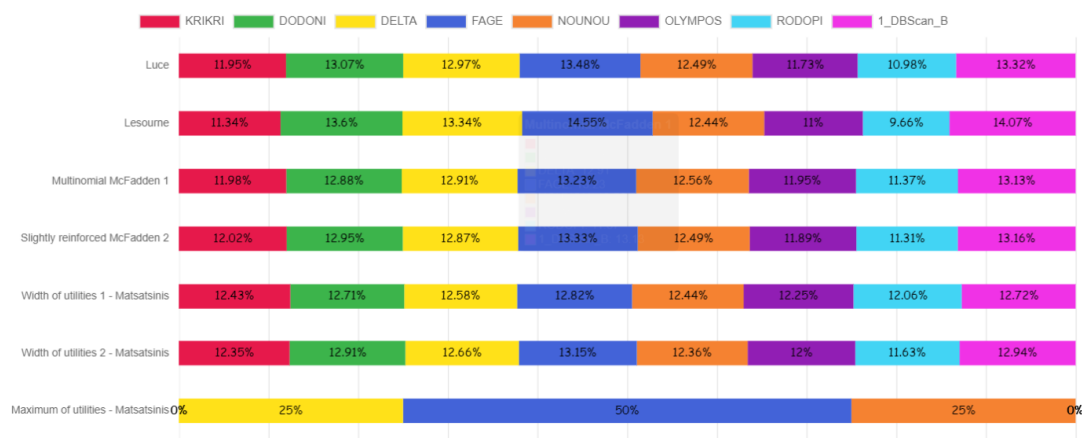


Σχήμα 5.51 Παρουσίαση μεριδίων για όλα τα μοντέλα για το δεύτερο Cluster.

Όπως γίνεται αντιληπτό, το προϊόν της εταιρείας «ΦΑΓΕ», κατέχει το μεγαλύτερο μερίδιο σε όλες τις περιπτώσεις. Αμέσως μετά ακολουθεί η εταιρεία «Δωδώνη», με εξαίρεση στο μοντέλο Maximum of utilities. Τρίτο έρχεται το προϊόν της εταιρείας «ΔΕΛΤΑ. Τελευταίο στην προτίμηση έρχεται το προϊόν της εταιρείας «ΡΟΔΟΠΗ». Όσον αφορά τώρα το πρώτο απλό σενάριο που δημιουργήθηκε, το νέο προϊόν (1_DBScan_B) που εισάγεται στην αγορά έχει τα εξής χαρακτηριστικά για την αύξουσα, πενταβάθμια κλίμακα των κριτηρίων.

- Επωνυμία: Αδιάφορη (3/5)
- Γεύση: Καλή (4/5)
- Προέλευση: Αδιάφορη (3/5)
- Υφή: Αδιάφορη (3/5)
- Άρωμα: Καλό (4/5)
- Συσκευασία: Καλή (4/5)
- Τιμή: Λογική (3/5)

Τα αποτελέσματα που προκύπτουν για τα μερίδια της αγοράς, φαίνονται στο επόμενο διάγραμμα.



Σχήμα 5.52 Παρουσίαση μεριδίων με την εισαγωγή του 1_DBScan_B για το Cluster B

Σε αντίθεση με τις προηγούμενες περιπτώσεις, το νέο προϊόν φαίνεται αρκετά ανταγωνιστικό και εντοπίζεται στην δεύτερη θέση για όλα σχεδόν τα μοντέλα, αμέσως μετά το προϊόν της «ΦΑΓΕ». Όσον αφορά τώρα τα μερίδια της αγοράς υπάρχουν αρκετά μικρές ποσοστιαίες διαφορές μεταξύ των προϊόντων.

Η διαδικασία αυτή, θα επαναληφθεί και για την εισαγωγή ενός δεύτερου νέου προϊόντος (2_DBScan_B). Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα διαστήματος, όπως αυτή διατυπώθηκε στον πίνακα 5.12, εμφανίζονται στην συνέχεια.

- Επωνυμία: Καλή (4/5)
- Γεύση: Αδιάφορη (3/5)
- Προέλευση: Αδιάφορη (3/5)
- Υφή: Αδιάφορη (3/5)
- Άρωμα: Καλό (4/5)
- Συσκευασία: Καλή (4/5)
- Τιμή: Λογική (3/5)

Όπως ήταν αναμενόμενο τα μερίδια της αγοράς για το νέο προϊόν δεν είναι βελτιωμένα. Αντιθέτως, τα ποσοστιαία μερίδια της αγοράς είναι πλέον μικρότερα για το νέο προϊόν. Τέλος, θα πρέπει να σημειωθεί πως το νέο προϊόν (2_DBScan_B) βρίσκεται στην τέταρτη θέση.

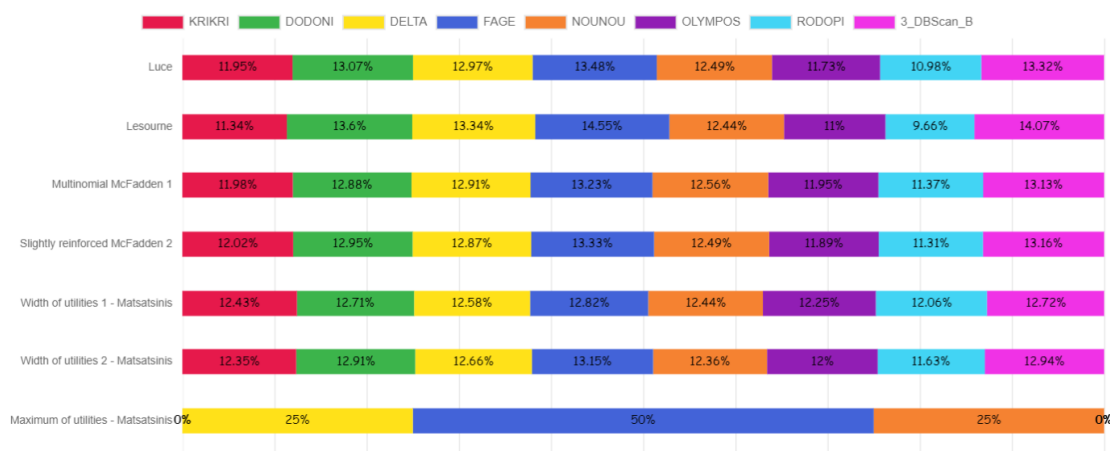


Σχήμα 5.53 Παρουσίαση μεριδίων με την εισαγωγή του 2_DBScan_B για το Cluster B

Τελευταία περίπτωση είναι η εισαγωγή ενός τρίτου προϊόντος (3_DBScan_B), το οποίο έχει τα ακόλουθα χαρακτηριστικά:

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Το τρίτο αυτό σενάριο φέρνει τα καλύτερα δυνατά αποτελέσματα στα ποσοστιαία μερίδια, τα οποία ωστόσο φαίνεται να είναι ακριβώς ίδια με αυτά του πρώτου ανταγωνιστικού προϊόντος. Ο λόγος για τον οποίον συμβαίνει αυτό οφείλεται στο γεγονός ότι το προϊόν της επωνυμίας όπου μεταβάλλεται η τιμή του δεν έχει ιδιαίτερη βαρύτητα για τους καταναλωτές αυτής της ομάδας.



Σχήμα 5.54 Παρουσίαση μεριδίων με την εισαγωγή του 3_DBScan_B για το Cluster B

5.10.5. Μελέτη τρίτης ομάδας (Συσταδοποίηση με τον DBScan)

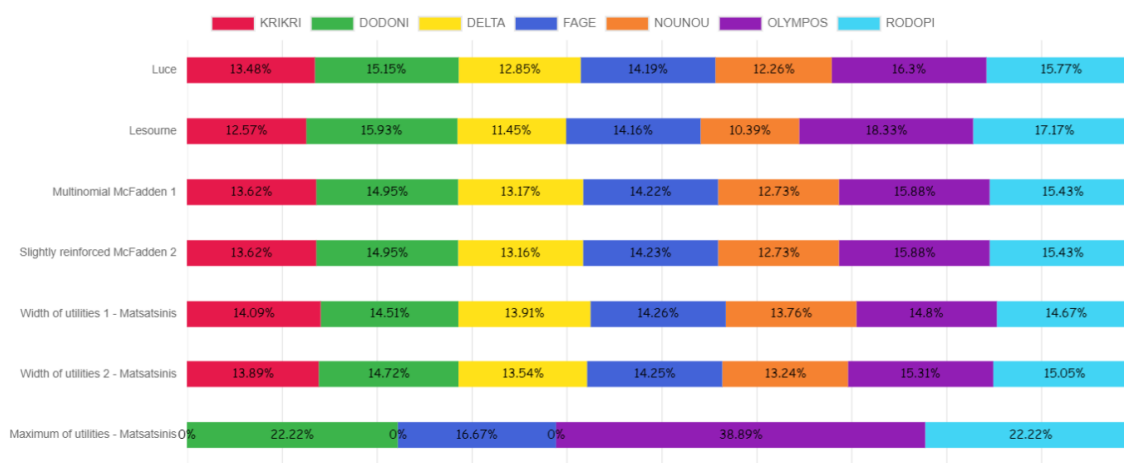
Όσον αφορά την τρίτη ομάδα που προέκυψε από την συσταδοποίηση με χρήση του αλγορίθμου DBScan, και αυτή, όπως και η δεύτερη ομάδα αποτελούνται από ελάχιστα στοιχεία. Παρόλο αυτά, αναλύοντας το προφίλ των καταναλωτών, φαίνεται πως το κριτήριο με το μεγαλύτερο βάρος είναι η επωνυμία και αμέσως μετά η προέλευση του γάλακτος. Επιπλέον, παρατηρούμε πως το πρώτο κριτήριο είναι κοινό με αυτό της πρώτης ομάδας. Συνεπώς, ενδέχεται η συσταδοποίηση να μην είναι αρκετά ικανοποιητική και να πρέπει να ληφθεί υπόψη κάποιος άλλος αλγόριθμος για την ομαδοποίηση των δεδομένων, όπως ο αλγόριθμος K-Means. Το προϊόν τώρα που έρχεται πρώτο σε προτίμηση είναι το γιαούρτι της εταιρείας «ΟΛΥΜΠΟΣ» και αμέσως μετά αυτό της εταιρείας «ΡΟΔΟΠΗ». Ως απόρροια αυτού γίνεται αντιληπτό πως οι καταναλωτές προτιμούν κατσικίσια γιαούρτια.

Average Criteria Weights - Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.171419	0.135728	0.169026	0.138084	0.140483	0.133336	0.111924

Average Alternatives Utilities - Post Optimization						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPUS	RODOPI
0.738093	0.830926	0.704783	0.776210	0.671435	0.892843	0.864259

Σχήμα 5.55 Μέσος όρος βαρών και ολικών χρησιμοτήτων τρίτου cluster

Τα μερίδια της αγοράς τώρα, βάσει των διαφορετικών προτιμσιακών μοντέλων, παρουσιάζονται στην εικόνα που ακολουθεί. Όπως αναφέρθηκε, πρώτο σε προτίμηση έρχεται το προϊόν της εταιρείας «ΟΛΥΜΠΟΣ», και μετά της εταιρείας «ΡΟΔΟΠΗ». Αμέσως μετά ακολουθεί η «ΔΩΔΩΝΗ» και η «ΦΑΓΕ» ενώ στις δύο τελευταίες θέσεις έρχεται η «ΔΕΛΤΑ» και η «ΝΟΥΝΟΥ» αντίστοιχα. Τέλος, ότι αφορά τα ποσοστιαία μερίδια οι αποκλίσεις είναι κοντά στο 1% μεταξύ των δύο συνεχόμενων, ως προς την κατάταξη, προϊόντων, τόσο δηλαδή του πρώτου με το δεύτερο, όσο και του δεύτερου με το τρίτο.



Σχήμα 5.56 Παρουσίαση μεριδίων για όλα τα μοντέλα για το τρίτο Cluster.

Όσον αφορά τώρα το πρώτο απλό σενάριο που δημιουργήθηκε, το νέο προϊόν (1_DBScan_C) που εισάγεται στην αγορά έχει τα εξής χαρακτηριστικά για την αύξουσα, πενταβάθμια κλίμακα των κριτηρίων.

- a) Επωνυμία: Αδιάφορη (3/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Αναλυτικότερα, το νέο προϊόν φαίνεται να καταλαμβάνει την πέμπτη θέση στην αγορά για όλα τα μοντέλα, εκτός του μοντέλου «Width of utilities 2 – Matsatsinis» όπου τόσο το νέο προϊόν, όσο και το προϊόν της «ΦΑΓΕ» ισοψηφούν. Τα ποσοστιαία μερίδια είναι μεταξύ του 12.17% έως 12.49% και απέχει από το προπορευόμενο σε κατάταξη προϊόν από 0.01% μέχρι 0.28%.

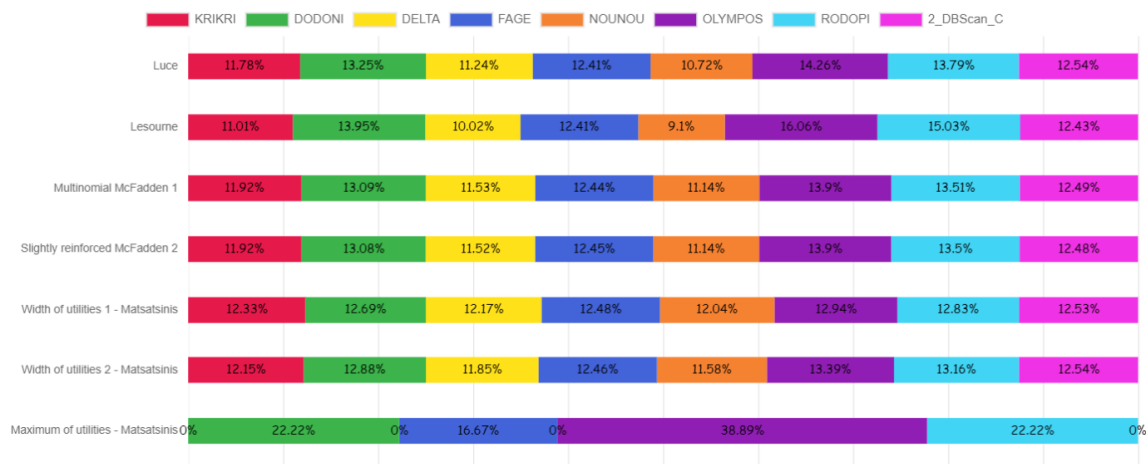


Σχήμα 5.57 Παρουσίαση μεριδίων με την εισαγωγή του 1_DBScan_C για το Cluster C

Η διαδικασία αυτή, επαναλήφθηκε και για την εισαγωγή ενός δεύτερου νέου προϊόντος (2_DBScan_C). Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα διαστήματος, όπως αυτή διατυπώθηκε στον πίνακα 5.12, εμφανίζονται στην συνέχεια.

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Αδιάφορη (3/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Μετά την βελτίωση του κριτηρίου της επωνυμίας τα μερίδια της αγοράς φαίνεται να έχουν αυξηθεί ελαφρώς για το νέο προϊόν. Το νέο προϊόν πλέον βρίσκεται στην τέταρτη θέση αν και απέχει ελάχιστα από το προϊόν της «ΦΑΓΕ» που το ακολουθεί.

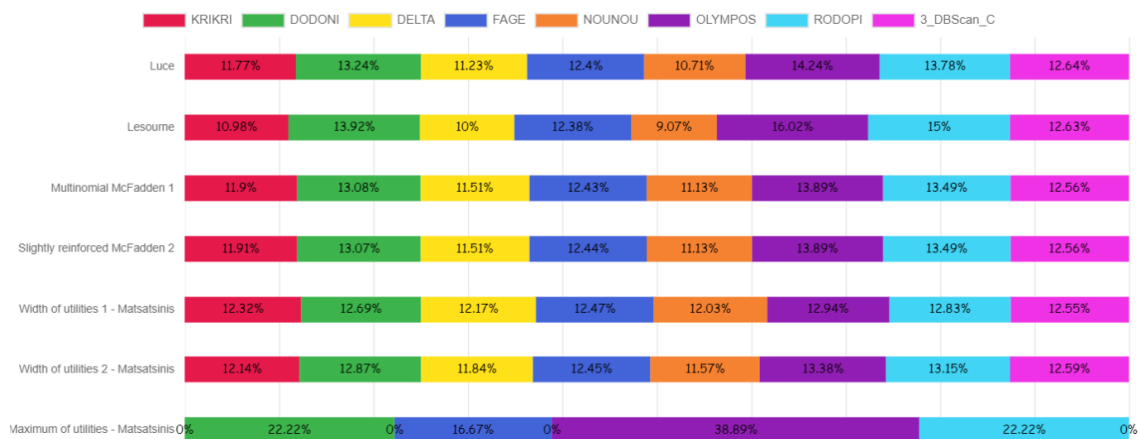


Σχήμα 5.58 Παρουσίαση μεριδίων με την εισαγωγή του 2_DBScan_C για το Cluster C

Τελευταία περίπτωση είναι η εισαγωγή ενός τρίτου προϊόντος (3_DBScan_C), το οποίο έχει τα ακόλουθα χαρακτηριστικά:

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Το τρίτο αυτό σενάριο φέρνει τα καλύτερα δυνατά αποτελέσματα στα ποσοστιαία μερίδια, τα οποία αυξήθηκαν. Παρόλο αυτά, η βελτίωση αυτή φαίνεται πως δεν ήταν αρκετή για να ανεβάσουν στην κατάταξη το νέο προϊόν. Η βασική διαφορά σε σχέση με τα προηγούμενα δύο σενάρια που δημιουργήθηκαν είναι οι ποσοστιαίες αυξήσεις των μεριδίων και το μεγαλύτερο εύρος στις τιμές μεταξύ των προϊόντων.



Σχήμα 5.59 Παρουσίαση μεριδίων με την εισαγωγή του 3_DBScan_C για το Cluster C

5.10.6. Μελέτη τέταρτης ομάδας (Συσταδοποίηση με τον DBScan)

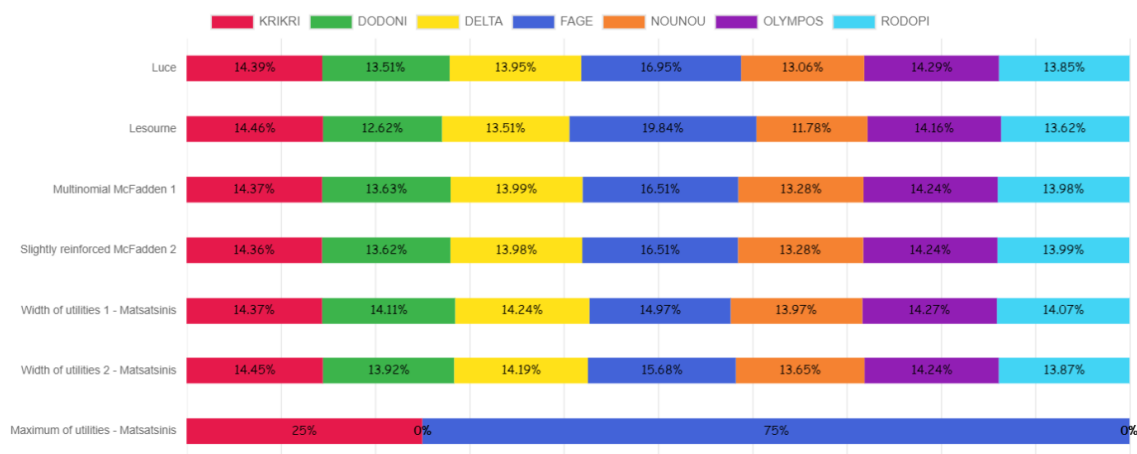
Η τέταρτη και τελευταία ομάδα που προέκυψε από τον αλγόριθμο DBScan, αποτελείται και αυτή μόλις από 2 στοιχεία. Και σε αυτή την περίπτωση ως το πλέον σημαντικότερο κριτήριο είναι η επωνυμία και αμέσως μετά η γεύση. Όπως είναι αναμενόμενο το προϊόν με την μεγαλύτερη απήχηση θα είναι αυτό της εταιρείας «ΦΑΓΕ», ενώ αμέσως μετά ακολουθεί αυτό της εταιρείας «ΚΡΙ-ΚΡΙ». Τα βάρη και οι ολικές χρησιμότητες των καταναλωτών παρουσιάζονται στο επόμενο σχήμα που ακολουθεί.

Average Criteria Weights – Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.253522	0.182143	0.096450	0.107179	0.110736	0.103600	0.146372

Average Alternatives Utilities – Post Optimization						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPOS	RODOPI
0.807243	0.757207	0.782236	0.950072	0.732236	0.800115	0.775072

Σχήμα 5.60 Μέσος όρος βαρών και ολικών χρησιμοτήτων τέταρτου cluster

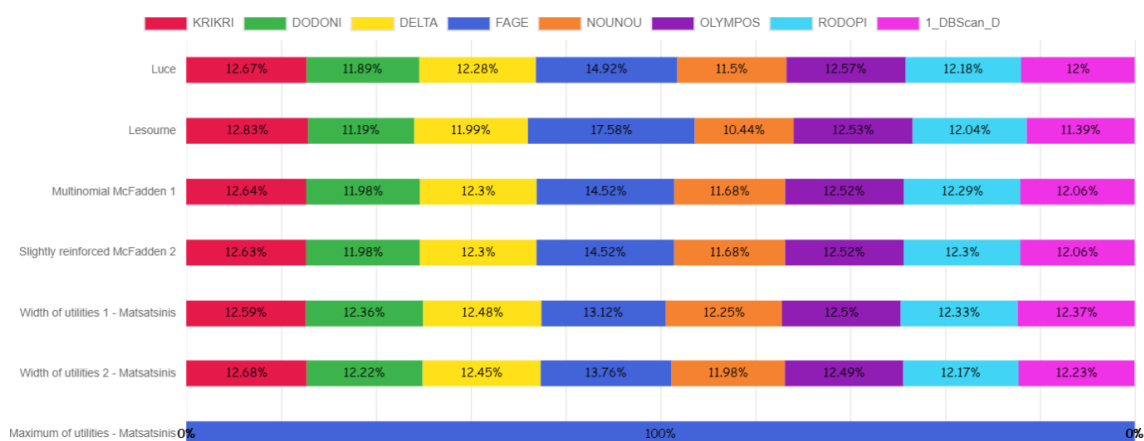
Παρατηρώντας τα μερίδια αγοράς για όλα τα μοντέλα καταναλωτικής συμπεριφοράς, γίνεται αντιληπτό πως το προϊόν της εταιρείας «ΦΑΓΕ», είναι αυτό που επικρατεί και με διαφορά. Ειδικότερα, η ελάχιστη ποσοστιαία διαφορά είναι 0.6% και εμφανίζεται στο μοντέλο «Width of utilities-1», ενώ η μεγαλύτερη διαφορά εμφανίζεται στο τελευταίο μοντέλο (Maximum of utilities) και ισούται με 50%.



Σχήμα 5.61 Παρουσίαση μεριδίων για όλα τα μοντέλα για το τέταρτο Cluster.

Όσον αφορά τώρα το πρώτο απλό σενάριο που δημιουργήθηκε, το νέο προϊόν (1_DBScan_D) που εισάγεται στην αγορά έχει τα εξής χαρακτηριστικά για την αύξουσα, πενταβάθμια κλίμακα των κριτηρίων.

- a) Επωνυμία: Αδιάφορη (3/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)



Σχήμα 5.62 Παρουσίαση μεριδίων με την εισαγωγή του 1_DBScan_D για το Cluster D.

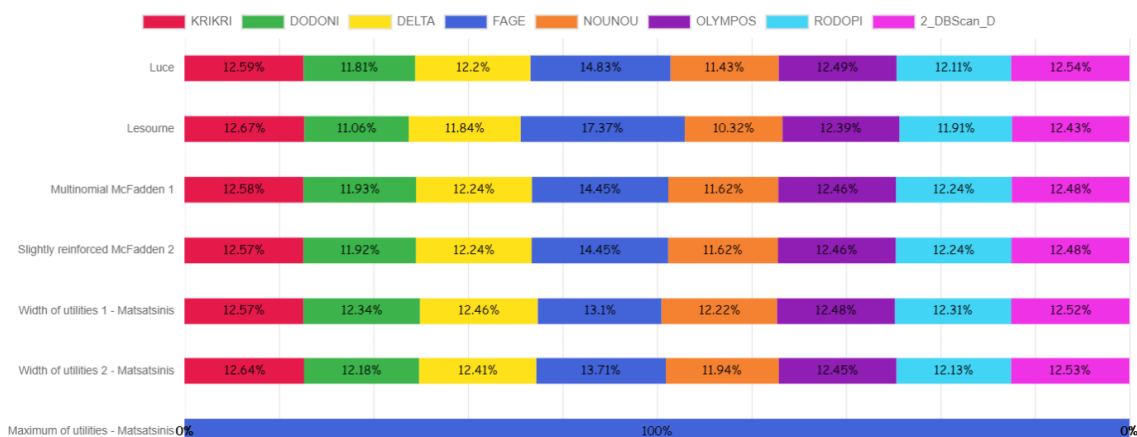
Το νέο προϊόν φαίνεται να καταλαμβάνει την έκτη θέση στα περισσότερα μοντέλα, εκτός από τα μοντέλα «Width of utilities 1,2», στα οποία κατέχουν την πέμπτη

θέση. Τα ποσοστιαία μερίδια της αγοράς, είναι μεταξύ 11.39% έως 12.38% και απέχουν ελάχιστα από τα προπορευόμενα προϊόντα.

Η διαδικασία αυτή, επαναλήφθηκε και για την εισαγωγή ενός δεύτερου νέου προϊόντος (2_DBScan_D). Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα διαστήματος, όπως αυτή διατυπώθηκε στον πίνακα 5.12, εμφανίζονται στην συνέχεια.

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Αδιάφορη (3/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Όπως ήταν αναμενόμενο, μετά την βελτίωση του κριτηρίου της επωνυμίας τα μερίδια της αγοράς φαίνεται να έχουν αυξηθεί ελαφρώς, σχεδόν 0.5%, για το νέο προϊόν. Το νέο προϊόν πλέον βρίσκεται στην τρίτη θέση αν και απέχει ελάχιστα από το προϊόν της «KPI-KPI» που προηγείται.

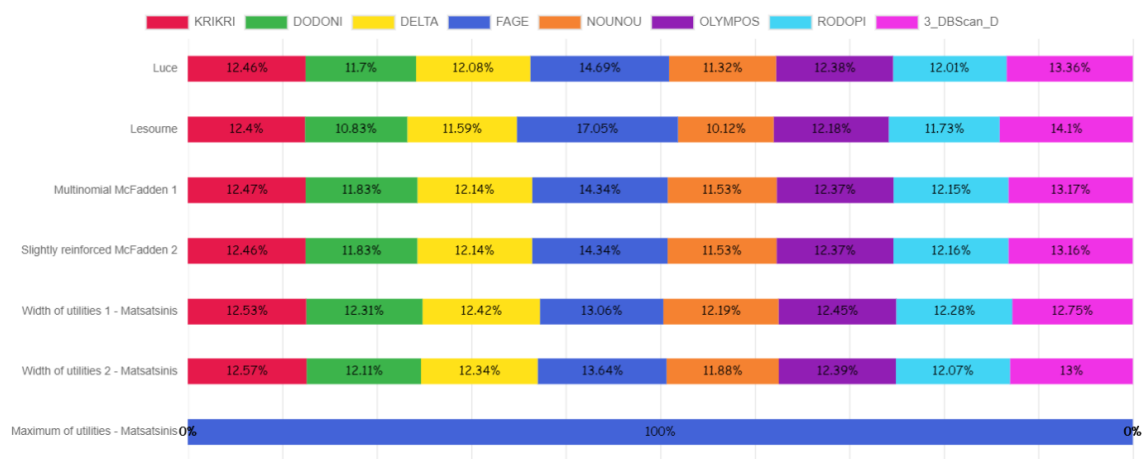


Σχήμα 5.63 Παρουσίαση μεριδίων με την εισαγωγή του 2_DBScan_D για το Cluster D.

Τελευταία περίπτωση είναι η εισαγωγή ενός τρίτου προϊόντος (3_DBScan_D), το οποίο έχει τα ακόλουθα χαρακτηριστικά:

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Το τρίτο αυτό σενάριο φέρνει τα καλύτερα δυνατά αποτελέσματα στα ποσοστιαία μερίδια, τα οποία αυξήθηκαν. Η βελτίωση αυτή φαίνεται πως ήταν αρκετή για να ανεβάσουν στην κατάταξη το νέο προϊόν σε κάποια από τα μοντέλα. Η βασική διαφορά σε σχέση με τα προηγούμενα δύο σενάρια που δημιουργήθηκαν είναι οι ποσοστιαίες αυξήσεις των μεριδίων και το μεγαλύτερο εύρος στις τιμές μεταξύ των προϊόντων.



Σχήμα 5.64 Παρουσίαση μεριδίων με την εισαγωγή του 3_DBScan_D για το Cluster D.

5.10.7. Μελέτη πρώτης ομάδας (Συσταδοποίηση με τον EM)

Η πρώτη ομάδα, όπως αυτή προέκυψε από την συσταδοποίηση του συγκεκριμένου αλγορίθμου αποτελείται από 108 στοιχεία, εκ των οποίων 76 από αυτά έχουν συντελεστή συσχέτισης μεγαλύτερο από 0.9. Πιο αναλυτικά, η ομάδα αυτή φαίνεται να θεωρεί ως πιο σημαντικό κριτήριο αυτό της επωνυμίας, ενώ αμέσως μετά αυτό της γεύσης. Το λιγότερο σημαντικό κριτήριο είναι αυτό της συσκευασίας. Όπως και στις προηγούμενες περιπτώσεις, έτσι και σε αυτή, βάσει του κριτηρίου με το μεγαλύτερο βάρος, ως πλέον προτιμότερο προϊόν είναι αυτό της εταιρείας «ΦΑΓΕ», ενώ αμέσως μετά ακολουθεί αυτό της εταιρείας «ΔΩΔΩΝΗ».

Average Criteria Weights - Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.177618	0.172252	0.136741	0.133603	0.137788	0.120831	0.121168
Average Alternatives Utilities - Post Optimization						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPOS	RODOPI
0.784538	0.829699	0.794254	0.844909	0.749954	0.801259	0.767767

Σχήμα 5.65 Μέσος όρος βαρών και ολικών χρησιμότητων πρώτου cluster

Τα αντίστοιχα μερίδια στην αγορά, για όλα τα μοντέλα καταναλωτικής συμπεριφοράς, παρουσιάζονται στην συνέχεια στο σχήμα 5.65. Γενικότερα, φαίνεται να υπάρχει

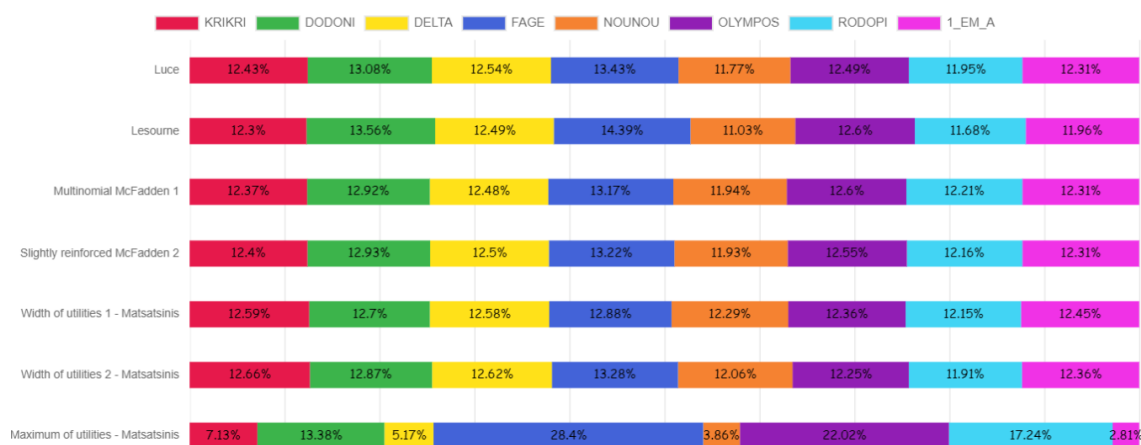
αρκετός ανταγωνισμός μεταξύ των προϊόντων, δεδομένου ότι υπάρχουν μικρές αποκλίσεις ανάμεσα στα προϊόντα των εταιρειών.



Σχήμα 5.66 Παρουσίαση μεριδίων για όλα τα μοντέλα για το πρώτο Cluster.

Όσον αφορά τώρα το πρώτο απλό σενάριο που δημιουργήθηκε, το νέο προϊόν (1_EM_A) που εισάγεται στην αγορά έχει τα εξής χαρακτηριστικά για την αύξουσα, πενταβάθμια κλίμακα των κριτηρίων.

- Επωνυμία: Αδιάφορη (3/5)
- Γεύση: Καλή (4/5)
- Προέλευση: Αδιάφορη (3/5)
- Υφή: Αδιάφορη (3/5)
- Άρωμα: Καλό (4/5)
- Συσκευασία: Καλή (4/5)
- Τιμή: Λογική (3/5)



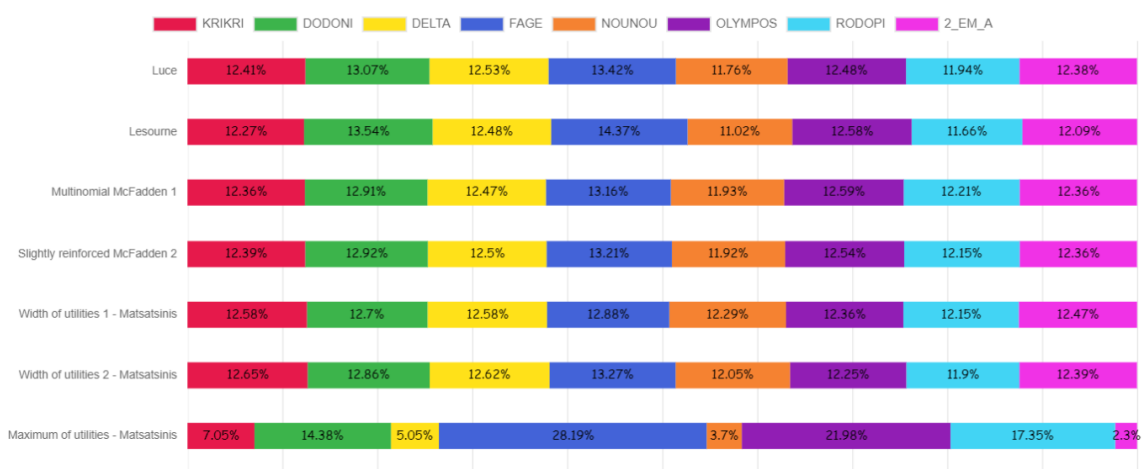
Σχήμα 5. 67 Παρουσίαση μεριδίων με την εισαγωγή του 1_EM_A για το Cluster A.

Το νέο προϊόν φαίνεται να καταλαμβάνει την έκτη θέση στα περισσότερα μοντέλα. Παρόλο αυτά, αξίζει να σημειωθεί πως είναι αρκετά ανταγωνιστικό σε σχέση με τα προπορευόμενα, ως προς την κατάταξη προϊόντα, καθώς οι ποσοστιαίες μεταβολές είναι αρκετά μικρές.

Η διαδικασία αυτή, επαναλήφθηκε και για την εισαγωγή ενός δεύτερου νέου προϊόντος (2_EM_A). Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα διαστήματος, όπως αυτή διατυπώθηκε στον πίνακα 5.12, εμφανίζονται στην συνέχεια.

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Αδιάφορη (3/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Όπως ήταν αναμενόμενο, δεδομένου ότι βελτιώθηκε η ποιότητα του σημαντικότερου, ως προς τους καταναλωτές, κριτηρίου, τα μερίδια της αγοράς για το νέο προϊόν αυξήθηκε. Η αύξηση αυτή ωστόσο δεν ήταν αρκετή για να ανέβει στην κατάταξη το νέο προϊόν. Συνεπώς, η μόνη αλλαγή έγινε στην μείωση του εύρους των τιμών σε σχέση με τα υπόλοιπα ανταγωνιστικά προϊόντα της αγοράς.



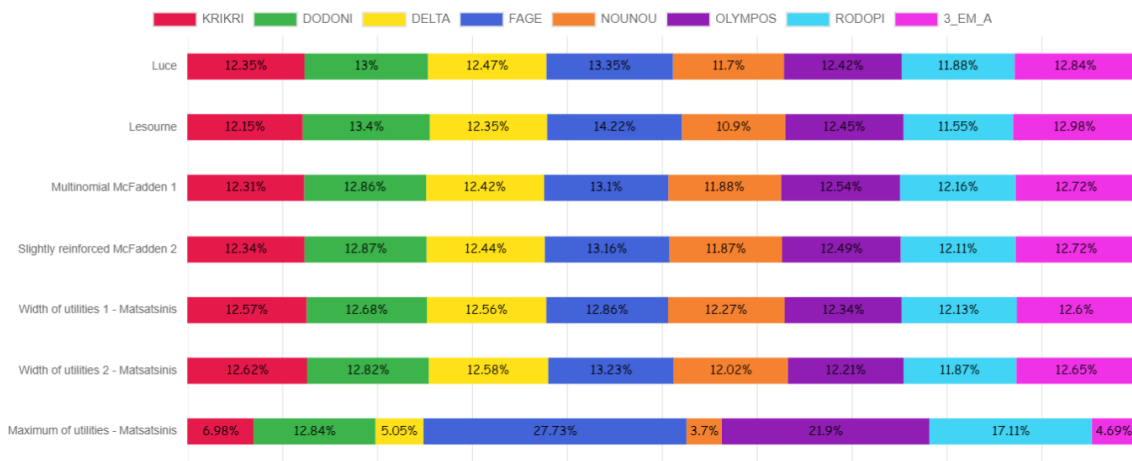
Σχήμα 5.68 Παρουσίαση μεριδίων με την εισαγωγή του 2_EM_A για το Cluster A.

Τελευταία περίπτωση είναι η εισαγωγή ενός τρίτου προϊόντος (3_EM_A), το οποίο έχει τα ακόλουθα χαρακτηριστικά:

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)

g) Τιμή: Λογική (3/5)

Το τρίτο αυτό σενάριο φέρνει τα καλύτερα δυνατά αποτελέσματα στα ποσοστιαία μερίδια, τα οποία αυξήθηκαν. Η βελτίωση αυτή φαίνεται πως ήταν αρκετή για να ανεβάσουν στην κατάταξη το νέο προϊόν στην τρίτη θέση σε όλα τα μοντέλα.



Σχήμα 5.69 Παρουσίαση μεριδίων με την εισαγωγή του 3_EM_A για το Cluster A.

5.10.8. Μελέτη δεύτερης ομάδας (Συσταδοποίηση με τον EM)

Όσον αφορά τώρα το δεύτερο σετ δεδομένων, αυτό αποτελείται από 39 στοιχεία, και μόλις 16 από αυτά έχουν συντελεστή συσχέτισης πάνω από 0.9. Σημαντικό είναι να αναφερθεί, πως τόσο στην πρώτη ομάδα που μελετήθηκε στις προηγούμενες σελίδες, όσο και στην δεύτερη, εμφανίζονται τα πρώτα δύο κριτήρια ίδια ως σημαντικά για τους καταναλωτές, διαφοροποιώντας από το τρίτο κριτήριο και μετά.

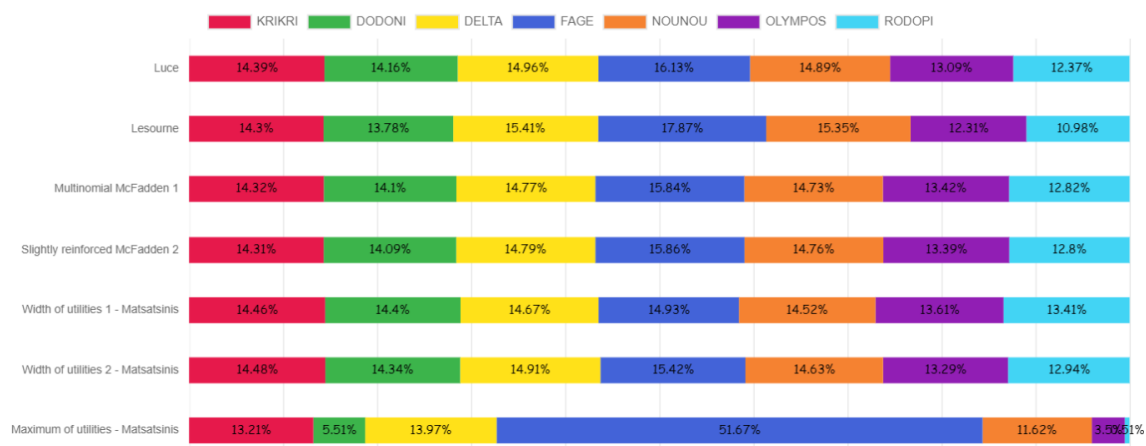
Average Criteria Weights – Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.179826	0.172335	0.122174	0.145973	0.139562	0.123820	0.116311
Average Alternatives Utilities – Post Optimization						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPOS	RODOPI
0.867799	0.853887	0.898568	0.969786	0.895456	0.797839	0.754078

Σχήμα 5.70 Βάρη και ολικές χρησιμότητες δεύτερου Cluster

Βάσει αυτών θα μπορούσε να χαρακτηριστεί ως μία όχι και τόσο επιτυχημένη συσταδοποίηση. Τα βάρη και οι ολικές χρησιμότητες, καθώς και τα προϊόντα που προτιμούν περισσότερο οι καταναλωτές φαίνονται στο σχήμα 5.69 που ακολουθεί. Τα αντίστοιχα μερίδια στην αγορά, για όλα τα μοντέλα καταναλωτικής συμπεριφοράς, παρουσιάζονται στην συνέχεια στο σχήμα 5.70. Γενικότερα, φαίνεται να υπάρχει αρκετός ανταγωνισμός μεταξύ των προϊόντων που βρίσκονται από την δεύτερη μέχρι την πέμπτη θέση, δεδομένου ότι υπάρχουν μικρές αποκλίσεις.

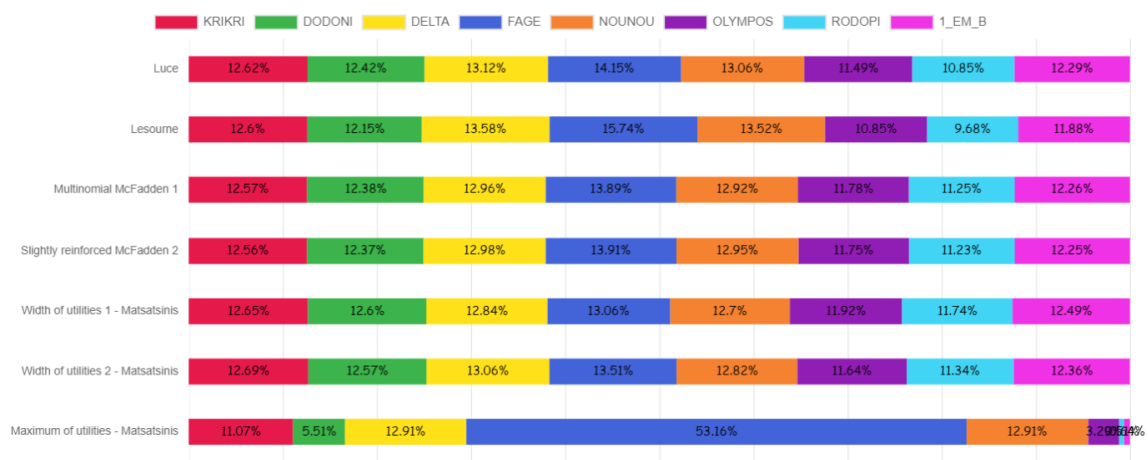
Όσον αφορά τώρα το πρώτο απλό σενάριο που δημιουργήθηκε, το νέο προϊόν (1_EM_B) που εισάγεται στην αγορά έχει τα εξής χαρακτηριστικά για την αύξουσα, πενταβάθμια κλίμακα των κριτηρίων.

- a) Επωνυμία: Αδιάφορη (3/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)



Σχήμα 5.71 Παρουσίαση μεριδίων για όλα τα μοντέλα για το δεύτερο Cluster.

Το νέο προϊόν φαίνεται να καταλαμβάνει την έκτη θέση στα περισσότερα μοντέλα. Παρόλο αυτά, αξίζει να σημειωθεί πως είναι αρκετά ανταγωνιστικό σε σχέση με τα προπορευόμενα, ως προς την κατάταξη προϊόντα, καθώς οι ποσοστιαίες μεταβολές είναι αρκετά μικρές.

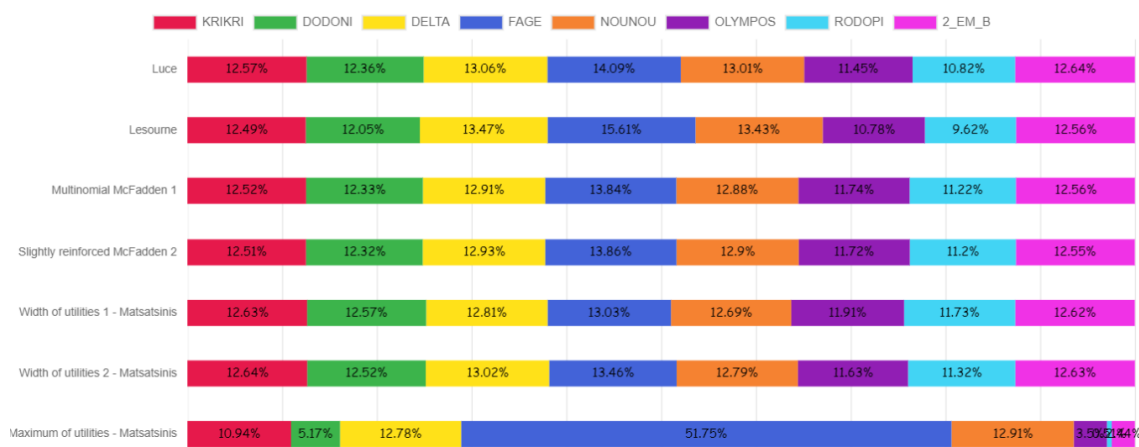


Σχήμα 5.72 Παρουσίαση μεριδίων με την εισαγωγή του 1_EM_A για το Cluster B.

Η διαδικασία αυτή, επαναλήφθηκε και για την εισαγωγή ενός δεύτερου νέου προϊόντος (2_EM_B). Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα διαστήματος, όπως αυτή διατυπώθηκε στον πίνακα 5.12, εμφανίζονται στην συνέχεια.

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Αδιάφορη (3/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Όπως ήταν αναμενόμενο με βελτίωση του κριτηρίου της επωνυμίας, το μερίδιο αγοράς του νέου ανταγωνιστικού προϊόντος βελτιώθηκε. Η βελτίωση αυτή είναι αρκετά ικανοποιητικό, καθώς το νέο προϊόν αυτό πλέον βρίσκεται στην τέταρτη από την έκτη θέση, και απέχει λιγότερο από 1% από τα δύο προπορευόμενα προϊόντα.



Σχήμα 5.73 Παρουσίαση μεριδίων με την εισαγωγή του 2_EM_B για το Cluster B.

Τελευταία περίπτωση είναι η εισαγωγή ενός τρίτου προϊόντος (3_EM_B), το οποίο έχει τα ακόλουθα χαρακτηριστικά:

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Το τρίτο αυτό σενάριο φέρνει τα καλύτερα δυνατά αποτελέσματα στα ποσοστιαία μερίδια, τα οποία αυξήθηκαν. Η βελτίωση αυτή φαίνεται πως δεν ήταν αρκετή για

περαιτέρω βελτίωση στην κατάταξη, και το νέο προϊόν παρέμεινε στην τρίτη θέση, με ελαφρώς μειωμένη την ποσοστιαία διαφορά των μεριδίων της αγοράς από τα δύο προπορευόμενα προϊόντα.



Σχήμα 5.74 Παρουσίαση μεριδίων με την εισαγωγή του 2_EM_B για το Cluster B.

5.10.9. Μελέτη πρώτης ομάδας (Συσταδοποίηση με τον Hierarchical Clusterer)

Η πρώτη ομάδα, όπως αυτή προέκυψε από τον αλγόριθμο Hierarchical Clusterer, φαίνεται ότι αποτελείται από 90 καταναλωτές. Οι μισοί από αυτούς έχουν δείκτη συσχέτισης μεγαλύτερο από 0.9 και τα βάρη με τις αντίστοιχες ολικές χρησιμότητες παρουσιάζονται στο σχήμα που ακολουθεί.

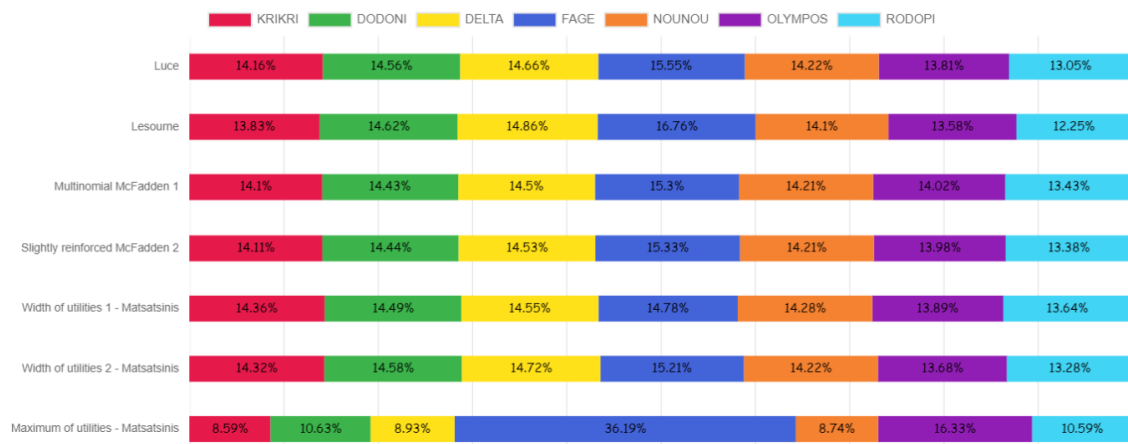
Average Criteria Weights - Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.185278	0.166069	0.130205	0.136989	0.138400	0.123606	0.119452
Average Alternatives Utilities - Post Optimization						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPOS	RODOPI
0.842102	0.865117	0.869161	0.920973	0.847818	0.831379	0.787581

Σχήμα 5.75 Μέσος όρος βαρών και ολικών χρησιμότητων πρώτου cluster

Το κριτήριο με την μεγαλύτερη σημαντικότητα για τους καταναλωτές φαίνεται να είναι η επωνυμία του προϊόντος και αμέσως μετά η γεύση. Όμοια με τις προηγούμενες περιπτώσεις, το προϊόν της εταιρείας «ΦΑΓΕ» φαίνεται να είναι το πλέον προτιμητέο, ενώ αμέσως μετά ακολουθεί αυτό της εταιρείας «ΔΕΛΤΑ» και της «ΔΩΔΩΝΗ».

Με χρήση όλων των μοντέλων καταναλωτικής συμπεριφοράς, τα μερίδια αγοράς διαμορφώνονται σύμφωνα με το σχήμα της εικόνας 5.75 που παρουσιάζεται στην συνέχεια. Πιο αναλυτικά, τα προϊόντα με εξαίρεση το πρώτο (ΦΑΓΕ) και τα 2

τελευταία (ΟΛΥΜΠΟΣ, ΡΟΔΟΠΗ) φαίνεται να είναι αρκετά ανταγωνιστικά μεταξύ τους και να απέχουν ελάχιστα.

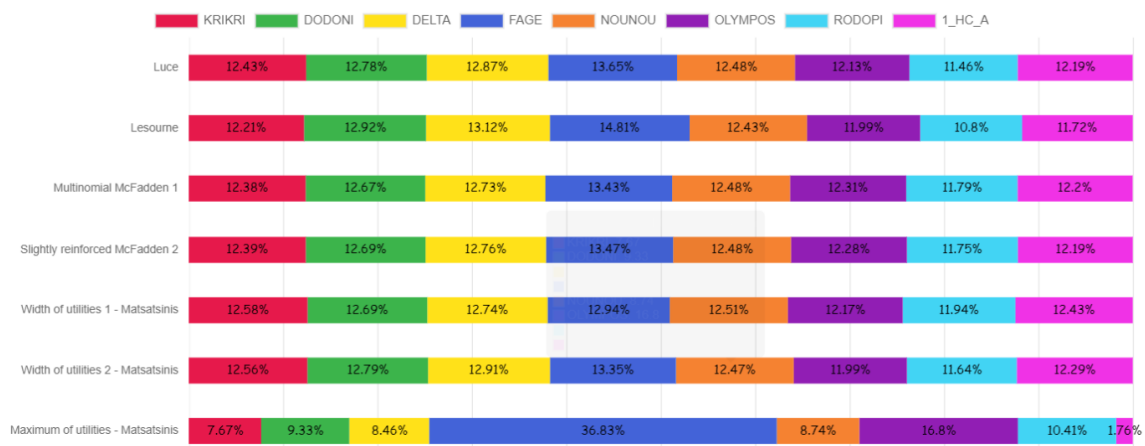


Σχήμα 5.76 Παρουσίαση μεριδίων για όλα τα μοντέλα για το πρώτο Cluster.

Όσον αφορά τώρα το πρώτο απλό σενάριο που δημιουργήθηκε, το νέο προϊόν (1_HC_A) που εισάγεται στην αγορά έχει τα εξής χαρακτηριστικά για την αύξουσα, πενταβάθμια κλίμακα των κριτηρίων.

- Επωνυμία: Αδιάφορη (3/5)
- Γεύση: Καλή (4/5)
- Προέλευση: Αδιάφορη (3/5)
- Υφή: Αδιάφορη (3/5)
- Άρωμα: Καλό (4/5)
- Συσκευασία: Καλή (4/5)
- Τιμή: Λογική (3/5)

Το νέο προϊόν που τοποθετήθηκε στην αγορά, φαίνεται να έρχεται στην πέμπτη θέση για τα περισσότερα μοντέλα, με εξαίρεση το μοντέλο Maximum of Utilities, και να απέχει περισσότερο από 0.5% από τα προπορευόμενα προϊόντα.

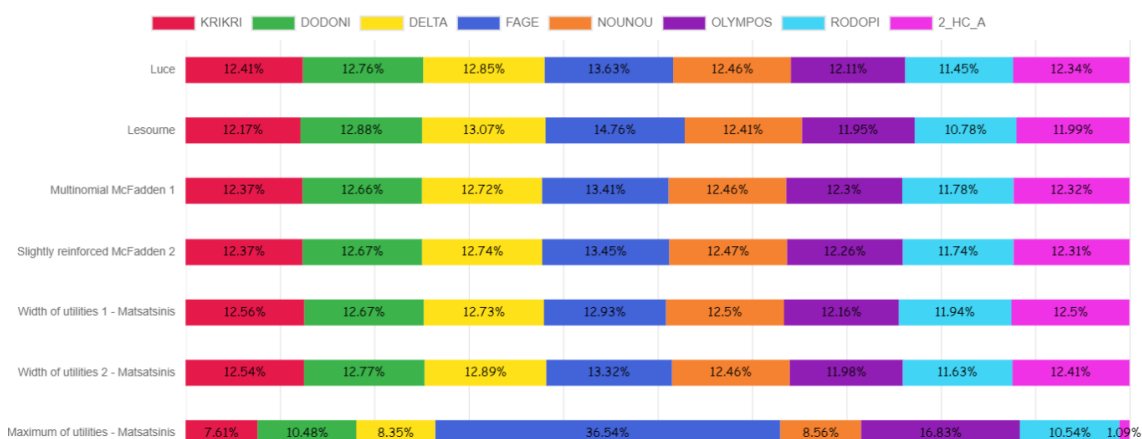


Σχήμα 5.77 Παρουσίαση μεριδίων με την εισαγωγή του 1_HC_A για το Cluster A.

Η διαδικασία αυτή, επαναλήφθηκε και για την εισαγωγή ενός δεύτερου νέου προϊόντος (2_HC_A). Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα διαστήματος, όπως αυτή διατυπώθηκε στον πίνακα 5.12, εμφανίζονται στην συνέχεια.

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Αδιάφορη (3/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Όπως ήταν αναμενόμενο με βελτίωση του κριτηρίου της επωνυμίας, το μερίδιο αγοράς του νέου ανταγωνιστικού προϊόντος βελτιώθηκε. Η βελτίωση αυτή δεν είναι αρκετά ικανοποιητική, καθώς το νέο προϊόν αυτό εξακολουθεί να βρίσκεται στην έκτη θέση.



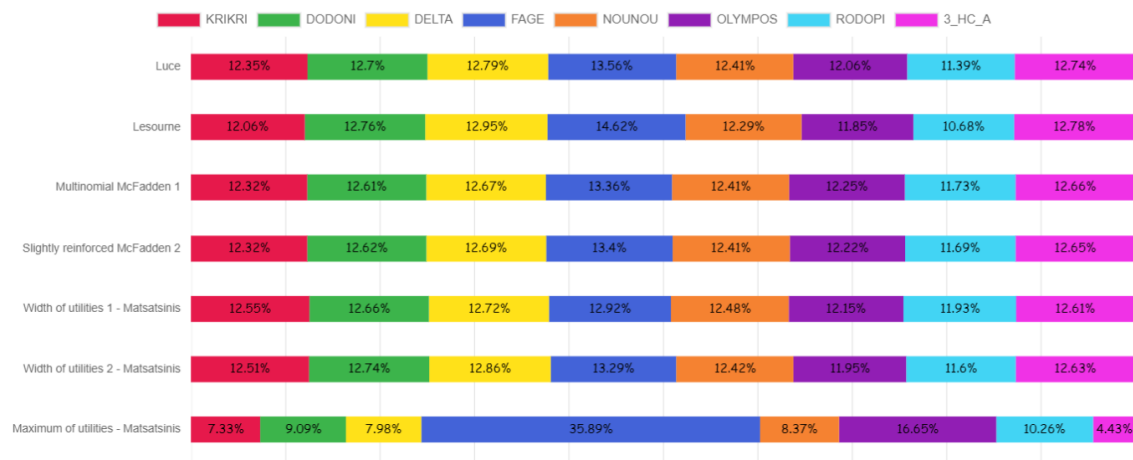
Σχήμα 5.78 Παρουσίαση μεριδίων με την εισαγωγή του 2_HC_A για το Cluster A.

Τελευταία περίπτωση είναι η εισαγωγή ενός τρίτου προϊόντος (3_HC_A), το οποίο έχει τα ακόλουθα χαρακτηριστικά:

- a) Επωνυμία: Καλή (4/5)
- b) Γεύση: Καλή (4/5)
- c) Προέλευση: Αδιάφορη (3/5)
- d) Υφή: Αδιάφορη (3/5)
- e) Άρωμα: Καλό (4/5)
- f) Συσκευασία: Καλή (4/5)
- g) Τιμή: Λογική (3/5)

Το τρίτο αυτό σενάριο φέρνει τα καλύτερα δυνατά αποτελέσματα στα ποσοστιαία μερίδια, τα οποία αυξήθηκαν. Η βελτίωση αυτή φαίνεται πως ήταν αρκετή για

περαιτέρω βελτίωση στην κατάταξη, και το νέο προϊόν φτάσει στην τρίτη θέση της αγοράς στα περισσότερα μοντέλα.



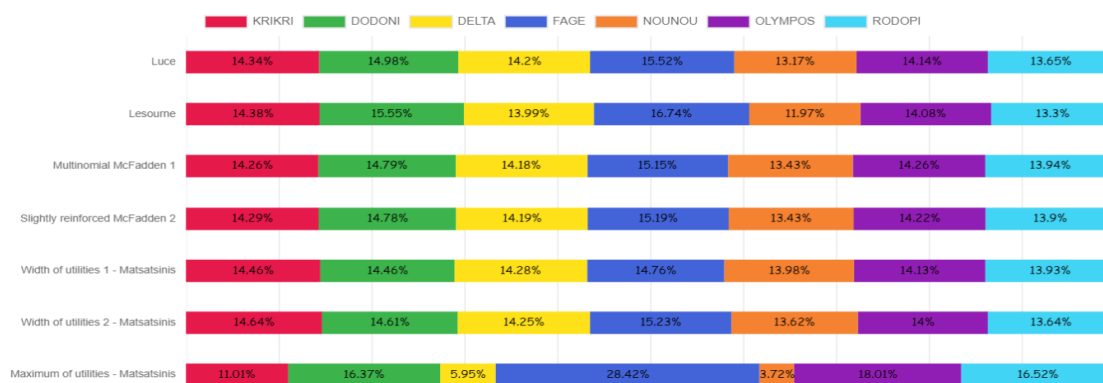
Σχήμα 5.79 Παρουσίαση μεριδίων με την εισαγωγή του 3_HC_A για το Cluster A.

5.10.10. Μελέτη δεύτερης ομάδας (Συσταδοποίηση με τον Hierarchical Clusterer)

Η δεύτερη ομάδα, όπως αυτή προέκυψε από τον αλγόριθμο αυτόν, αποτελείται από 56 στοιχεία, εκ των οποίων 46 από αυτά έχουν συντελεστή συσχέτισης μεγαλύτερο από 0.9. Και σε αυτήν την περίπτωση, το κριτήριο με την μεγαλύτερη σπουδαιότητα για τους καταναλωτές είναι η επωνυμία και αμέσως μετά το κριτήριο της γεύσης. Το προϊόν που έρχεται πρώτο σε προτίμηση αντίστοιχα είναι αυτό της εταιρείας «ΦΑΓΕ» και ακολουθεί της «ΔΩΔΩΝΗ» και της «ΟΛΥΜΠΟΣ».

Criteria Weights						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.287162	0.208586	0.128756	0.090742	0.056190	0.090729	0.137834
Alternatives Utilities						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPOS	RODOPI
0.530198	0.583612	0.557594	0.626596	0.478561	0.570859	0.524936

Σχήμα 5.80 Μέσος όρος βαρών και ολικών χρησιμοτήτων δεύτερου cluster.

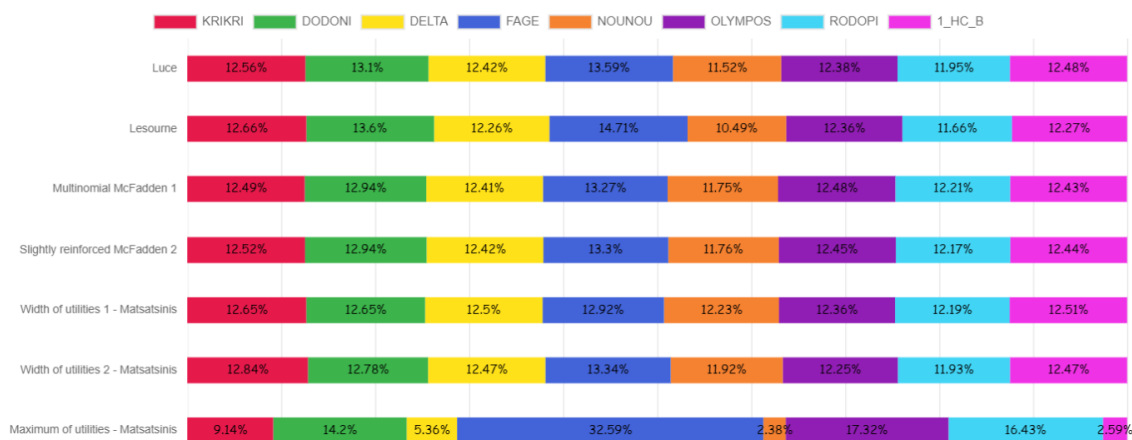


Σχήμα 5.81 Παρουσίαση μεριδίων για όλα τα μοντέλα για το δεύτερο Cluster.

Με εξαίρεση το προϊόν της «ΦΑΓΕ», όλα τα υπόλοιπα προϊόντα φέρουν αρκετά παρόμοια ποσοστιαία μερίδια, και είναι αρκετά ανταγωνιστικά μεταξύ τους. Αυτό γίνεται ευκόλως αντιληπτό από το σχήμα 5.80 που ακολουθεί. Όσον αφορά τώρα το πρώτο απλό σενάριο που δημιουργήθηκε, το νέο προϊόν (1_HC_B) που εισάγεται στην αγορά έχει τα εξής χαρακτηριστικά για την αύξουσα, πενταβάθμια κλίμακα των κριτηρίων.

- h) Επωνυμία: Αδιάφορη (3/5)
- i) Γεύση: Καλή (4/5)
- j) Προέλευση: Αδιάφορη (3/5)
- k) Υφή: Αδιάφορη (3/5)
- l) Άρωμα: Καλό (4/5)
- m) Συσκευασία: Καλή (4/5)
- n) Τιμή: Λογική (3/5)

Το νέο προϊόν που τοποθετήθηκε στην αγορά, φαίνεται να έρχεται στην τέταρτη θέση για τα περισσότερα μοντέλα, με εξαίρεση το μοντέλο Maximum of Utilities, και να απέχει ελάχιστα τόσο από το προπορευόμενο προϊόν, όσο και από αυτό που βρίσκεται στην πέμπτη θέση.

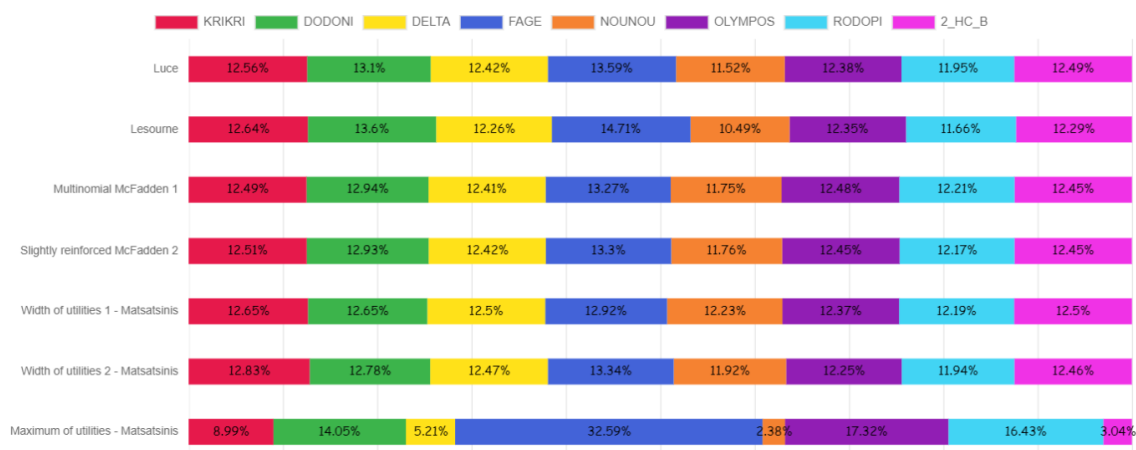


Σχήμα 5.82 Παρουσίαση μεριδίων με την εισαγωγή του 1_HC_B για το Cluster B.

Η διαδικασία αυτή, επαναλήφθηκε και για την εισαγωγή ενός δεύτερου νέου προϊόντος (2_HC_B). Τα αντίστοιχα χαρακτηριστικά του κριτηρίου στην πενταβάθμια κλίμακα διαστήματος, όπως αυτή διατυπώθηκε στον πίνακα 5.12, εμφανίζονται στην συνέχεια.

- h) Επωνυμία: Καλή (4/5)
- i) Γεύση: Αδιάφορη (3/5)
- j) Προέλευση: Αδιάφορη (3/5)
- k) Υφή: Αδιάφορη (3/5)
- l) Άρωμα: Καλό (4/5)
- m) Συσκευασία: Καλή (4/5)
- n) Τιμή: Λογική (3/5)

Όπως ήταν αναμενόμενο με βελτίωση του κριτηρίου της επωνυμίας, το μερίδιο αγοράς του νέου ανταγωνιστικού προϊόντος βελτιώθηκε. Η βελτίωση αυτή δεν είναι αρκετά ικανοποιητική, καθώς το νέο προϊόν αυτό εξακολουθεί να βρίσκεται στην ίδια θέση στην κατάταξη με σχεδόν μηδενική ποσοστιαία αύξηση.

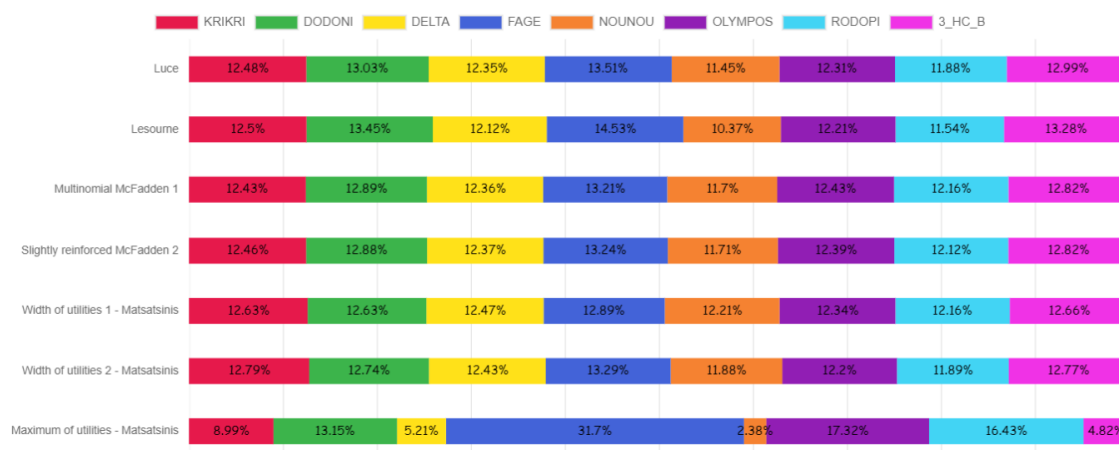


Σχήμα 5.83 Παρουσίαση μεριδίων με την εισαγωγή του 2_HC_B για το Cluster B.

Τελευταία περίπτωση είναι η εισαγωγή ενός τρίτου προϊόντος (3_HC_B), το οποίο έχει τα ακόλουθα χαρακτηριστικά:

- h) Επωνυμία: Καλή (4/5)
- i) Γεύση: Καλή (4/5)
- j) Προέλευση: Αδιάφορη (3/5)
- k) Υφή: Αδιάφορη (3/5)
- l) Άρωμα: Καλό (4/5)
- m) Συσκευασία: Καλή (4/5)
- n) Τιμή: Λογική (3/5)

Το τρίτο αυτό σενάριο φέρνει τα καλύτερα δυνατά αποτελέσματα στα ποσοστιαία μερίδια, τα οποία αυξήθηκαν. Η βελτίωση αυτή φαίνεται πως ήταν αρκετή για περαιτέρω βελτίωση στην κατάταξη, και το νέο προϊόν φτάσει στην τρίτη θέση της αγοράς στα περισσότερα μοντέλα.



Σχήμα 5.84 Παρουσίαση μεριδίων με την εισαγωγή του 3_HC_A για το Cluster A.

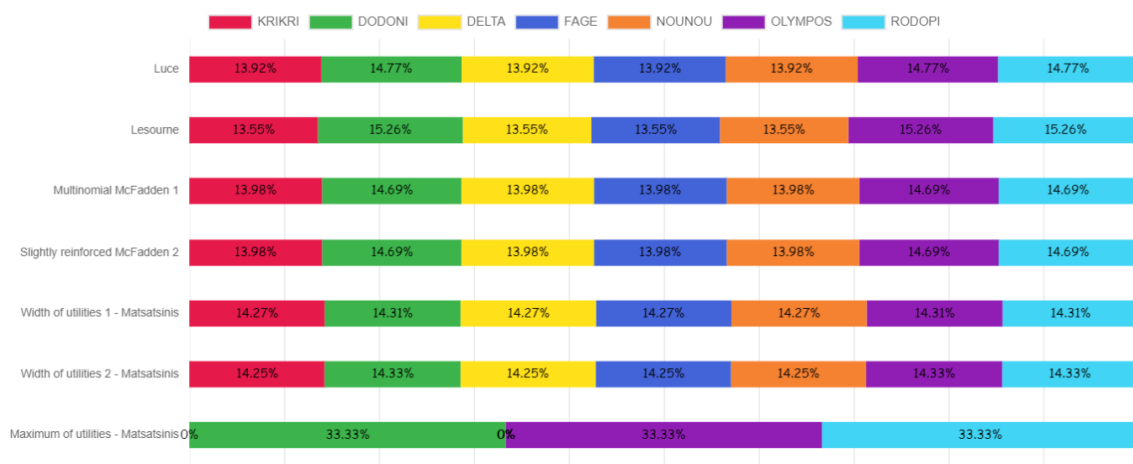
5.10.11. Μελέτη τρίτης ομάδας (Συσταδοποίηση με τον Hierarchical Clusterer)

Η ομάδα αυτή αποτελείται μόλις από 1 στοιχείο. Στην συνέχεια παρουσιάζονται συνοπτικά τα αποτελέσματα, όπως αυτά προέκυψαν από το MARKEX, καθαρά για λόγους πληρότητας. Ως πλέον σημαντικό κριτήριο εμφανίζεται αυτό της επωνυμίας. Επιπλέον, αξίζει να σημειωθεί πως τόσο το προϊόν της εταιρείας «ΔΩΔΩΝΗ», όσο και της εταιρείας «ΟΛΥΜΠΟΣ» και «ΡΟΔΟΠΗ» έρχονται πρώτα σε προτίμηση.

Average Criteria Weights - Post Optimization						
Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0.171371	0.135729	0.142857	0.135729	0.142857	0.135729	0.135729
Average Alternatives Utilities - Post Optimization						
KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPOS	RODOPI
0.814374	0.864273	0.814374	0.814374	0.814374	0.864273	0.864273

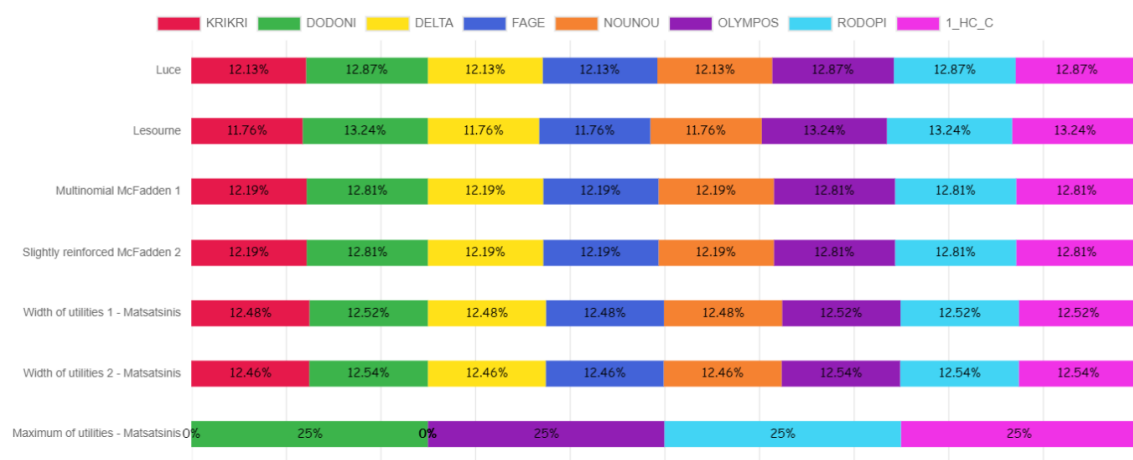
Σχήμα 5.85 Μέσος όρος βαρών και ολικών χρησιμότητων τρίτου cluster.

Με χρήση όλων των μοντέλων καταναλωτικής συμπεριφοράς, τα μερίδια της αγοράς διαμορφώνονται ακριβώς όπως παρουσιάζονται στο σχήμα που ακολουθεί. Δηλαδή, οι τρεις εταιρείες που προαναφέρθηκαν έχουν μερίδιο ίσο με 14.77% και οι υπόλοιπες τέσσερις με ποσοστό 13.92%.



Σχήμα 5.86 Παρουσίαση μεριδίων για όλα τα μοντέλα για το τρίτο Cluster.

Εφαρμόζοντας και τα τρία σενάρια, για την εισαγωγή των νέων προϊόντων στην αγορά, αξίζει να σημειωθεί πως δεν γίνεται κάποια περαιτέρω μεταβολή στα μερίδια της αγοράς με την βελτίωση των κριτηρίων. Έτσι, εισάγοντας ένα νέο ανταγωνιστικό προϊόν στην αγορά τα μερίδια μετασχηματίζονται ως εξής: 12.87% για τις τρεις πρώτες εταιρείες και το νέο, ανταγωνιστικό και 12.13% για τις υπόλοιπες εταιρείες.



Σχήμα 5.87 Παρουσίαση μεριδίων με την εισαγωγή του 1,2,3_HC_A για το Cluster C.

Κεφάλαιο 6 Σύγκριση μεθόδων συσταδοποίησης

Στο κεφάλαιο αυτό θα γίνει μία σύνοψη των μεθόδων, καθώς και ποια από αυτές έδωσε τα καλύτερα αποτελέσματα για το συγκεκριμένο δείγμα που μελετάται στην παρούσα έρευνα. Τα κριτήρια τα οποία χρησιμοποιήθηκαν για να συγκριθούν οι συσταδοποιήσεις μεταξύ τους ήταν, τόσο το πλήθος των κριτηρίων που θεώρησαν σημαντικά οι καταναλωτές, έτσι ώστε να διαφοροποιηθούν οι προτιμήσεις τους, όσο και το πλήθος των εγγραφών που ανήκουν σε κάθε ομάδα, λαμβάνοντας υπόψη πάντα και τις τιμές των δεικτών της κάθε συσταδοποίησης στον εκάστοτε αλγόριθμο, όπως είναι για παράδειγμα ο δείκτης silhouette. Αρχικά, θα πρέπει να αναφερθεί πως σε καμία περίπτωση δεν θα πρέπει να εφαρμόζεται, αποκλειστικά και μόνο, μία μέθοδος συσταδοποίησης για την ομαδοποίηση του δείγματος, αλλά περισσότερες έτσι ώστε να προκύπτουν όλα τα δυνατά σενάρια και στην συνέχεια να επιλέγεται το πλέον προτιμότερο. Στην συνέχεια ακολουθεί η παρουσίαση των σεναρίων αυτών, τόσο για το αρχείο των ολικών χρησιμότητων, όσο και για το αρχείο των βαρών.

Από τους πέντε αλγόριθμους συσταδοποίησης που χρησιμοποιήθηκαν στην συγκεκριμένη μεθοδολογία, δύο από αυτούς έφεραν τα καλύτερα αποτελέσματα όταν προέκυψαν τέσσερις συστάδες. Οι αλγόριθμοι αυτοί είναι ο DBSCAN και ο Canopy. Στην περίπτωση του DBSCAN, η πρώτη ομάδα αποτελείται από 90 εγγραφές, η δεύτερη και η τέταρτη από 2, ενώ η τρίτη από 3 εγγραφές. Παρότι οι εγγραφές στις υπόλοιπες ομάδες είναι αρκετά λιγότερες από αυτές της πρώτης, το αποτέλεσμα είναι αρκετά ικανοποιητικό, καθώς εντοπίστηκαν πλήρως οι διάφορες προτιμήσεις των καταναλωτών. Ειδικότερα, εμφανίστηκε και το μικρότερο μερίδιο αγοράς που προτιμά το κατσικίσιο γιαούρτι από το αγελαδινό, που εστιάζει δηλαδή στην γεύση και την προέλευση της πρώτης ύλης. Όμοια, εντοπίστηκαν και οι καταναλωτές που ενδιαφέρονται περισσότερο για την γεύση και το άρωμα του γιαουρτιού. Αντίστοιχα, στην περίπτωση του αλγόριθμου Canopy εντοπίστηκε η πλειοψηφία των προτιμήσεων των καταναλωτών που εντοπίστηκαν και στον DBSCAN. Τέλος, οι άλλοι δύο αλγόριθμοι, K - Means και EM, φέρουν καλύτερα αποτελέσματα όταν προκύπτουν δύο ομάδες. Ωστόσο, δεδομένου ότι η συγκεκριμένη έρευνα έχει ως στόχο την τμηματοποίηση της αγοράς, είναι ευκόλως αντιληπτό πως η αγορά γιαουρτιού δεν μπορεί να χωριστεί μόνο σε δύο ομάδες. Όλα όσα προαναφέρθηκαν συνοψίζονται στον πίνακα που ακολουθεί.

Πίνακας 6.1 Καλύτερες συσταδοποιήσεις για το αρχείο των ολικών χρησιμότητων.

Αλγόριθμος	K – Means	DBScan	EM	Hierarchical Clusterer	Canopy
Πλήθος Συστάδων	2	4	2	3	4

Αντίστοιχα για το αρχείο των βαρών, παρατηρείται παρόμοια συμπεριφορά με αυτή που περιεγράφηκε στο αρχείο των ολικών χρησιμότητων. Πιο αναλυτικά, και σε αυτήν την περίπτωση, δύο αλγόριθμοι φέρουν τα καλύτερα αποτελέσματα με την δημιουργία τεσσάρων συστάδων. Αυτοί είναι ο EM και ο Canopy. Σε αντίθεση με το αρχείο των ολικών χρησιμότητων, ο αλγόριθμος DBSCAN και ο Hierarchical Clusterer παρουσιάζουν την καλύτερη συσταδοποίηση τους για 3 και 2 συστάδες αντίστοιχα. Όσο για το αρχείο του K – Means και σε αυτή την περίπτωση η καλύτερη συσταδοποίηση προκύπτει για 2 συστάδες. Στον πίνακα που ακολουθεί παρουσιάζονται όλα όσα προαναφέρθηκαν για το αρχείο των βαρών.

Πίνακας 6.2 Καλύτερες συσταδοποιήσεις για το αρχείο των βαρών.

Αλγόριθμος	K – Means	DBScan	EM	Hierarchical Clusterer	Canopy
Πλήθος Συστάδων	2	3	4	2	4

Κεφάλαιο 7 Συμπεράσματα

Συνοψίζοντας, έχοντας πλέον ολοκληρωθεί η παρούσα ανάλυση φαίνεται πως η μελέτη τμηματοποίησης της αγοράς του γιαουρτιού δεν είναι και τόσο απλή υπόθεση. Εκτός από τον έντονο ανταγωνισμό που υπάρχει, λόγω της ποικιλίας των προϊόντων, η τμηματοποίηση καθίσταται ακόμα δυσκολότερη λόγω της υποκειμενικότητας των απόψεων των καταναλωτών. Για να αντιμετωπιστούν αυτά τα προβλήματα, κρίθηκε απαραίτητη η ομαδοποίηση των δεδομένων με ποικίλους αλγορίθμους, ο καθορισμός των σημαντικότερων κριτηρίων για τους καταναλωτές και η μελέτη της κάθε ομάδας ξεχωριστά μέσω μοντέλων προτίμησης για την εισαγωγή τριών νέων ανταγωνιστικών προϊόντων.

Όσον αφορά την ομαδοποίηση, τα αποτελέσματα για τους πέντε αλγορίθμους διαφέρουν ως έναν βαθμό μεταξύ τους. Πιο αναλυτικά, στην περίπτωση του K –Means, όπως φαίνεται και στον πίνακα 5.2 τα καλύτερα αποτελέσματα, βάσει του δείκτη σκιαγράφησης, προέκυψαν για τη δημιουργία 2 συστάδων. Αντίθετα, στην περίπτωση του DBScan, καλύτερα αποτελέσματα προέκυψαν για το αρχείο των ολικών χρησιμότητων όταν δημιουργήθηκαν τέσσερις συστάδες. Για το αρχείο των βαρών το καλύτερο αποτέλεσμα προκύπτει για 3 συστάδες. Όπως ήταν αναμενόμενο στην περίπτωση του αλγορίθμου EM, η συμπεριφορά είναι παρόμοια με αυτή του K –Means, και τα καλύτερα αποτελέσματα προκύπτουν για 2 συστάδες. Τέλος, στην περίπτωση του Canopy η καλύτερη συσταδοποίηση προέκυψε για τέσσερις συστάδες.

Έχοντας πλέον χωρίσει το δείγμα σε τέσσερις ομάδες, βάσει του αλγορίθμου DBSCAN, και εντοπίζοντας ποια κριτήρια είναι πιο σημαντικά για την κάθε ομάδα, παρατηρούμε ότι το κριτήριο της τιμής δεν έρχεται πρώτο σε σημασία σε καμία από τις τέσσερις ομάδες, αλλά το κριτήριο της γεύσης, της επωνυμίας και του αρώματος αντίστοιχα. Σύμφωνα με τα δεδομένα της κλαδικής έρευνας αυτό είναι το βασικότερο χαρακτηριστικό το οποίο διαφοροποιείται και αυτό ίσως οφείλεται στο δείγμα της παρούσας έρευνας, καθώς το μεγαλύτερο μέρος αυτού ανήκει στην δεκαετία των τριάντα και κάτω. Ακόμα, μεταξύ των συστάδων το προϊόν της εταιρείας «ΦΑΓΕ» και «ΔΩΔΩΝΗ» φαίνεται να είναι το πλέον προτιμητέο. Τέλος, τα νέα προϊόντα που εισάγονται φαίνεται πως είναι αρκετά ανταγωνιστικά.

Βιβλιογραφία

Ελληνική

- Αγγελίδης, Α., Μάντης, Α., Παπαγεωργίου, Δ., Φλετούρης, Δ., (2015). Υγιεινή και Τεχνολογία του γάλακτος και των προϊόντων του. Θεσσαλονίκη: Αφοί Κυριακίδη.
- Βούλγαρη, Φ., Ζοπουνίδης, Κ., (2000). Χρηματοοικονομική Στρατηγική Μικρομεσαίων Επιχειρήσεων Στην Ελλάδα. Αθήνα: Κλειδάριθμος.
- Δούμπος, Μ., Ζοπουνίδης, Κ., (2001). Πολυκριτήριες Τεχνικές Ταξινόμησης: Θεωρία και Εφαρμογές. Αθήνα: Κλειδάριθμος.
- Ζυγομήτρος, Α., (2008). Εξόρυξη γνώσης στην μέτρηση ικανοποίησης. Μεταπτυχιακή Διατριβή Πανεπιστήμιο Πειραιώς, Αθήνα.
- Ιωαννίδης, Η., (2016). Ο Συντελεστής Συσχέτισης τ του Kendall. Ανακτήθηκε από <https://docplayer.gr/654-3-4-2-o-syntelestis-syshetisis-t-toy-kendall.html>
- Κερδίζουν σημαντικό έδαφος τα προϊόντα ιδιωτικής ετικέτας (2019, Αύγουστος 20). Ανακτήθηκε από <https://m.naftemporiki.gr/story/1507688/kerdizoun-simantiko-edafos-ta-proionta-idiotikis-etiketas>.
- Ματσατσίνης, Ν. (2021), Επιχειρηματική Ευφυΐα: Αναλυτική και Ανάλυση Μεγάλων Δεδομένων για Λήξη Αποφάσεων. Αθήνα: Νέες Τεχνολογίες.
- Ματσατσίνης Ν., 1995. Ένα Έμπειρο Σύστημα Υποστήριξης Αποφάσεων Μάρκετινγκ: Μεθοδολογία Υποστήριξης και Ολοκληρωμένη Αρχιτεκτονική. Διδακτορική Διατριβή Πολυτεχνείο Κρήτης, Χανιά.
- Ματσατσίνης Ν., 2022 Συστήματα Υποστήριξης Αποφάσεων. Αθήνα: Νέες Τεχνολογίες.
- Μπούρου, Δ., (2016) Στατιστική Ανάλυση Δεδομένων με Ακραίες και Ελλιπείς Τιμές (Αδημοσίευτη Μεταπτυχιακή Εργασία). Πανεπιστημίου Πατρών, Πάτρα.
- Πεπινίδης, Σ., Λαζάρου, Α., Χατζηδάκης, Ι., (2015) Λογισμικό εξόρυξης δεδομένων WEKA: Αναλυτικό εγχειρίδιο χρήσης και εφαρμογές. Τεχνολογικό Εκπαιδευτικό Ίδρυμα Δυτικής Ελλάδας, Πάτρα.
- Σίσκος, Ι. (2008), Μοντέλα Αποφάσεων, Νέες Τεχνολογίες.
- Τσακανίκας, Α., Danchev, S., Γιωτόπουλος, Γ., Κόρρα, Ε., Παύλου, Γ., (2014). Υιοθέτηση των ΤΠΕ και ψηφιακή ανάπτυξη στην Ελλάδα. Ίδρυμα Οικονομικών και Βιομηχανικών Ερευνών (IOBE).
- Τσαφράκης, Σ., (2007) Ανάπτυξη ευρετικού αλγορίθμου με χρήση πολλαπλών κριτηρίων για την επιλογή μοντέλων προσωπικής επιλογής καταναλωτή στο μάρκετινγκ σε πρόβλημα ανάπτυξης νέων προϊόντων. Πολυτεχνείο Κρήτης, Χανιά.

Ξένη

- An Introduction to Expectation-Maximization. Ανακτήθηκε από https://courses.csail.mit.edu/6.867/wiki/images/b/b5/Em_tutorial.pdf
- Canopy with k-means clustering algorithm for big data analytics, (2021). Ανακτήθηκε από <https://aip.scitation.org/doi/pdf/10.1063/5.0042398>
- DBSCAN Clustering Algorithm in Machine Learning, (2022). Ανακτήθηκε από <https://www.kdnuggets.com/2020/04/DBScan-clustering-algorithm-machine-learning.html>

- Dempster, A.P., N.M. Laird, D.B. Rubin (1977), Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 39, no. 1, pp. 1–38.
- Digital Economy and Society Index (DESI), (2022). Ανακτήθηκε από <https://digital-strategy.ec.europa.eu/en/policies/desi>
- El Falah, Z., Rafalia, N., Abouchabaka, J., (2021). An Intelligent Approach for Data Analysis and Decision Making in Big Data: A Case Study on E – commerce Industry. *International Journal of Advances Computer Science and Applications*, 12(7), 723-727.
- Ester, M., H-P. Kriegel, J. Sander, X. Xu (1996), A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Knowledge Discovery and Data Mining*, pp. 226–231.
- Ester, M. (2009), Density-based Clustering, in: Liu L., Özsu M.T. (eds), *Encyclopedia of Database Systems*. Springer, Boston, MA. (https://doi.org/10.1007/978-0-387-39940-9_605).
- Expectation-Maximization (2022, April 10). Ανακτήθηκε από https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm
- Greco, S., M. Ehr Gott, J. Figueira (eds.), *Multiple Criteria Decision Analysis, - State of the Art – Surveys* (2nd Edition), *International Series in Operations Research and Management Science*, Springer.
- Hanselman, D., Littlefield, B., (2005). *Μάθε το MATLAB 7*. Αθήνα: Εκδόσεις Κλειδάριθμος.
- Hierarchical Clustering (2022, February 28). Ανακτήθηκε από https://en.wikipedia.org/wiki/Hierarchical_clustering
- Hyun Kang, (2013). The prevention and handling of the missing data, *Korean Journal of Anesthesiology*, 64(5), 402-406
- Jacquet-Lagrange, E. and Y. Siskos (1982). Assessing a set of additive utility functions for multicriteria decision making: The UTA method, *European Journal of Operational Research*, 10 (2), 151–164.
- J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*", Berkeley, University of California Press, 1:281-297
- Lloyd's algorithm(2022, October 31). Ανακτήθηκε από https://en.wikipedia.org/wiki/Lloyd%27s_algorithm
- Matsatsinis, N.F., Y. Siskos., (1999). MARKEX: An intelligent decision support system for product development decisions, *European Journal of Operational Research*, 113(2), pp. 336-354.
- McCallum, A.; Nigam, K.; Ungar L.H. (2000) "Efficient Clustering of High Dimensional Data Sets with Application to Reference Matching", *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 169-178 (doi:10.1145/347090.347123).
- Rousseeuw, P.J. (1987). "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". *Υπολογιστικά και Εφαρμοσμένα Μαθηματικά*. 20: 53–65. doi: 10.1016/0377-0427(87)90125-7
- 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.
- Silhouette (clustering) (2021, December 2). Ανακτήθηκε από [https://en.wikipedia.org/wiki/Silhouette_\(clustering\)](https://en.wikipedia.org/wiki/Silhouette_(clustering)).
- Siskos, Y. and N.F. Matsatsinis (1993), A DSS for market analysis and new product design, *Journal of Decision Systems*, vol. 2, no. 1, pp. 35-60.
- Siskos, J., D. Yannakopoulos (1985), UTASTAR: An ordinal regression method for building additive value functions, *Investigação Operational*, vol. 5, no. 1, pp. 39-53.
- Tan, P. N., M. Steinbach, V. Kumar (2014), *Introduction to Data Mining*, Pearson.
- Wang, H., Wang, J., Zhong, Z., (2020). Research on Precision Marketing Strategy Based on Cluster Analysis Algorithm, *International Conference on E-Commerce and Internet Technology (ECIT)*, pp. 208-211, doi: 10.1109/ECIT50008.2020.00054.
- Ziafat, H., Shakeri, M., (2014). Using Data Mining Techniques in Customer Segmentation. *International Journal of Engineering Research and Applications*, 4(9), 70-79. Ανακτήθηκε από IJERA.

Παράρτημα

1. Ερωτηματολόγιο

- 1. Ποιο είναι το φύλο σας;**
 - α) Άνδρας
 - β) Γυναίκα
- 2. Ποια είναι η ηλικία σας;**
- 3. Από πόσα μέλη αποτελείται συνολικά η οικογένειά σας για την οποία κάνετε την αγορά του γιαουρτιού;**
- 4. Όσον αφορά το μορφωτικό επίπεδο σε ποια κατηγορία ανήκετε;**
 - α) Απόφοιτος Λυκείου/Ι.Ε.Κ
 - β) Φοιτητής Α.Ε.Ι/Τ.Ε.Ι
 - γ) Πτυχιούχος Α.Ε.Ι/Τ.Ε.Ι
 - δ) Κάτοχος Μεταπτυχιακού/Διδακτορικού
 - ε) Άλλο
- 5. Ποιο είναι το επάγγελμά σας;**
 - α) Αγρότης
 - β) Άνεργος
 - γ) Αυτοαπασχολούμενος
 - δ) Δημόσιος Υπάλληλος
 - ε) Εισοδηματίας
 - στ) Ελεύθερος Επαγγελματίας
 - ζ) Ιδιωτικός Υπάλληλος
 - η) Συνταξιούχος
 - θ) Φοιτητής
 - ι) Άλλο
- 6. Πόσο συχνά αγοράζετε γιαούρτι;**
 - α) 2 με 3 φορές την εβδομάδα
 - β) 1 φορά την εβδομάδα
 - γ) 2 με 3 φορές τον μήνα
 - δ) 2 με 3 φορές το εξάμηνο
- 7. Ποιο είναι το είδος του γιαουρτιού που προτιμάτε κυρίως να καταναλώνετε;**
 - α) Αγελαδινό
 - β) Κατσικίσιο
 - γ) Πρόβειο
 - δ) Άλλο
- 8. Ποιον τύπο γιαουρτιού προτιμάτε να καταναλώνετε συνήθως;**

- α) Στραγγιστό (Είναι το προϊόν που προκύπτει από την στράγγιση του γιαουρτιού μέσα από υφασμάτινους σάκους)
- β) Παραδοσιακό (Δημιουργία υμένιου/πέτσας)
- γ) Set type (το γάλα ομογενοποιείται και γι αυτό δεν σχηματίζεται υμένιο/πέτσα)
- δ) Επιδόρπιο γιαουρτιού με κομμάτια φρούτων
- ε) Γιαούρτι με προβιοτικά
- στ) Αναμιγμένο(το γάλα επωάζεται σε δεξαμενές και συσκευάζεται σε ερμητικά κλειστούς περιέκτες)
- ζ) Άλλο

9. Από πού αγοράζετε συνήθως το συγκεκριμένο προϊόν; (πολλαπλές απαντήσεις)

- α) Supermarket
- β) Online
- γ) Λιανοπωλητές
- δ) Χονδρέμπορους
- ε) Άλλο

10. Θα σας παρακαλούσαμε να περιγράψετε τα προϊόντα (όνομα εταιρείας), τα είδη (αγελαδινό, κατσικίσιο, πρόβειο κτλ), την ποσότητα(ατομική μερίδα, πολυσυσκευασία, οικογενειακή/1Kg) και τη συσκευασία του γιαουρτιού/κεσεδάκι(πλαστικό, πήλινο) που αγοράζετε συνήθως.

11. Θα σας παρακαλούσαμε να περιγράψετε τους λόγους κατανάλωσης του γιαουρτιού(π.χ δίαιτα, συνήθεια,...), την συχνότητα κατανάλωσης, όπως επίσης και που το χρησιμοποιείτε το προϊόν αυτό(π.χ στα γλυκά, στο φαγητό,...).

12. Σημειώστε, στο αντίστοιχο τετράγωνο, το βαθμό σημαντικότητας που αποδίδετε σε κάθε ένα από τα παρακάτω κριτήρια-χαρακτηριστικά, κατά τη διαδικασία επιλογής του γιαουρτιού σας, χρησιμοποιώντας την κλίμακα 0-10 (Όπου με 0 σημειώνετε όταν δεν το λαμβάνετε καθόλου υπόψη ενώ με 10 αυτό με την μέγιστη σημαντικότητα)

	1	2	3	4	5	6	7	8	9	10
Επωνυμία Εταιρείας (Brand Name)										
Γεύση										
Προέλευση Γάλακτος(τόπος παραγωγής)										
Υφή										
Άρωμα										
Συσκευασία										
Τιμή										

13. Θα σας παρακαλούσαμε να μας αναφέρετε αν, κατά την αγορά του γιαουρτιού, υπάρχει κάποιο επιπλέον κριτήριο που λαμβάνετε υπόψη σας. Ποιο είναι το κριτήριο αυτό, και πόσο το βαθμολογείτε βάσει της προηγούμενης κλίμακας;
14. Ποια είναι η γνώμη σας για το κριτήριο 'Επωνυμία εταιρίας (Brand name)' των παρακάτω προϊόντων γιαουρτιού;

	Πολύ κακή	Κακή	Αδιάφορη	Καλή	Πολύ καλή
Κρι-Κρι αγελαδινό στραγγιστό 'αγαπημένο'					
Δωδώνη Ηπειρωτικό αγελαδινό στραγγιστό					
ΔΕΛΤΑ διπλοστραγγιστό 'Μικρές οικογενειακές Φάρμες'					
ΦΑΓΕ Total					
ΝΟΥΝΟΥ αγελαδινό στραγγιστό γιαούρτι					
Όλυμπος κατσικίσιο γιαούρτι					
Ροδόπη κατσικίσιο γιαούρτι					

15. Ποια είναι η γνώμη σας για το κριτήριο 'Γεύση' των παρακάτω προϊόντων γιαουρτιού;

	Άσχημη	Περίεργη	Αδιάφορη	Ωραία	Πολύ ωραία
Κρι-Κρι αγελαδινό στραγγιστό 'αγαπημένο'					

Δωδώνη Ηπειρωτικό αγελαδινό στραγγιστό					
ΔΕΛΤΑ διπλοστραγγιστό 'Μικρές οικογενειακές Φάρμες'					
ΦΑΓΕ Total					
ΝΟΥΝΟΥ αγελαδινό στραγγιστό γιαούρτι					
Όλυμπος κατσικίσιο γιαούρτι					
Ροδόπη κατσικίσιο γιαούρτι					

16. Ποια είναι η γνώμη σας για το κριτήριο 'Προέλευση γάλακτος(τόπος παραγωγής) των παρακάτω προϊόντων γιαουρτιού;

	Πολύ κακή	Κακή	Αδιάφορη	Καλή	Πολύ καλή
Κρι-Κρι αγελαδινό στραγγιστό 'αγαπημένο'					
Δωδώνη Ηπειρωτικό αγελαδινό στραγγιστό					
ΔΕΛΤΑ διπλοστραγγιστό 'Μικρές οικογενειακές Φάρμες'					
ΦΑΓΕ Total					
ΝΟΥΝΟΥ αγελαδινό στραγγιστό γιαούρτι					

Όλυμπος κατσικίσιο γιαούρτι					
Ροδόπη κατσικίσιο γιαούρτι					

17. Ποια είναι η γνώμη σας για το κριτήριο 'Υφή' των παρακάτω προϊόντων γιαουρτιού;

	Άσχημη	Περίεργη	Αδιάφορη	Ωραία	Πολύ ωραία
Κρι-Κρι αγελαδινό στραγγιστό 'αγαπημένο'					
Δωδώνη Ηπειρωτικό αγελαδινό στραγγιστό					
ΔΕΛΤΑ διπλοστραγγιστό 'Μικρές οικογενειακές Φάρμες'					
ΦΑΓΕ Total					
ΝΟΥΝΟΥ αγελαδινό στραγγιστό γιαούρτι					
Όλυμπος κατσικίσιο γιαούρτι					
Ροδόπη κατσικίσιο γιαούρτι					

18. Ποια είναι η γνώμη σας για το κριτήριο 'Αρωμα' των παρακάτω προϊόντων γιαουρτιού;

	Δυσάρεστο	Περίεργο	Αδιάφορο	Ευχάριστο	Πολύ Ευχάριστο
Κρι-Κρι αγελαδινό					

στραγγιστό 'αγαπημένο'					
Δωδώνη Ηπειρωτικό αγελαδινό στραγγιστό					
ΔΕΛΤΑ διπλοστραγγιστό 'Μικρές οικογενειακές Φάρμες'					
ΦΑΓΕ Total					
ΝΟΥΝΟΥ αγελαδινό στραγγιστό γιαούρτι					
Όλυμπος κατσικίσιο γιαούρτι					
Ροδόπη κατσικίσιο γιαούρτι					

19. Ποια είναι η γνώμη σας για το κριτήριο 'Συσκευασία' των παρακάτω προϊόντων γιαουρτιού;

	Πολύ κακή	Κακή	Αδιάφορη	Καλή	Πολύ καλή
Κρι-Κρι αγελαδινό στραγγιστό 'αγαπημένο'					
Δωδώνη Ηπειρωτικό αγελαδινό στραγγιστό					
ΔΕΛΤΑ διπλοστραγγιστό 'Μικρές οικογενειακές Φάρμες'					
ΦΑΓΕ Total					
ΝΟΥΝΟΥ αγελαδινό					

στραγγιστό γιαούρτι					
Όλυμπος κατσικίσιο γιαούρτι					
Ροδόπη κατσικίσιο γιαούρτι					

20. Ποια είναι η γνώμη σας για το κριτήριο 'Τιμή' των παρακάτω προϊόντων γιαουρτιού;

	Πολύ υψηλή	Υψηλή	Λογική	Χαμηλή	Πολύ χαμηλή
Κρι-Κρι αγελαδινό στραγγιστό 'αγαπημένο'					
Δωδώνη Ηπειρωτικό αγελαδινό στραγγιστό					
ΔΕΛΤΑ διπλοστραγγιστό 'Μικρές οικογενειακές Φάρμες'					
ΦΑΓΕ Total					
ΝΟΥΝΟΥ αγελαδινό στραγγιστό γιαούρτι					
Όλυμπος κατσικίσιο γιαούρτι					
Ροδόπη κατσικίσιο γιαούρτι					

21. Σας παρακαλούμε, λαμβάνοντας υπόψη σας όλες τις παραπάνω εκτιμήσεις, να κατατάξετε τα παρακάτω προϊόντα γιαουρτιού, κατά σειρά προτίμησης-αγοράς. Ποιο ή ποια από τα παρακάτω προϊόντα γιαουρτιού θα επιλέγατε για να αγοράσετε (χρησιμοποιήστε κλίμακα από το 1 έως το 7; όπου με 1 σημειώνουμε την πρώτη προτίμηση, με 2 την επόμενη, κοκ Υπάρχει

δυνατότητα έκφρασης της ίδιας σειράς προτίμησης με απόδοση της ίδιας τιμής);

	1 επιλογή	2 στην σειρά	3 στην σειρά	4 ως επιλογή	6 ως επιλογή	Τελευταίο σαν επιλογή
Κρι-Κρι αγελαδινό στραγγιστό 'αγαπημένο'						
Δωδώνη Ηπειρωτικό αγελαδινό στραγγιστό						
ΔΕΛΤΑ διπλοστραγγιστό 'Μικρές οικογενειακές Φάρμες'						
ΦΑΓΕ Total						
ΝΟΥΝΟΥ αγελαδινό στραγγιστό γιαούρτι						
Όλυμπος κατσικίσιο γιαούρτι						
Ροδόπη κατσικίσιο γιαούρτι						

22. Θα σας παρακαλούσαμε να εκφράσετε τη γνώμη σας για τα προϊόντα γιαουρτιού.

2. Αποτελέσματα UTASTAR για τις ολικές χρησιμότητες των εναλλακτικών

KRIKRI	DODONI	DELTA	FAGE	NOUNOU	OLYMPOS	RODOPI
0,81437	0,86427	0,81437	0,81437	0,81437	0,86427	0,86427
0,8429	0,8429	0,99997	0,99999	0,83576	1	0,99282
0,85009	0,80007	0,75	0,70009	0,65009	0,60009	0,60009
0,8786	0,77864	0,97857	0,97859	0,97859	0,77864	0,77864
0,70713	0,85002	0,56427	0,89999	0,31426	0,25713	0,10714
0,90006	1	0,85007	0,95006	0,80006	0,74291	0,74291
0,80007	0,80007	0,80007	1	0,95009	0,80007	0,80007

1	0,94999	1	1	0,74294	0,05	0
0,79291	1	0,90006	0,80006	0,79291	1	1
0,67862	0,67862	0,67862	0,7714	0,67862	0,55006	0,55006
1	0,7929	0,79292	0,95	0,79289	0,90002	0,7929
0,75003	0,75003	0,95001	1	0,85001	0,90001	0,7001
0,90009	0,90009	0,80009	1	0,90009	0,80009	0,7501
0,80007	0,80007	0,95009	1	0,80007	0,80007	0,80007
0,82146	0,72137	0,77146	0,87146	0,62147	0,67146	0,57146
1	1	0,85007	0,85007	0,85007	0,85007	0,85007
0,90719	0,80007	0,90719	1	0,90004	0,73579	0,73579
0,60016	0,80009	0,7501	0,90002	0,85009	0,65017	0,55017
0,79294	0,85723	0,85723	0,90009	0,79294	1	0,95007
0,7501	0,7501	0,90001	1	0,85009	0,73581	0,73581
0,50004	0,55004	0,85004	0,80007	1	0,42147	0,42147
0,58578	0,68578	0,7357	0,83577	0,78579	0,63578	0,53578
0,8643	0,81428	0,76432	0,66432	0,71434	0,61432	0,56435
0,89296	0,6359	0,84296	0,7501	0,7501	0,64306	0,6359
0,85006	0,85001	0,90006	0,85006	1	0,80006	0,75006
0,76436	0,76436	0,90007	0,79293	0,79293	0,95004	1
0,80004	0,85004	0,85003	0,95007	1	0,75006	0,70004
0,49292	0,5429	0,84292	0,99286	0,59289	0,34289	0,29289
0,81433	0,81431	0,90003	1	0,95006	0,75004	0,70006
0,77148	0,87141	0,82146	0,82146	0,77148	0,77148	0,77148
0,79284	0,87138	0,74285	0,6429	0,69286	0,89284	0,87138
0,75004	0,90006	0,75004	0,90006	0,90006	1	0,75007
0,97146	0,95008	0,90009	0,78584	0,85009	1	0,97146
0,95007	0,85009	0,75723	1	0,80008	0,75723	0,70008
0,71438	0,71438	0,71438	0,95007	0,71438	0,85013	0,90006
1	0,95003	0,80004	0,75006	0,70006	0,8572	0,8572
0,75006	0,70004	0,95004	0,80005	1	0,55007	0,50004
0,90006	0,80004	0,95007	1	0,85004	0,74291	0,74291
0,30024	0,5002	0,45014	0,4002	0,4002	0,6002	0,5502
0,7929	1	1	0,84287	0,84287	0,99284	0,84289
0,97142	1	0,93566	0,89287	0,85715	0,78573	0,78573
0,72856	0,72856	0,62861	0,89284	0,67854	0,89284	0,82856
0,85	0,75003	0,82141	1	1	0,94999	0,82141
1	1	1	1	1	0,90004	0,90004
1	1	0,8286	0,98569	0,93574	0,82859	0,82859
0,75721	0,77861	0,85714	1	0,8286	0,78575	0,72865
0,79292	0,79292	0,84287	0,95003	0,90007	1	0,84287
0,89999	0,95	1	1	0,9	0,6214	0,57146
0,80006	0,80006	0,80006	1	0,80006	0,95006	0,80006

0,97144	0,48576	0,73576	0,58576	0,53572	0,3929	0,3929
0,70717	0,65721	1	0,55719	0,85004	0,50719	0,45721
0,8286	0,8286	0,8286	0,99999	1	0,8286	0,8286
0,92856	0,82863	0,95004	0,95004	0,82861	0,57146	0,43576
0,6072	0,65724	0,7572	0,89289	0,70721	0,5572	0,5072
0,9	0,29999	0,29999	0,85005	0,39999	0,25	0,2
0,8929	0,8929	1	1	0,8929	1	0,8929
0,73579	0,80008	0,85008	0,93579	0,93579	1	0,73579
0,93571	0,98569	0,88574	0,83576	0,83576	0,90004	1
0,87863	0,82864	0,7072	0,92861	0,7072	0,97857	0,77864
0,61434	0,51434	0,77149	0,97859	0,56436	0,92146	0,51434
1	0,95007	1	1	0,85009	0,7858	0,7858
0,80009	0,84999	0,84999	1	1	0,83571	0,83571
0,80001	0,80716	1	0,70002	0,94993	0,80716	0,80001
0,95004	0,95004	0,95004	1	1	0,90004	0,90004
0,79293	0,79293	0,9072	0,99286	0,78579	0,90007	0,79293
0,95001	0,95001	0,90001	0,99999	0,90001	1	0,99999
0,85004	0,80004	1	0,95004	1	0,90006	0,75011
1	0,80007	0,80004	0,90716	0,80004	0,90718	0,80004
0,80007	0,80007	0,80007	0,95003	0,80007	1	0,80007
0,90004	0,90004	0,90003	1	0,95002	0,95	0,95001
1	1	0,78572	0,95	0,62145	0,99999	0,95003
0,86432	0,86432	0,80004	0,7429	0,7429	1	0,95
0,75722	0,92152	0,75722	0,75722	0,75722	1	0,92152
0,75007	0,75007	0,90007	1	0,95006	0,85006	0,75007
0,75009	0,86436	0,75009	0,86436	0,90007	1	0,75009
0,89291	0,84293	0,99286	0,99286	0,99286	0,79289	0,74291
0,78577	0,68577	0,62865	0,83577	0,52873	0,73579	0,57865
0,80001	0,89999	0,73574	0,84999	0,57859	0,89996	0,80002
0,80001	0,85002	0,84998	1	1	0,95002	0,75003
0,77151	0,77151	0,90009	1	0,90009	0,77151	0,77151
0,90001	0,99999	0,90001	0,90001	0,90001	1	1
0,99996	0,80009	0,95007	1	0,85004	0,75007	0,70009
0,89294	0,94294	0,84294	0,79294	0,74296	0,84301	0,94294
0,94994	0,9	0,85004	1	0,85	0,75	0,6501
0,6572	0,6572	0,76437	0,90714	0,71434	0,56433	0,51434
0,87146	1	0,86432	0,99286	0,82147	0,75717	0,75717
0,74291	1	0,91434	0,91434	0,74291	0,8072	0,80717
0,44284	0,90713	0,49286	0,56427	0,56426	0,64279	0,90713
0,8143	0,8143	0,8143	1	0,8143	1	0,8143
0,78577	0,78577	0,95004	1	0,90006	0,78577	0,78577
0,84999	0,93572	1	0,93572	1	0,80001	0,80002

0,85007	0,85007	0,90006	1	0,95006	0,85007	0,85007
0,76437	0,76437	0,95007	1	0,90009	0,76437	0,76437
0,99999	1	0,85004	0,85004	0,85004	1	1
0,9001	0,80003	0,85009	0,95007	0,75009	0,7501	0,65009
0,72147	0,82147	1	0,95	0,90006	0,82147	0,72147
0,75003	1	0,80004	0,85006	0,50004	0,56433	0,56433
0,63574	0,68574	0,53574	0,53574	0,53574	0,73572	0,78569
0,69289	0,89284	0,74289	0,79289	0,6429	0,89284	0,89284
0,87147	0,92146	0,87145	0,97144	0,82152	0,97146	0,97146
0,85719	0,85719	0,94999	1	0,80004	0,42146	0,42146
0,82857	0,83576	0,78571	0,95714	0,73574	0,93571	1
0,95006	1	0,63573	0,53577	0,58577	1	1
0,94294	0,94294	0,56441	0,61434	0,66433	0,94294	0,94294
0,67149	0,82149	0,72144	0,87143	0,67149	0,82149	0,82149
0,70717	0,80717	0,75716	0,90712	0,65716	0,85714	0,85714
0,84289	0,89286	0,99286	0,89286	0,99286	0,94286	0,89282
0,78579	0,83577	0,83577	0,8857	0,73579	0,57149	0,57149
0,84646	0,99644	0,99638	0,89644	0,84644	0,94643	0,94644
0,74296	0,94294	0,89296	1	0,84296	0,74296	0,74296
0,65006	0,75004	0,70004	0,79997	0,65004	0,60004	0,60004
0,60003	0,69287	0,89286	0,69287	0,7	0,55003	0,50006
0,9572	0,9572	0,78577	0,80003	0,94289	0,9572	1
0,83573	0,98573	0,78572	0,63576	0,83573	0,93569	0,88574
0,99999	1	0,95	0,95	0,95	1	0,99999
0,77143	0,77143	0,77143	0,67152	0,5715	0,5715	0,5715
0,80004	0,85007	0,80004	1	0,90717	0,90717	0,75004
0,72143	0,82143	0,7714	0,82143	0,72146	0,87142	0,87142
0,83572	0,93572	0,83573	0,88572	0,78577	0,99286	0,99286
0,85005	0,94999	0,90001	0,94999	0,90001	1	1
0,76444	0,76444	0,66454	0,76444	0,71444	0,44293	0,44293
0,80003	0,94996	0,85713	0,89999	0,85714	1	1
0,90001	0,85004	0,75003	0,74287	0,7429	0,94999	1
0,90001	0,8501	0,7501	0,6501	0,70011	0,8001	1
0,75719	0,75719	0,73577	0,91434	0,53583	0,58577	0,63576
0,9501	1	0,9001	0,9501	0,9001	0,8501	0,8501
0,66433	0,76424	0,61433	0,71433	0,61433	0,56433	0,51434
0,56434	0,80721	0,75721	0,85722	0,61434	0,51433	0,46443
0,90009	0,90009	0,90009	1	0,95006	0,85007	0,85007
0,80006	0,80006	1	1	1	0,85002	0,80006
0,79291	0,89289	0,79291	0,84293	0,79294	0,74294	0,74294
0,67146	0,8214	0,67861	0,77141	0,67861	0,87143	0,87143
0,67874	0,87864	0,72867	0,77874	0,62877	0,92864	0,92864

0,61433	0,81435	0,66425	0,86433	0,61432	0,56432	0,56432
0,75004	0,84286	0,75004	0,89286	0,6929	0,6429	0,59292
0,75007	0,90007	0,80007	0,85006	0,70008	1	0,95007
0,92148	0,9714	0,82149	0,97147	0,87148	0,77148	0,77148
0,72859	0,77861	0,72859	0,82855	0,72859	0,87857	0,87857
0,57153	0,62153	0,67153	0,52153	0,4716	0,7715	0,72153
0,70003	0,79997	0,70003	0,84997	0,70003	0,89999	0,89999
0,80002	0,95001	0,85001	0,90002	0,75005	1	1
0,70009	0,8358	0,70009	0,7858	0,63587	0,88581	0,93577
0,72146	0,82146	0,82144	0,77147	0,67147	0,87144	0,87144
0,89285	0,89286	0,84286	0,94284	0,79292	0,99287	0,99287
0,80004	0,85007	0,74293	0,90007	0,74293	1	0,95007
0,89288	0,94996	0,80002	1	0,89288	0,75001	0,70001
0,63579	0,73575	0,63579	0,78574	0,63579	0,83575	0,98929

3. Αποτελέσματα UTASTAR για τα βάρη των κριτηρίων

Brand_name	Taste	Milk_origin	Texture	Odor	Packaging	Price
0,17137	0,13573	0,14286	0,13573	0,14286	0,13573	0,13573
0,27134	0,12144	0,12144	0,12144	0,12144	0,12144	0,12144
0,1786	0,22139	0,11424	0,07149	0,10716	0,09289	0,21424
0,21426	0,17854	0,13571	0,1143	0,1143	0,12859	0,1143
0,22853	0,14286	0,1286	0,12144	0,12859	0,14997	0,10001
0,23566	0,17143	0,15	0,1143	0,10716	0,1143	0,10716
0,26421	0,14286	0,1143	0,13573	0,1143	0,1143	0,1143
0,27134	0,12144	0,12144	0,12144	0,12144	0,12144	0,12144
0,19994	0,18571	0,14286	0,12859	0,1143	0,1143	0,1143
0,1143	0,26421	0,1143	0,13573	0,14286	0,1143	0,1143
0,12857	0,2	0,14286	0,17853	0,1143	0,12144	0,1143
0,2785	0,12144	0,10001	0,2	0,10001	0,10001	0,10001
0,15714	0,2428	0,10716	0,12144	0,15714	0,10716	0,10716
0,17141	0,14286	0,14286	0,14286	0,14286	0,14286	0,1143
0,21424	0,10721	0,08573	0,1357	0,12861	0,09286	0,23564
0,12144	0,12144	0,12144	0,27134	0,12144	0,12144	0,12144
0,1143	0,22851	0,10716	0,12859	0,20713	0,10716	0,10716
0,24994	0,12859	0,10001	0,18573	0,17853	0,0786	0,0786
0,22136	0,15716	0,16429	0,1143	0,1143	0,1143	0,1143
0,25707	0,10716	0,10716	0,10716	0,20714	0,10716	0,10716
0,19283	0,17144	0,10716	0,12859	0,17139	0,12144	0,10716
0,06431	0,27859	0,21421	0,14289	0,17137	0,06431	0,06431
0,13571	0,20714	0,12856	0,11429	0,12144	0,20713	0,08573

0,09289	0,34279	0,09289	0,19279	0,09289	0,09289	0,09289
0,25231	0,17854	0,10717	0,10717	0,12382	0,12382	0,10717
0,15709	0,12144	0,16427	0,1143	0,1143	0,1143	0,2143
0,19994	0,10716	0,10001	0,18571	0,17143	0,12144	0,1143
0,15713	0,14286	0,12144	0,12144	0,17856	0,16429	0,11429
0,15	0,18572	0,10001	0,24279	0,12144	0,10001	0,10002
0,12859	0,15711	0,16425	0,14289	0,14285	0,13572	0,12859
0,16431	0,1143	0,16425	0,12858	0,16428	0,15712	0,10716
0,12144	0,12144	0,18569	0,14286	0,14286	0,14286	0,14286
0,1143	0,12143	0,22855	0,12143	0,1143	0,17142	0,12857
0,20714	0,23564	0,10001	0,12859	0,10001	0,12859	0,10001
0,22139	0,13573	0,10716	0,12146	0,13573	0,12146	0,15709
0,27856	0,10001	0,12853	0,10001	0,19286	0,10001	0,10001
0,16784	0,24286	0,11786	0,1143	0,13927	0,11785	0,10001
0,22136	0,12858	0,12144	0,15716	0,15714	0,10716	0,10716
0,12141	0,05004	0,23571	0,0857	0,28571	0,03576	0,18566
0,1143	0,2428	0,1143	0,13573	0,16427	0,1143	0,1143
0,2	0,13573	0,19994	0,1143	0,1143	0,12144	0,1143
0,12144	0,11431	0,21424	0,16423	0,15001	0,12144	0,11431
0,19283	0,18569	0,10716	0,14286	0,12859	0,13573	0,10716
0,18567	0,14286	0,12859	0,14286	0,14286	0,12859	0,12859
0,14286	0,19282	0,07144	0,20714	0,12859	0,12858	0,12858
0,17853	0,12859	0,12859	0,19283	0,12859	0,12144	0,12144
0,2143	0,1571	0,1143	0,1143	0,12144	0,16426	0,1143
0,1714	0,14284	0,14289	0,13573	0,14284	0,15	0,1143
0,21421	0,13573	0,13573	0,13573	0,14286	0,1143	0,12144
0,17143	0,16426	0,1714	0,14287	0,12859	0,1143	0,10716
0,29286	0,10716	0,07855	0,12857	0,19283	0,10001	0,10001
0,12144	0,22851	0,14286	0,12144	0,14286	0,12144	0,12144
0,16427	0,13573	0,1143	0,16429	0,13573	0,12143	0,16426
0,15711	0,31427	0,08573	0,10716	0,10716	0,14284	0,08573
0,19284	0,15709	0,12858	0,13573	0,13573	0,13573	0,1143
0,12858	0,22853	0,12858	0,12858	0,12858	0,12858	0,12858
0,24278	0,12859	0,10716	0,15716	0,15001	0,10716	0,10716
0,12144	0,22139	0,12146	0,12144	0,17141	0,12144	0,12144
0,20709	0,15714	0,12857	0,13573	0,11431	0,15	0,10716
0,22142	0,16426	0,12144	0,15713	0,1143	0,11429	0,10716
0,19279	0,20714	0,1143	0,13573	0,12144	0,1143	0,1143
0,1143	0,24287	0,1143	0,18563	0,1143	0,1143	0,1143
0,11429	0,2143	0,10716	0,10716	0,10716	0,24279	0,10716
0,19996	0,14286	0,12858	0,13572	0,13572	0,12858	0,12858
0,29277	0,13573	0,1143	0,1143	0,1143	0,1143	0,1143

0,19281	0,14286	0,12859	0,14286	0,13572	0,12859	0,12858
0,15709	0,15716	0,15	0,10717	0,21424	0,10717	0,10717
0,17854	0,12859	0,1143	0,12858	0,2214	0,1143	0,1143
0,15713	0,12144	0,2428	0,13573	0,1143	0,1143	0,1143
0,17855	0,12859	0,12858	0,17854	0,12859	0,12858	0,12858
0,12857	0,12144	0,22142	0,14285	0,12144	0,14284	0,12144
0,22135	0,16429	0,17144	0,1143	0,1143	0,10716	0,10716
0,1928	0,14286	0,13573	0,13573	0,13573	0,14286	0,1143
0,19286	0,24279	0,10716	0,1143	0,10716	0,12859	0,10716
0,27134	0,15001	0,10716	0,12859	0,10716	0,12859	0,10716
0,10716	0,24995	0,10716	0,13572	0,13572	0,10716	0,15714
0,2429	0,07147	0,16423	0,06433	0,07861	0,16423	0,21423
0,15717	0,16429	0,15001	0,16426	0,1143	0,10716	0,14282
0,21428	0,10716	0,10715	0,12859	0,15713	0,17854	0,10715
0,22853	0,14286	0,1143	0,14286	0,14286	0,1143	0,1143
0,17141	0,14286	0,14286	0,14286	0,14286	0,12859	0,12858
0,2214	0,1643	0,09289	0,09293	0,19281	0,09289	0,14279
0,15001	0,11434	0,10717	0,10716	0,20708	0,15715	0,15709
0,22853	0,19999	0,11434	0,09287	0,16424	0,09287	0,10716
0,20714	0,15001	0,14287	0,17136	0,12144	0,09287	0,1143
0,23567	0,17143	0,11429	0,13573	0,11429	0,11429	0,11429
0,19286	0,12859	0,10716	0,1643	0,19278	0,10716	0,10716
0,19997	0,12144	0,13573	0,12144	0,13573	0,1143	0,17139
0,18569	0,14286	0,14286	0,14286	0,14286	0,12144	0,12144
0,23566	0,17141	0,1143	0,13573	0,1143	0,1143	0,1143
0,15	0,22852	0,12859	0,13573	0,11429	0,12858	0,11429
0,19993	0,17143	0,12144	0,12859	0,12859	0,12859	0,12144
0,20707	0,17857	0,1143	0,12859	0,12859	0,12859	0,1143
0,18569	0,14286	0,14286	0,14286	0,14286	0,12144	0,12144
0,28566	0,22856	0,08574	0,0929	0,08574	0,08574	0,13566
0,17854	0,12859	0,12859	0,19287	0,1571	0,10716	0,10716
0,20709	0,12144	0,12501	0,17141	0,13573	0,10001	0,1393
0,22858	0,14283	0,12859	0,15711	0,1143	0,1143	0,11429
0,2285	0,17857	0,12859	0,12144	0,10716	0,12144	0,1143
0,12144	0,17856	0,13571	0,15716	0,12859	0,15709	0,12144
0,24994	0,14286	0,12144	0,1143	0,15	0,1143	0,10716
0,19999	0,14286	0,17141	0,1143	0,12859	0,12856	0,1143
0,18567	0,12859	0,12859	0,15001	0,18569	0,1143	0,10716
0,19997	0,12859	0,12859	0,12859	0,12859	0,11431	0,17137
0,17857	0,16426	0,12856	0,13573	0,15714	0,1143	0,12144
0,22143	0,14283	0,15714	0,10716	0,10001	0,12144	0,14999
0,12144	0,12859	0,12859	0,14286	0,22134	0,12147	0,13571

0,1143	0,22136	0,1143	0,19286	0,12859	0,1143	0,1143
0,15359	0,17854	0,16069	0,14284	0,12504	0,11786	0,12144
0,10716	0,3	0,10716	0,1143	0,10716	0,10716	0,15707
0,12143	0,18575	0,11431	0,17141	0,10716	0,14284	0,15709
0,2214	0,17143	0,1143	0,1143	0,10714	0,10716	0,16427
0,12143	0,26427	0,1143	0,1143	0,1571	0,1143	0,1143
0,2	0,12144	0,17143	0,14999	0,11429	0,14284	0,10001
0,17853	0,13573	0,13572	0,14286	0,13572	0,13572	0,13573
0,28563	0,12859	0,1143	0,12859	0,1143	0,1143	0,1143
0,19284	0,12146	0,20712	0,12858	0,14997	0,10001	0,10001
0,15002	0,17141	0,17853	0,12144	0,12144	0,12857	0,12859
0,19999	0,18564	0,12857	0,12861	0,12859	0,1143	0,1143
0,13572	0,17144	0,14998	0,12144	0,12859	0,14998	0,14285
0,08576	0,21433	0,08576	0,20709	0,18566	0,08576	0,13566
0,18567	0,16429	0,12144	0,1143	0,12144	0,17856	0,1143
0,22851	0,12146	0,12144	0,19283	0,10716	0,12144	0,10716
0,20719	0,09289	0,10717	0,14279	0,26421	0,12144	0,06431
0,10714	0,23571	0,16784	0,2	0,10001	0,09644	0,09287
0,15717	0,23566	0,10717	0,12859	0,15707	0,10717	0,10717
0,14287	0,32137	0,10003	0,1143	0,12853	0,10001	0,09289
0,14636	0,31423	0,09289	0,15716	0,09289	0,1036	0,09289
0,18566	0,17856	0,12859	0,13573	0,12859	0,12144	0,12144
0,1143	0,1143	0,16426	0,22139	0,13573	0,13573	0,1143
0,1143	0,28923	0,14644	0,11431	0,1143	0,1143	0,10711
0,19284	0,1714	0,12859	0,1143	0,13573	0,14285	0,1143
0,15713	0,22857	0,18566	0,1571	0,10003	0,08576	0,08576
0,10003	0,33566	0,14993	0,1143	0,10003	0,10003	0,10003
0,07859	0,3071	0,1643	0,12144	0,14999	0,07859	0,1
0,13571	0,15003	0,2	0,10001	0,1143	0,19993	0,10001
0,16427	0,24996	0,12141	0,15716	0,10001	0,11429	0,09289
0,18571	0,15715	0,12857	0,12856	0,12859	0,14998	0,12144
0,05001	0,10716	0,17861	0,13567	0,1786	0,17853	0,17141
0,14999	0,15	0,12144	0,1143	0,23567	0,1143	0,1143
0,20001	0,24994	0,10715	0,10715	0,1143	0,11074	0,11071
0,14285	0,21427	0,16429	0,14996	0,1143	0,12144	0,09289
0,1143	0,15713	0,16427	0,12856	0,20715	0,11429	0,11429
0,15	0,20716	0,12144	0,1143	0,16423	0,12144	0,12144
0,15	0,15719	0,19994	0,10003	0,14286	0,10003	0,14996
0,14285	0,30712	0,09287	0,12859	0,14281	0,09289	0,09287
0,14283	0,20001	0,1643	0,11429	0,15712	0,10716	0,11429