



Πολυτεχνείο Κρήτης

Σχολή Μηχανικών Παραγωγής και Διοίκησης

***Ανάπτυξη πολυκριτήριου συστήματος
συστάσεων αγοράς ηλεκτρονικών προϊόντων
βασισμένο σε μεθόδους εξόρυξης δεδομένων και
ανάλυσης συμπεριφοράς καταναλωτών***

Διπλωματική Εργασία

Καλαϊτζάκης Εμμανουήλ

Επιβλέπων

Ματσατσίνης Νικόλαος, Καθηγητής

Χανιά, Μάιος 2022

ΕΥΧΑΡΙΣΤΙΕΣ

Με αφορμή την ολοκλήρωση της διπλωματικής μου εργασίας, θα ήθελα να ευχαριστήσω όλους όσους με στήριξαν. Ιδιαίτερα, θα ήθελα να ευχαριστήσω την κυρία Καλαφάτη Φωτεινή, για την πολύτιμη βοήθειά της. Τέλος, θα ήθελα να ευχαριστήσω τον καθηγητή και επιβλέπων της παρούσας διπλωματικής εργασίας, κύριο Ματσατσίνη Νικόλαο για την καθοδήγησή του καθ' όλη τη διάρκεια της εργασίας.

ΠΕΡΙΛΗΨΗ

Με την ταχεία πρόοδο του ηλεκτρονικού εμπορίου, η γνώση της συμπεριφοράς των καταναλωτών κατά την αγορά ενός προϊόντος ή υπηρεσίας καθώς και το πως επιλέγουν ένα προϊόν ανάμεσα από ένα σύνολο προϊόντων, είναι σημαντική τόσο για τους υπεύθυνους μάρκετινγκ των επιχειρήσεων όσο και τους πωλητές. Αν τα στελέχη μάρκετινγκ μιας επιχείρησης, μπορούσαν να γνωρίζουν τις προτιμήσεις των καταναλωτών (χαρακτηριστικά, κριτήρια, σημαντικότητες κριτηρίων επιλογής, κλπ.), θα ήταν σε θέση να τμηματοποιήσουν την αγορά με βάση τα προφίλ των καταναλωτών, να σχεδιάσουν νέα προϊόντα, να εφαρμόσουν προσωποποιημένο μάρκετινγκ και να αναπτύξουν πιο αξιόπιστες στρατηγικές μάρκετινγκ. Ένα προϊόν του ηλεκτρονικού εμπορίου που παρουσιάζει πολύ υψηλό ενδιαφέρον και ραγδαία ανάπτυξη, είναι τα κινητά τηλέφωνα. Σε αυτό το χώρο θα εφαρμόσουμε την μεθοδολογία που θα αναπτυχθεί στα πλαίσια αυτής της εργασίας. Στα πλαίσια της εργασίας αρχικά θα αναπτυχθεί η μεθοδολογία συστάσεων η οποία θα βασίζεται στην ανάλυση συμπεριφοράς με χρήση μεθόδων πολυκριτήριας ανάλυσης αποφάσεων και μεθόδους εξόρυξης γνώσης, ενώ εν συνεχεία θα αναπτυχθεί ένα σύστημα συστάσεων για την αγορά ηλεκτρονικών προϊόντων. Για την διαμόρφωση των κριτηρίων, εκτός των πληροφοριών από την έρευνα αγοράς, θα γίνει ανάλυση και των σχολίων των αξιολογήσεων που θα αντληθούν από κατάλληλες ιστοσελίδες, με εφαρμογή της μεθόδου Latent Dirichlet Allocation, για να δημιουργηθούν συστάδες (υπόκριτήρια) από λέξεις και εκφράσεις. Εν συνεχεία θα εφαρμοστούν μέθοδοι πολυκριτήριας αξιολόγησης των εναλλακτικών επιλογών (UTASTAR) για να διαμορφωθεί η σύσταση. Για την συγκέντρωση των απαραίτητων πληροφοριών, θα συλλέξουμε αξιολογήσεις κινητών από σχετικές ιστοσελίδες (πχ. skroutz.gr) καθώς επίσης και από έρευνες αγοράς με ανάπτυξη ειδικού ερωτηματολογίου για συγκέντρωση πρόσθετων πληροφοριών. Αναλύοντας τα δεδομένα αυτά θα είμαστε σε θέση να ανιχνεύσουμε ποια χαρακτηριστικά (κριτήρια) των κινητών προτιμούν οι καταναλωτές, καθώς και το πόσο σημαντικά είναι για αυτούς. Με αυτόν τον τρόπο θα διαμορφώνονται αφενός τα προφίλ των προϊόντων της αγοράς (χαρακτηριστικά προϊόντων) και αφετέρου τα προφίλ διαφόρων καταναλωτών (απαιτήσεις αγοραστών). Τα προφίλ αυτά θα αξιολογούνται και θα ταιριάζουν μεταξύ τους με τη βοήθεια της πολυκριτήριας μεθοδολογίας και θα διαμορφώνεται η σύσταση.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΥΧΑΡΙΣΤΙΕΣ	1
ΠΕΡΙΛΗΨΗ.....	2
ΠΕΡΙΕΧΟΜΕΝΑ.....	3
ΚΕΦΑΛΑΙΟ 1	8
1.1 ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ	8
1.2 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ	8
ΚΕΦΑΛΑΙΟ 2	9
2.1 ΕΙΣΑΓΩΓΗ	9
2.2 ΥΦΙΣΤΑΜΕΝΗ ΚΑΤΑΣΤΑΣΗ.....	9
ΚΕΦΑΛΑΙΟ 3	14
3.1 ΕΙΣΑΓΩΓΗ	14
3.2 ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ	14
3.3 ΒΑΣΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΗΜΑΤΩΝ ΣΥΣΤΑΣΗΣ.....	14
3.3.1 ΣΥΝΕΡΓΑΤΙΚΟ ΦΙΛΤΡΑΡΙΣΜΑ	14
3.3.2 ΦΙΛΤΡΑΡΙΣΜΑ ΜΕ ΒΑΣΗ ΤΟ ΠΕΡΙΕΧΟΜΕΝΟ	16
3.3.3 ΔΗΜΟΓΡΑΦΙΚΟ ΦΙΛΤΡΑΡΙΣΜΑ	18
3.3.4 ΦΙΛΤΡΑΡΙΣΜΑ ΜΕ ΒΑΣΗ ΤΗ ΓΝΩΣΗ	19
3.3.5 ΠΟΛΥΚΡΙΤΗΡΙΑ ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ	19
3.3.6 ΥΒΡΙΔΙΚΑ ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ.....	20
3.4 ΠΡΟΒΛΗΜΑΤΑ ΣΤΑ ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ	21
3.4.1 ΠΡΟΒΛΗΜΑ ΤΗΣ ΨΥΧΡΗΣ ΕΚΚΙΝΗΣΗΣ.....	21
3.4.2 ΑΡΑΙΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ	21
3.4.3 GREY SHEEP (ΔΥΣΚΟΛΟΙ ΠΕΛΑΤΕΣ).....	22
3.4.4 ΣΥΝΩΝΥΜΕΣ ΛΕΞΕΙΣ	22
3.5 ΑΝΙΧΝΕΥΤΗΣ ΙΣΤΟΥ	23
3.6 ΑΠΟΣΠΑΣΗ ΔΕΔΟΜΕΝΩΝ	23
3.7 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ ΜΕΣΩ ΕΡΩΤΗΜΑΤΟΛΟΓΙΩΝ	23
3.8 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ	24
3.9 ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ	26

3.10	LATENT DIRICHLET ALLOCATION	27
3.10.1	ΣΥΝΟΨΗ.....	27
3.10.2	ΜΟΝΤΕΛΟ.....	28
3.10.3	ΚΑΤΑΝΟΜΗ DIRICHLET.....	30
3.10.4	ΥΠΕΡΠΑΡΑΜΕΤΡΟΙ	30
3.10.5	ΠΑΡΑΔΕΙΓΜΑ ΕΚΤΕΛΕΣΗΣ ΑΛΓΟΡΙΘΜΟΥ LDA	31
3.11	AGENT ALLOCATOR	36
3.11.1	ΒΑΣΙΚΕΣ ΘΕΩΡΗΣΕΙΣ ΤΗΣ ΜΕΘΟΔΟΥ	36
3.11.2	ΠΑΡΑΔΟΧΕΣ ΤΗΣ ΜΕΘΟΔΟΥ.....	37
3.11.3	ΜΕΘΟΔΟΛΟΓΙΑ.....	37
3.12	Η ΜΕΘΟΔΟΣ UTASTAR	39
ΚΕΦΑΛΑΙΟ 4	43
4.1	ΕΙΣΑΓΩΓΗ	43
4.2	ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ.....	45
4.3	LATENT DIRICHLET ALLOCATION	52
4.4	ΑΠΟΤΕΛΕΣΜΑΤΑ LDA	55
4.5	ΑΝΑΘΕΣΗ ΑΠΑΙΤΗΣΕΩΝ ΚΑΤΑΝΑΛΩΤΗ ΣΕ ΚΑΘΕ ΘΕΜΑΤΙΚΗ ΕΝΟΤΗΤΑ.....	59
4.6	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΚΙΝΗΤΩΝ	61
4.7	ΔΥΝΑΤΟΤΗΤΕΣ ΚΙΝΗΤΩΝ	61
4.8	AGENT ALLOCATOR	63
4.8.1	ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΡΙΤΗΡΙΩΝ	63
4.8.2	ΑΠΟΔΟΣΗ ΤΙΜΩΝ ΥΠΟΚΡΙΤΗΡΙΩΝ	65
4.8.3	ΑΠΟΔΟΣΗ ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΚΑΙ ΚΛΙΜΑΚΑΣ ΜΕΤΡΗΣΗΣ ΣΤΑ ΚΡΙΤΗΡΙΑ.....	68
4.9	ΠΟΛΥΚΡΙΤΗΡΙΟΣ ΠΙΝΑΚΑΣ	76
4.10	UTASTAR.....	78
4.11	ΑΠΟΤΕΛΕΣΜΑΤΑ UTASTAR- ΣΥΣΤΑΣΕΙΣ	79
ΚΕΦΑΛΑΙΟ 5	89
5.1	ΣΥΜΠΕΡΑΣΜΑΤΑ	89
ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ ΑΓΟΡΑΣ ΚΙΝΗΤΟΥ ΤΗΛΕΦΩΝΟΥ	91
ΠΑΡΑΡΤΗΜΑ Α	91
ΠΑΡΑΡΤΗΜΑ Β	93
ΠΑΡΑΡΤΗΜΑ Γ	96

ΠΑΡΑΡΤΗΜΑ Δ	97
ΒΙΒΛΙΟΓΡΑΦΙΑ	99
ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ	
Σχήμα 4.1 Διάγραμμα ροής μεθοδολογίας	43
Σχήμα 4.2 Ηλικιακή Κατανομή	45
Σχήμα 4.3 Ποσοστιαία Κατανομή Φύλων	46
Σχήμα 4.4 Οικογενειακή κατάσταση	46
Σχήμα 4.5 Μορφωτικό επίπεδο	46
Σχήμα 4.6 Επάγγελμα	47
Σχήμα 4.7 Ετήσιο Εισόδημα	47
Σχήμα 4.8 Ώρες χρήσης κινητού τηλεφώνου ημερησίως	48
Σχήμα 4.9 Λειτουργικό Σύστημα Κινητού Τηλεφώνου	48
Σχήμα 4.10 Αποδοχή διαφορετικού λειτουργικού	48
Σχήμα 4.11 Μέγεθος Οθόνης	49
Σχήμα 4.12 Μνήμη RAM	49
Σχήμα 4.13 Χωρητικότητα Μπαταρίας	50
Σχήμα 4.14 Αποθηκευτικός Χώρος	50
Σχήμα 4.16 Συχνότητα Επεξεργαστή	50
Σχήμα 4.15 Ανάλυση Κάμερας	50
Σχήμα 4.17 Ανάλυση Οθόνης	51
Σχήμα 4.18 Εμφάνιση	51
Σχήμα 4.19 Δυνατότητες Σύνδεσης	52
Σχήμα 4.20 Ποσό για την αγορά κινητού τηλεφώνου	52
Σχήμα 4.21 Διάγραμμα συχνοτήτων λέξεων	54
Σχήμα 4.22 Διάγραμμα Perplexity ανά αριθμό θεματικών ενότητων	55
Σχήμα 4.23 Γραφική Απεικόνιση θεματικών ενότητων	59
Σχήμα 4.24 Απαιτήσεις Καταναλωτών	61
Σχήμα 4.25 Ολικές Χρησιμότητες συνόλου αναφοράς κινητών τηλεφώνων	83
Σχήμα 4.26 Βάρη Κριτηρίων Χρήστη4	87
ΠΙΝΑΚΑΣ ΠΙΝΑΚΩΝ	
Πίνακας 3.1 Παράμετροι Latent Dirichlet Allocation	29
Πίνακας 3.2 Πιθανότητα τυχαίας θεματικής ενότητας	33
Πίνακας 3.3 Θεματικές ενότητες της δειγματοληψίας	33
Πίνακας 3.4 Πιθανότητα τυχαίας λέξης	34
Πίνακας 3.5 Λέξεις της δειγματοληψίας	35
Πίνακας 3.6 Νέο Εγγραφο	35
Πίνακας 3.7 Κατανομή λέξεων στις θεματικές ενότητες	36
Πίνακας 4.1 Θεματικές ενότητες LDA	56
Πίνακας 4.2 Θεματικές ενότητες LDA	57
Πίνακας 4.3 Θεματικές ενότητες LDA	57

Πίνακας 4.4 Θεματική Περιγραφή Συστάδας	60
Πίνακας 4.5 Χαρακτηριστικά Κινητών	61
Πίνακας 4.6 Κινητά συνόλου αγοράς	62
Πίνακας 4.7 Κινητά συνόλου αναφοράς	62
Πίνακας 4.8 Πίνακας Κριτηρίων	65
Πίνακας 4.9 Πίνακας τιμών 1ου Υποκριτηρίου Επιδόσεων	67
Πίνακας 4.10 Πίνακας τιμών 2ου Υποκριτηρίου Επιδόσεων	67
Πίνακας 4.11 Πίνακας τιμών 3ου Υποκριτηρίου Επιδόσεων	67
Πίνακας 4.12 Πίνακας τιμών 4 ^{ου} Υποκριτηρίου Επιδόσεων	68
ΠΙΝΑΚΑΣ 4.13 : Κλίμακα και Κωδικοποίηση κριτηρίου Επιδόσεις	68
Πίνακας 4.14 Πίνακας τιμών 1 ^{ου} Υποκριτηρίου Εμφάνισης	70
Πίνακας 4.15: Κλίμακα και Κωδικοποίηση κριτηρίου Εμφάνιση	70
Πίνακας 4.16 Πίνακας τιμών 1 ^{ου} Υποκριτηρίου Απεικόνισης	71
Πίνακας 4.17 Πίνακας τιμών 2 ^{ου} Υποκριτηρίου Απεικόνισης	71
Πίνακας 4.18 Πίνακας τιμών 3 ^{ου} Υποκριτηρίου Απεικόνισης	71
Πίνακας 4.19: Κλίμακα και Κωδικοποίηση κριτηρίου Απεικόνιση	71
Πίνακας 4.20 Πίνακας τιμών Υποκριτηρίου Διάρκειας Χρήσης	72
Πίνακας 4.21: Κλίμακα και Κωδικοποίηση κριτηρίου Διάρκεια Χρήσης	72
Πίνακας 4.22 Πίνακας τιμών Υποκριτηρίου Σχέσης ποιότητας-τιμής	74
Πίνακας 4.23: Κλίμακα και Κωδικοποίηση κριτηρίου Σχέση ποιότητας-τιμής	74
Πίνακας 4.24 Πίνακας τιμών Υποκριτηρίου Προτίμησης Μάρκας	74
Πίνακας 4.25 Κλίμακα και Κωδικοποίηση κριτηρίου Μάρκας	75
Πίνακας 4.26 Πίνακας τιμών Υποκριτηρίου Λειτουργικού συστήματος	75
Πίνακας 4.27 Κλίμακα και Κωδικοποίηση κριτηρίου Λειτουργικό Σύστημα	75
Πίνακας 4.28 Πολυκριτήριος Πίνακας για τον Χρήστη 1	76
Πίνακας 4.29 Πολυκριτήριος Πίνακας για τον Χρήστη 2	77
Πίνακας 4.30 Πολυκριτήριος Πίνακας για τον Χρήστη 3	78
Πίνακας 4.31 Πίνακας μεταδεδομένων Utastar	79
Πίνακας 4.32 Συντελεστής Συσχέτισης t του Kendall για Χρήστες 1,2,3	79
Πίνακας 4.33 Ολικές χρησιμότητες νέου συνόλου κινητών τηλεφώνων	84
Πίνακας 4.34 Πίνακας συστάσεων για χρήστες 1,2,3	85
Πίνακας 4.35 Συντελεστής συσχέτισης Kendall για χρήστη 4	85
Πίνακας 4.36 Πολυκριτήριος Πίνακας Χρήστη 4	86
Πίνακας 4.37 Ολικές Χρησιμότητες εναλλακτικών Χρήστη 4	87
Πίνακας 4.38 Συστάσεις Χρήστη 4	87

ΠΙΝΑΚΑΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

Διάγραμμα 4.1 Μερικές Χρησιμότητες κριτηρίου «Επιδόσεις»	80
Διάγραμμα 4.2 Μερικές Χρησιμότητες κριτηρίου "Απεικόνιση"	80
Διάγραμμα 4.3 Μερικές Χρησιμότητες κριτηρίου "Διάρκεια Χρήσης"	81
Διάγραμμα 4.4 Μερικές Χρησιμότητες "Εμφάνισης"	81

Διάγραμμα 4.5 Μερικές Χρησιμότητες “Λειτουργικού”	81
Διάγραμμα 4.6 Μερικές Χρησιμότητες "Σχέσης Ποιότητας-Τιμής"	81
Διάγραμμα 4.7 Μερικές Χρησιμότητες κριτηρίου "Προτίμηση μάρκας"	81
Διάγραμμα 4.8 Βάρη κριτηρίων	82

ΚΕΦΑΛΑΙΟ 1

1.1 ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ

Σκοπός αυτής της εργασίας είναι η ανάπτυξη ενός πολύ-κριτήριου συστήματος συστάσεων κινητών τηλεφώνων, που εντοπίζει τις απαιτήσεις του καταναλωτή και προβλέπει ποια κινητά τηλέφωνα τις ικανοποιούν και σε ποιο βαθμό. Για τον καλύτερο εντοπισμό των απαιτήσεων των καταναλωτών, πραγματοποιήθηκε διανομή έντυπων ερωτηματολογίων στην πόλη των Χανίων, καθώς και ηλεκτρονικά σε τυχαίους καταναλωτές.

1.2 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ

Το πρώτο κεφάλαιο εισάγει τον αναγνώστη στον σκοπό της εργασίας ενώ στο δεύτερο γίνεται μία ανασκόπηση της υφιστάμενης κατάστασης. Το τρίτο κεφάλαιο παρουσιάζει τις θεωρητικές βάσεις, απαραίτητες για την κατανόηση της μεθοδολογίας της εργασίας. Στο τέταρτο κεφάλαιο γίνεται αναλυτική εφαρμογή της προτεινόμενης προσέγγισης, ώστε στο πέμπτο και τελευταίο κεφάλαιο, να εξαχθούν τα συμπεράσματα της εργασίας.

ΚΕΦΑΛΑΙΟ 2

2.1 ΕΙΣΑΓΩΓΗ

Με την ραγδαία αύξηση των κινητών τηλεφώνων, καθώς και των αλλαγών στα χαρακτηριστικά τους, αποτελεί πρόκληση για τους υποψήφιους καταναλωτές να εντοπίσουν τα προϊόντα που θα ικανοποιήσουν τις ανάγκες τους. Παράλληλα η γνώση της συμπεριφοράς των καταναλωτών κατά την αγορά ενός προϊόντος καθώς και το πως επιλέγουν ένα προϊόν ανάμεσα από ένα σύνολο προϊόντων, είναι σημαντική τόσο για τους υπεύθυνους μάρκετινγκ των επιχειρήσεων όσο και τους πωλητές. Συνεπώς ανά τα χρόνια έχουν αναπτυχθεί συστήματα που επιχειρούν να δώσουν λύση στις παραπάνω προκλήσεις. Σε αυτήν την ενότητα θα παρουσιαστούν αξιοσημείωτες προσεγγίσεις στον τομέα των συστημάτων συστάσεων.

2.2 ΥΦΙΣΤΑΜΕΝΗ ΚΑΤΑΣΤΑΣΗ

Υπάρχουν κάποιες μελέτες που έχουν χρησιμοποιήσει πολυκριτήριες αναλύσεις αποφάσεων για να βοηθήσουν πωλητές στο ηλεκτρονικό εμπόριο. Για παράδειγμα ο Rekik (Denguir-Rekik, 2009) παρουσίασε ένα σύστημα αξιολόγησης πολλαπλών κριτηρίων ηλεκτρονικού εμπορίου που μπορούσαν να χρησιμοποιήσουν ομάδες μάρκετινγκ και να επωφεληθούν από τις διαθέσιμες πληροφορίες των καταναλωτών (συμπεριλαμβανομένων και των σχολίων) των καταναλωτών, ώστε να είναι σε θέση να σχεδιάσουν καλύτερες στρατηγικές.

Στο ίδιο μήκος κύματος, ο Lee (2010) πρότεινε ένα σύστημα συστάσεων σχετικά με την μουσική στα κινητά με ονομασία CoFoSIM, που μέσω πολυκριτήριας ανάλυσης αποφάσεων κατέληγε στις προτιμήσεις του χρήστη, λαμβάνοντας υπόψιν την ανατροφοδότηση των καταναλωτών.

Ο Zheng (2019) ανέπτυξε ένα πολύκριτήριο σύστημα συστάσεων βασισμένο στη χρησιμότητα κατά το οποίο τα αντικείμενα συστήνονται σε έναν χρήστη με βάση τη συνάρτηση χρησιμότητας κάθε αντικειμένου για τον χρήστη. Η συνάρτηση χρησιμότητας δημιουργείται με την βοήθεια των βαθμολογιών του χρήστη στα πολλαπλά κριτήρια. Το σύστημα αυτό συγκρίνει τον πίνακα των αξιολογήσεων του χρήστη και τον πίνακα των προσδοκιών του χρήστη και βρίσκει την ομοιότητα μεταξύ αυτών των πινάκων. Χρησιμοποιούνται τρεις δείκτες ομοιότητας για τον υπολογισμό της χρησιμότητας: η συσχέτιση Pearson, η ομοιότητα συνημίτονου και η ευκλείδεια απόσταση. Ο πίνακας των προσδοκιών του χρήστη δημιουργείται με τρεις ιεραρχικές μεθόδους βελτιστοποίησης (Pointwise ranking, Pairwise Ranking, and Listwise Ranking). Η συγκεκριμένη μεθοδολογία, για σκοπούς ελέγχου, χρησιμοποίησε δεδομένα από το Tripadvisor και το Yahoo movies. Για το Tripadvisor 14,300 ξενοδοχεία, 1502 Χρήστες και 22,130 αξιολογήσεις οι οποίες εμπεριέχουν επτά κριτήρια όπως η τιμή, η τοποθεσία, η ποιότητα δωματίων και η άνεση του ξενοδοχείου, ενώ για το yahooovies η αντίστοιχη βάση δεδομένων περιείχε 2,162 χρήστες με 62,739 βαθμολογίες σε 3,078 ταινίες σε τέσσερα κριτήρια (ιστορία, σκηνοθεσία, οπτικά εφέ και ηθοποιία). Τα αποτελέσματα αυτής της μεθοδολογίας έδειξαν, με γνώμονα την ακρίβεια, πως το μέτρο της συσχέτισης Pearson βρέθηκε να δίνει καλύτερα αποτελέσματα και το Listwise ranking την καλύτερη επίδοση από τις υπόλοιπες μεθόδους.

Οι Wasid και Ali (2018) ανέπτυξαν ένα πολυκριτήριο σύστημα χρησιμοποιώντας προσέγγιση με βάση την ομαδοποίηση. Η κύρια ιδέα αυτής της προσέγγισης είναι η εύρεση παρόμοιων γειτόνων ενός χρήστη με στόχο τις καλύτερες συστάσεις. Σε αυτή τη μεθοδολογία οι προτιμήσεις των χρηστών βρίσκονται από τις πολυκριτήριες βαθμολογήσεις των χρηστών και τα κέντρα των συστάδων, διαμορφώνονται με βάση τις αντίστοιχες προτιμήσεις των χρηστών. Στη συνέχεια χρησιμοποιείται η ευκλείδεια απόσταση για να ανατεθεί κάθε χρήστης στην πλησιέστερη συστάδα, ενώ η απόσταση Mahalanobis χρησιμοποιείται για να υπολογιστούν οι πλησιέστεροι N γείτονες ενός χρήστη, στην ίδια συστάδα. Αυτή η μεθοδολογία προβλέπει τη βαθμολόγηση ενός αντικειμένου για έναν χρήστη με βάση τη βαθμολογία που έδωσαν σε αυτό πλησιέστεροι γείτονες από την ίδια συστάδα με τον χρήστη. Αυτή η προσέγγιση ελέγχθηκε με μία βάση δεδομένων από το Yahoo movies, όπου επιλέχθηκαν χρήστες οι οποίοι έχουν βαθμολογήσει τουλάχιστον 20 ταινίες. (Συνολικά 484 χρήστες, 945 ταινίες and 19,050 βαθμολογήσεις). Η μεθοδολογία ελέγχθηκε με την χρήση συσταδοποίησης και χωρίς, με μέτρο σύγκρισης το μέσο απόλυτο σφάλμα και κατέληξε στο αποτέλεσμα πώς με τη βοήθεια αυτής της μεθόδου συσταδοποίησης υπάρχει το καλύτερο αποτέλεσμα μέσου απόλυτου σφάλματος, ίσο με 2,175.

Σύμφωνα με τους Adomavicius και Kwon το 2007, τα παραδοσιακά συστήματα συστάσεων με βάση το συνεργατικό φιλτράρισμα χαρακτηρίζονται ως συστήματα τα οποία προσπαθούν να υπολογίσουν μία συνάρτηση R τέτοια ώστε R : χρήστες X αντικείμενα $\rightarrow R_0$. Αυτή η συνάρτηση προσπαθεί να προβλέψει μία βαθμολογία για ένα δοσμένο ζευγάρι χρήστη αντικειμένου, με το r_0 να ανήκει στους πραγματικούς αριθμούς σε μία κλίμακα, όπως για παράδειγμα το 1-10. Αυτή η συνάρτηση στα πολυκριτηρια συστήματα συστάσεων όμως έχει διαφορετική δομή.

Χρήστες X Αντικείμενα $\rightarrow R_0 \times R_1 \times \dots \times R_k$.

Συνεπώς πρέπει να γίνει πρόβλεψη της τιμής R_0 και για τα k επιμέρους κριτήρια. Στα κριτήρια, η βάση δεδομένων των βαθμολογιών περιέχει τις συνολικές βαθμολογίες καθώς και βαθμολογίες για τα επιμέρους κριτήρια για κάθε χρήστη.

Στο **Σχήμα 2.1** εμφανίζεται ένα παράδειγμα βάσης δεδομένων ενός πολύκριτρητου συστήματος για το Yahoo movies.

Table 1. Multi-criteria rating database fragment for Yahoo!Movies.

Row	User	Item	Overall	Acting	Story	Visuals	Directing
1	u1	i1	4	3	3	5	5
2	u1	i2	3	2	2	4	2
3	u2	i1	4	5	5	3	4
4	u2	i2	5	5	4	3	5

Σχήμα 2.1 Τμήμα βάσης δεδομένων για το Yahoo!Movies

Πηγή: D. Jannach, Z. Karakaya, Fatih Gedikli(2012)

Οι Adomavicius και Kwon (2007), πρότειναν δύο μεθοδολογίες: μία με βάση την ομοιότητα και μία με βάση τη συνάρτηση συνάθροισης. Στην πρώτη μεθοδολογία προσπαθούν να βρουν ένα μέτρο ομοιότητας μεταξύ χρηστών με βάση τις βαθμολογίες τους. Οι βαθμολογίες ενός χρήστη για ένα αντικείμενο παρουσιάζονται σαν διανύσματα K -διαστάσεων, με το K να αποτελεί τον αριθμό των κριτηρίων, ώστε να μπορεί να υπολογιστεί η απόσταση με την βοήθεια μετρικών εργαλείων όπως της Ευκλείδειας απόστασης ή της απόστασης Chebyshev. Στη δεύτερη μεθοδολογία γίνεται η γενική υπόθεση πως υπάρχει μία σχέση μεταξύ των συνολικών βαθμολογιών των αντικειμένων και των ατομικών βαθμολογιών των κριτηρίων και για αυτό αντιμετωπίζουν το πρόβλημα των K -κριτηρίων σαν K διαφορετικά προβλήματα, στο καθένα από τα οποία προσπαθούν να υπολογίσουν την ομοιότητα μεταξύ χρηστών μόνο σε ένα κριτήριο.

Το σύστημα συστάσεων που παρουσιάζεται στην παρούσα εργασία χρησιμοποιεί μεθόδους εξόρυξης χρήσιμης γνώσης σε κριτικές κινητών τηλεφώνων, για τον καθορισμό των απαιτήσεων των καταναλωτών κινητών τηλεφώνων, καθώς και χαρακτηριστικών κινητών που καλύπτουν αυτές τις απαιτήσεις. Στην συνέχεια εφαρμόζονται μέθοδοι πολυκριτήριας ανάλυσης που αφενός αξιολογούν τον βαθμό που καλύπτει τις απαιτήσεις ενός χρήστη ένα σύνολο αξιολογημένων από τον χρήστη κινητών, και αφετέρου ανιχνεύουν τα κριτήρια που προτιμάει, ώστε να υπολογιστεί ο βαθμός κάλυψης των αναγκών του από ένα μεγαλύτερο σύνολο κινητών τηλεφώνων.

ΚΕΦΑΛΑΙΟ 3

3.1 ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο παρουσιάζονται οι θεωρητικές έννοιες, πάνω στις οποίες βασίστηκε η μεθοδολογία για την ανάπτυξη του συστήματος συστάσεων. Γίνεται αναφορά στις μεθόδους συλλογής δεδομένων μέσω ερωτηματολογίων καθώς και στις μεθόδους που χρησιμοποιήθηκαν για την οργάνωση και επεξεργασία αυτών των δεδομένων (Latent Dirichlet Allocation, Agent Allocator, Utastar).

3.2 ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ

Τα συστήματα συστάσεων έχουν ως πρωταρχικό σκοπό την πρόβλεψη της εκτίμησης ενός χρήστη σε ένα αντικείμενο. Το πετυχαίνουν αυτό χρησιμοποιώντας συστήματα που αντλούν πληροφορία από άλλους χρήστες με παρόμοιες προτιμήσεις ή/και άλλα αντικείμενα με παρόμοια χαρακτηριστικά. Ανάλογα με τον τρόπο που αντλούν τις πληροφορίες τα συστήματα, έχουν καταταχθεί σε 4 βασικές κατηγορίες: συνεργατικό φιλτράρισμα, φιλτράρισμα βασισμένο στο περιεχόμενο, φιλτράρισμα βασισμένο στη γνώση και υβριδικό φιλτράρισμα.

3.3 ΒΑΣΙΚΟΙ ΑΛΓΟΡΙΘΜΟΙ ΣΥΣΤΗΜΑΤΩΝ ΣΥΣΤΑΣΗΣ

3.3.1 ΣΥΝΕΡΓΑΤΙΚΟ ΦΙΛΤΡΑΡΙΣΜΑ

Τα συνεργατικά συστήματα βρίσκουν ομοιότητες και διαφορές μεταξύ χρηστών και προβλέπουν την εκτίμηση ενός χρήστη με βάση άλλους χρήστες που εμφανίζουν κοινές προτιμήσεις με αυτόν. Για να το πετύχουν αυτό, βασίζονται μόνο σε προηγούμενες αλληλεπιδράσεις μεταξύ χρηστών και αντικειμένων, δεδομένα τα οποία αποθηκεύονται σε έναν πίνακα K , διαστάσεων $N \times M$ όπου N ο αριθμός των χρηστών και M ο αριθμός των αντικειμένων. Κάθε στοιχείο του πίνακα K_{ij} αντιστοιχεί στην αξιολόγηση του χρήστη i στο αντικείμενο j .

Τα συνεργατικά μοντέλα φιλτραρίσματος διαχωρίζονται σε δύο τύπους.

Συνεργατικό φιλτράρισμα βασισμένο στην μνήμη (Memory-based)

Υπολογίζει την ομοιότητα μεταξύ χρηστών ή αντικειμένων, χρησιμοποιώντας τα αντικείμενα που έχει βαθμολογήσει ο χρήστης στο παρελθόν και στοχεύει στον υπολογισμό του βαθμού ομοιότητας μεταξύ χρηστών ή αντικειμένων, για να προχωρήσει στην σύσταση.

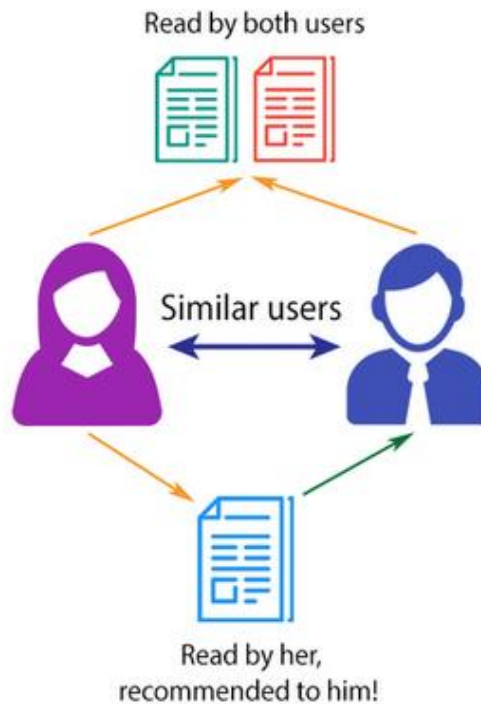
Τα μοντέλα βασισμένα στην μνήμη χωρίζονται με την σειρά τους σε δύο κατηγορίες: Με βάση τον χρήστη και με βάση το αντικείμενο. Στο συνεργατικό φιλτράρισμα με βάση τον χρήστη(user-based), για να προβλεφθεί η βαθμολογία ενός άγνωστου αντικειμένου ενός χρήστη, ο αλγόριθμος ψάχνει να βρει έναν άλλον χρήστη που να έχει παρόμοιες αξιολογήσεις και αφού εξερευνήσει τα κοινά αντικείμενα, προχωράει στην πρόβλεψη της αξιολόγησης του αντικειμένου. Ενώ σε αυτά που βασίζονται στο αντικείμενο(item-based), ο αλγόριθμος ψάχνει να βρει αντικείμενα που να έχουν ίδιες αξιολογήσεις με τα αντικείμενα του χρήστη, και προχωράει στην πρόβλεψη των αξιολογήσεων.

Συνεργατικό φιλτράρισμα βασισμένο στο μοντέλο (Model-based)

Αυτά τα συστήματα χρησιμοποιούνται με στόχο να προβλέψουν και να υπολογίσουν την διαδικασία με την οποία ένας χρήστης δίνει μια προσωπική του αξιολόγηση σε κάθε αντικείμενο. Αυτοί οι αλγόριθμοι είναι βασισμένοι στην μηχανική εκμάθηση στοχεύοντας στην πρόβλεψη αξιολογήσεων μη αξιολογημένων αντικειμένων, από τα αξιολογημένα αντικείμενα του χρήστη. Οι αλγόριθμοι αυτοί χωρίζονται σε κατηγορίες όπως βαθιά εκμάθηση(deep learning methods) και στοιβάδων(clustering algorithms)

Το κύριο πλεονέκτημα του συνεργατικού φιλτραρίσματος είναι πως δεν απαιτεί ανάλυση του περιεχομένου ενός αντικειμένου, συνεπώς βρίσκεται σε θέση να προτείνει αντικείμενα περίπλοκου περιεχομένου, χωρίς να χρειάζεται να το κατανοήσει. Επιπλέον προσαρμόζονται και βελτιώνονται σε βάθος χρόνου. Στο Σχήμα 3.1 παρουσιάζεται ένα παράδειγμα σύστασης με χρήση συνεργατικού φιλτραρίσματος.

COLLABORATIVE FILTERING



Σχήμα 3.1 Παράδειγμα Συνεργατικού φιλτραρίσματος

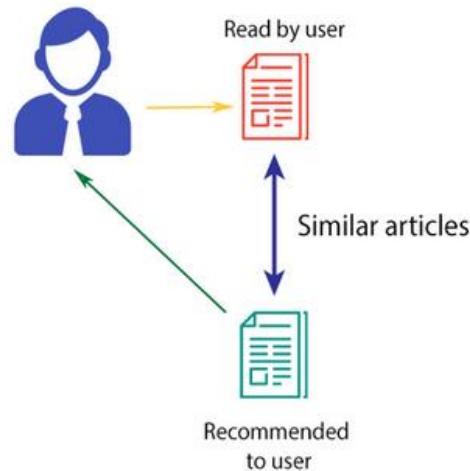
Πηγή: www.marutitech.com/recommendation-engine-benefits/

3.3.2 ΦΙΛΤΡΑΡΙΣΜΑ ΜΕ ΒΑΣΗ ΤΟ ΠΕΡΙΕΧΟΜΕΝΟ

Το φιλτράρισμα με βάση το περιεχόμενο δίνει τη δυνατότητα στα συστήματα να δημιουργήσουν προφίλ των χρηστών. Για να το επιτύχουν αυτό χρησιμοποιούν δεδομένα τα οποία αντλούν από τον χρήστη έμμεσα, μέσω των ιδιοτήτων των αντικειμένων με τα οποία αλληλοεπιδράσει.

Τα δεδομένα που χρησιμοποιούνται στηρίζονται στα επιμέρους δεδομένα των αντικειμένων, όπως το είδος ή η διάρκεια μιας ταινίας, καθώς και στις προηγούμενες ενέργειες ενός χρήστη, όπως τα αντικείμενα στα οποία έκανε κλικ σε μια ιστοσελίδα. Η λογική αυτών των μοντέλων βασίζεται στο ότι ένας χρήστης τείνει να αγοράσει αντικείμενα τα οποία παρουσιάζουν ομοιότητες με αντικείμενα που έχει αγοράσει. Η ομοιότητα των αντικειμένων υπολογίζεται έχοντας ως επίκεντρο τα χαρακτηριστικά τους. Αυτό το είδος συστημάτων θεωρείται πιο αποτελεσματικό για να προτείνει αντικείμενα τα οποία βρίσκονται σε μορφή κειμένου (Folasade Isinkaye, 2015). Στο φιλτράρισμα με βάση το περιεχόμενο χρησιμοποιούνται τεχνικές μέτρησης της ομοιότητας, όπως η ομοιότητα συνημιτόνου (Cosine Similarity) και η Ευκλείδεια απόσταση (Euclidean distance), καθώς και τεχνικές μηχανικής εκμάθησης. Οι τελευταίες δημιουργούν συστάσεις με την βοήθεια ενός διανύσματος, το οποίο λαμβάνει τιμή 0 ή 1, ανάλογα με την παρουσία ή απουσία μίας λέξης σε ένα έγγραφο. Όμως, προκύπτει πως με αυτόν τον τρόπο ευνοούνται έγγραφα με περισσότερες λέξεις. Λύση σε αυτό το πρόβλημα δίνουν οι τεχνικές συχνότητας όρων τύπου TF (term Frequency) και IDF (Inverse Document Frequency). Με τις πρώτες λαμβάνεται υπόψη η συχνότητα ενός όρου σε ένα έγγραφο και υπολογίζεται η σημαντικότητα του μέσα σε ένα έγγραφο. Αντίστροφα, το IDF στοχεύει στην μοναδικότητα των όρων που εμφανίζονται μόνο σε ένα έγγραφο, επομένως εμφανίζεται η ομοιότητα και η διαφορά μεταξύ δύο εγγράφων. Επιπροσθέτως άλλοι τρόποι για να υπολογισθεί η ομοιότητα μεταξύ αντικειμένων περιλαμβάνουν τον Naïve Bayes classifier (Friedman, 1997), δέντρα αποφάσεων (Pawlicka, 2021) και νευρωνικά δίκτυα κατά το οποίο το μοντέλο μαθαίνει χρησιμοποιώντας μηχανική εκμάθηση και τεχνικές στατιστικής ανάλυσης. Στο Σχήμα 3.2 παρουσιάζεται ένα παράδειγμα φιλτραρίσματος με βάση το περιεχόμενο.

CONTENT-BASED FILTERING



Σχήμα 3.2 Παράδειγμα φιλτραρίσματος με βάση το περιεχόμενο
Πηγή: www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-recommendation-engine-python/

3.3.3 ΔΗΜΟΓΡΑΦΙΚΟ ΦΙΛΤΡΑΡΙΣΜΑ

Το δημογραφικό φιλτράρισμα στοχεύει στο να δημιουργήσει προφίλ χρηστών ανάλογα με τα δημογραφικά τους δεδομένα, όπως είναι η ηλικία, το φύλο και το επάγγελμα. Μέσα από αυτή την διαδικασία προσπαθεί να δημιουργήσει δημογραφικά προφίλ που παρουσιάζουν προτίμηση κάποιο αντικείμενο. Αν οι χρήστες που ανήκουν στο δημογραφικό προφίλ u προτιμούν το αντικείμενο i και ένας χρήστης που ανήκει στο δημογραφικό προφίλ u δεν έχει αγοράσει το αντικείμενο i , το αντικείμενο i θα του συσταθεί.

Τα κύρια πλεονεκτήματα του δημογραφικού φιλτραρίσματος είναι πως τα δεδομένα αυτά τυπικά παρέχονται μέσω ερωτηματολογίων από τους χρήστες, συνεπώς λαμβάνονται υπόψιν αλλαγές στις προτιμήσεις των χρηστών. Αντιθέτως, ενώ αυτή η διαδικασία δεν χρειάζεται περίπλοκα δεδομένα, όπως αξιολόγησης προϊόντων από τους χρήστες, μπορεί να καταλήξει σε γενικευμένες συστάσεις οι οποίες να μην ανταποκρίνονται στις προσωπικές προτιμήσεις των χρηστών.

3.3.4 ΦΙΛΤΡΑΡΙΣΜΑ ΜΕ ΒΑΣΗ ΤΗ ΓΝΩΣΗ

Το φιλτράρισμα με βάση τη γνώση χρησιμοποιείται για να γίνει σύσταση αντικειμένων, τα οποία δεν αγοράζονται πολύ συχνά λόγω μεγάλης αξίας όπως διαμερίσματα και αυτοκίνητα. Αυτή η μέθοδος προσπαθεί να εντοπίσει τα χαρακτηριστικά ενός προϊόντος για το οποίο ενδιαφέρεται ένας χρήστης. Αφού εντοπίσει ποιο χαρακτηριστικό ενδιαφέρει τον χρήστη μπορεί να του προτείνει παρόμοια προϊόντα. Ουσιαστικά αυτή η μέθοδος βασίζεται στην σχέση μεταξύ αναγκών του χρήστη και των αντικειμένων που ανταποκρίνονται σε αυτές. Αν για παράδειγμα ένας χρήστης αγοράζει ένα αυτοκίνητο γιατί ενδιαφέρεται περισσότερο για τον σχεδιασμό, από ότι για την ποιότητα κατασκευής ή την ταχύτητα, ο χρήστης θα οδηγηθεί σε αυτοκίνητα που κύριο χαρακτηριστικό τους αποτελεί ο υψηλός σχεδιασμός.

Τα θετικά χαρακτηριστικά αυτής της μεθόδου περιλαμβάνουν την αποτελεσματικότητα της σε περίπλοκα αντικείμενα καθώς και την αξιοπιστία των αποτελεσμάτων. Επειδή η γνώση του τομέα, στην οποία βασίζονται οι συστάσεις, δεν είναι θορυβώδης, προκύπτουν αξιόπιστες συστάσεις (Bouraga, 2014). Σε αντίθεση με τις προηγούμενες μεθόδους δεν απαιτούνται μεγάλες βάσεις δεδομένων και δεν χρειάζεται να αντιμετωπιστούν προβλήματα όπως της ψυχρής εκκίνησης (cold start) και των δύσκολων πελατών (grey sheep). Επομένως υπάρχει προτίμηση αυτής της μεθόδου όταν υπάρχει έλλειψη στις διαθέσιμες αξιολογήσεις στα προϊόντα ή όταν αυτά αγοράζονται σπάνια. Ωστόσο τα συστήματα βασισμένα στη γνώση απαιτούν λεπτομερή γνώση για το κάθε αντικείμενο, τον χρήστη καθώς και για τον τρόπο με τον οποίο τα αντικείμενα καλύπτουν τις ανάγκες του.

3.3.5 ΠΟΛΥΚΡΙΤΗΡΙΑ ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ

Με τη ραγδαία εξέλιξη στις πληροφορίες έχει γίνει επιτακτική ανάγκη τα συστήματα να μπορούν να λαμβάνουν όσο το δυνατόν περισσότερες πληροφορίες. Σε αντίθεση με τα συστήματα που εξετάσαμε έως τώρα τα οποία εξέταζαν ένα χαρακτηριστικό για να προχωρήσουν στη σύσταση τα πολυκριτήρια συστήματα λαμβάνουν υπόψιν όσο το δυνατόν περισσότερα κριτήρια.

3.3.6 ΥΒΡΙΔΙΚΑ ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ

Τα υβριδικά συστήματα συστάσεων χρησιμοποιούνται σε μία προσπάθεια καλύτερων συστάσεων από τα υπόλοιπα συστήματα. Χρησιμοποιούν συνδυασμούς των παραπάνω συστημάτων ώστε να είναι σε θέση να αντιμετωπίσουν τα προβλήματα και τους περιορισμούς που θέτονται. Βασίζονται στους τρεις κύριους τύπους των συστημάτων συστάσεων (τα συνεργατικά, τα συστήματα με βάση του περιεχομένου και τα συστήματα με βάση τη γνώση) και προσπαθούν να προβούν σε καλύτερες προβλέψεις συνδυάζοντας διαφορετικές προσεγγίσεις για τη συλλογή πληροφοριών καθώς και συστάσεων, οι οποίες βασίζονται σημαντικά στο είδος υβριδισμού. Συνήθως η μέθοδος του συνεργατικού φιλτραρίσματος συνδυάζεται μαζί με μία άλλη τεχνική σε μία προσπάθεια αποφυγής του προβλήματος της ψυχρής εκκίνησης

Οι πιο γνωστοί συνδυασμοί μεθόδων που έχουν εφαρμοστεί περιλαμβάνουν:

- σταθμισμένος συνδυασμός: οι βαθμολογίες όλων των διαθέσιμων τεχνικών συστάσεων συνδυάζονται μαζί ώστε να παράγουν μία σύσταση.
- εναλλαγή: το σύστημα να εναλλάσσεται μεταξύ τεχνικών συστάσης ανάλογα με την τρέχουσα κατάσταση.
- ανάμεικτα: συστάσεις από διαφορετικά συστήματα συστάσεων χρησιμοποιούνται μαζί.
- συνδυασμός χαρακτηριστικών: χαρακτηριστικά από διαφορετικά συστήματα συστάσεων χρησιμοποιούνται μαζί σε έναν αλγόριθμο.
- αλληλουχίας: ένα σύστημα συστάσεων διυλίζει τις συστάσεις του από ένα άλλο σύστημα συστάσεων.
- Firefly: Ένας αλγόριθμος που αποτελείται από δύο παραπάνω αλγορίθμους με το σημείο εξόδου (output) του ενός να αποτελεί το σημείο εισόδου (input) του επόμενου.
- Metalevel: το μοντέλο που δημιουργήθηκε από ένα σύστημα σύστασης χρησιμοποιείται σαν είσοδος για ένα άλλο.

3.4 ΠΡΟΒΛΗΜΑΤΑ ΣΤΑ ΣΥΣΤΗΜΑΤΑ ΣΥΣΤΑΣΕΩΝ

3.4.1 ΠΡΟΒΛΗΜΑ ΤΗΣ ΨΥΧΡΗΣ ΕΚΚΙΝΗΣΗΣ

Αυτό το πρόβλημα αφορά την περίπτωση που λείπουν αρκετές πληροφορίες σχετικά με ένα αντικείμενο ή έναν χρήστη και εμφανίζεται όταν ένας νέος χρήστης εισέρχεται στο σύστημα ή όταν νέα αντικείμενα προστίθενται σε αυτό. Οι νέοι χρήστες δεν έχουν βαθμολογήσει κανένα από τα αντικείμενα, συνεπώς δεν μπορούν να γίνουν ακριβείς προβλέψεις. Το πρόβλημα της ψυχρής εκκίνησης εντοπίζεται κυρίως στα συστήματα που χρησιμοποιούν συνεργατικό φιλτράρισμα, και λιγότερο συχνά στα συστήματα με βάση το περιεχόμενο, και αντιμετωπίζεται με διάφορες τεχνικές. Όταν η πληροφορία για έναν καινούργιο χρήστη λείπει, το σύστημα μπορεί να αντλήσει πληροφορίες από άλλες πηγές όπως κοινωνικά δίκτυα και να χρησιμοποιήσει την πληροφορία αυτή, στοχεύοντας στην άντληση βασικών πληροφοριών για τον χρήστη. Ένας άλλος τρόπος επίλυσης αυτού του προβλήματος είναι η χρήση των δημογραφικών στοιχείων του νέου χρήστη ώστε να βρεθούν χρήστες με παρόμοια δημογραφικά στοιχεία (Pawlicka, 2021). Μία άλλη μέθοδος επίλυσης αυτού του προβλήματος αποτελεί η ρητή απαίτηση από το νέο χρήστη να βαθμολογήσει αντικείμενα ώστε να παραχθεί η ελλείπουσα πληροφορία. Για παράδειγμα σε μία ιστοσελίδα ζητείται από τον νέο χρήστη να βαθμολογήσει τα πιο δημοφιλή προϊόντα.

3.4.2 ΑΡΑΙΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ

Οι περισσότεροι χρήστες δεν βαθμολογούν τα περισσότερα αντικείμενα, συνεπώς υπάρχει το φαινόμενο της αραιότητας του πίνακα με δομή: Χρήστης x Αντικείμενο. Αποτελεί ιδιαίτερο πρόβλημα στα συστήματα που χρησιμοποιούν συνεργατικό φιλτράρισμα, καθώς χρήστες με παρόμοιες βαθμολογήσεις είναι δυσεύρετοι. Το πρόβλημα αυτό παρατηρείται συνήθως στα αρχικά στάδια των συστημάτων, καθώς και σε επόμενα στάδια, όταν υπάρχουν πολύ περισσότερα αντικείμενα από ότι χρήστες.

Κοινό παρονομαστή στα προβλήματα της ψυχρής εκκίνησης και της αραιότητας των δεδομένων, αποτελεί η δυσκολία αντιστοίχισης χρηστών με μικρό αριθμό αξιολογήσεων αντικειμένων με άλλους χρήστες. Συνεπώς μία λύση αποτελεί η μετατροπή του πίνακα χρηστών και αντικειμένων σε μορφή γραφήματος, για να ακολουθήσει η διαδικασία εύρεσης συσχετίσεων μεταξύ χρηστών και αντικειμένων, κατά την οποία η σύσταση πραγματοποιείται με βάση την απόσταση αντικειμένων και χρηστών. Μια εναλλακτική λύση βασίζεται στην ομοιότητα μεταξύ χρηστών αλλά και αντικειμένων. Για τον υπολογισμό της πρόβλεψης συνδυάζονται οι βαθμολογήσεις αντικειμένων, χρηστών όμοιων με τον χρήστη, οι βαθμολογίες του χρήστη σε όμοια αντικείμενα και βαθμολογίες όμοιων αντικειμένων από όμοιους χρήστες (Jun Wang, 2006).

3.4.3 GREY SHEEP (ΔΥΣΚΟΛΟΙ ΠΕΛΑΤΕΣ)

Το πρόβλημα των 'γκρίζων προβάτων' εντοπίζεται κυρίως σε συστήματα συνεργατικού φιλτραρίσματος στα οποία η γνώμη του χρήστη δεν ταιριάζει με κανένα προφίλ. Πρόκειται για πελάτες οι οποίοι έχουν μοναδικές προτιμήσεις με αποτέλεσμα οι συστάσεις προς αυτούς να είναι χαμηλής ποιότητας. Συνήθως σε αυτούς τους χρήστες προτείνονται τα πιο δημοφιλή προϊόντα. Επίσης αξίζει να σημειωθεί πως η γνώμη αυτών των χρηστών επηρεάζει αρνητικά τα συστήματα συστάσεων για τους υπόλοιπους χρήστες.

3.4.4 ΣΥΝΩΝΥΜΕΣ ΛΕΞΕΙΣ

Το πρόβλημα των συνωνύμων αναφέρεται στην τάση αντικειμένων να έχουν διαφορετικούς τίτλους με άλλα αντικείμενα, ενώ είναι ταυτόσημα. Εν γένει τα συστήματα συστάσεων και ειδικότερα τα συστήματα με βάση το περιεχόμενο χρειάζεται να αναλύσουν μεγάλο όγκο αδόμητης πληροφορίας σε μορφή κειμένου. Λόγω των ασαφειών που περιλαμβάνουν οι μέθοδοι επεξεργασίας της φυσικής γλώσσας, τα συστήματα συστάσεων αδυνατούν να εντοπίσουν την ομοιότητα αντικειμένων και τα διαχειρίζονται ως διαφορετικά. Ένας τρόπος αντιμετώπισης αυτού του προβλήματος είναι η θεματική μοντελοποίηση (όπως η τεχνική του Latent Dirichlet Allocation), κατά την οποία διαφορετικές λέξεις ή εκφράσεις με συναφή έννοια κατηγοριοποιούνται σε θεματικές ενότητες.

3.5 ΑΝΙΧΝΕΥΤΗΣ ΙΣΤΟΥ

Το web crawler (“ανιχνευτής ιστού”) ή spider (αράχνη) ή spiderbot που είναι ένα πρόγραμμα που κυκλοφορεί στο διαδίκτυο ανακαλύπτοντας ιστοσελίδες. Ένα τέτοιο πρόγραμμα ακολουθεί επαναληπτικά κάθε σύνδεσμο μέσα σε μια ιστοσελίδα, αποθηκεύοντας κάθε σύνδεσμο και σελίδα, έως ότου καλύψει όλους τους συνδέσμους της ιστοσελίδας. Συνήθως χρησιμοποιείται από μηχανές αναζήτησης όπως της Google και Bing, για την αύξηση και επικαιροποίηση της βάσης δεδομένων τους.

3.6 ΑΠΟΣΠΑΣΗ ΔΕΔΟΜΕΝΩΝ

Ο όρος «απόξεση» ή «απόσπαση» δεδομένων (data scraping) αναφέρεται στο σύνολο μεθόδων που έχουν ως στόχο την απόσπαση και συλλογή δεδομένων από το διαδίκτυο. Ένας χρήστης μπορεί χειροκίνητα να προχωρήσει σε απόσπαση δεδομένων, όμως ο πιο διαδεδομένος τρόπος συλλογής πληροφοριών είναι ο αυτοματοποιημένος, με χρήση ειδικού λογισμικού. Η απόσπαση δεδομένων χρησιμοποιείται για ακαδημαϊκές έρευνες και μελέτες καθώς και στην οικονομία και στο εμπόριο, προς όφελος των καταναλωτών. Για παράδειγμα, η ιστοσελίδα skroutz.gr συνδυάζει μεθόδους ανίχνευσης ιστού και απόσπασης δεδομένων, συλλέγοντας πληροφορίες σχετικά με προϊόντα από όλα τα καταστήματα που τα διαθέτουν και τα κατατάσσουν με βάση κάποια κριτήρια όπως τιμή ή αξιοπιστία καταστήματος.

3.7 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ ΜΕΣΩ ΕΡΩΤΗΜΑΤΟΛΟΓΙΩΝ

Το ερωτηματολόγιο αποτελεί ένα ακόμα εργαλείο συλλογής δεδομένων. Αποτελείται από μία σειρά σαφών ερωτήσεων, ίδιων για όλους τους ερωτώμενους, κατά τις οποίες καλούνται να δώσουν την γνώμη τους. Σε αντίθεση με κλασικές μεθόδους συλλογής δεδομένων, όπως της συνέντευξης, ο διεξάγων της έρευνας δεν χρειάζεται να είναι παρών στην συμπλήρωση του ερωτηματολογίου. Τα ερωτηματολόγια μπορεί να έχουν έντυπη ή ηλεκτρονική μορφή, ανάλογα με τις ανάγκες και τις συνθήκες της εκάστοτε έρευνας. Άλλα θετικά στοιχεία περιλαμβάνουν την ευκολία διανομής τους, καθώς και την ανωνυμία που καθιστά την μέθοδο των ερωτηματολογίων, κατάλληλη για την συλλογή δεδομένων που αφορούν ευαίσθητα θέματα.

3.8 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Ως προ-επεξεργασία δεδομένων στην εξόρυξη γνώσης και ανάλυση δεδομένων, εννοείται το σύνολο των εργασιών μετατροπής των ακατέργαστων δεδομένων σε μία μορφή που θα εξασφαλίσει ή θα ενισχύσει τις επιδόσεις των μεθόδων που θα τα χρησιμοποιήσουν ως είσοδο. Η εργασία αυτή είναι απαραίτητη καθώς η πλειοψηφία των ακατέργαστων δεδομένων υπό μορφής κειμένου, φωτογραφίας και βίντεο είναι επιρρεπή σε διάφορα είδη προβλημάτων.

3.8.1.1 Ελλιπείς τιμές

Η ύπαρξη ελλιπών τιμών αποτελεί το συνηθέστερο πρόβλημα των ακατέργαστων δεδομένων, κυρίως λόγω των περιορισμών της διαδικασίας της συλλογής τους. Για παράδειγμα μία κάμερα ασφαλείας που καταγράφει μπορεί να υποστεί βλάβη για ένα χρονικό διάστημα. Αντίστοιχα κατά την συμπλήρωση ενός ερωτηματολογίου, κάποιες ερωτήσεις μπορεί να παραληφθούν λόγω ανθρώπινου λάθους.

Λόγω της εντατικής εμφάνισης αυτού του προβλήματος, έχουν προταθεί πολλοί τρόποι επίλυσης, όπως η αντικατάσταση των χαμένων τιμών με μία σταθερά όπως την τιμή μηδέν, ή την λέξης 'άγνωστο'. Όμως αυτός ο τρόπος δεν προτείνεται, καθώς πιθανότατα θα δημιουργήσει προβλήματα στους αλγορίθμους που θα χρησιμοποιήσουν αυτά τα δεδομένα. Μία καλύτερη εναλλακτική αποτελεί η αντικατάσταση της ελλείπουσας τιμής, με την μέση τιμή. Λύση στο πρόβλημα μπορεί επίσης να δώσει η αντικατάσταση με κάθε πιθανή τιμή. Αν η ελλείπουσα τιμή μπορεί να λάβει για παράδειγμα N πιθανές τιμές, τότε προστίθενται $N-1$ γραμμές στα δεδομένα, οι στήλες των οποίων θα είναι ίδιες, εκτός από αυτής της ελλείπουσας τιμής, όπου θα τοποθετούν οι πιθανές τιμές. Τέλος υπάρχει η λύση της διαγραφής ολόκληρης της γραμμής που προτείνεται σε δεδομένα με μεγάλο αριθμό δείγματος ή ελλείπων τιμών. Με αυτούς τους τρόπους περιορίζεται το αρνητικό αντίκτυπο στις μεθόδους που θα χρησιμοποιήσουν αυτά τα δεδομένα.

3.8.1.2 Θορυβώδη Δεδομένα

Θόρυβος στα δεδομένα μπορεί να παρουσιαστεί λόγω σφάλματων είτε στη συλλογή δεδομένων, είτε κατά την εισαγωγή δεδομένων ή στην μετάδοση. Παραδείγματα τέτοιου τύπου θορύβου αποτελούν η άγνωστη κωδικοποίηση, οι τιμές εκτός εύρους, και τα ασυνεπή δεδομένα. Ο όρος «ασυνεπή δεδομένα» αποδίδεται σε ίδια δεδομένα, που βρίσκονται σε πολλαπλές θέσεις μίας βάσης δεδομένων. Ο θόρυβος μπορεί να γίνει αντιληπτός από την γραφική απεικόνιση των δεδομένων. Για παράδειγμα αν τα δεδομένα απεικονίζονται με στοιβάδες ή ακολουθούν μία γραμμική παρεμβολή, ο θόρυβος θα εμφανίζεται ως τιμές εκτός στοιβάδων και τιμές που δεν παρεμβάλλονται γραμμικά αντίστοιχα. Αυτό το πρόβλημα αντιμετωπίζεται είτε στοχεύοντας σε όλες τις αριθμητικές τιμές και με την αντικατάστασή τους με άλλες τιμές, ή στοχεύοντας μόνο στις εγγραφές με ακραίες τιμές και την τροποποίηση αυτών των τιμών, ή την διαγραφή ολόκληρης της εγγραφής.

3.8.1.3 Μεγάλος όγκος δεδομένων

Ο αριθμός των χαρακτηριστικών εισόδου, των μεταβλητών ή των στηλών που υπάρχουν σε μία βάση δεδομένων είναι γνωστός ως διαστάσεις δεδομένων. Όσο οι διαστάσεις των δεδομένων αυξάνονται, τόσο πιο πολύ αυξάνεται η περιπλοκότητα των αλγορίθμων που θα τα χρησιμοποιήσουν ως είσοδο, καθώς και το υπολογιστικό κόστος.

3.8.1.4 Ύπαρξη πολλών χαρακτηριστικών

Οι στήλες αντιπροσωπεύουν χαρακτηριστικά(features) των δεδομένων. Η ύπαρξη πολλών όμως χαρακτηριστικών απαιτεί μεγάλη υπολογιστική ισχύ, συνεπώς δημιουργείται η ανάγκη μείωσης του πλήθους των χαρακτηριστικών τους. Κάποια χαρακτηριστικά ενδέχεται να μην προσφέρουν σημαντική πληροφορία ή κάποια χαρακτηριστικά να σχετίζονται, καθιστώντας την παρουσία τους περιττή. Η σωστή επιλογή χαρακτηριστικών προσφέρει πληθώρα πλεονεκτημάτων στους αλγορίθμους που θα χρησιμοποιήσουν αυτά τα δεδομένα, όπως γρηγορότερη εκτέλεση, μείωση της περιπλοκότητας και αύξηση της ακρίβειας του μοντέλου. Προτείνονται δύο ειδών μέθοδοι για την επίλυση αυτού του προβλήματος. Οι μέθοδοι τύπου Filter και τύπου Wrapper, σύμφωνα με τους πρώτους τα χαρακτηριστικά επιλέγονται σύμφωνα με τις επιδόσεις τους σε διάφορα στατιστικά τεστ, με γνώμονα την συσχέτισή τους με τη μεταβλητή έκβασης(outcome variable). Κύρια στατιστικά τεστ αποτελούν η συσχέτιση Pearson, η Γραμμική Διακριτική Ανάλυση, η Ανάλυση Διακύμανσης ANOVA, και το τεστ του Χι(X square test). Οι μέθοδοι τύπου Wrapper δοκιμάζουν επαναληπτικά υποσύνολα των χαρακτηριστικών και εντοπίζουν το υποσύνολο με τις καλύτερες επιδόσεις.

3.9 ΣΤΑΤΙΣΤΙΚΑ ΜΕΤΡΑ

Μετά την συλλογή και προ-επεξεργασία των δεδομένων, σειρά έχει η οργάνωση, σύνοψη και παρουσίαση των δεδομένων με έναν εύχρηστο και κατανοητό τρόπο. Αυτά μας τα προσφέρει η περιγραφική στατιστική με τεχνικές που περιλαμβάνουν γραφήματα που οπτικοποιούν τα δεδομένα και αριθμητικούς δείκτες που αντιπροσωπεύουν τις ιδιότητες των δεδομένων. Οι αριθμητικοί αυτοί δείκτες περιλαμβάνουν την Μέση τιμή ή τον αριθμητικό μέσο, τον σταθμικό μέσο, την διάμεσο, το Εύρος, η Διακύμανση, και η Τυπική απόκλιση.

3.10 LATENT DIRICHLET ALLOCATION

3.10.1 ΣΥΝΟΨΗ

Ο αλγόριθμος LDA αποτελεί έναν από τους πιο συνήθεις αλγορίθμους θεματικών μοντελοποιήσεων, για την ανακάλυψη κρυφών δομών πίσω από μία συλλογή κειμένων.

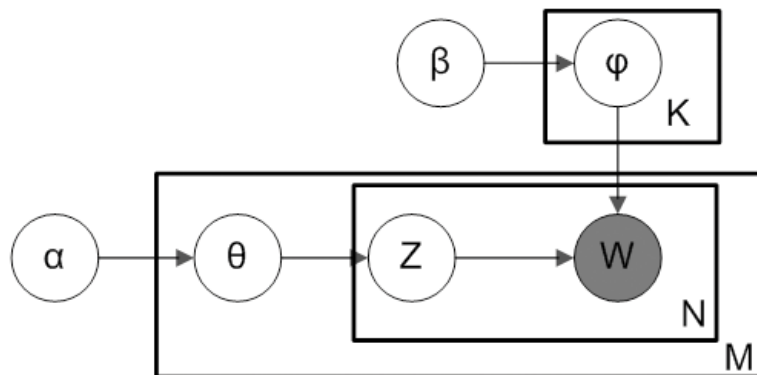
Σε αντίθεση με την δομή δεδομένων που προβλέπει συχνότητα λέξεων σε μία συλλογή λέξεων, το LDA δημιουργεί θέματα με αντίστοιχα βάρη για κάθε θέμα. Συνεπώς μειώνει τις διαστάσεις των δεδομένων. Ο αλγόριθμος LDA μπορεί να συγκριθεί με κλασσικούς αλγορίθμους δημιουργίας συστάδων, με το πλεονέκτημα της δημιουργίας συστάδων που περιέχουν λέξεις, και όχι χαρακτήρες, ενώ παράλληλα αναδεικνύει αφηρημένες και υποβόσκουσες έννοιες που υπάρχουν σε μία συλλογή κειμένων.

3.10.2 MONTELO

Το Latent Dirichlet Allocation (LDA) είναι ένα γεννητικό πιθανοτικό μοντέλο ενός συνόλου εγγράφων (Blei David, Ng Andrew & Jordan Michael, 2001). Η βασική ιδέα είναι ότι τα έγγραφα αντιπροσωπεύονται ως τυχαία μείγματα σε κρυμμένες θεματικές ενότητες, όπου κάθε θέμα χαρακτηρίζεται από μία κατανομή λέξεων (Lirong Qiu, 2018)

Το LDA αναλαμβάνει την ακόλουθη διαδικασία για κάθε έγγραφο w σε ένα συνολικό έργο D :

- 1) Διάλεξε $\theta_i \sim \text{Dir}(\alpha)$, όπου $i \in \{1, \dots, M\}$
- 2) Διάλεξε $\varphi_k \sim \text{Dir}(\beta)$, όπου $k \in \{1, \dots, K\}$
- 3) Για κάθε μία από τις N λέξεις w_n :
 - (a) διάλεξε ένα θέμα $z \sim \text{Πολυωνυμική}(\theta)$
 - (b) διάλεξε μια λέξη w_n από $p(w_n|z_n, \beta)$, μια πολυωνυμική πιθανότητα υπό όρους του θέματος z_n



Σχήμα 3.3 Γραφική απεικόνιση του αλγορίθμου LDA

Πηγή: https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation

Η παραπάνω διαδικασία αναπαρίσταται μαθηματικά με τον παρακάτω τύπο όπου κάθε όρος αντιπροσωπεύει 1 από τα 4 βήματα, και εκφράζει την συνολική πιθανότητα του μοντέλου.

$$P(\mathbf{W}, \mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\varphi}; \alpha, \beta) = \prod_{i=1}^K P(\varphi_i; \beta) \prod_{j=1}^M P(\theta_j; \alpha) \prod_{t=1}^N P(Z_{j,t} | \theta_j) P(W_{j,t} | \varphi_{Z_{j,t}})$$

Πίνακας 3.1 Παράμετροι Latent Dirichlet Allocation

N	Συνολικός αριθμός λέξεων σε όλα τα έγγραφα
D	Corpus
M	Αριθμός εγγράφων (στο corpus)
V	Αριθμός λέξεων (στο corpus)
$N_{d=1..M}$	Αριθμός λέξεων σε ένα έγγραφο d
$w_{d=1..M, w=1..N_d}$	Η ταυτότητα την λέξης w στο έγγραφο d
w	Ένα έγγραφο ($w=w_1, w_2, \dots, w_n$)
α	Παράμετρος της Κατανομής Dirichlet Προκαθορισμένο βάρος ενός θέματος σε ένα έγγραφο
$B_{w=1..V}$	Παράμετρος της Κατανομής Dirichlet Προκαθορισμένο βάρος μίας λέξης σε ένα θέμα
$\theta_{d=1..M}$	Η κατανομή των θεμάτων σε ένα έγγραφο d
ϕ	Η κατανομή των λέξεων σε ένα θέμα k
$Z_{d=1..M, w=1..N_d}$	Ταυτότητα Θέματος ενός εγγράφου d
K	Αριθμός θεμάτων
$Z_{d=1..M, w=1..N_d}$	Ταυτότητα του θέματος της λέξης w στο έγγραφο d

3.10.3 ΚΑΤΑΝΟΜΗ DIRICHLET

Στη θεωρία πιθανοτήτων και την στατιστική, κατανομή Dirichlet (εκ του Peter Gustav Lejeune Dirichlet) είναι μία οικογένεια συνεχών πολύ-μεταβλητών πιθανοτικών κατανομών με παράμετρο ένα διάνυσμα α θετικών πραγματικών αριθμών.

Η κατανομή Dirichlet, τάξεως $K \geq 2$ με παραμέτρους $\alpha_1, \dots, \alpha_K > 0$ έχει συνάρτηση πυκνότητας πιθανότητας:

$$f(x_1, \dots, x_K; \alpha_1, \dots, \alpha_K) = \frac{1}{B(\alpha)} \prod_{i=1}^K x_i^{\alpha_i-1}$$

Όπου, για το $\{x_k\}_{k=1}^{k=K}$ ισχύει : $\sum_{i=1}^K x_i = 1$ και $x_i \geq 0$ για όλα τα $i \in$

$\{1, \dots, K\}$ Η σταθερά εξομάλυνσης είναι η συνάρτηση Beta πολλαπλών μεταβλητών, η οποία μπορεί να εκφραστεί με τη συνάρτηση Gamma:

$$B(\alpha) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K \alpha_i)}, \alpha = (\alpha_1, \dots, \alpha_K)$$

3.10.4 ΥΠΕΡΠΑΡΑΜΕΤΡΟΙ

Το LDA χαρακτηρίζεται από 3 κύριες παραμέτρους.

- α Χαρακτηρίζει την πυκνότητα μεταξύ ενός εγγράφου και ενός θέματος
- β χαρακτηρίζει την πυκνότητα μεταξύ ενός θέματος και μιας λέξης
- K ο αριθμός των θεμάτων που δημιουργεί

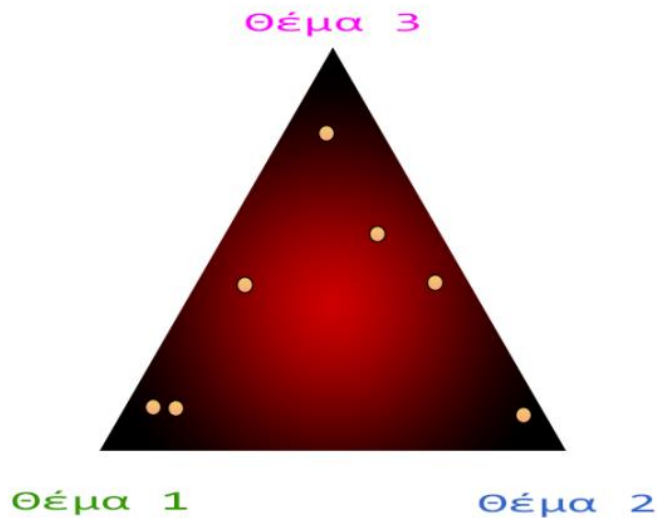
Η παράμετρος α ελέγχει τον αριθμό των θεμάτων που αναμένεται να εμφανιστούν σε ένα έγγραφο. Μικρή τιμή του α χρησιμοποιείται για να υπονοήσει πως προσδοκούνται λιγότερα θέματα, ενώ μεγάλη τιμή σημαίνει πως προσδοκούνται περισσότερα.

Αντίστοιχα η παράμετρος β ελέγχει την κατανομή των λέξεων σε ένα θέμα, με μικρότερες τιμές του β να υποδηλώνουν πως αναμένονται τα θέματα να έχουν λιγότερες λέξεις, ενώ σε μεγαλύτερες τιμές του προσδοκούνται να υπάρχουν περισσότερες λέξεις στα θέματα.

Τέλος, η παράμετρος **K** ρυθμίζει τον αριθμό των θεμάτων και μια μεγαλύτερη τιμή για το K οδηγεί σε χαμηλότερη βαθμολογία perplexity, υποδεικνύοντας καλύτερη γενίκευση αλλά περισσότερα θέματα καθώς και υπολογιστικό κόστος.

3.10.5 ΠΑΡΑΔΕΙΓΜΑ ΕΚΤΕΛΕΣΗΣ ΑΛΓΟΡΙΘΜΟΥ LDA

Για παράδειγμα έστω πως εκτελείται η παραπάνω διαδικασία για 3 θέματα(**K=3**). Ο αλγόριθμος **LDA** θα τοποθετήσει τα έγγραφα στον χώρο, ανάλογα με τις θεματικές ενότητες που ανήκουν. Στο παράδειγμα τα έγγραφα απεικονίζονται γραφικά με ένα τρίγωνο, όπου κάθε γωνία αντιπροσωπεύει 1 θεματική ενότητα. Τα έγγραφα αναπαρίστανται ως κίτρινοι κύκλοι(Σχήμα 3.4).

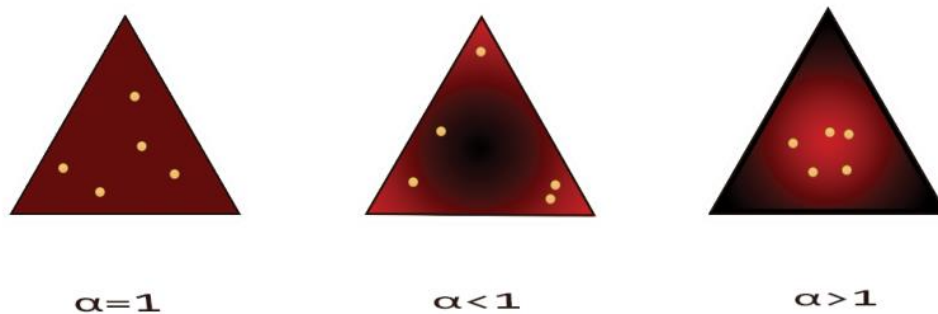


Έγγραφα- Θέματα

Σχήμα 3.4 Απεικόνιση εγγράφων σε κατανομή θεμάτων

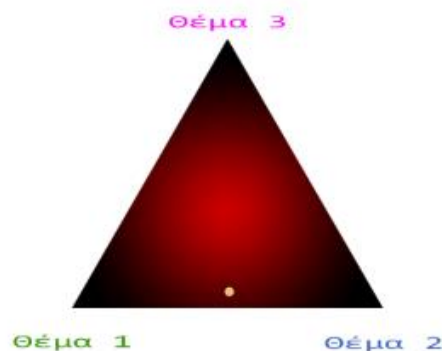
Κάποια έγγραφα αναφέρονται κυρίως σε ένα θέμα, όπως το έγγραφο κοντά στην κορυφή του **Θέματος 3**, ενώ άλλα αναφέρονται σε περισσότερα από ένα θέματα, όπως το έγγραφο μεταξύ των κορυφών του **Θέματος 3** και **Θέματος 1**. Αυτό το είδος κατανομής εγγράφων, ονομάζεται κατανομή Dirichlet, και έχει ως βασική παράμετρο το α .

Τιμή του α ίση με **1**, σηματοδοτεί πως τα έγγραφα κατανέμονται ομοιόμορφα στον χώρο, ενώ τιμές **μικρότερες του 1** δείχνουν συγκέντρωση των εγγράφων στην μέση του χώρου. Για τιμές **μεγαλύτερες του 1**, τα έγγραφα θα βρίσκονται στις γωνίες του τριγώνου. Η κατανομή των εγγράφων ανάλογα με την τιμή της παραμέτρου α , φαίνεται στο παρακάτω σχήμα(Σχήμα 3.5).



Σχήμα 3.5 Κατανομή εγγράφων ανάλογα με το α

Το έγγραφο στο Σχήμα 3.6 αναφέρεται κατά **55%** στο **Θέμα 1**, **45%** στο **Θέμα 3** και **5%** στο **Θέμα 2**.



Σχήμα 3.6 Απεικόνιση Εγγράφου 1 στην κατανομή 3 θεμάτων

Η θέση ενός εγγράφου σε ένα τρίγωνο μπορεί να αναπαρασταθεί ως ένα διάνυσμα πιθανοτήτων που αθροίζουν στην μονάδα (Σχήμα 3.7). Η κατανομή των θεμάτων σε ένα έγγραφο μπορεί να οριστεί ως μια **πολυωνυμική κατανομή** με παράμετρο το θ . **(3.10.2)**



Σχήμα 3.7 Διάνυσμα πιθανοτήτων Εγγράφου 1- θεμάτων

Μέσω της **πολυωνυμικής κατανομής** γίνεται δειγματοληψία **N** τυχαίων θεματικών ενοτήτων, με την πιθανότητα η τυχαία θεματική ενότητα να είναι ένα θέμα να παρουσιάζεται στον Πίνακα 3.2 **(3.10.2)**.

Πίνακας 3.2 Πιθανότητα τυχαίας θεματικής ενότητας

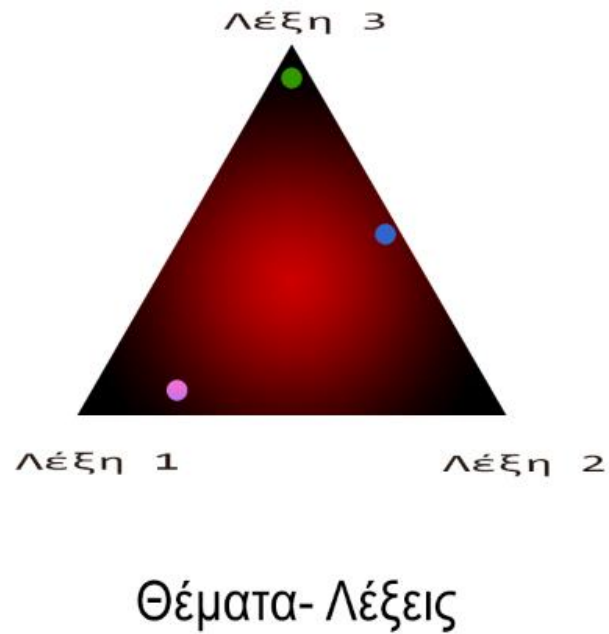
Θεματική Ενότητα	Θέμα 1	Θέμα 2	Θέμα 3
$P(Z_j, t \theta_j)$	0.55	0.45	0.05

Έστω πως σε μία δειγματοληψία 10 θεματικών ενοτήτων, επιλέχθηκαν τα θέματα που φαίνονται στον Πίνακα 3.3:

Πίνακας 3.3 θεματικές ενότητες της δειγματοληψίας

Θέμα 1	Θέμα 2	Θέμα 3	Θέμα 1	Θέμα 2
Θέμα 2	Θέμα 1	Θέμα 3	Θέμα 2	Θέμα 3

Αντίστοιχα, χρησιμοποιείται η **πολυωνυμική κατανομή**, καθώς και η κατανομή **Dirichlet** με παράμετρο το β , η οποία χαρτογραφεί θεματικές ενότητες στον χώρο των λέξεων. Συνεπώς σε αυτήν την κατανομή, οι γωνίες των τριγώνων θα είναι λέξεις, και οι κύκλοι στο τρίγωνο αναπαριστούν θεματικές ενότητες, όπως φαίνεται στο Σχήμα 3.8 Απεικόνιση θεματικών ενότητων σε κατανομή λέξεων για 3 λέξεις και 3 θέματα. (**3.10.2**)



Σχήμα 3.8 Απεικόνιση θεματικών ενότητων σε κατανομή λέξεων

Οι πολυωνυμικές κατανομές των λέξεων για κάθε θέμα αποτελούνται από ποσοστά των λέξεων 1,2,3 που αθροίζουν στο 100% (Πίνακας 3.4)

Πίνακας 3.4 Πιθανότητα τυχαίας λέξης

Λέξη	Λέξη 1	Λέξη 2	Λέξη 3
$P(W_{j,t} \phi_{zj,t})$	0.02	0.49	0.49
Λέξη	Λέξη 1	Λέξη 2	Λέξη 3
$P(W_{j,t} \phi_{zj,t})$	0.1	0.1	0.8
Λέξη	Λέξη 1	Λέξη 2	Λέξη 3
$P(W_{j,t} \phi_{zj,t})$	0.7	0.2	0.1

Για κάθε μία από τις θεματικές ενότητες της τυχαίας δειγματοληψίας (**Error! Reference source not found.**) ανατίθεται, μέσω της **πολυωνυμικής κατανομής** της εκάστοτε θεματικής ενότητας, μία τυχαία λέξη(**3.10.2**). Η πιθανότητα στην θεματική ενότητα i να ανατεθεί η λέξη j , $P(W_{j,t}|\phi_{zj,t})$ παρουσιάζεται στον Πίνακα 3.4 .

Πίνακας 3.5 Λέξεις της δειγματοληψίας

Λέξη 3	Λέξη 3	Λέξη 1	Λέξη 2	Λέξη 3	Λέξη 3	Λέξη 2	Λέξη 1	Λέξη 1	Λέξη 2
--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

Το σύνολο των λέξεων που ανατέθηκαν μαζί με την θεματική ενότητα από την οποία προήλθαν(

Πίνακας 3.5), θεωρείται ένα νέο έγγραφο(Πίνακας 3.6). Συνεπώς από ένα έγγραφο, το LDA δημιούργησε ένα καινούριο. Αυτή η διαδικασία επαναλαμβάνεται για κάθε έγγραφο. Το σύνολο αυτών των εγγράφων αποτελεί ένα νέο **corpus**. Η πιθανότητα που περιγράφεται στην εξίσωση, είναι η πιθανότητα το νέο *corpus* που παράχθηκε να είναι ίδιο με αυτό που χρησιμοποιήθηκε ως είσοδος και στόχος του LDA είναι η μεγιστοποίηση αυτής της πιθανότητας.

Πίνακας 3.6 Νέο Έγγραφο

Θέμα 1	Λέξη 3
Θέμα 2	Λέξη 3
Θέμα 3	Λέξη 1
Θέμα 1	Λέξη 2
Θέμα 2	Λέξη 3
Θέμα 2	Λέξη 3
Θέμα 1	Λέξη 2
Θέμα 3	Λέξη 1
Θέμα 2	Λέξη 1
Θέμα 3	Λέξη 2

Τα δεδομένα του σχήματος μπορούν να ανασχηματιστούν όπως παρουσιάζεται στον Πίνακα 3.7:

Πίνακας 3.7 Κατανομή λέξεων στις θεματικές ενότητες

Θέμα 1	P	Θέμα 2	P	Θέμα 3	P
Λέξη 2	0.666	Λέξη 3	0.750	Λέξη 1	0.666
Λέξη 3	0.250	Λέξη 1	0.333	Λέξη 2	0.333

Για παράδειγμα η **λέξη 2** στο **θέμα 1** εμφανίζεται στο νέο έγγραφο:

- **2** φορές στο **θέμα 1**
- **1** φορά στο **θέμα 3**

Άρα η πιθανότητα η λέξη 2, να εμφανιστεί στο θέμα 2 έχει **P=0.666**, ενώ να εμφανιστεί στο θέμα 3 **P=0.333**.

Ο Πίνακας 3.7 Κατανομή λέξεων στις θεματικές ενότητες αφορά ένα έγγραφο, αλλά διατηρεί την δομή του και για **N** έγγραφα και αποτελεί την έξοδο του LDA.

3.11 AGENT ALLOCATOR

Η μεθοδολογία Agent Allocator έχει ως στόχο την κατανομή των εργασιών, σε ένα περιβάλλον με πολλούς πράκτορες, έχοντας ως βάση τα χαρακτηριστικά του κάθε πράκτορα και την ικανότητα του να εκτελέσει μία εργασία. Η κατανομή προκύπτει από κριτήρια αξιολόγησης που επιλέγει ο αποφασίζοντας.

3.11.1 ΒΑΣΙΚΕΣ ΘΕΩΡΗΣΕΙΣ ΤΗΣ ΜΕΘΟΔΟΥ

Η ανάθεση των εργασιών προκύπτει από το ίδιο το σύστημα. Συνεπώς μεταξύ των υποψήφιων πρακτόρων δεν υπάρχει ιεραρχία, ούτε διαπραγμάτευση καθώς επίσης οι πράκτορες δεν δηλώνουν προτιμήσεις για τις εργασίες. Οι υποψήφιοι πράκτορες είναι προκαθορισμένοι και διαθέτουν ικανότητες, ενώ οι εργασίες διαθέτουν απαιτήσεις. Το σύστημα αξιολογεί όλους τους δυνατούς συνδυασμούς και επιλέγει το βέλτιστο σχέδιο.

3.11.2 ΠΑΡΑΔΟΧΕΣ ΤΗΣ ΜΕΘΟΔΟΥ

Ο στόχος της μεθοδολογίας είναι η ανάθεση k πρακτόρων σε m εργασίες. Το πλήθος των εργασιών μπορεί να είναι μεγαλύτερο, ίσο ή μικρότερο με τον αριθμό των πρακτόρων, ενώ παράλληλα κάθε εργασία ανατίθεται σε έναν και μόνο πράκτορα. Επιπροσθέτως, οι εργασίες περιγράφονται μέσω ενός κοινού συνόλου αναγκών-απαιτήσεων (demands) ενώ οι πράκτορες μπορούν να περιγράφονται μέσω ενός κοινού συνόλου χαρακτηριστικών (attributes).

3.11.3 ΜΕΘΟΔΟΛΟΓΙΑ

Η μεθοδολογία παρουσιάζεται συνοπτικά με τα παρακάτω βήματα:

Βήμα 1 Ανάθεση ενός συνόλου εργασιών $T\{t_1, \dots, t_k\}$ σε ένα σύνολο πρακτόρων $A\{a_1, \dots, a_m\}$

Βήμα 2 Καθορισμός των απαιτήσεων των εργασιών $v(T)$

Βήμα 3 Καθορισμός των χαρακτηριστικών των πρακτόρων $v(A)$

Βήμα 4 Καθορισμός ενός συνεπούς συνόλου κριτηρίων αξιολόγησης $F\{g_1, \dots, g_n\}$, ώστε να υπολογιστεί η τιμή όλων των συνδυασμών ανάθεσης (t_i, a_l) για $i=1, \dots, k$ και $l=1, \dots, m$ και τίθενται $T_0=T$ και $q=0$.

Να σημειωθεί σε αυτό το βήμα πως κάθε κριτήριο αποτελείται από ένα σύνολο υπό-κριτηρίων, τα οποία με την σειρά τους αποτελούνται από συνδυασμούς των απαιτήσεων των εργασιών με τα χαρακτηριστικά των πρακτόρων, ή μεμονωμένα χαρακτηριστικά των πρακτόρων ή απαιτήσεων των εργασιών. Σε κάθε τέτοιο υπό-κριτήριο προσδίδουμε μία αριθμητική βαθμολογία (κανονικοποιημένη στο $[0,1]$) για κάθε πιθανό συνδυασμό επίδοσης (βαθμολογία χαρακτηριστικού) κάποιου πράκτορα με την απαίτηση (βαθμολογία απαίτησης) κάποιας εργασίας. Η αριθμητική βαθμολογία περιορίζεται από τους παρακάτω κανόνες:

- Πράκτορες με υψηλό επίπεδο ικανοτήτων δεν πρέπει να καταναλώνονται σε εργασίες χαμηλών απαιτήσεων.
- Δεν πρέπει να δίνεται μεγάλο πλεονέκτημα σε συνδυασμούς, οι οποίοι δεν είναι απόλυτα ικανοποιητικοί.

- Μη εφικτοί συνδυασμοί αναθέσεων εργασιών βαθμολογούνται με την ελάχιστη δυνατή αριθμητική τιμή(0)

Βήμα 5 Απόδοση βαθμολογίας των κριτηρίων για κάθε συνδυασμό (t_i, a_l) για $i=1, \dots, k$ και $l=1, \dots, m$ όπως φαίνεται στο Σχήμα 3.9 Δομή πίνακα απόδοσης βαθμολογίας συνδυασμών πρακτόρων-εργασιών για κάθε κριτήριο

	g_1			g_i			g_n
t_1, a_1							
...							
...							
...							
t_i, a_l	$g_i(t_i, a_l)$...
...							
...							
...							
t_k, a_m							

Σχήμα 3.9 Δομή πίνακα απόδοσης βαθμολογίας συνδυασμών πρακτόρων-εργασιών για κάθε κριτήριο

Πηγή: Matsatsinis Nikolaos, Delias Pavlos (2003)

Βήμα 6 Σε αυτό το βήμα γίνεται προσδιορισμός της ολικής συνάρτησης χρησιμότητας

$$U(g) = \sum_{j=1}^n p_j u_j(g_j)$$

Καθώς και χρήση αυτής για τον υπολογισμό της συνάρτησης χρησιμότητας για κάθε κριτήριο. Επιπλέον δίνεται μια ενδεικτική διάταξη προτίμησης των εναλλακτικών, καθώς και προσδιορισμός των βαρών με την χρήση της UTA.

Βήμα 7 Σε αυτό το βήμα οι συνδυασμοί εργασιών κατατάσσονται ανάλογα με την επίδοσή τους στην ολική συνάρτηση χρησιμότητας και διαλέγεται ο καλύτερος συνδυασμός (t_i, a_l) .

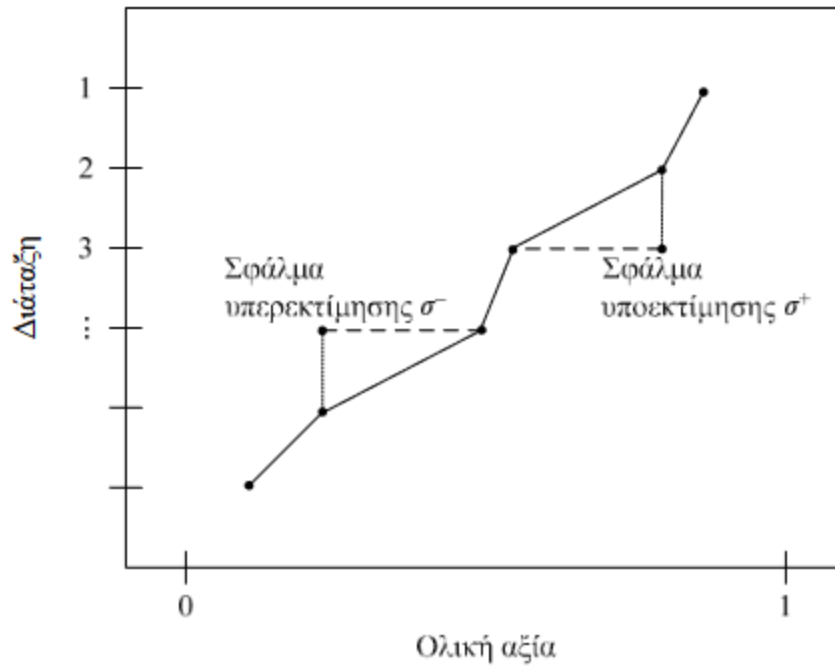
Βήμα 8 Ανάθεση της εργασίας t_i στον πράκτορα a_l .

Βήμα 9 Η εργασία i διαγράφεται από το σύνολο των εργασιών, και αν δεν υπάρχουν άλλες εργασίες στο σύνολο των εργασιών, ο αλγόριθμος τερματίζεται.

Βήμα 10 Τα χαρακτηριστικά του πράκτορα a_i επαναπροσδιορίζονται και γίνεται εκ νέου απόδοση βαθμολογίας κριτηρίων για κάθε συνδυασμό (t_i, a_i) (Βήμα 5)

3.12 Η ΜΕΘΟΔΟΣ UTASTAR

Η μέθοδος UTASTAR προτάθηκε από τους Siskos & Yannacopoulos (1985) και αποτελεί μια βελτιωμένη έκδοση της πρωτότυπης μεθόδου UTA. Στην αρχική έκδοση της μεθόδου UTA (Jacquet-Lagrèze and Siskos, 1982), για κάθε μία δράση $A_R \in A$ ορίζεται ένα μοναδικό σφάλμα, $\sigma(a)$. Αυτή η συνάρτηση σφάλματος δεν είναι επαρκής για την ελαχιστοποίηση της ολικής διασποράς των σημείων στη μονότονη καμπύλη του **Σχήμα 3.10** Καμπύλη ποιοτικής παλινδρόμησης. Το πρόβλημα αφορά τα σημεία που βρίσκονται δεξιά της καμπύλης, από τα οποία θα ήταν προτιμότερο να αφαιρεθεί μια ποσότητα αξίας χωρίς να αυξηθούν οι αξίες των άλλων. Οι Siskos & Yannacopoulos απέδειξαν, σε ένα σύνολο πειραματικών δεδομένων, ότι η UTASTAR δίνει καλύτερα αποτελέσματα από τον πρωτότυπο αλγόριθμο UTA.



Σχήμα 3.10 Καμπύλη ποιοτικής παλινδρόμησης

Πηγή: Siskos, Grigoroudis, Matsatsinis, Figueira, 2005)

Στη μέθοδο UTASTAR, οι Siskos & Yannacopoulos εισάγουν μια διπλή θετική συνάρτηση σφάλματος έτσι ώστε:

$$u'[g(\alpha)] = \sum_{i=1}^n u_i[g_i(\alpha)] - \sigma^+(\alpha) + \sigma^-(\alpha) \forall \alpha \in A_R$$

όπου σ^+ και σ^- είναι τα σφάλματα υποεκτίμησης και υπερεκτίμησης, αντίστοιχα.

Επιπρόσθετα, μια άλλη σημαντική τροποποίηση αφορά τους περιορισμούς μονοτονίας των κριτηρίων, οι οποίοι μοντελοποιούνται με τη βοήθεια των ακόλουθων μετασχηματισμών των μεταβλητών:

$$w_{ij} = u_i(g_i^{j+1}) - u_i(g_i^j) \geq 0 \forall i = 1, 2, \dots, n \text{ και } j = 1, 2, \dots, \alpha_i - 1$$

Με αυτό τον τρόπο, οι συνθήκες μονοτονίας μπορούν να αντικατασταθούν από
 περιορισμούς μη αρνητικότητας των μεταβλητών w_{ij} , άρα ο αλγόριθμος
 UTASTAR συνοψίζεται στα ακόλουθα βήματα:

Βήμα 1: Η ολική αξία των δράσεων $u[g(\alpha_k)]$, $k = 1, 2, \dots, m$, εκφράζεται αρχικά
 ως
 συνάρτηση των περιθωρίων αξιών (marginal utilities) $u_i(g_i)$ και στη συνέχεια
 των μεταβλητών w_{ij} σύμφωνα με την προηγούμενη εξίσωση, μέσω των
 ακόλουθων σχέσεων:

$$\begin{cases} u_i(g_i^1) = 0 \forall i = 1, 2, \dots, n \\ u_i(g_i^j) = \sum_{i=1}^{j-1} w_{ij} \forall i = 1, 2, \dots, n \text{ και } j = 2, 3, \dots, \alpha_i - 1 \end{cases}$$

Βήμα 2: Εισάγονται δύο συναρτήσεις σφάλματος $+\sigma$ και $-\sigma$ στο AR, γράφοντας
 για
 κάθε ζεύγος διαδοχικών δράσεων στην προ διάταξη τις αναλυτικές εκφράσεις:

$$\begin{aligned} \Delta(\alpha_k, \alpha_{k+1}) = & u[g(\alpha_k)] - \sigma^+(\alpha_k) + \sigma^-(\alpha_k) \\ & - u[g(\alpha_{k+1})] + \sigma^+(\alpha_{k+1}) - \sigma^-(\alpha_{k+1}) \end{aligned}$$

Βήμα 3: Επιλύεται το ακόλουθο γραμμικού προγραμματισμού:

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

τους περιορισμούς:

$$\left. \begin{aligned} \Delta(\alpha_k, \alpha_{k+1}) &\geq \delta \text{ αν } \alpha_k > \alpha_{k+1} \\ \Delta(\alpha_k, \alpha_{k+1}) &= 0 \text{ αν } \alpha_k \sim \alpha_{k+1} \end{aligned} \right\} \forall k$$

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

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

Βήμα 4: Ελέγχεται η ύπαρξη πολλαπλών βέλτιστων ή ημιβέλτιστων λύσεων στο γ.π., υπολογίζοντας το βαρύκεντρο των προσθετικών συναρτήσεων αξίας που μεγιστοποιούν τις ακόλουθες αντικειμενικές συναρτήσεις:

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

στο υπέρ-πολύεδρο των περιορισμών του γ.π. που περιορίζεται από τον επόμενο νέο περιορισμό:

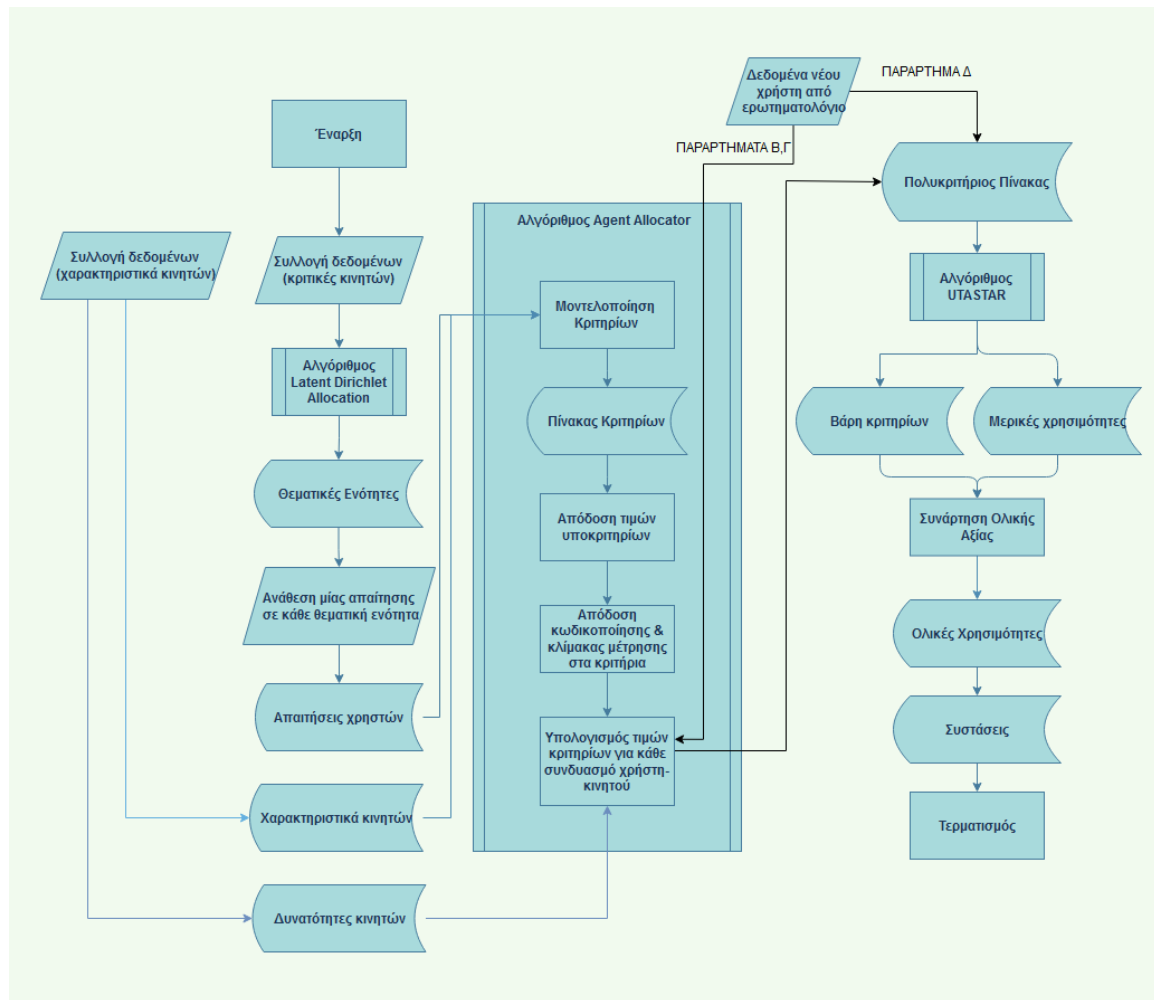
$$\sum_{k=1}^m [\sigma^+(\alpha_k) + \sigma^-(\alpha_k)] \leq z^* + \varepsilon$$

όπου z^* : η βέλτιστη τιμή (σφάλμα) του γ.π. του βήματος 3 και ε : είναι ένας πολύ μικρός θετικός αριθμός ή μηδέν.

ΚΕΦΑΛΑΙΟ 4

ΜΕΘΟΔΟΛΟΓΙΑ ΠΟΛΥΚΡΙΤΗΡΙΟΥ ΣΥΣΤΗΜΑΤΟΣ ΣΥΣΤΑΣΕΩΝ

4.1 ΕΙΣΑΓΩΓΗ



Σχήμα 4.1 Διάγραμμα ροής μεθοδολογίας

Στο Κεφάλαιο 4 παρουσιάζεται η μεθοδολογία που αναπτύχθηκε για την ανάπτυξη του πολυκριτήριου συστήματος συστάσεων, όπως παρουσιάζεται στο Σχήμα 4.1.

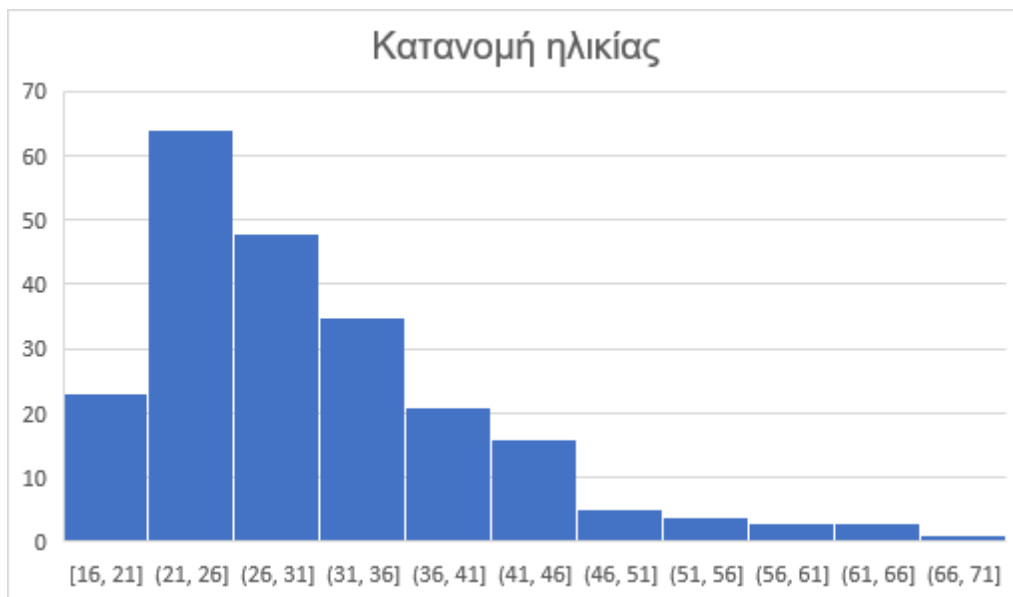
Η μεθοδολογία ξεκινάει με την δημιουργία ενός web crawler, με σκοπό την συλλογή κριτικών που αφορούν κινητά τηλέφωνα. Οι κριτικές χρησιμοποιούνται ως είσοδος για τον αλγόριθμο Latent Dirichlet Allocation, ο οποίος βρίσκει σχέσεις μεταξύ των λέξεων των κριτικών και τις χωρίζει σε θεματικές ενότητες. Οι θεματικές ενότητες ερμηνεύονται ως απαιτήσεις των καταναλωτών κατά την επιλογή κινητού τηλεφώνου. Στην συνέχεια βρίσκονται τα χαρακτηριστικά των κινητών που καλύπτουν αυτές τις απαιτήσεις, καθώς και οι τιμές των χαρακτηριστικών αυτών για ένα σύνολο κινητό τηλεφώνων (δυνατότητες κινητών). Για την καλύτερη κατανόηση των απαιτήσεων των καταναλωτών, δημιουργήθηκε ερωτηματολόγιο από το οποίο αντλήθηκαν πολύτιμες πληροφορίες σχετικά τον βαθμό των απαιτήσεων των καταναλωτών, καθώς και τις προτιμήσεις τους.

Έπειτα για την εφαρμογή της μεθοδολογίας Agent Allocator, κάθε απαίτηση των καταναλωτών αντιστοιχίζεται με ένα χαρακτηριστικό κινητών τηλεφώνων, δημιουργώντας υποκριτήρια, τα οποία ομαδοποιούνται σε κριτήρια. Για κάθε ένα υποκριτήριο προσδίδεται μία αριθμητική βαθμολογία, από 0 μέχρι και 1, για κάθε πιθανό συνδυασμό επίδοσης, σύμφωνα με την μεθοδολογία Agent Allocator. Ο Agent Allocator λαμβάνει από τον χρήστη τον βαθμό των απαιτήσεων του, και υπολογίζει το πόσο κάθε ένα από τα κινητά τηλέφωνα του συνόλου αναφοράς καλύπτει τις ανάγκες τους, για κάθε κριτήριο.

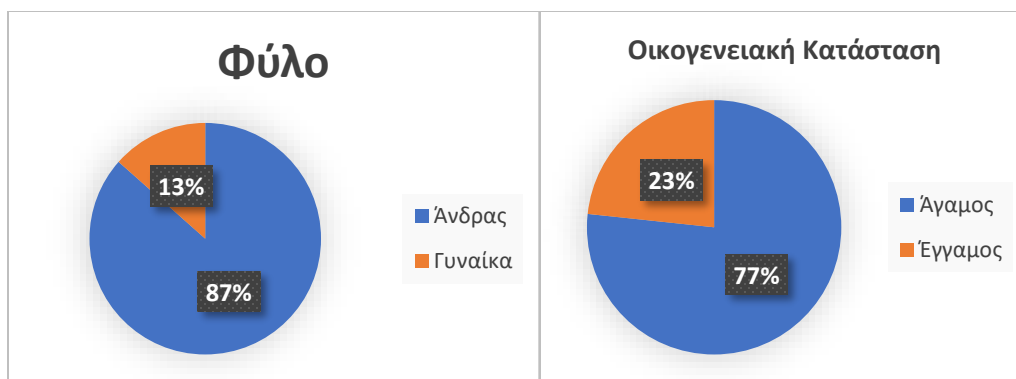
Στην συνέχεια διαμορφώνεται ο πολυκριτήριος πίνακας και εφαρμόζεται η UTASTAR με σκοπό τον υπολογισμό των βαρών των κριτηρίων, και των μερικών χρησιμοτήτων του χρήστη. Μέσω την συνάρτησης ολικής αξίας προκύπτουν οι ολικές αξίες για ένα μεγαλύτερο σύνολο κινητών τηλεφώνων, το σύνολο των κινητών τηλεφώνων της αγοράς, σύμφωνα με τις οποίες δημιουργούνται οι συστάσεις για τον χρήστη.

4.2 ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Το δείγμα της έρευνας αποτελείται από 223 ανθρώπους που ζουν στην Ελλάδα και των οποίων τα δημογραφικά χαρακτηριστικά παρουσιάζονται παρακάτω. Το μεγαλύτερο ποσοστό των ερωτηθέντων(87%) είναι άντρες, ενώ το 77% του δείγματος δήλωσε πως είναι έγγαμο. Το **Σχήμα 4.2** Ηλικιακή Κατανομή εμφανίζει την κατανομή της ηλικίας, και δείχνει πως το μεγαλύτερο μέρος του δείγματος έχει ηλικία από 22 μέχρι 36 έτη.



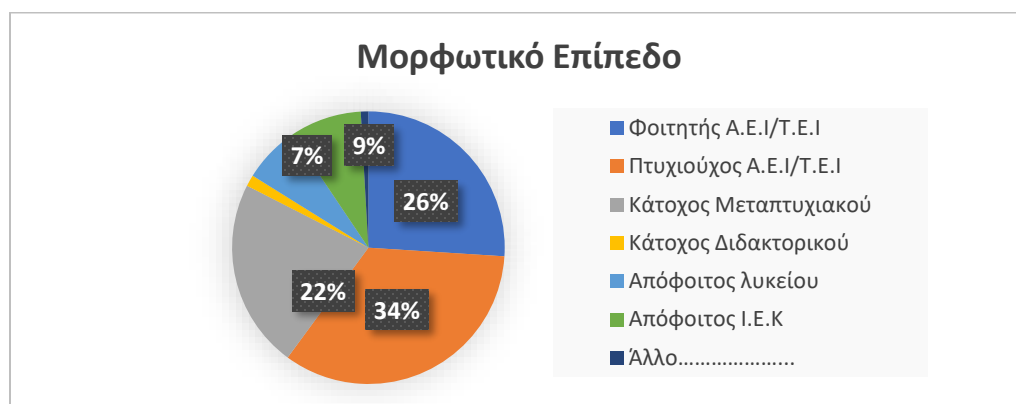
Σχήμα 4.2 Ηλικιακή Κατανομή



Σχήμα 4.3 Ποσοστιαία Κατανομή Φύλων

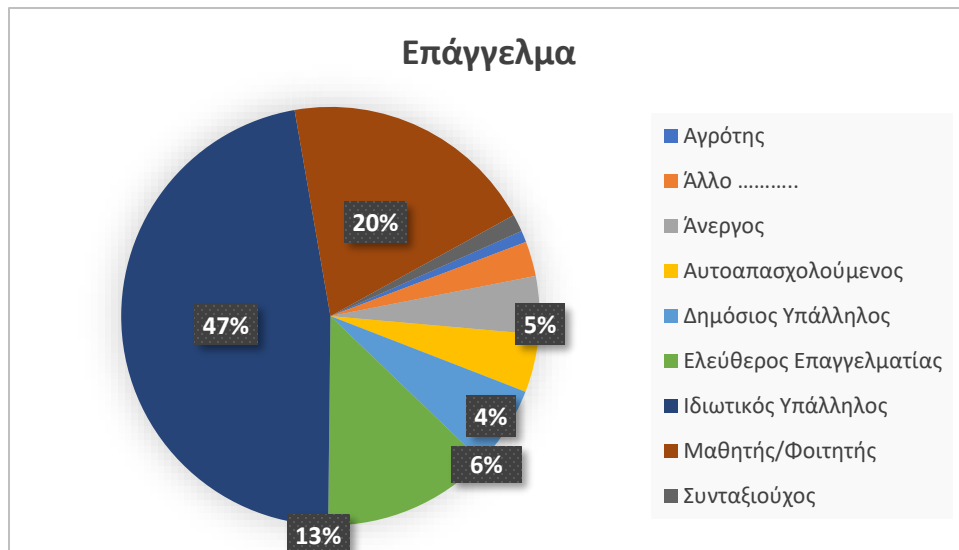
Σχήμα 4.4 Οικογενειακή κατάσταση

Σύμφωνα με το **Σχήμα 4.5**, το 60% των ερωτώμενων ή βρίσκεται στην διάρκεια των σπουδών του, ή έχει ολοκληρώσει τις σπουδές του σε κάποιο ΑΕΙ ή ΤΕΙ, ενώ το 22% είναι κάτοχος μεταπτυχιακού διπλώματος.

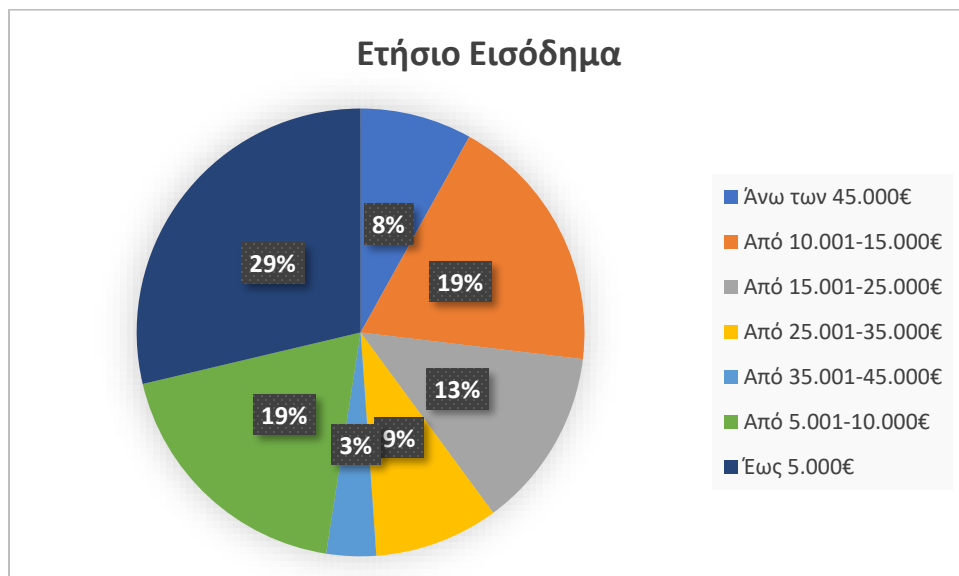


Σχήμα 4.5 Μορφωτικό επίπεδο

Στα πλαίσια της απασχόλησης τους, το 47% δήλωσε πως εργάζεται στον ιδιωτικό τομέα, το 20% είναι μαθητές ή φοιτητές και το 13% ελεύθεροι επαγγελματίες(**Σχήμα 4.6**). Το μεγαλύτερο ποσοστό των ερωτηθέντων(47%) λαμβάνει ετήσιες αποδοχές έως 5.000 ευρώ, το 38% από 5.000 έως 15.000 ευρώ, ενώ για το 15% το ποσό του εισοδήματος ξεπερνάει τις 15.000 ευρώ.(**Σχήμα 4.7**)

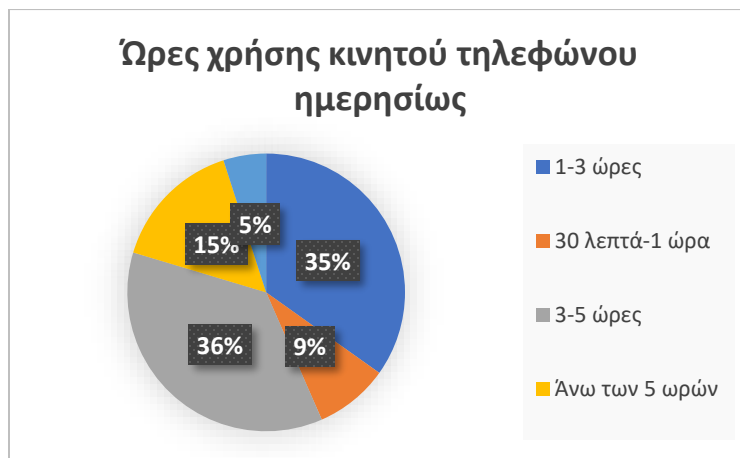


Σχήμα 4.6 Επάγγελμα

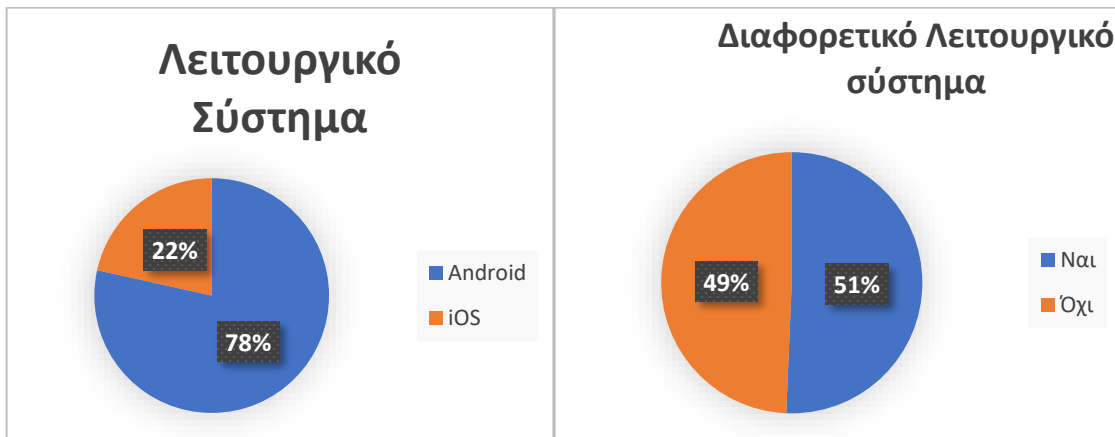


Σχήμα 4.7 Ετήσιο Εισόδημα

Ζητήθηκε από τους ερωτώμενους να δηλώσουν τις προτιμήσεις τους σε μία σειρά από χαρακτηριστικά κινητών. Από τις απαντήσεις τους προέκυψε πως οι περισσότεροι (78%) προτιμούν το λειτουργικό σύστημα “Android”, ενώ οι υπόλοιποι διαλέγουν λογισμικό σύστημα “IOS”(Σχήμα 4.9). Ένα ενδιαφέρον στοιχείο της έρευνας αφορά την επιλογή ενός διαφορετικού λειτουργικού. Πάνω από τους μισούς ερωτώμενους(51%) είναι πρόθυμοι να αγοράσουν ένα κινητό τηλέφωνο με διαφορετικό λειτουργικό σύστημα(Σχήμα 4.10).



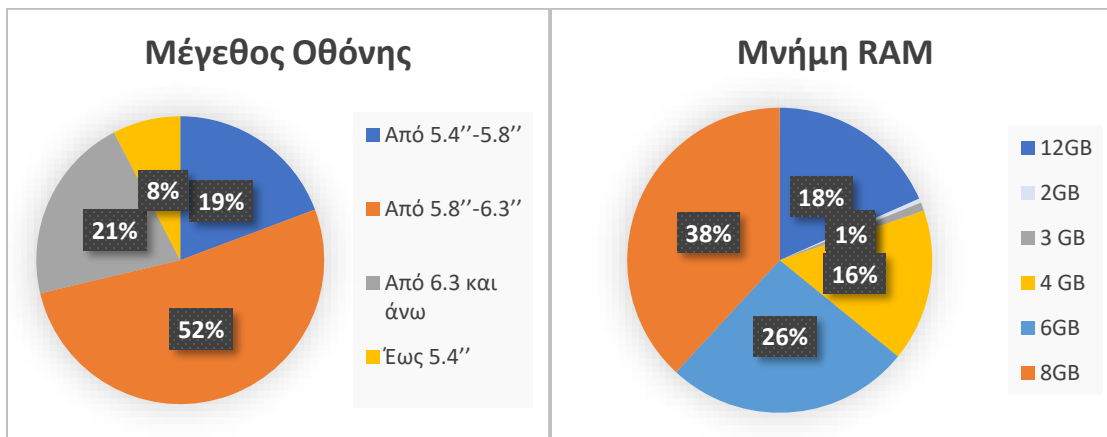
Σχήμα 4.8 Ώρες χρήσης κινητού τηλεφώνου ημερησίως



Σχήμα 4.9 Λειτουργικό Σύστημα Κινητού Τηλεφώνου

Σχήμα 4.10 Αποδοχή διαφορετικού λειτουργικού

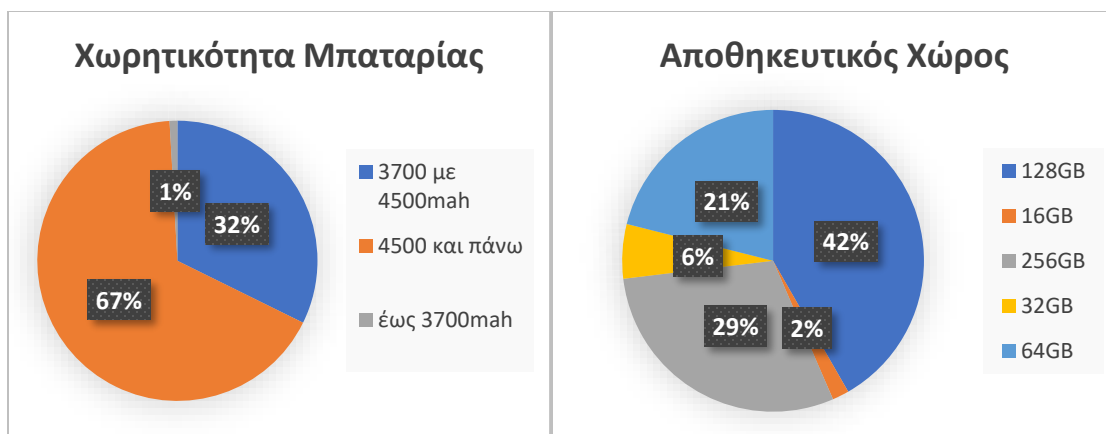
Η έρευνα έδειξε πως διαφορετικές τιμές των χαρακτηριστικών καλύπτουν τις ανάγκες του κάθε ερωτώμενου. Ενδεικτικά, στα πλαίσια της επιλογής μεγέθους οθόνης, το μεγαλύτερο ποσοστό (52%) δήλωσε πως προτιμάει μέγεθος οθόνης 5.8"-6.3", με το 19% να προτιμούν μικρότερο μέγεθος(5.4"-5.8") και 21% να δηλώνει πως το μέγεθος οθόνης που θα καλύψει τις ανάγκες τους θα πρέπει να είναι πάνω από 6.3"(Σχήμα 4.11). Όσον αφορά την μνήμη Ram, το 38% επέλεξε μνήμη 8GB, και το 26% μνήμη 6GB(Σχήμα 4.12). Στα επόμενα σχήματα παρουσιάζονται οι τιμές των χαρακτηριστικών των κινητών τηλεφώνων που οι ερωτώμενοι προτιμούν.



Σχήμα 4.11 Μέγεθος Οθόνης

Σχήμα 4.12 Μνήμη RAM

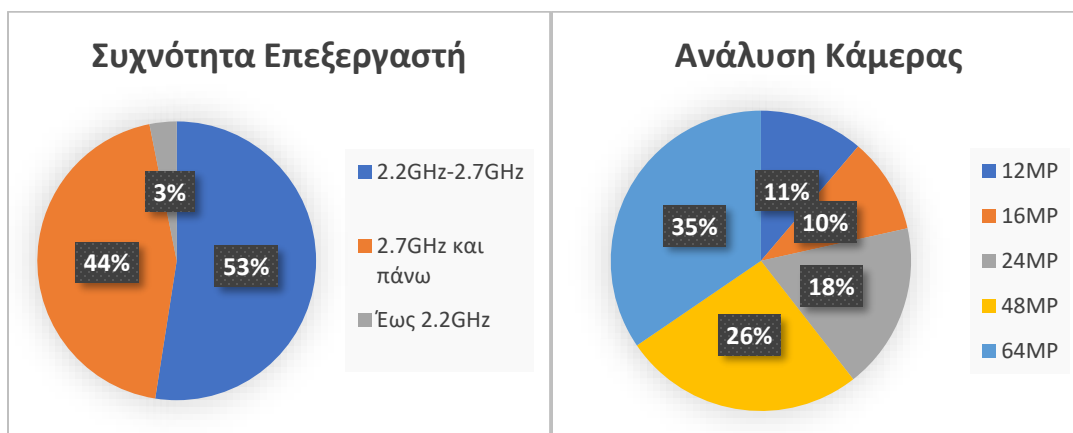
Η έρευνα έδειξε πως οι περισσότεροι προτιμούν κινητά τηλέφωνα με χωρητικότητα μπαταρίας μεγαλύτερη των 4.500 mah(67%), ενώ για το 32% αρκεί μία μπαταρία με χωρητικότητα από 3700mah μέχρι 4500mah. Αντιθέτως χωρητικότητα μπαταρίας έως 3700mah φαίνεται πως καλύπτει ένα πολύ μικρό μέρος των ερωτώμενων(Σχήμα 4.13).



Σχήμα 4.13 Χωρητικότητα Μπαταρίας

Σχήμα 4.14 Αποθηκευτικός Χώρος

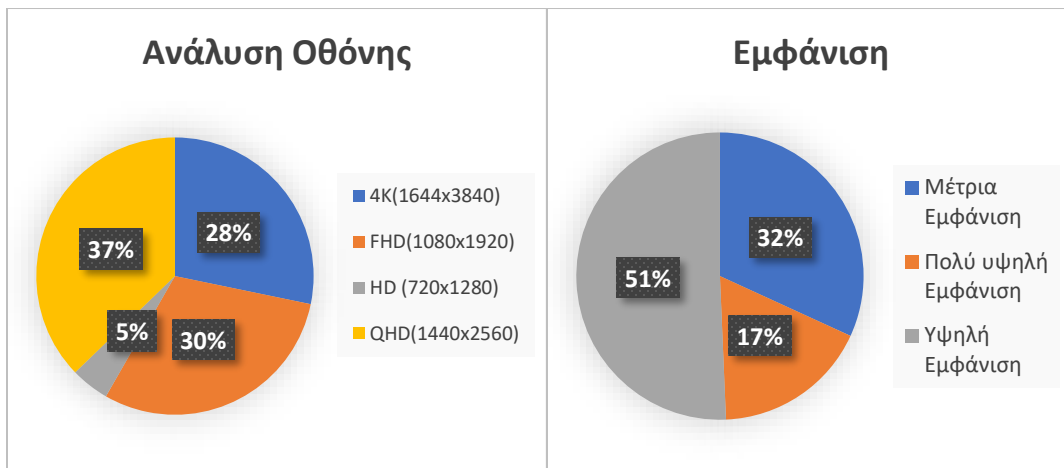
Σχετικά με την συχνότητα του επεξεργαστή, σύμφωνα με το δείγμα το 53% θεωρεί πως κινητά τηλέφωνα με συχνότητα 2.2-2.7GHz θα ικανοποιήσουν τις ανάγκες τους και το 44% χρειάζεται συχνότητα μεγαλύτερη των 2.7GHz. Κινητά τηλέφωνα με συχνότητα επεξεργαστή μικρότερη των 2.2GHz δεν φαίνεται να καλύπτουν τις ανάγκες περισσότερων από 3% των ανθρώπων που συμμετείχαν στην έρευνα(Σχήμα 4.15).



Σχήμα 4.15 Συχνότητα Επεξεργαστή

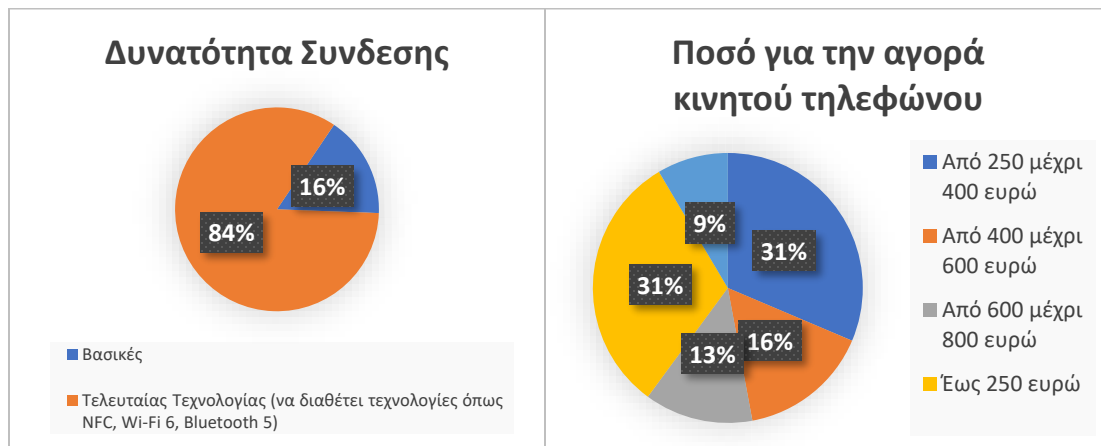
Σχήμα 4.16 Ανάλυση Κάμερας

Όσον αφορά τα χαρακτηριστικά της Ανάλυσης Κάμερας, Ανάλυσης Οθόνης, του Αποθηκευτικού Χώρου και του ποσού διάθεσης για την αγορά κινητού τηλεφώνου, τα σχήματα **Σχήμα 4.14**, **Σχήμα 4.16**, **Σχήμα 4.17** και **Σχήμα 4.20** αποτυπώνουν πως διαφορετικές τιμές των χαρακτηριστικών καλύπτουν τον διαφορετικό βαθμό απαιτήσεων των καταναλωτών. Στα πλαίσια της συνδεσιμότητας, το μεγαλύτερο ποσοστό του δείγματος φαίνεται να βασίζεται στις τελευταίες τεχνολογίες όπως (wi-fi 6 και Bluetooth 5, NFC) για ταχύτερη και ασφαλέστερη μεταφορά δεδομένων. (Σχήμα 4.19). Για το χαρακτηριστικό της εμφάνισης, το 68% επέλεξε από υψηλή, έως πολύ υψηλή εμφάνιση, σε αντίθεση με το 32% που προτίμησε μία μέτρια εμφάνιση, σύμφωνα με το **Σχήμα 4.18**.



Σχήμα 4.17 Ανάλυση Οθόνης

Σχήμα 4.18 Εμφάνιση



Σχήμα 4.19 Δυνατότητες Σύνδεσης

Σχήμα 4.20 Ποσό για την αγορά κινητού τηλεφώνου

4.3 LATENT DIRICHLET ALLOCATION

Χρησιμοποιήθηκε η βιβλιοθήκη **Gensim** της Python για την εφαρμογή του αλγορίθμου του LDA, ενώ για την γραφική απεικόνιση των αποτελεσμάτων, χρησιμοποιήθηκε η βιβλιοθήκη **pyLDavis**.

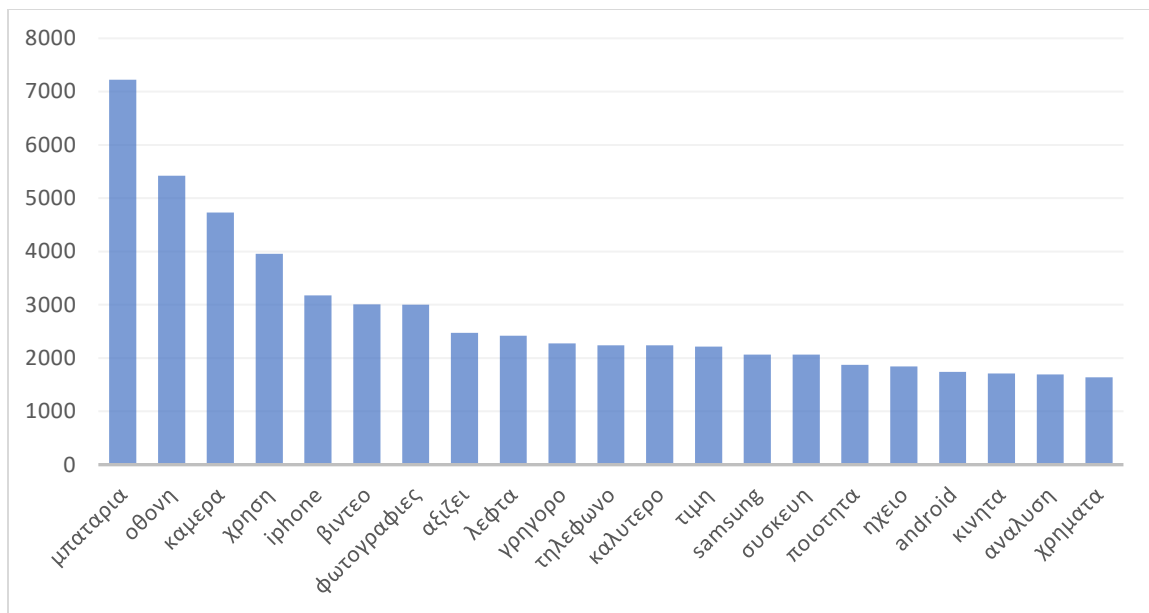
Το μοντέλο της Gensim για τον αλγόριθμο του LDA, λαμβάνει ως παραμέτρους το σύνολο των εγγράφων (**corpus**), τον αριθμό των θεμάτων K (**num_topics**), την θέση της κάθε λέξης σε κάθε έγγραφο (**id2word**), την παράμετρο α (**alpha**), και την παράμετρο β (**eta**).

Ως έγγραφα για την δημιουργία του corpus, χρησιμοποιήθηκαν τα κείμενα από 6073 κριτικές από την ιστοσελίδα skroutz.gr για 25 κινητά τηλέφωνα. Στην συνέχεια ακολούθησε ο καθαρισμός των δεδομένων αυτών, για να χρησιμοποιηθούν ως είσοδος στον αλγόριθμο LDA. Ο καθαρισμός των δεδομένων αποτελεί μία απαραίτητη και κρίσιμη διαδικασία, καθώς η ποιότητα των δεδομένων που θα εισέλθουν στον αλγόριθμο, κρίνει την ποιότητα των εξόδων, δηλαδή των θεματικών ενοτήτων. Η διαδικασία αυτή περιλαμβάνει τα συνήθη βήματα στις μεθόδους επεξεργασίας της φυσικής γλώσσας:

- Μετατροπή των κειμένων των κριτικών σε λέξεις.

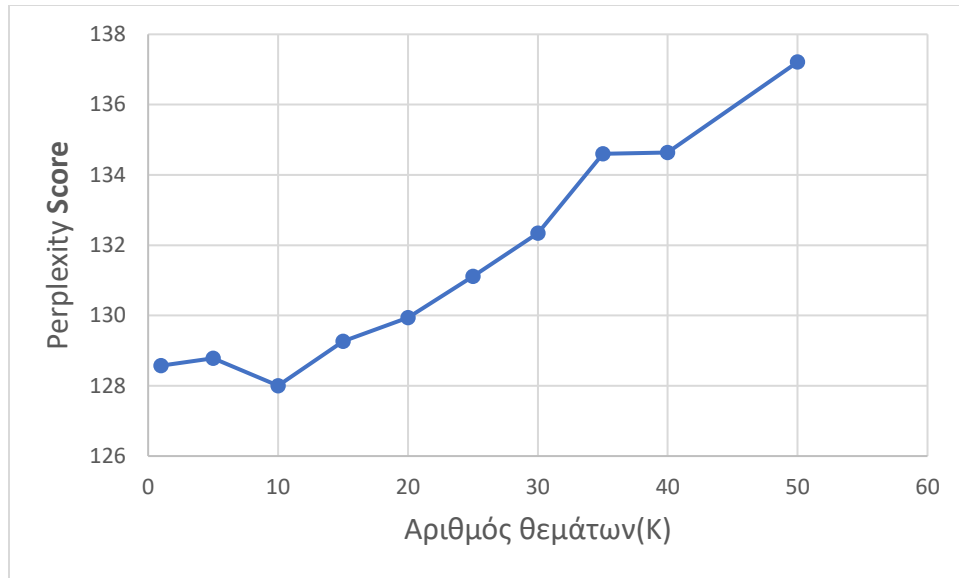
- Αντικατάσταση όλων των κεφαλαίων χαρακτήρων των λέξεων των κειμένων των κριτικών με μικρά γράμματα.
- Αφαίρεση των σημείων στίξης, καθώς και των τόνων από τις λέξεις.
- Αφαίρεση συμβόλων και ειδικών χαρακτήρων.
- Μετατροπή όλων των χαρακτήρων σε κωδικοποίηση UTF-8.
- Ομαλοποίηση των λέξεων. Πραγματοποιήθηκε μετατροπή ρημάτων από διάφορους χρόνους στον Ενεστώτα. Για παράδειγμα η λέξη 'άξιζε' αντικαταστάθηκε από την λέξη 'αξίζει'.
- Αφαίρεση των πιο κοινών λέξεων της ελληνικής γλώσσας. Αυτές οι λέξεις (stop words) δεν προσθέτουν σημαντική πληροφορία, συνεπώς το φιλτράρισμα αυτών βοηθάει στην δημιουργία καλύτερων θεματικών ενοτήτων(clusters). Χρησιμοποιήθηκε μια λίστα με τις πιο γνωστές ελληνικές λέξεις, όπως 'μας', 'είναι', 'εγώ'.

Στο σύνολο των εγγράφων(corpus) που δημιουργήθηκε, κάποιες λέξεις εμφανίζονται συχνότερα από άλλες. Ακολουθεί το γράφημα(**Σχήμα 4.21**) με τις 21 λέξεις με την μεγαλύτερη συχνότητα εμφάνισης.



Σχήμα 4.21 Διάγραμμα συχνοτήτων λέξεων

Για την επιλογή των κατάλληλων παραμέτρων του LDA για το συγκεκριμένο μοντέλο, έγινε δοκιμή τιμών για τις παραμέτρους α και β , ενώ για τον προσδιορισμό του βέλτιστου αριθμού θεματικών ενοτήτων (K), υπολογίστηκε η αμηχανία (perplexity score) για διάφορες τιμές της παραμέτρου K . Ο βαθμός αμηχανίας είναι η μέτρηση την ικανότητας πρόβλεψης ενός πιθανολογικού μοντέλου, με χαμηλότερες τιμές να εκφράζουν ακριβέστερη ικανότητα πρόβλεψης.



Σχήμα 4.22 Διάγραμμα *Perplexity* ανά αριθμό θεματικών ενοτήτων

Από το **Σχήμα 4.22** συμπεραίνεται πως η επιλογή 10 θεματικών ενοτήτων θα ευνοήσει το μοντέλο, καθώς μετά από αυτόν τον αριθμό αυτό αυξάνεται το υπολογιστικό κόστος καθώς και η αμηχανία(perplexity).

Για τον καθορισμό των παραμέτρων α και β , έγινε χρήση της βαθμολογίας συνοχής (coherence score), η οποία χρησιμοποιείται για να μετρηθεί το μέτρο της ερμηνευσιμότητας των θεματικών ενοτήτων. Σύμφωνα με το οποίο, για τις παραμέτρους α, β επιλέχθηκαν $\alpha=0.7$ και $\beta=0.1$.

4.4 ΑΠΟΤΕΛΕΣΜΑΤΑ LDA

Ακολουθεί ο πίνακας με την κατανομή των λέξεων στις 10 θεματικές ενότητες που προέκυψε από τον αλγόριθμο LDA, όπως περιγράφεται στην ενότητα **3.10.5**, με τις 25 λέξεις που έχουν την μεγαλύτερη πιθανότητα εμφάνισης σε κάθε θεματική ενότητα, καθώς και την πιθανότητα τους να εμφανιστούν σε αυτή.

Πίνακας 4.1Θεματικές ενότητες LDA

Θέματική Ενότητα 1: Οθόνη		Θέματική Ενότητα 2: Μπαταρία		Θέματική Ενότητα 3: Συνδεσιμότητα		Θέματική Ενότητα 4: Επεξεργαστής	
Λέξεις	P	Λέξεις	P	Λέξεις	P	Λέξεις	P
οθονη	0.06	χρηση	0.089	εφαρμογες	0.027	γρηγορο	0.022
χερια	0.019	μερα	0.082	google	0.021	οθονη	0.019
διαφορα	0.019	μπαταρια	0.069	android	0.014	φωτογραφια	0.015
χρωματ α	0.016	βγαζει	0.053	εκδοση	0.011	καλυτερη	0.014
καμερα	0.016	ανετα	0.028	εφαρμογη	0.011	μπαταρια	0.013
μεγεθος	0.015	καμερα	0.026	youtube	0.01	καμερα	0.012
max	0.012	βαρια	0.019	συσκευη	0.01	ποιοτητα	0.01
αναλυσ η	0.011	κραται	0.019	app	0.01	προβλημα	0.01
mode	0.011	δυνατο	0.017	huawei	0.009	λειτουργια	0.009
xs	0.011	τιμη	0.015	bluetooth	0.009	αρνητικα	0.008
μπαταρι α	0.01	ωρες	0.014	τηλεφωνο	0.009	iphone	0.008
μεγαλη	0.01	μετρια	0.013	services	0.009	αναλυση	0.008
φορτιση	0.01	βιντεο	0.013	ειδικα	0.007	βιντεο	0.007
καλυτερ η	0.009	ηχος	0.012	gps	0.006	εξαιρετικη	0.007
night	0.009	ηχειο	0.012	wifi	0.006	συσκευη	0.007
μερες	0.008	id	0.011	updates	0.006	χερι	0.007
βιντεο	0.008	νυχτα	0.011	επαγγελματικη	0.006	χρηστης	0.007
συσκευ η	0.008	οθονη	0.009	προσοχη	0.006	φορτισω	0.006
Φωτογρα -φια	0.008	αφογο	0.009	μουσικη	0.006	απιστευτο	0.006
pro	0.007	ποιοτητας	0.009	μπορειτε	0.005	ηχειο	0.006
διπλα	0.007	κολλαει	0.009	εγκατασταση	0.005	τιμη	0.006
φωτεινοτη α	0.007	ταχυτητα	0.008	χρειασει	0.005	xiaomi	0.006
καμερες	0.006	παραπονο	0.008	gb	0.005	χρωματα	0.005
φακο	0.006	face	0.008	σωστα	0.005	επεξεργαστη	0.005
premiu m'	0.006	νομιζω	0.008	play	0.005	χρονου	0.005

Πίνακας 4.2 Θεματικές ενότητες LDA

Θέματική Ενότητα 5:Εμφάνιση		Θέματική Ενότητα 6:Ποιότητα Κατασκευής		Θέματική Ενότητα 7: Μάρκα		Θέματική Ενότητα 8: Λογισμικό	
Λέξεις	P	Λέξεις	P	Λέξεις	P	Λέξεις	P
οθονη	0.036	συσκευη	0.023	note	0.034	χρονια	0.054
θηκη	0.026	τηλεφωνο	0.022	samsung	0.031	iphones	0.052
αποτυπω μα	0.019	τιμη	0.016	pro	0.029	apple	0.04
πισω	0.015	οθονη	0.015	καλυτερη	0.024	android	0.029
φορτιση	0.012	επιλογη	0.013	redmi	0.02	ios	0.024
ακουστικ α	0.011	χαρακτηριστικ α	0.012	9	0.019	κινητα	0.021
s7	0.009	μοντελο	0.01	καλυτερο	0.013	for	0.019
μεγαλο	0.009	λειτουργιες	0.01	κινητα	0.013	value	0.019
τζαμακι	0.009	εξαιρετο	0.01	s10	0.012	money	0.019
αρνητικο	0.009	συσκευες	0.009	xiaomi	0.012	μπαταρια	0.017
δακτυλικο	0.008	εξαιρετικα	0.009	καμερες	0.011	plus	0.016
αμεσα	0.008	προβληματα	0.009	huawei	0.01	6s	0.014
samsung	0.007	λειτουργικου	0.009	τιμη	0.01	καλυτερη	0.011
γρηγορα	0.007	ποιοτητας	0.008	galaxy	0.009	μπαταριες	0.01
edge	0.007	γενικες	0.008	plus	0.009	λογισμικο	0.008
θετικα	0.007	αγορα	0.008	ευρω	0.008	χρηση	0.008
ηχειο	0.007	ιδιαιτερα	0.007	λογισμικ ο	0.008	τελευταια	0.008
γρηγορος	0.006	μπορω	0.007	snapdrag on	0.007	αλλαξω	0.008
ωραια	0.006	κληση	0.007	s9	0.007	στιγμη	0.007
λειτουργι ες	0.006	220	0.007	p30	0.007	επομενα	0.007
κατω	0.006	κατασκευη	0.007	lite	0.007	se	0.007
αναγνωρι ση	0.006	υπολοιπα	0.007	ναυαρχιδ ες	0.007	καμερα	0.006
καμερες	0.006	δυνατοτητες	0.006	mi	0.006	παρω	0.006
καμερα	0.006	προηγουμενο	0.006	mate	0.006	μικρη	0.006
σημα	0.006	2019	0.006	καμερα	0.006	220	0.006

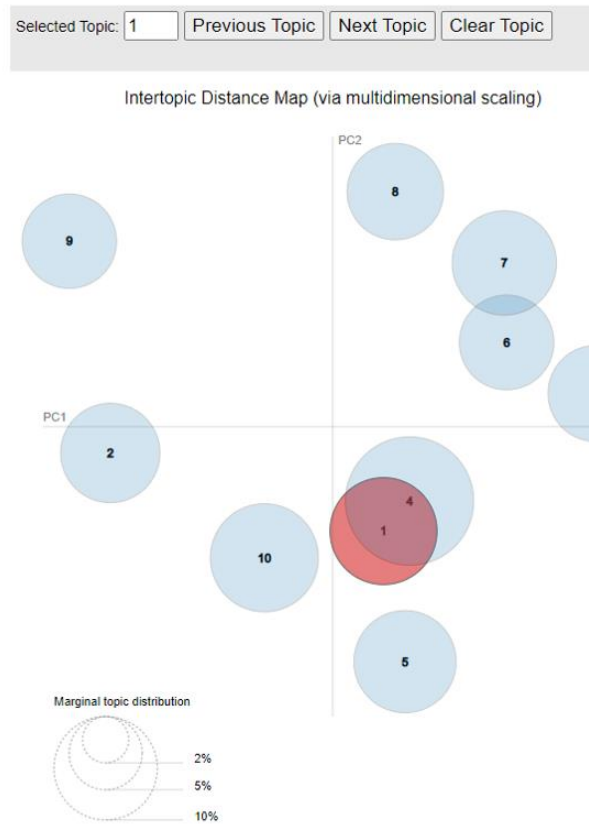
Πίνακας 4.3 Θεματικές ενότητες LDA

Θέματική Ενότητα 9: Τιμή/Κόστος

Θέματική Ενότητα 10: Ποιότητα λήψης
Φωτογραφίας/Βίντεο

Λέξεις	P	Λέξεις	P
αξιζει	0.07	βιντεο	0.029
λεφτα	0.064	φωτογραφια	0.025
μπαταρια	0.051	παιχνιδια	0.02
χρηματα	0.044	καμερα	0.019
μηνες	0.033	χρονο	0.019
τηλεφωνο	0.033	γρηγορος	0.018
κολλαει	0.031	μπορω	0.015
μερες	0.025	απιστευτα	0.014
ευρω	0.024	μπαταρια	0.014
τελειο	0.022	ηχειο	0.013
καμερα	0.018	γρηγορα	0.013
ευχαριστημενος	0.018	κινητου	0.012
ανεπιφυλακτα	0.018	προβληματα	0.011
τελεια	0.017	αναλυση	0.011
κραταει	0.016	πιστευω	0.009
αγορασα	0.015	μουσικη	0.009
καλυτερη	0.012	social	0.009
προτεινω	0.012	ικανοποιημενη	0.008
μεγαλη	0.012	επεξεργαστης	0.008
συνιστω	0.009	ηχεια	0.007
μηνα	0.009	ωρα	0.007
εβδομαδα	0.008	100	0.007
χαρα	0.008	media	0.007
200	0.007	αρνητικο	0.006
συστηνω	0.007	καταπληκτικο	0.006

Για την καλύτερη κατανόηση των αποτελεσμάτων, δημιουργήθηκε γραφική απεικόνιση των θεματικών ενοτήτων, με την χρήση της βιβλιοθήκης **pyldavis**.



Σχήμα 4.23 Γραφική Απεικόνιση θεματικών ενοτήτων

Στην παραπάνω γραφική απεικόνιση (**Σχήμα 4.23**), κάθε κύκλος συμβολίζει μία θεματική ενότητα. Το μέγεθος τους καθορίζει το ποσοστό των αναφορών στο αρχείο λέξεων, σχετικά με το θέμα που αντιπροσωπεύουν. Η απόσταση των θεματικών ενοτήτων, συμβολίζει την σχέση μεταξύ των θεμάτων. Για παράδειγμα υπάρχει σχετικά μεγάλη συσχέτιση μεταξύ της 1^{ης} και 4^{ης} θεματικής ενότητας, ενώ πολύ μικρότερη μεταξύ της 6^{ης} και της 7^{ης}.

4.5 ΑΝΑΘΕΣΗ ΑΠΑΙΤΗΣΕΩΝ ΚΑΤΑΝΑΛΩΤΗ ΣΕ ΚΑΘΕ ΘΕΜΑΤΙΚΗ ΕΝΟΤΗΤΑ

Οι λέξεις με την μεγαλύτερη συχνότητα κάθε θεματικής ενότητας, παρουσιάζουν μία απαίτηση των καταναλωτών. Βρίσκεται η απαίτηση των καταναλωτών που περιγράφουν οι λέξεις κάθε θεματικής ενότητας. Για παράδειγμα η 9^η θεματική ενότητα περιγράφει την αξία του κινητού σε σχέση με τα χρήματα που δαπάνησε ο καταναλωτής. Οι απαιτήσεις των χρηστών που περιγράφονται στις θεματικές ενότητες παρουσιάζονται στον **Πίνακας 4.4**.

Πίνακας 4.4 Θεματική Περιγραφή Συστάδας

Θεματική Ενότητα	Απαίτηση
1	Οθόνη
2	Διάρκεια Μπαταρίας
3	Συνδεσιμότητα
4	Ταχύτητα Επεξεργασίας
5	Εμφάνιση
6	Ποιότητα Κατασκευής
7	Προτίμηση Μάρκας
8	Δυνατότητες Λογισμικού
9	Τιμή/Κόστος
10	Ποιότητα λήψης Φωτογραφίας/Βίντεο

Για να είμαστε σε θέση να κατανοήσουμε σε μεγαλύτερο βάθος τις απαιτήσεις ενός καταναλωτή και να καταλήξουμε σε καλύτερες συστάσεις, κάποιες απαιτήσεις χωρίζονται σε επιμέρους απαιτήσεις.

Η απαίτηση της οθόνης σε:

- Ανάλυση οθόνης
- Μέγεθος Οθόνης

Η απαίτηση της ταχύτητας επεξεργασίας σε:

- Ταχύτητα προσωρινής μνήμης
- Ταχύτητα επεξεργαστή

Η απαίτηση που περιγράφεται από τις λέξεις της 6ης συστάδας, παραλήφθηκε λόγω έλλειψης δεδομένων σχετικά με την ποιότητα κατασκευής των κινητών τηλεφώνων.

Οι τελικές απαιτήσεις των καταναλωτών όπως προέκυψαν στο Σχήμα 4.24 Απαιτήσεις Καταναλωτών

Απαιτήσεις Καταναλωτών
<ul style="list-style-type: none"> • Κόστος • Αποθηκευτικός χώρος • Λειτουργικό σύστημα • Ταχύτητα προσωρινής μνήμης • Διάρκεια Μπαταρίας • Ποιότητα λήψης Φωτογραφίας & Βίντεο • Ταχύτητα επεξεργασίας • Ανάλυση οθόνης • Εμφάνιση • Μέγεθος Οθόνης

- Δυνατότητα Συνδεσιμότητας
- Προτίμηση Μάρκας

Σχήμα 4.24 Απαιτήσεις Καταναλωτών

4.6 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΚΙΝΗΤΩΝ

Στην συνέχεια ακολούθησε συλλογή δεδομένων με την βοήθεια ενός νέου web crawler που αναπτύχθηκε στην γλώσσα προγραμματισμού Python με την βοήθεια των βιβλιοθηκών **Selenium** και **BeautifulSoup** για τα χαρακτηριστικά των κινητών. Επιλέχθηκε η ιστοσελίδα Skroutz.gr αφού διαθέτει μεγάλη ποικιλία τεχνικών χαρακτηριστικών κινητών. Στόχος ήταν να βρεθούν μόνο τα χαρακτηριστικά κινητών που καλύπτουν τις απαιτήσεις των καταναλωτών.

Ακολουθεί ο πίνακας με συγκεντρωμένα τα χαρακτηριστικά των κινητών(Πίνακας 4.5 Χαρακτηριστικά Κινητών).

Πίνακας 4.5 Χαρακτηριστικά Κινητών

Χαρακτηριστικά Κινητών
<ul style="list-style-type: none"> • Μέγεθος οθόνης • Λειτουργικό Σύστημα • Τιμή • Μνήμη RAM • Χωρητικότητα ROM • Χωρητικότητα Μπαταρίας • Ανάλυση Κάμερας • Συχνότητα Επεξεργαστή • Ανάλυση οθόνης • Βαθμολογία Εμφάνισης • Βαθμολογία Δικτύου & Συνδεσιμότητας • Μάρκα

4.7 ΔΥΝΑΤΟΤΗΤΕΣ ΚΙΝΗΤΩΝ

Έγινε συλλογή των τιμών των χαρακτηριστικών για κάθε ένα από 33 κινητά από την ιστοσελίδα skroutz.gr, και κατόπιν οι τιμές χωρίστηκαν σε διαστήματα με σκοπό να κατηγοριοποιηθούν.

Για την μέτρηση της **Δυνατότητας Συνδεσιμότητας** και της **Εμφάνισης** των κινητών, έγινε συλλογή των αντίστοιχων βαθμολογιών από την www.dxomark.com.

Χωρίζονται τα κινητά τηλέφωνα σε δύο σύνολα. Το σύνολο αναφοράς(Πίνακας 4.7) και το σύνολο κινητών αγοράς (Πίνακας 4.6)

Πίνακας 4.6Κινητά συνόλου αγοράς

Όνομα	Ανάλυση Κάμερας	Μνήμη RAM	Αποθηκευτικός Χώρος(GB)	Μπαταρία (mAh)	CPU (GHz)	Ανάλυση Οθόνης (Px)	Βαθμολογία Συνδεσιμότητας	Βαθμολογία Εμφάνισης	Τιμή (ευρώ)
Xiaomi Redmi 5 Plus	12	4	64	4000	2.00	2160	6	5.9	149
Samsung Galaxy A71	64	6	128	4500	2.2	2400	7.4	7.6	320
Xiaomi Redmi Note 9	48	3	64	5020	2.00	2340	7.6	6.8	137
Xiaomi Redmi Note 8 Pro	64	6	64	4500	2.05	2340	7.3	7.5	174
Xiaomi Redmi Note 7	48	4	64	4000	2.20	2340	6.5	6.4	179
Xiaomi Poco M4 Pro	50	6	128	5000	2.4	2400	8.2	8.6	207
Xiaomi Redmi 8	12	4	64	5000	1.95	1520	6	5.7	169
Xiaomi Redmi Note 9 Pro	64	6	64	5020	2.30	2400	7.4	7.3	192
Xiaomi Mi A2	12	4	64	3010	2.20	2160	6.1	6.3	184
Xiaomi Redmi Note 8T	48	4	64	4000	2.00	2340	7.3	6.5	154
Samsung Galaxy A7	24	4	64	3300	2.20	2220	6.7	7.3	279
Apple iPhone SE	12	2	64	1642	1.84	1136	7.1	5.7	719
Xiaomi Redmi 7	48	4	64	5020	2.30	2400	5.8	5.5	169
Apple iPhone 11	12	4	64	3110	2.65	1792	8.1	7.9	655
Huawei P20 Lite Dual	16	4	64	3000	2.36	2280	5.8	6	199
Samsung Galaxy S9+	12	6	64	3500	2.70	2960	7.9	7.4	958
Samsung Galaxy S10+	12	8	512	4100	2.70	3040	8.3	7.6	846
Samsung Galaxy S10+	12	12	1024	4100	2.70	3040	8.3	8.8	1629
Huawei Mate 20 Lite	20	4	64	3750	2.20	2340	5.9	5.4	189
Xiaomi Mi 8	12	6	64	3400	2.80	2248	7.3	7	251
Samsung Galaxy S8+	12	4	64	3500	2.30	2960	7.7	7.6	560
Apple iPhone X	12	3	64	2716	2.39	2436	7.4	7	850
Huawei Nova 5T	48	6	128	3750	2.60	2340	7.3	6.8	258
Apple iPhone XR	12	3	64	2942	2.50	1792	7.5	6.9	581
Xiaomi Mi 9T	48	6	64	4000	2.20	2340	7.2	7.9	337

Πίνακας 4.7 Κινητά συνόλου αναφοράς

Όνομα	Ανάλυση Κάμερας	Μνήμη RAM	Αποθηκευτικός Χώρος(GB)	Μπαταρία (mAh)	CPU (GHz)	Ανάλυση Οθόνης (Px)	Βαθμολογία Συνδεσιμότητας	Βαθμολογία Εμφάνισης	Τιμή (ευρώ)
Apple iPhone 12 Pro	12	6	128	2815	3.1	2532	8.5	9.1	1230
Samsung Galaxy A71	64	6	128	4500	2.2	2400	7.4	7.6	320
OnePlus Nord	48	12	256	4115	2.4	2400	7.8	8.4	450
Samsung Galaxy A51	48	4	128	4000	2.3	2340	7.6	7	240
Xiaomi Poco X3	64	6	64	5160	2.3	2400	7.7	7	200
Xiaomi Redmi Note 9	13	3	32	5020	2	2340	6.8	7	140
Apple iPhone 12 Mini	12	4	64	2227	3.1	2340	8.1	9	760
Samsung Galaxy S20+	64	8	128	4500	2.73	3200	9.3	8	730

4.8 AGENT ALLOCATOR

Ο αλγόριθμος Agent Allocator απαιτεί ως είσοδο τις απαντήσεις που έδωσε ο χρήστης στο ΠΑΡΑΡΤΗΜΑ Β, τα χαρακτηριστικά των κινητών, τις απαιτήσεις των χρηστών και τις δυνατότητες των κινητών που χωρίστηκαν σε 2 σύνολα. Το πρώτο σύνολο αποτελεί το σύνολο αναφοράς(**Πίνακας 4.7**) και περιέχει 9 γνωστά στον χρήστη κινητά τηλέφωνα, τα οποία κατέταξε κατά την συμπλήρωση του Ερωτηματολογίου. Το δεύτερο σύνολο αποτελείται από 25 κινητά τηλέφωνα(**Πίνακας 4.6**) τα οποία ο χρήστης ενδέχεται να μην γνωρίζει, και αποκαλείται «σύνολο κινητών τηλεφώνων αγοράς». Μέσω του αλγορίθμου Agent Allocator θα προκύψει η βαθμολογία που αποδίδεται στον συνδυασμό χρήστης-κινητό, για έναν χρήστη και για κάθε κινητό τηλέφωνο των δύο συνόλων.

4.8.1 ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΡΙΤΗΡΙΩΝ

Για την μοντελοποίηση των κριτηρίων, απαιτούνται τα χαρακτηριστικά-δυνατότητες των κινητών, καθώς και οι απαιτήσεις των καταναλωτών.

Για την μοντελοποίηση αρχικά γίνεται αντιστοίχιση κάθε μίας **απαίτησης** με ένα **χαρακτηριστικό**, θεωρώντας αυτόν τον συνδυασμό ως ένα **υποκριτήριο** και κατόπιν γίνεται κατηγοριοποίηση των συνδυασμών απαιτήσεων-χαρακτηριστικών σε 7 κριτήρια που θα χρησιμοποιηθούν για την εκτίμηση κάθε ζεύγους χρήστη-κινητού. Οι συνδυασμοί θα αξιολογηθούν στα κριτήρια εμφάνισης, επιδόσεων, απεικόνισης, διάρκειας χρήσης, προτίμησης μάρκας, λειτουργικού και σχέσης ποιότητας τιμής, όπως παρουσιάζεται στον Πίνακα(**Πίνακας 4.8**)

Πίνακας 4.8 Πίνακας Κριτηρίων

Πίνακας Κριτηρίων						
Επιδόσεις	Εμφάνιση	Απεικόνιση	Διάρκεια Χρήσης	Προτίμηση Μάρκας	Σχέση ποιότητας -τιμής	Λειτουργικό
Χωρητικότητα προσωρινής μνήμης Μνήμη RAM	Εμφάνιση & Design Βαθμολογία Εμφάνισης	Ανάλυση οθόνης Ανάλυση οθόνης	Διάρκεια μπαταρίας Χωρητικότητα Μπαταρίας	Προτίμηση Μάρκας Μάρκα	Κόστος Τιμή	Λειτουργικό σύστημα Λειτουργικό Σύστημα
Ταχύτητα επεξεργασίας Συχνότητα Επεξεργαστή		Μέγεθος Οθόνης Μέγεθος οθόνης				
Δυνατότητα Συνδεσιμότητας Βαθμολογία Δικτύου & Συνδεσιμότητας		Ποιότητα λήψης Φωτογραφίας & Βίντεο Ανάλυση Κάμερας				
Αποθηκευτικός χώρος Χωρητικότητα ROM						

4.8.2 ΑΠΟΔΟΣΗ ΤΙΜΩΝ ΥΠΟΚΡΙΤΗΡΙΩΝ

Για κάθε ένα υποκριτήριο προσδίδεται μία αριθμητική βαθμολογία, από 0 μέχρι και 1, για κάθε πιθανό συνδυασμό επίδοσης (βαθμολογία χαρακτηριστικού) ενός κινητού με την απαίτηση (βαθμολογία απαίτησης) ενός χρήστη, τηρώντας τους εξής περιορισμούς:

- Αποδίδεται τιμή 0 σε συνδυασμούς που αφορούν κινητά με καλύτερα χαρακτηριστικά από αυτά που απαιτούνται.
- Δεν αποδίδεται βαθμολογία 1 σε συνδυασμούς που δεν είναι απόλυτα ικανοποιητικοί.
- Όλες οι μη εφικτοί συνδυασμοί αναθέσεων βαθμολογούνται με 0.

- Βαθμολογία 1 αποδίδεται σε συνδυασμούς που είναι απολύτως ικανοποιητικοί.

Στους πίνακες τιμών υποκριτηρίων, όπως ο **Πίνακας 4.9**, οι **γραμμές αφορούν απαιτήσεις του χρήστη, ενώ οι στήλες χαρακτηριστικά κινητών.**

Αυτοί οι πίνακες θα χρειαστούν στη συνέχεια για τον υπολογισμό βαθμολογιών κριτηρίων για κάθε συνδυασμό χρήστη-κινητού.

Ένας νέος χρήστης εισέρχεται στο σύστημα, συμπληρώνοντας το ΠΑΡΑΡΤΗΜΑ Β του ερωτηματολογίου, δίνει την τιμή που προτιμά για κάθε χαρακτηριστικό κινητού (απαίτηση χρήστη).

Αυτή η ποσοτική τιμή, θα μετατρέπεται σε ποιοτική, ανάλογα με το εύρος που ανήκει και θα αντιστοιχεί σε μία απαίτηση των καταναλωτών. Για παράδειγμα αν δηλώσει πως προτιμά 2GB μνήμη RAM, αυτό θα τροποποιείται σε Χαμηλή (1GB-3GB) Ταχύτητα προσωρινής μνήμης.

Παράλληλα κατηγοριοποιούνται και οι τιμές των χαρακτηριστικών των κινητών π.χ. στην Μνήμη RAM σε 4 διαστήματα [Έως 4GB, 4GB με 6GB, 6GB με 8GB, 8GB και πάνω] Αυτή η διαδικασία πραγματοποιείται, λαμβάνοντας υπόψιν τα διαστήματα που διαλέγουν sites όπως το www.skroutz.gr και το www.gsmarena.com.

Για κάθε κριτήριο, οι τιμές σε κάθε υποκριτήριο αθροίζονται, και ανάλογα το διάστημα που ανήκει αυτό το άθροισμα, λαμβάνει και την αντίστοιχη κωδικοποίηση.

Πραγματοποιείται απόδοση τιμών για κάθε κριτήριο και για κάθε συνδυασμό απαιτήσεων-χαρακτηριστικών από τα οποία αποτελούνται τα κριτήρια, καθώς και οι κωδικοποιήσεις των κριτηρίων, ξεκινώντας από το 1^ο κριτήριο, τις «**επιδόσεις**».

ΕΠΙΔΟΣΕΙΣ

A1:ΥΠΟΚΡΙΤΗΡΙΟ 1^ο

Πίνακας 4.9 Πίνακας τιμών 1ου Υποκριτηρίου Επιδόσεων

		Χαρακτηριστικά Κινητών			
		Χωρητικότητα προσωρινής μνήμης /Μνήμη RAM			
		Έως 4GB	4GB με 6GB	6GB με 8GB	8GB και πάνω
Απαιτήσεις Χρήστη	Χαμηλή (Έως 4GB)	1	0	0	0
	Μέτρια (4GB με 6GB)	0.5	1	0	0
	Υψηλή (6GB με 8GB)	0	0.5	1	0
	Πολύ Υψηλή(8GB και πάνω)	0	0.1	0.5	1

A2:ΥΠΟΚΡΙΤΗΡΙΟ 2^ο

Πίνακας 4.10 Πίνακας τιμών 2ου Υποκριτηρίου Επιδόσεων

Ταχύτητα επεξεργασίας /Συχνότητα Επεξεργαστή	Έως 2.2GHz	2.2GHz-2.7GHz	2.7GHz και πάνω
Χαμηλή(Έως 2.2GHz)	1	0	0
Μέτρια (2.2GHz-2.7GHz)	0.4	1	0
Υψηλή(2.7GHz και πάνω)	0	0.4	1

A3:ΥΠΟΚΡΙΤΗΡΙΟ 3^ο

Πίνακας 4.11 Πίνακας τιμών 3ου Υποκριτηρίου Επιδόσεων

Δυνατότητα Συνδεσιμότητας /Δίκτυο & Συνδεσιμότητα Score	Έως 7.8/10	(8-10)/10
Μέτρια(Έως 7.8/10)	1	0
Υψηλή((8-10)/10)	0.3	1

A4:ΥΠΟΚΡΙΤΗΡΙΟ 4^ο

Πίνακας 4.12 Πίνακας τιμών 4^{ου} Υποκριτηρίου Επιδόσεων

Αποθηκευτικός χώρος /Χωρητικότητα ROM	Έως 64GB	128GB	πανω από 128GB
Μέτρια (Έως 64GB)	1	0	0
Υψηλή (128GB)	0.3	1	0
Πολύ Υψηλή (πανω από 128GB)	0	0.4	1

Η τιμή του κριτηρίου «Επιδόσεις» προκύπτει:

$$K1=(A1+A2+A3+A4)/4$$

4.8.3 ΑΠΟΔΟΣΗ ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΚΑΙ ΚΛΙΜΑΚΑΣ ΜΕΤΡΗΣΗΣ ΣΤΑ ΚΡΙΤΗΡΙΑ

Ακολουθεί ο πίνακας(ΠΙΝΑΚΑΣ 4.13) με την **κωδικοποίηση** του κριτηρίου 'Επιδόσεις' σύμφωνα με την κλίμακα μέτρησης του ανάλογα την κλίμακα που θα ανήκει το K1, λαμβάνει και την αντίστοιχη κωδικοποίηση. Συνεπώς ένας συνδυασμός χρήστη-κινητού που συγκέντρωσε $K1 = 0.80$ θα βαθμολογείται ως προς το κριτήριο 'Επιδόσεις' με την κωδικοποίηση 4.

ΠΙΝΑΚΑΣ 4.13 : Κλίμακα και Κωδικοποίηση κριτηρίου Επιδόσεις

Επιδόσεις	Κλίμακα Μέτρησης	Κωδικοποίηση
Κακές	0-0,39	1
Μέτριες	0.4-0.65	2
Καλές	0.66-0.8	3
Πολύ καλές	0.81-1	4

ΕΜΦΑΝΙΣΗ

Για το 1^ο υποκριτήριο του κριτηρίου της Εμφάνισης, διαμορφώνεται ο Πίνακας 4.14, ο οποίος περιέχει την απαίτηση του χρήστη για την εμφάνιση ενός κινητού(Μέτρια, Υψηλή, Πολύ Υψηλή), καθώς και την βαθμολογία που κατέχει ένα κινητό ως προς την εμφάνισή του, σε μία κλίμακα από το 1 μέχρι το 10, ενώ η κωδικοποίηση που θα λάβει το κριτήριο της Εμφάνισης, παρουσιάζεται στον **Πίνακας 4.15**

A1:ΥΠΟΚΡΙΤΗΡΙΟ 1^ο

Πίνακας 4.14 Πίνακας τιμών 1^{ου} Υποκριτηρίου Εμφάνισης

Εμφάνιση & Design/ Βαθμολογία Εμφάνισης	Έως 7	Από 7 μέχρι 9	Από 9 μέχρι 10
Μέτρια Εμφάνιση	1	0	0
Υψηλή Εμφάνιση	0.3	1	0
Πολύ υψηλή Εμφάνιση	0	0.6	1

Πίνακας 4.15: Κλίμακα και Κωδικοποίηση κριτηρίου Εμφάνιση

Εμφάνιση	Κλίμακα Μέτρησης	Κωδικοποίηση
Μέτρια	0-0.3	1
Καλή	0.31-0.79	1
Πολύ καλή	0.8-1	2

ΑΠΕΙΚΟΝΙΣΗ

Για το κριτήριο της Απεικόνισης, αποδίδονται τιμές για το υποκριτήριο της Ανάλυσης οθόνης (**Πίνακας 4.16**), το υποκριτήριο του μεγέθους της οθόνης (Πίνακας 4.17), καθώς και το υποκριτήριο σχετικά με την Ανάλυση της Κάμερας (Πίνακας 4.18). Η κλίμακα μέτρησης και η κωδικοποίηση του κριτηρίου «Εμφάνιση», φαίνεται στον **Πίνακας 4.19**

A1:ΥΠΟΚΡΙΤΗΡΙΟ 1^ο

Πίνακας 4.16 Πίνακας τιμών 1^{ου} Υποκριτηρίου Απεικόνισης

Ανάλυση οθόνης	Τουλάχιστον Full HD	Full HD -QHD	QHD και πάνω
Μέτρια(Τουλάχιστον Full HD)	1	0	0
Υψηλή(Full HD - QHD)	0.3	1	0
Πολύ Υψηλή(QHD και πάνω)	0	0.5	1

A2:ΥΠΟΚΡΙΤΗΡΙΟ 2^ο

Πίνακας 4.17 Πίνακας τιμών 2^{ου} Υποκριτηρίου Απεικόνισης

Μέγεθος οθόνης	Από 5.4"-5.8"	Από 5.8"-6.3"	Από 6.3" και άνω
Μικρό(Από 5.4"-5.8")	1	0	0
Μέτριο (Από 5.8"-6.3")	0.4	1	0
Μεγάλο (Από 6.3 και άνω)	0	0.4	1

A3:ΥΠΟΚΡΙΤΗΡΙΟ 3^ο

Πίνακας 4.18 Πίνακας τιμών 3^{ου} Υποκριτηρίου Απεικόνισης

Ποιότητα λήψης Φωτογραφίας & Βίντεο/Ανάλυση Κάμερας	εως 12MP	12 έως 24MP	24-48MP	48MP και πάνω
Χαμηλή(έως 12MP)	1	0	0	0
Μέτρια(12-36MP)	0.4	1	0	0
Υψηλή(36-48MP)	0	0.6	1	0
Πολύ υψηλή (48MP+)	0	0.1	0.5	1

Πίνακας 4.19: Κλίμακα και Κωδικοποίηση κριτηρίου Απεικόνιση

Απεικόνιση	Κλίμακα Μέτρησης	Κωδικοποίηση
-------------------	------------------	--------------

Κακή	0-0,39	1
Μέτρια	0.4-0.65	2
Καλή	0.66-1	3

ΔΙΑΡΚΕΙΑ ΧΡΗΣΗΣ

Το κριτήριο της Διάρκειας χρήσης αποτελείται από το υποκριτήριο της Χωρητικότητας μπαταρίας και αποδίδονται οι τιμές σύμφωνα με τους πίνακες **Πίνακας 4.20** και **Πίνακας 4.21**.

A1:ΥΠΟΚΡΙΤΗΡΙΟ

Πίνακας 4.20 Πίνακας τιμών Υποκριτηρίου Διάρκειας Χρήσης

Διάρκεια Χρήσης	Κλίμακα Μέτρησης	Κωδικοποίηση	
Κακή	0-0,3	1	
Μέτρια	0.31-0.4	2	
Καλή	0.41-1	3	
Χωρητικότητα Μπαταρίας	έως 3700mah	3700 με 4500mah	4500 και πάνω
Μικρή (έως 3700mah)	1	0	0
Μέτρια (3700 με 4500mah)	0.3	1	0
Μεγάλη (4500 και πάνω)	0	0.4	1

Πίνακας 4.21: Κλίμακα και Κωδικοποίηση κριτηρίου Διάρκειας Χρήσης

Για το κριτήριο της «Σχέσης ποιότητας-τιμής» ο συνδυασμός των χρημάτων που είναι διαθέσιμος να δαπανήσει ο χρήστης, με την τιμή των κινητών τηλεφώνων θα αξιολογηθεί σύμφωνα με τους Πίνακες Πίνακας 4.22 και Πίνακας 4.23.

A1:ΥΠΟΚΡΙΤΗΡΙΟ

Πίνακας 4.22 Πίνακας τιμών Υποκριτηρίου Σχέσης ποιότητας-τιμής

Τιμή Κινητού-κόστος	Έως 250 Ευρώ	250-400 Ευρώ	400-600 Ευρώ	600-800 Ευρώ	πάνω από 800 Ευρώ
Πολύ Μικρό (Έως 250 Ευρώ)	1	0	0	0	0
Μικρό (250-400 Ευρώ)	0.3	1	0	0	0
Μέτριο (400-600 Ευρώ)	0.1	0.3	1	0	0
Μεγάλο (Από 600 μέχρι 800 ευρώ)	0	0.1	0.4	1	0
Πολύ Μεγάλο (Πάνω από 800 ευρώ)	0	0	0.2	0.4	1

Πίνακας 4.23: Κλίμακα και Κωδικοποίηση κριτηρίου Σχέση ποιότητας-τιμής

Σχέση ποιότητας-τιμής	Κλίμακα Μέτρησης	Κωδικοποίηση
Κακή	0-0,12	1
Μέτρια	0.13-0.4	2
Καλή	0.41-0.7	3
Πολύ Καλή	0.71-1	4

ΠΡΟΤΙΜΙΣΗ ΜΑΡΚΑΣ

Στο ερωτηματολόγιο έχει ζητηθεί από τον χρήστη να κατατάξει 8 μάρκες κινητών. Στο στάδιο της σύγκρισης των συνδυασμών χρήστη-κινητού, η μάρκα του κινητού θα βαθμολογείται ανάλογα με την θέση που κατέχει στην κατάταξη του χρήστη, όπως παρουσιάζεται στον **Πίνακας 4.24**. Για παράδειγμα, αν ένα κινητό είναι μάρκας 'Χίαomi', και ο χρήστης έχει κατατάξει 2^η αυτή τη μάρκα, τότε ο συνδυασμός αυτός θα βαθμολογείται με 0.8.

A1:ΥΠΟΚΡΙΤΗΡΙΟ

Πίνακας 4.24 Πίνακας τιμών Υποκριτηρίου Προτίμησης Μάρκας

Προτίμηση Μάρκας/Μάρκα	Μάρκα Κινητού
1 ^η	1
2 ^η	0.7
3 ^η	0.4
4 ^η	0.1
5 ^η	0
6 ^η	0
7 ^η	0
8 ^η	0

Πίνακας 4.25 Κλίμακα και Κωδικοποίηση κριτηρίου Μάρκας

Μάρκα	Κλίμακα Μέτρησης	Κωδικοποίηση
Κακή	0-0.1	1
Μέτρια	0.2-0.4	2
Καλή	0.41-0.7	3
Πολύ καλή	0.71-1	4

ΛΕΙΤΟΥΡΓΙΚΟ ΣΥΣΤΗΜΑ

Στον

Πίνακας 4.27 παρουσιάζονται οι τιμές που αποδόθηκαν στους πιθανούς συνδυασμούς στο υποκριτήριο του Λειτουργικού συστήματος και ακολουθεί η κλίμακα μέτρησης στον

Πίνακας 4.27

A1:ΥΠΟΚΡΙΤΗΡΙΟ 1^ο

Πίνακας 4.26 Πίνακας τιμών Υποκριτηρίου Λειτουργικού συστήματος

Λειτουργικό Σύστημα	IOS	Android
IOS	1	0.2
Android	0.4	1

Πίνακας 4.27 Κλίμακα και Κωδικοποίηση κριτηρίου Λειτουργικό Σύστημα

Λειτουργικό σύστημα	Κλίμακα Μέτρησης	Κωδικοποίηση
Κακή	0-0,2	1
Μέτρια	0.21-0.4	2
Καλή	0.41-01	3

4.9 ΠΟΛΥΚΡΙΤΗΡΙΟΣ ΠΙΝΑΚΑΣ

Από την μεθοδολογία του Agent Allocator, προκύπτει ένας πίνακας όπου κάθε γραμμή αντιστοιχεί σε έναν συνδυασμό χρήστη-κινητού για κάθε ένα από τα κινητά του συνόλου αναφοράς, ενώ οι στήλες αντιστοιχούν στα κριτήρια. Κάθε στοιχείο του πίνακα a_{ij} περιέχει την βαθμολογία του συνδυασμού i , στο κριτήριο j . Χρήστης έχει κατατάξει τα κινητά τηλέφωνα του συνόλου αναφοράς, συνεπώς η προσθήκη του πίνακα της κατάταξης του Χρήστη, στον πίνακα Agent Allocator θα έχει ως αποτέλεσμα τον πολυκριτήριο πίνακα του Χρήστη.

Στους παρακάτω πίνακες(Πίνακας 4.28, Πίνακας 4.29, Πίνακας 4.30), παρουσιάζονται οι πολυκριτήριοι πίνακες για 3 χρήστες, για τα 8 κινητά τηλέφωνα του συνόλου αναφοράς.

Πίνακας 4.28 Πολυκριτήριος Πίνακας για τον Χρήστη 1

Χρήστης	Εναλλακτική	Επιδόσεις	Απεικόνιση	Διάρκεια Χρησης	Εμφάνιση	Λειτουργικό	Σχέση Ποιότητας-Τιμής	Προτίμηση Μάρκας	Κατάταξη
Χρήστης 1	Apple iPhone 12 Pro	2	1	2	1	1	1	1	4
Χρήστης 1	Samsung Galaxy A71	2	1	3	1	3	1	1	4
Χρήστης 1	OnePlus Nord	2	1	3	1	3	1	3	1
Χρήστης 1	Samsung Galaxy A51	2	1	3	3	3	4	1	4
Χρήστης 1	Xiaomi Poco X3	3	1	1	3	3	4	1	2
Χρήστης 1	Xiaomi Redmi Note 9	3	1	1	3	3	4	1	3
Χρήστης 1	Apple iPhone 12 Mini	2	2	2	1	1	1	1	4
Χρήστης 1	Samsung Galaxy S20+	2	1	3	1	3	1	1	4

Πίνακας 4.29 Πολυκριτήριο Πίνακας για τον Χρήστη 2

Χρήστης	Εναλλακτική	Επιδόσεις	Απεικόνιση	Διάρκεια Χρησης	Εμφάνιση	Λειτουργικό	Σχέση Ποιότητας-Τιμής	Προτίμηση Μάρκας	Κατάταξη
Χρήστης2	Apple iPhone 12 Pro	3	2	1	1	1	1	1	5
Χρήστης2	Samsung Galaxy A71	4	2	2	1	3	2	4	1
Χρήστης2	OnePlus Nord	3	2	2	1	3	3	3	3
Χρήστης2	Samsung Galaxy A51	4	2	2	3	3	1	4	2
Χρήστης2	Xiaomi Poco X3	4	2	3	3	3	1	3	4
Χρήστης2	Xiaomi Redmi Note 9	2	1	3	3	3	1	3	4
Χρήστης2	Apple iPhone 12 Mini	2	2	1	1	1	2	1	5
Χρήστης2	Samsung Galaxy S20+	3	3	2	1	3	4	4	1

4.10 UTASTAR

Ο Πολυκριτήριο πίνακας για έναν χρήστη μαζί με τον πίνακα μεταδεδομένων των κριτηρίων χρησιμοποιούνται ως είσοδος στην Utastar, η οποία θα υπολογίσει για τον Χρήστη τις μερικές χρησιμότητες των κριτηρίων, καθώς και τα βάρη των κριτηρίων. Μέσω της συνάρτησης ολικής αξίας, υπολογίζονται οι ολικές χρησιμότητες ενός ευρύτερου συνόλου κινητών τηλεφώνων (Πίνακας 4.7). Για την Utastar χρησιμοποιήθηκε ο κώδικας που αναπτύχθηκε από τον Νικόλαο Βαβουλιώτη (2021), στην γλώσσα προγραμματισμού Python.

Ο πίνακας μεταδεδομένων (Πίνακας 4.31) στην πρώτη στήλη εμπεριέχει το όνομα του κάθε κριτηρίου, ενώ στην δεύτερη στήλη περιέχει το είδος της μονοτονίας του κάθε κριτηρίου, όπου αναπαρίσταται με την τιμή 0 για φθίνουσα μονοτονία, και την τιμή 1 για αύξουσα μονοτονία. Ακολουθεί το

Χρήστης	Εναλλακτική	Επιδόσεις	Απεικόνιση	Διάρκεια Χρήσης	Εμφάνιση	Λειτουργικό	Σχέση Ποιότητας-Τιμής	Προτίμηση Μάρκας	Κατάταξη
Χρήστης3	Apple iPhone 12 Pro	4	2	1	1	1	1	1	7
Χρήστης3	Samsung Galaxy A71	3	2	2	1	3	1	4	4
Χρήστης3	OnePlus Nord	2	2	2	1	3	1	2	3
Χρήστης3	Samsung Galaxy A51	3	2	2	1	3	4	4	5
Χρήστης3	Xiaomi Poco X3	3	2	3	1	3	4	3	2
Χρήστης3	Xiaomi Redmi Note 9	1	1	3	1	3	4	3	8
Χρήστης3	Apple iPhone 12 Mini	3	3	1	1	1	1	1	6
Χρήστης3	Samsung Galaxy S20+	4	3	2	1	3	1	4	1

είδος των κριτηρίων με την τιμή 0 να αφορά ποιοτικό κριτήριο και την τιμή 1 να αφορά ποσοτικό κριτήριο. Τα στοιχεία των επόμενων δύο στηλών περιέχουν τις χειρότερες και τις καλύτερες τιμές που μπορεί να λάβει το κάθε κριτήριο και η τελευταία στήλη δηλώνει τον αριθμό των διαστημάτων στα οποία χωρίζονται οι τιμές του κάθε κριτηρίου.

4.11 ΑΠΟΤΕΛΕΣΜΑΤΑ UTASTAR- ΣΥΣΤΑΣΕΙΣ

Η εφαρμογή της Utastar για 3 χρήστες, απέδωσε τον συντελεστή συσχέτισης τ του Kendall, τα βάρη και τις μερικές χρησιμότητες των κριτηρίων καθώς και τις ολικές χρησιμότητες των εναλλακτικών(**Πίνακας 4.32**).

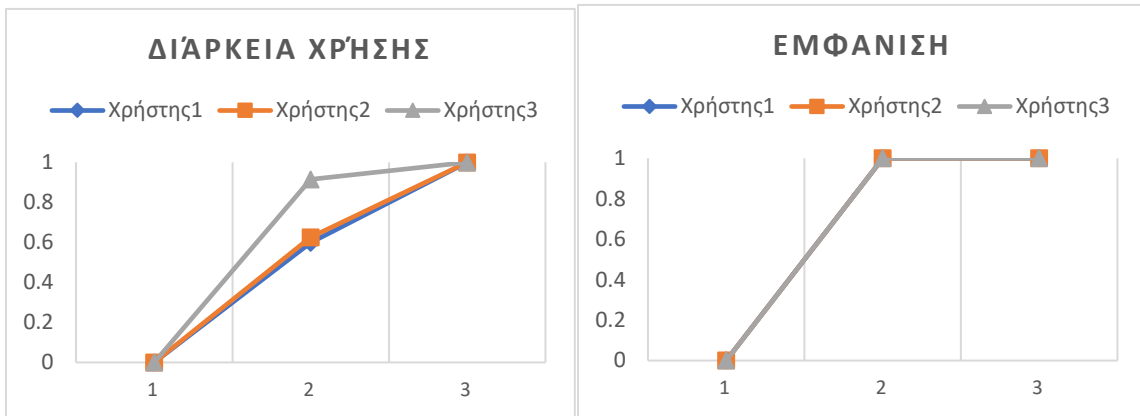
Πίνακας 4.32 Συντελεστής Συσχέτισης τ του Kendall για Χρήστες 1,2,3

Χρήστης	Kendall τ
Χρήστης1	1
Χρήστης2	0.96
Χρήστης3	0.97

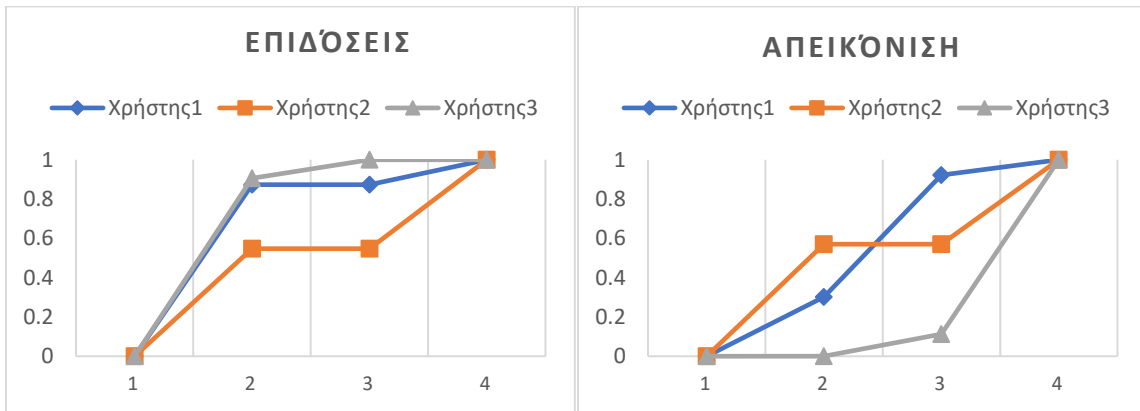
Criteria	Mon/ty	Type	Worst	Best	a
Επιδόσεις	1	1	1	4	4
Απεικόνιση	1	1	1	4	4
Διάρκεια Χρήσης	1	1	1	3	3
Εμφάνιση	1	1	1	3	3
Λειτουργικό	1	1	1	3	3
Σχέση ποιότητας-τιμής	1	1	1	4	4
Προτίμηση Μάρκας	1	1	1	4	4

Ο συντελεστής τ χρησιμοποιείται για τη μέτρηση της συσχέτισης μεταξύ δύο ποσοτήτων, στην προκειμένη περίπτωση οι δύο ποσότητες είναι η κατάταξη των εναλλακτικών από τον Χρήστη, και η κατάταξη των εναλλακτικών από την Utastar. Εφόσον ο συντελεστής κυμαίνεται κοντά στην μονάδα για τους 3 χρήστες, τα αποτελέσματα της Utastar θα είναι αξιόπιστα.

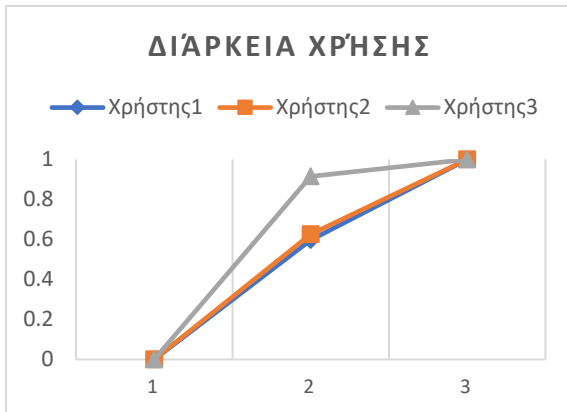
Ακολουθούν οι μερικές χρησιμότητες των χρηστών(Διάγραμμα 4.1,Διάγραμμα 4.2,



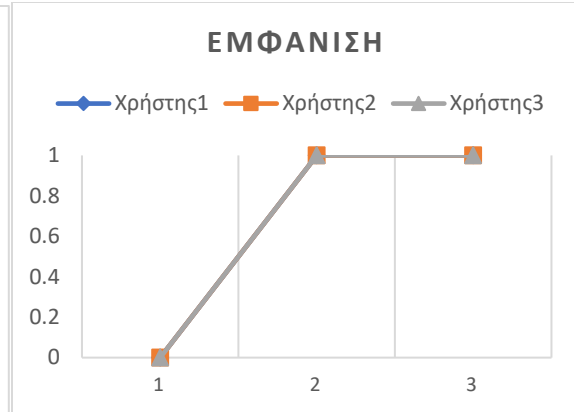
Διάγραμμα 4.3.Διάγραμμα 4.4.Διάγραμμα 4.5.Διάγραμμα 4.6.Διάγραμμα 4.7), όπως προέκυψαν από την εφαρμογή της Utastar. Δηλώνουν την επιπλέον αξία που θα προσέφερε στον χρήστη η απόκτηση μίας επιπλέον μονάδας σε ένα κριτήριο.



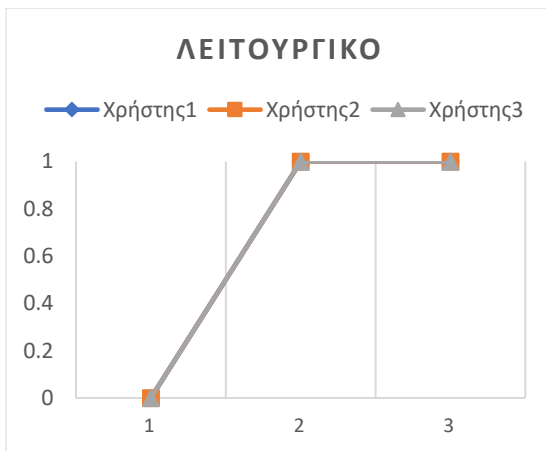
Διάγραμμα 4.1Μερικές Χρησιμότητες κριτηρίου «Επιδόσεις» Διάγραμμα 4.2 Μερικές Χρησιμότητες κριτηρίου "Απεικόνιση"



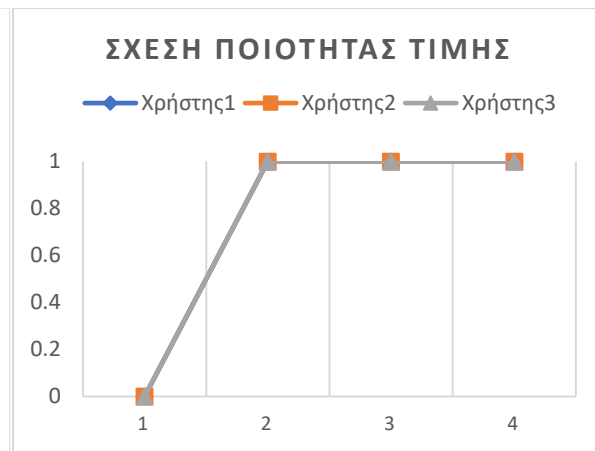
Διάγραμμα 4.3 Μερικές Χρησιμότητες κριτηρίου "Διάρκεια Χρήσης"



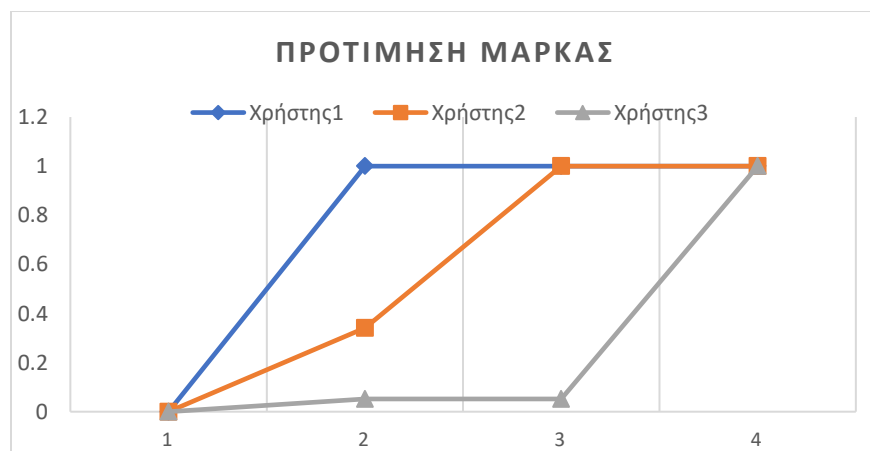
Διάγραμμα 4.4 Μερικές Χρησιμότητες κριτηρίου "Εμφάνιση"



Διάγραμμα 4.5 Μερικές Χρησιμότητες κριτηρίου "Λειτουργικό"

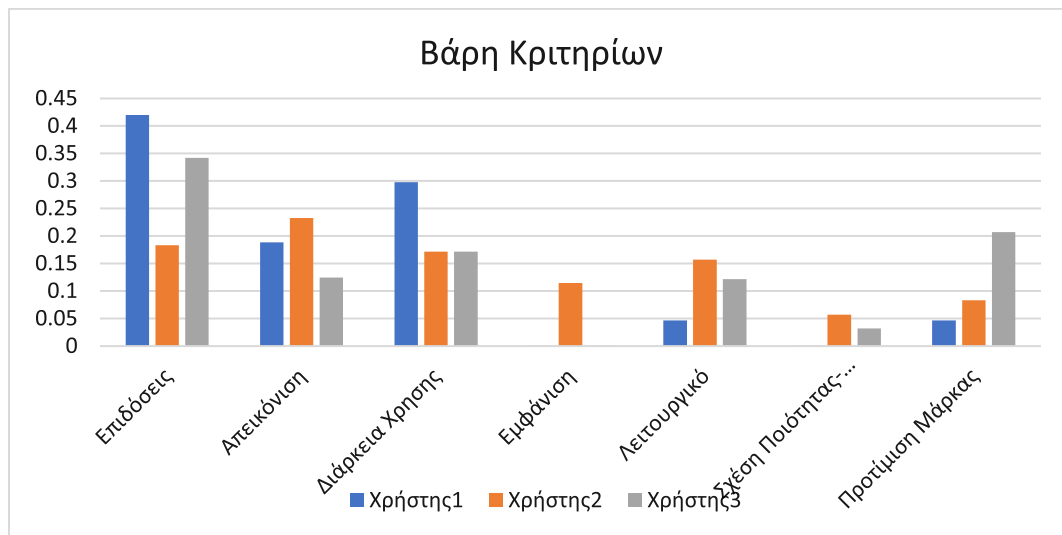


Διάγραμμα 4.6 Μερικές Χρησιμότητες κριτηρίου "Σχέση Ποιότητας-Τιμής"



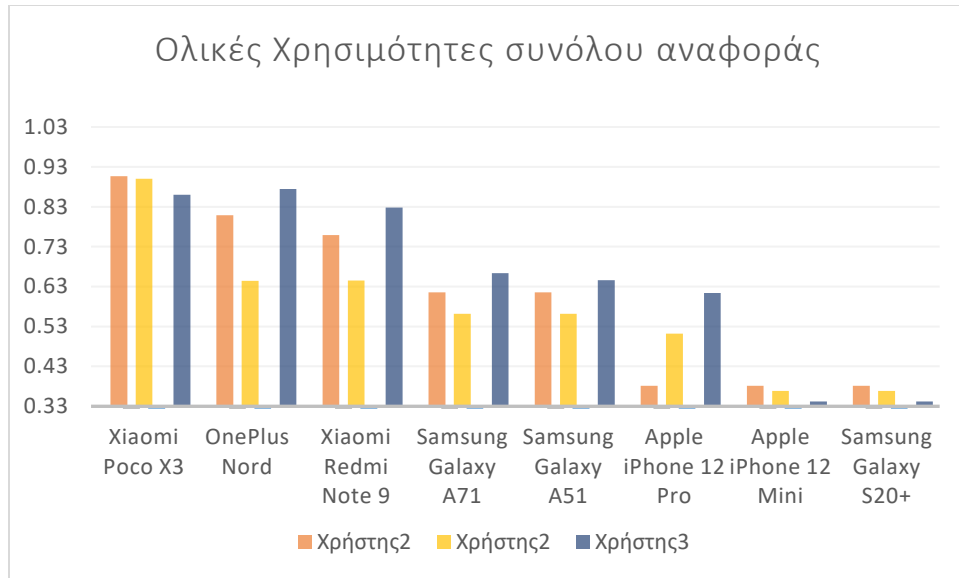
Διάγραμμα 4.7 Μερικές Χρησιμότητες κριτηρίου "Προτίμηση μάρκας"

Τα βάρη των κριτηρίων για κάθε χρήστη δηλώνουν ποια κριτήρια προτιμούν οι χρήστες καθώς και τον βαθμό που τα προτιμούν από τα υπόλοιπα. Στο Διάγραμμα 4.8 φαίνεται πως ο χρήστης1 θεωρεί τα κριτήρια των επιδόσεων, της διάρκειας χρήσης και της απεικόνισης τα σημαντικότερα, και καθόλου σημαντικά τα κριτήρια της εμφάνισης, της σχέσης ποιότητας-τιμής. Ο χρήστης2 θεωρεί σημαντικότερο το κριτήριο της απεικόνισης, ενώ περίπου ίση βαρύτητα έχουν για αυτόν τα κριτήρια των επιδόσεων, της διάρκειας χρήσης και του λειτουργικού. Τέλος ο χρήστης3 δίνει μεγάλη έμφαση στο κριτήριο των επιδόσεων και κατόπιν στην Μάρκα.



Διάγραμμα 4.8 Βάρη κριτηρίων

Ο βαθμός που καλύπτουν τις απαιτήσεις του χρήστη τα κινητά τηλέφωνα του συνόλου αναφοράς, εκφράζονται μέσω της ολικής χρησιμότητας τους και παρατίθενται στο επόμενο σχήμα(**Σχήμα 4.8** Ώρες χρήσης κινητού τηλεφώνου ημερησίως)για τους 3 χρήστες. Λόγω των διαφορετικών απαιτήσεων των χρηστών, καθώς και της βαρύτητας που αποδίδουν σε αυτές, οι ολικές χρησιμότητες κάθε εναλλακτικές διαφέρουν για κάθε χρήστη.



Σχήμα 4.25 Ολικές Χρησιμότητες συνόλου αναφοράς κινητών τηλεφώνων

Οι ολικές χρησιμότητες του συνόλου αναφοράς των κινητών τηλεφώνων (Σχήμα 4.25 Ολικές Χρησιμότητες συνόλου αναφοράς κινητών τηλεφώνων), αφορούν κινητά τηλέφωνα που ο κάθε χρήστης βαθμολόγησε, γνωρίζοντας τις τιμές των χαρακτηριστικών τους. Ενδέχεται να υπάρχουν κινητά τηλέφωνα στην αγορά τα οποία να καλύπτουν εξίσου ή και περισσότερο τις απαιτήσεις του χρήστη. Για αυτόν τον λόγο επαναλαμβάνουμε τον αλγόριθμο του Agent Allocator για τους συνδυασμούς χρήστης-κινητό, για κάθε κινητό του συνόλου 25 κινητών τηλεφώνων (σύνολο κινητών αγοράς). Η συνάρτηση ολικής αξίας, με την χρήση των βαρών των κριτηρίων και των μερικών χρησιμοτήτων ενός χρήστη, υπολογίζει την ολική χρησιμότητα ενός συνδυασμού χρήστης-κινητό. Μέσω της συνάρτησης ολικής αξίας υπολογίστηκε η ολική αξία των 25 συνδυασμών χρήστη-κινητού τηλεφώνου, όπως παρουσιάζεται στο **Πίνακας 4.33** Ολικές χρησιμότητες νέου συνόλου κινητών τηλεφώνων.

Πίνακας 4.33 Ολικές χρησιμότητες νέου συνόλου κινητών τηλεφώνων

Χρήστης1		Χρήστης2		Χρήστης3	
Alternative	Global Utilities	Alternative	Global Utilities	Alternative	Global Utilities
Xiaomi Poco M4 Pro	0.868153061	Xiaomi Redmi Note 7	0.833929	Samsung Galaxy A52s	0.860571429
Xiaomi Redmi Note 9	0.815306122	Xiaomi Redmi Note 8T	0.833929	Samsung Galaxy S10+	0.828142857
Xiaomi Redmi Note 9 Pro	0.815163265	Huawei Mate 20 Lite	0.816929	Samsung Galaxy S10+	0.828142857
Xiaomi Redmi Note 8T	0.812602041	Huawei P20 Lite Dual	0.752643	Huawei Mate 20 Lite	0.796142857
Xiaomi Redmi 8	0.758306122	Xiaomi Redmi 5 Plus	0.733857	Samsung Galaxy S9+ Dual	0.685142857
Samsung Galaxy A52s	0.748173469	Samsung Galaxy A52s	0.727786	Samsung Galaxy S8+	0.685142857
Xiaomi Redmi Note 8 Pro	0.695326531	Xiaomi Redmi Note 8 Pro	0.702	Apple iPhone 13 5G	0.671357143
Xiaomi Mi 9T	0.695183673	Apple iPhone 13 5G	0.695357	Samsung Galaxy A7	0.671214286
Huawei Mate 20 Lite	0.648862245	Xiaomi Mi 8	0.695286	Xiaomi Mi 9T	0.663857143
Samsung Galaxy S10+	0.638183673	Xiaomi Mi A2	0.669571	Xiaomi Redmi Note 9	0.646428571
Samsung Galaxy S10+	0.638183673	Samsung Galaxy S9+ Dual	0.6635	Xiaomi Poco M4 Pro	0.646285714
Samsung Galaxy S9+ Dual	0.634397959	Xiaomi Redmi 7	0.662214	Xiaomi Redmi Note 9 Pro	0.646285714
Samsung Galaxy S8+	0.634397959	Xiaomi Redmi Note 9	0.645143	Huawei P20 Lite Dual	0.639071429
Xiaomi Redmi 7	0.565387755	Samsung Galaxy A7	0.5805	Xiaomi Redmi Note 8 Pro	0.631571429
Apple iPhone 13 5G	0.523790816	Samsung Galaxy S8+	0.5805	Xiaomi Redmi 7	0.628071429
Xiaomi Mi A2	0.51755102	Apple iPhone X	0.566714	Xiaomi Redmi 8	0.614
Xiaomi Mi 8	0.517408163	Xiaomi Mi 9T	0.561714	Xiaomi Redmi Note 7	0.613357143
Samsung Galaxy A7	0.517265306	Xiaomi Poco M4 Pro	0.530286	Xiaomi Redmi Note 8T	0.613357143
Huawei P20 Lite Dual	0.470943878	Apple iPhone 11	0.525786	Xiaomi Redmi 5 Plus	0.599428571
Apple iPhone 11	0.470658163	Samsung Galaxy S10+	0.511857	Xiaomi Mi 8	0.506928571
Xiaomi Redmi Note 7	0.44555102	Samsung Galaxy S10+	0.511857	Apple iPhone 11	0.463857143
Apple iPhone X	0.424193878	Apple iPhone SE	0.491071	Xiaomi Mi A2	0.442357143
Apple iPhone XR	0.42405102	Apple iPhone XR	0.451857	Apple iPhone X	0.342428571
Apple iPhone SE	0.367336735	Xiaomi Redmi Note 9 Pro	0.447286	Apple iPhone SE	0.342285714
Xiaomi Redmi 5 Plus	0.328418367	Xiaomi Redmi 8	0.429214	Apple iPhone XR	0.342285714

Τα κινητά τηλέφωνα με ολική χρησιμότητα κοντά στην μονάδα, καλύπτουν τις απαιτήσεις του χρήστη περισσότερο σε σχέση με τα υπόλοιπα. Συνεπώς εκείνα τα κινητά θα αποτελούν συστάσεις τους χρήστες. Οι συστάσεις του συστήματος συστάσεων για τους χρήστες 1,2,3 παρουσιάζεται στο παρακάτω πίνακα (**Πίνακας 4.34**).

Πίνακας 4.34 Πίνακας συστάσεων για χρήστες 1,2,3

Χρήστης 1	Χρήστης 2	Χρήστης 3
Xiaomi Poco M4 Pro	Xiaomi Redmi Note 7	Samsung Galaxy A52s
Xiaomi Redmi Note 9	Xiaomi Redmi Note 8T	Samsung Galaxy S10+ (1TB)
Xiaomi Redmi Note 9 Pro	Huawei Mate 20 Lite	Samsung Galaxy S10+ (512GB)
Xiaomi Redmi Note 8T	Huawei P20 Lite Dual	Huawei Mate 20 Lite
Xiaomi Redmi 8	Xiaomi Redmi 5 Plus	Samsung Galaxy S9+

Κάποιοι χρήστες ενδέχεται να απαιτούν συγκεκριμένες τιμές για κάποια χαρακτηριστικά των κινητών. Για παράδειγμα μπορεί να θέλουν να αγοράσουν κινητά τηλέφωνα μίας συγκεκριμένης εταιρίας ή κινητά που να διαθέτουν ένα συγκεκριμένο λειτουργικό σύστημα. Σε αυτήν την περίπτωση το σύστημα αντιμετωπίζει αυτά τα χαρακτηριστικά ως κριτήρια αποκλεισμού και δεν τα λαμβάνει υπόψιν. Ο χρήστης4 θέλει κινητά τηλέφωνα που διαθέτουν λογισμικό Android της εταιρίας Xiaomi. Επομένως ο πολυκριτήριος πίνακας για τον Χρήστη4(**Πίνακας 4.36** Πολυκριτήριος Πίνακας Χρήστη4) δεν διαθέτει τα χαρακτηριστικά «Λειτουργικό» και «Προτίμηση μάρκας»

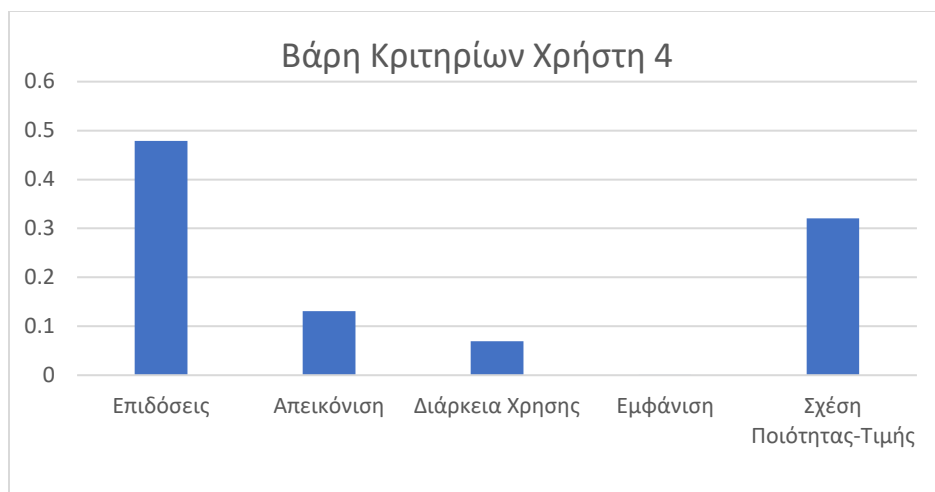
Πίνακας 4.35 Kendall τ για χρήστης4

Χρήστης	Kendall τ
Χρήστης4	0.8693

Πίνακας 4.36 Πολυκριτήριος Πίνακας Χρήστη4

Χρήστης	Εναλλακτική	Επιδόσεις	Απεικόνιση	Διάρκεια Χρησης	Εμφάνιση	Σχέση Ποιότητας-Τιμής	Κατάταξη
Χρήστης4	Apple iPhone 12 Pro	2	1	2	1	1	8
Χρήστης4	Samsung Galaxy A71	3	1	3	3	1	4
Χρήστης4	OnePlus Nord	2	1	3	3	1	5
Χρήστης4	Samsung Galaxy A51	4	1	3	3	4	3
Χρήστης4	Xiaomi Poco X3	4	1	1	1	4	2
Χρήστης4	Xiaomi Redmi Note 9	3	1	1	1	4	1
Χρήστης4	Apple iPhone 12 Mini	3	2	2	3	1	6
Χρήστης4	Samsung Galaxy S20+	2	1	3	3	1	7

Τα βάρη του Χρήστη 4 παρουσιάζονται στο(Σχήμα 4.26 Βάρη Κριτηρίων Χρήστη4). Παρατηρείται πως ο Χρήστης δίνει μεγάλη έμφαση στο κριτήριο των Επιδόσεων, καθώς και στο κριτήριο «σχέση Ποιότητας-Τιμής»



Σχήμα 4.26 Βάρη Κριτηρίων Χρήστη4

Οι ολικές χρησιμότητες υπολογίζονται μόνο για τα κινητά τηλέφωνα που τηρούν τους περιορισμούς του Χρήστη4, και εμφανίζονται στον Πίνακα 4.37

Πίνακας 4.37Ολικές Χρησιμότητες εναλλακτικών Χρήστη4

Alternative	Global Utilites
Xiaomi Redmi 5 Plus	0.8694
Xiaomi Redmi Note 8 Pro	0.8694
Xiaomi Redmi Note 7	0.8694
Xiaomi Redmi Note 8T	0.8694
Xiaomi Poco M4 Pro	0.7998
Xiaomi Redmi Note 9 Pro	0.7998
Xiaomi Mi A2	0.7998
Xiaomi Redmi Note 9	0.7996
Xiaomi Redmi 8	0.7996
Xiaomi Redmi 7	0.7996
Xiaomi Mi 9T	0.5488
Xiaomi Mi 8	0.4792

Οι συστάσεις για τον Χρήστη4 παρουσιάζονται στον Πίνακα 4.38.

Πίνακας 4.38 Συστάσεις Χρήστη4

Χρήστης 4
Xiaomi Redmi 5 Plus

Xiaomi Redmi Note 8 Pro
Xiaomi Redmi Note 7
Xiaomi Redmi Note 8T
Xiaomi Poco M4 Pro

ΚΕΦΑΛΑΙΟ 5

5.1 ΣΥΜΠΕΡΑΣΜΑΤΑ

Στα πλαίσια της ανάπτυξης του πολυκριτήριου συστήματος συστάσεων, έγινε συλλογή αξιολογήσεων κινητών τηλεφώνων, οι οποίες χρησιμοποιήθηκαν ως είσοδος στον αλγόριθμο LDA. Από την εφαρμογή του LDA προέκυψαν οι απαιτήσεις των χρηστών κατά την επιλογή κινητών τηλεφώνων. Τα διαφορετικά χαρακτηριστικά των κινητών τηλεφώνων καλύπτουν αυτές τις απαιτήσεις σε διαφορετικό βαθμό, ο οποίος υπολογίστηκε με την μεθοδολογία Agent Allocator κατά την οποία καθεμία από τις απαιτήσεις των χρηστών αντιστοιχήθηκαν σε ένα χαρακτηριστικό των κινητών, και σχηματίστηκαν τα υποκριτήρια. Κατόπιν τα υποκριτήρια χωρίστηκαν σε κατηγορίες, τα κριτήρια, πάνω στα οποία αξιολογήθηκε ο κάθε συνδυασμός χρήστη-κινητού. Κατόπιν μέσω εφαρμογής του αλγορίθμου πολυκριτήριας ανάλυσης Utastar, υπολογίστηκαν τα βάρη για κάθε κριτήριο για έναν χρήστη, καθώς και η ολική χρησιμότητα ενός ευρύτερου συνόλου κινητών για τον χρήστη. Οι συνδυασμοί χρήστη-κινητών τηλεφώνων με την μεγαλύτερη ολική χρησιμότητα ταίριαζαν περισσότερο στις ανάγκες του χρήστη και αποτέλεσαν τις συστάσεις του συστήματος προς αυτόν.

Η επεξεργασία των κριτικών των κινητών τηλεφώνων μέσω του αλγορίθμου LDA έδειξε πως οι απαιτήσεις των χρηστών εντοπίζονται στα κείμενα των κριτικών τους, οι οποίες μπορούν να κατηγοριοποιηθούν. Συμπεραίνεται πως οι κριτικές με την κατάλληλη επεξεργασία μπορούν να αποτελέσουν χρήσιμη πληροφορία σε ένα σύστημα συστάσεων. Από την έρευνα μέσω ερωτηματολογίων, προέκυψε πως υπάρχουν διάφοροι συνδυασμοί απαιτήσεων των χρηστών, συνεπώς απαιτείται η ανάπτυξη ενός συστήματος συστάσεων ικανό να ανταποκρίνεται στις ανάγκες του κάθε χρήστη.

Από την εφαρμογή της παραπάνω μεθοδολογίας συμπεραίνεται πως το παρόν σύστημα συστάσεων καταφέρνει να εντοπίσει τις ανάγκες του χρήστη, να τις κατατάξει και να προτείνει κινητά τηλέφωνα της αγοράς που τις καλύπτουν. Συνεπώς οι συστάσεις του συστήματος, δεν υπολογίζουν μόνο ποια κινητά ταιριάζουν περισσότερο σε έναν χρήστη, αλλά και γιατί του ταιριάζουν.

Οι συστάσεις του πολυκριτήριου συστήματος συστάσεων μπορούν να βελτιωθούν με την επέκταση του συνόλου των κινητών τηλεφώνων. Επιπλέον κάποια χαρακτηριστικά των κινητών όπως είναι η τιμή, μεταβάλλονται πολύ συχνά, συνεπώς θα ήταν χρήσιμο το σύστημα να επικαιροποιεί αυτά τα χαρακτηριστικά για μεγαλύτερη ακρίβεια στις συστάσεις.

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ ΑΓΟΡΑΣ ΚΙΝΗΤΟΥ ΤΗΛΕΦΩΝΟΥ

Η συγκεκριμένη έρευνα πραγματοποιείται στα πλαίσια της διπλωματικής μου εργασίας με στόχο τη μελέτη και την ανάλυση της συμπεριφοράς των καταναλωτών κατά την αγορά ενός νέου κινητού τηλεφώνου.

ΠΑΡΑΡΤΗΜΑ Α

Φύλο:

- ☐ Γυναίκα
- ☐ Άνδρας

Ηλικία: (αριθμητικά):

Μορφωτικό Επίπεδο:

- ☐ Απόφοιτος λυκείου
- ☐ Απόφοιτος Ι.Ε.Κ
- ☐ Φοιτητής Α.Ε.Ι/Τ.Ε.Ι
- ☐ Πτυχιούχος Α.Ε.Ι/Τ.Ε.Ι
- ☐ Κάτοχος Μεταπτυχιακού
- ☐ Κάτοχος Διδακτορικού
- ☐ Άλλο.....

Οικογενειακή κατάσταση:

- ☐ Έγγαμος
- ☐ Άγαμος

Αριθμός τέκνων (αριθμητικά):

Επάγγελμα:

- ☐ Αγρότης
- ☐ Άνεργος
- ☐ Αυτοαπασχολούμενος
- ☐ Δημόσιος Υπάλληλος
- ☐ Εισοδηματίας
- ☐ Ελεύθερος Επαγγελματίας
- ☐ Ιδιωτικός Υπάλληλος
- ☐ Μαθητής/Φοιτητής
- ☐ Οικιακά
- ☐ Συνταξιούχος
- ☐ Άλλο

Ετήσιο ατομικό εισόδημα:

- ☐ Έως 5.000€
- ☐ Από 5.001-10.000€
- ☐ Από 10.001-15.000€
- ☐ Από 15.001-25.000€
- ☐ Από 25.001-35.000€
- ☐ Από 35.001-45.000€
- ☐ Άνω των 45.000€

Πόση ώρα χρησιμοποιείτε το κινητό σας τηλέφωνο ημερησίως:

- ☐ Έως 30 λεπτά της ώρας
- ☐ 30 λεπτά-1 ώρα
- ☐ 1-3 ώρες
- ☐ 3-5 ώρες
- ☐ Άνω των 5 ωρών

Για ποιους λόγους αντικαθιστάτε το κινητό σας; (πολλαπλής επιλογής)

- ☐ Για να διαθέτω την τελευταία 'λέξη' της τεχνολογίας
- ☐ Για να μπορώ να χρησιμοποιώ τις νέες υπηρεσίες-δυνατότητες που υποστηρίζουν οι νέες συσκευές
- ☐ Γιατί προέκυψε μια συμφέρουσα προσφορά σε συνδυασμό με τις δυνατότητες που μου προσφέρει η νέα συσκευή κινητού
- ☐ Όταν η παλιά μου συσκευή καταστρέφεται, χαλάει ή δεν λειτουργεί σωστά
- ☐ Άλλο...

Πότε αγοράζετε το κινητό σας τηλέφωνο;

- ☐ Με το που εμφανίζεται το κινητό στην αγορά.
- ☐ Λίγο καιρό μετά την εμφάνισή του.
- ☐ Όταν βρίσκεται στην αγορά για αρκετό διάστημα και αφότου έχει αγοραστεί από έναν ικανοποιητικό ποσοστό αγοραστών και έχουν υπάρξει θετικές κριτικές για αυτό.
- ☐ Όταν έχει σταθεροποιηθεί στην αγορά και είναι ευρύτερα αναγνωρίσιμο προϊόν στην αγορά και η τιμή του είναι συμφέρουσα.
- ☐ Όταν πλέον το προϊόν καλύπτει άμεσες ανάγκες και προσφέρεται σε μια συμφέρουσα τιμή.

Πως θα χαρακτηρίζατε τον εαυτό σας ως καταναλωτή; (Πολλαπλής επιλογής)

- ☐ Καινοτόμοι (Innovators), άτομο που θέλει να έχει ότι καινούργιο προϊόν ή τεχνολογία κυκλοφορεί στην αγορά.
- ☐ Πρώιμοι Αποδέκτες (Early Adopters), άτομο που υιοθέτησαν ένα νέο προϊόν ή τεχνολογία πριν από την πλειονότητα του πληθυσμού.
- ☐ Πρώιμη πλειοψηφία (Early Majority), άτομο που πριν αγοράσει ένα καινούργιο προϊόν θέλει να έχει σιγουρευτεί για την αγορά του.
- ☐ Όψιμη Πλειοψηφία (Late Majority), άτομο που σκέφτεται και αναλύει την αγορά ενός καινούργιου προϊόντος και αναμένει την πιο συμφέρουσα προσφορά.
- ☐ Αργοπορημένοι (Laggards), άτομο που για να αγοράσει ένα νέο προϊόν της αγοράς θα πρέπει να καλύπτει τις άμεσες ανάγκες του.

ΠΑΡΑΡΤΗΜΑ Β

Εάν αγοράζατε σήμερα κινητό:

Ποιο λειτουργικό σύστημα θα επιλέγατε να έχει το κινητό;

- ☐ Android
- ☐ iOS

Θα αγοράζατε κινητό που διαθέτει διαφορετικό λειτουργικό σύστημα από αυτό που επιλέξατε;

- ☐ Ναι
- ☐ Όχι

Ποιο μέγεθος οθόνης θα επιλέγατε;

- ☐ Έως 5.4"
- ☐ Από 5.4"-5.8"
- ☐ Από 5.8"-6.3"
- ☐ Από 6.3 και άνω

Ποια μνήμη RAM θα επιλέγατε;

- ☐ 2GB
- ☐ 3 GB
- ☐ 4 GB
- ☐ 6GB
- ☐ 8GB
- ☐ 12GB

Πόση χωρητικότητα μπαταρίας θα θέλατε να διαθέτει;

- ☐ έως 3700mah
- ☐ 3700 με 4500mah
- ☐ 4500 και πάνω

Ποια μνήμη ROM θα θέλατε να διαθέτει;

- ☐ 16GB
- ☐ 32GB
- ☐ 64GB
- ☐ 128GB
- ☐ 256GB

Ποια η συχνότητα του επεξεργαστή που θα επιλέγατε;

- ☐ Έως 2.2GHz
- ☐ 2.2GHz-2.7GHz
- ☐ 2.7GHz και πάνω

Ποια η Ανάλυση Κάμερας που προτιμάτε;

- ☐ 12MP
- ☐ 16MP
- ☐ 24MP
- ☐ 48MP
- ☐ 64MP

Τι ανάλυση οθόνης θα θέλατε να έχει;

- ☐ HD (720x1280)
- ☐ FHD(1080x1920)
- ☐ QHD(1440x2560)
- ☐ 4K(1644x3840)
- ☐

Οι δυνατότητες συνδεσιμότητας του κινητού θα θέλατε να είναι:

- ☐ Βασικές
- ☐ Τελευταίας Τεχνολογίας (να διαθέτει τεχνολογίες όπως NFC, Wi-Fi 6, Bluetooth 5)

Θα προτιμούσατε να ανήκει στα κινητά με:

- ☐ Μέτρια Εμφάνιση
- ☐ Υψηλή Εμφάνιση
- ☐ Πολύ υψηλή Εμφάνιση

Πόσα χρήματα είστε διατεθειμένοι να δώσετε για την αγορά ενός νέου κινητού;

- ☐ Έως 250 ευρώ

- ☐ Από 250 μέχρι 400 ευρώ
- ☐ Από 400 μέχρι 600 ευρώ
- ☐ Από 600 μέχρι 800 ευρώ
- ☐ Πάνω από 800 ευρώ

Κατατάξτε τις εταιρείες κατά σειρά προτίμησης:

Η κλίμακα της σειράς κατάταξης ορίζεται από το 1 έως το 7. Το 1 θεωρείται η καλύτερη επιλογή και το 7 η χειρότερη (Υπάρχει δυνατότητα έκφρασης της ίδιας προτίμησης με απόδοση της ίδιας σειράς προτίμησης).

Μάρκα	Κατάταξη
Apple	
Xiaomi	
Huawei	
Samsung	
OnePlus	
LG	
Nokia	

Θα αγοράζατε κινητό Μάρκας διαφορετικής από αυτήν/αυτές που κατατάξατε ως καλύτερη/καλύτερες;

- ☐ Ναι
- ☐ Όχι

ΠΑΡΑΡΤΗΜΑ Γ

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

Χαρακτηριστικό	Βαθμός Σημαντικότητας
----------------	--------------------------

Τιμή	
Επωνυμία εταιρίας κινητού (Brand name)	
Λειτουργικό σύστημα(Android, iOS)	
Αποθηκευτικός χώρος (ROM)	
Προσωρινός αποθηκευτικός χώρος (RAM)	
Μέγεθος οθόνης	
Μπαταρία(Διάρκεια)	
Κάμερα(Ανάλυση)	
Επεξεργαστής	
Ανάλυση οθόνης	
Εμφάνιση(Σχεδιασμός, Χρώμα)	
Δυνατότητα Συνδεσιμότητας	
Άλλο	

ΠΑΡΑΡΤΗΜΑ Δ

Δίνονται τα παρακάτω αντιπροσωπευτικά μοντέλα κινητών τηλεφώνων μέσω διαφορετικών συνδυασμών των αντίστοιχων χαρακτηριστικών τους: RAM/ROM/Μέγεθος Οθόνης/Κάμερα/Μέγεθος/Μπαταρίας/Λειτουργικό Σύστημα/ Τιμή/ Συχνότητα Επεξεργαστή/Ανάλυση Οθόνης/ Βαθμολογία Εμφάνισης / Δυνατότητες Συνδεσιμότητας και σας παρακαλούμε να συμπληρώσετε τη σειρά προτίμησής σας για τις εναλλακτικές επιλογές του παρακάτω πίνακα, απαντώντας στις ερωτήσεις:

Ερώτηση: Από τα παρακάτω μοντέλα κινητών τηλεφώνων ποιο ή ποια θα επιλέγατε πρώτο/α για να αγοράσετε (Χρησιμοποιείτε τη κλίμακα 1 έως 9, όπου με 1 θα σημειώσετε το πλέον προτιμώμενο, με 2 το αμέσως επόμενο, κοκ., ενώ σε περίπτωση που προτιμάτε το ίδιο δύο ή περισσότερα από τα κινητά, τότε αυτό μπορείτε να το εκφράσετε αποδίδοντας σε αυτά την ίδια σειρά προτίμησης);

Επαναλαμβανόμενη ερώτηση: Αν δεν υπήρχε η προηγούμενη επιλογή σας, ποιο/α από τα υπόλοιπα μοντέλα κινητών τηλεφώνων θα επιλέγατε να αγοράσετε;

Εναλλακτικές Επιλογές	Σειρά προτίμησης
Apple iPhone 12 Pro: 6GB/ 128GB/ Από 5.9"- 6.3"/ 12MP/ 2.815mAh/ iOS / 1230€/ 3.1/ FHD+/ 8.5/ 9.1	
Samsung Galaxy A71: 6GB/ 128GB/ Από 6.4" και άνω / 64MP/ 4.500mAh/ Android/ 320€/ 2.2/ FHD+/ 7.4/ 7.6	
OnePlus Nord: 12GB/ 256GB/ Από 6.4" και άνω / 48MP/ 4.115mAh/ Android / 450€/ 2.4/ FHD+/ 7.8/ 8.4	

Samsung Galaxy A51: 4GB/ 128GB/ Από 6.4" και άνω/ 48MP/ 4.000mAh/ Android/ 240€/ 2.3/ FHD+/ 7.6/ 7.7	
Xiaomi Poco X3: 6GB/ 64GB/ Από 6.4" και άνω/ 64MP/ 5.160mAh/ Android/ 200€/ 2.3/ FHD+/ 7.7/ 7.4	
Xiaomi Redmi Note 9: 3GB/ 32GB/ Από 6.4" και άνω/ 13MP/ 5.020mAh/ Android/ 140€/ 2/ FHD+/ 6.8/ 7.5	
Apple iPhone 12 Mini: 4GB/ 64GB/ Από 5.4"-5.8"/ 12MP/ 2.227mAh/ iOS/ 760€/ 3.1/ FHD+/ 8.1/ 9	
Samsung Galaxy S20+: 8GB/ 128GB/ Από 6.4" και άνω/ 64MP/ 4.500mAh/ Android/ 730€/ 2.73/ WQHD+/ 9.3/ 8	
Apple iPhone 12 Pro: 6GB/ 128GB/ Από 5.9"- 6.3"/ 12MP/ 2.815mAh/ iOS / 1230€/ 3.1/ FHD+/ 8.5/ 9.1	

BIBΛΙΟΓΡΑΦΙΑ

1. Denguir-Rekik, A., Montmain, J., & Mauris, G. (2009). A possibilistic-valued multi-criteria decision-making support for marketing activities in e-commerce: Feedback Based Diagnosis System. *European Journal of Operational Research*, 195(3), 876–888.
<https://doi.org/10.1016/j.ejor.2007.11.020>
2. Lee, S. K., Cho, Y. H., & Kim, S. H. (2010). Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations. *Information Sciences*, 180(11), 2142–2155.
<https://doi.org/10.1016/j.ins.2010.02.004>
3. Zheng, Y. (2019, April 8). Utility-based multi-criteria recommender systems. *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*.
<https://dl.acm.org/doi/pdf/10.1145/3297280.3297641>
4. Wasid, M., & Ali, R. (2018). An Improved Recommender System based on Multi-criteria Clustering Approach. *Procedia Computer Science*, 131, 93–101. <https://doi.org/10.1016/j.procs.2018.04.190>
5. G. Adomavicius and Y. Kwon, "New Recommendation Techniques for Multicriteria Rating Systems," in *IEEE Intelligent Systems*, vol. 22, no. 3, pp. 48-55, May-June 2007, doi: 10.1109/MIS.2007.58.
6. F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, "Recommendation systems: Principles, methods and evaluation", *Egyptian Informatics Journal*, Volume 16, Issue 3, 2015, Pages 261-273, ISSN 1110-8665, <https://doi.org/10.1016/j.eij.2015.06.005>.
7. Friedman, Geiger, & Goldszmidt. (n.d.). Bayesian Network Classifiers. *Machine Learning*, 29(2), 131–163.
<https://doi.org/10.1023/A:1007465528199>
8. Pawlicka, A., Pawlicki, M., Kozik, R., & Choraś, R. S. (2021). A Systematic Review of Recommender Systems and Their Applications in Cybersecurity. *Sensors*, 21(15), 5248. <https://doi.org/10.3390/s21155248>

9. Bouraga, Sarah, Jureta, Ivan, Faulkner, Stéphane, Herssens, & Caroline. (1 C.E.). Knowledge-Based Recommendation Systems: A Survey. *International Journal of Intelligent Information Technologies (IJIT)*, 10(2), 1–19.
<https://doi.org/10.4018/ijit.2014040101>
10. Jun Wang, Arjen P. de Vries, and Marcel J. T. Reinders. 2006. Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '06)*. Association for Computing Machinery, New York, NY, USA, 501–508.
DOI:<https://doi.org/10.1145/1148170.1148257>
11. Burke, R. Hybrid Recommender Systems: Survey and Experiments. *User Model User-Adap Inter* 12, 331–370 (2002).
<https://doi.org/10.1023/A:1021240730564>
12. Νικόλαος Βαβουλιώτης, "Ανάπτυξη εργαλειοθήκης μεθόδων πολυκριτήριας ανάλυσης σε γλώσσα Python", Διπλωματική Εργασία, Σχολή Μηχανικών Παραγωγής και Διοίκησης, Πολυτεχνείο Κρήτης, Χανιά, Ελλάς, 2021 <https://doi.org/10.26233/heallink.tuc.89663>
13. Guo, Mengzhuo & Liao, Xiuwu & Liu, Jiapeng & Zhang, Qingpeng, 2020. "[**Consumer preference analysis: A data-driven multiple criteria approach integrating online information**](#)," *Omega*, Elsevier, vol. 96(C). DOI: 10.1016/j.omega.2019.05.010
14. Siskos, Yannis & Grigoroudis, Evangelos & Matsatsinis, Nikolaos & Figueira, José & Greco, Salvatore & Ehrogott, Matthias. (2005). UTA methods. 10.1007/0-387-23081-5_8.
15. E. Jacquet-Lagrange, J. Siskos, Assessing a set of additive utility functions for multicriteria decision-making, the UTA method, *European Journal of Operational Research*, Volume 10, Issue 2, 982, Pages 151-164, ISSN 0377-2217, [https://doi.org/10.1016/0377-2217\(82\)90155-2](https://doi.org/10.1016/0377-2217(82)90155-2).

16. Matsatsinis, N. F., & Delias, P. (2003). AgentAllocator: An Agent-Based Multi-criteria Decision Support System for Task Allocation. *Holonic and Multi-Agent Systems for Manufacturing*. Springer. doi: 10.1007/978-3-540-45185-3_21
17. Pazzani, M.J. (2004). *A Framework for Collaborative, Content-Based and Demographic Filtering*. *Artificial Intelligence Review*, 13, 393-408.
18. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2001). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan), 601–608. Retrieved from https://www.researchgate.net/publication/221620547_Latent_Dirichlet_Allocation
19. Campbell, J. C., Hindle, A., & Stroulia, E. (2015). Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data. *The Art and Science of Analyzing Software Data*. Morgan Kaufmann. doi: 10.1016/B978-0-12-411519-4.00006-9
20. Rieger, J., Rahnenführer, J., & Jentsch, C. (2020). Improving Latent Dirichlet Allocation: On Reliability of the Novel Method LDAPrototype. *Natural Language Processing and Information Systems*, 118–125. https://doi.org/10.1007/978-3-030-51310-8_11
21. Qiu, L., & Yu, J. (2018). CLDA: An Effective Topic Model for Mining User Interest Preference under Big Data Background. *Complexity*, 2018, 1–10. <https://doi.org/10.1155/2018/2503816>
22. Hoffman, M., Bach, F., & Blei, D. (2022). Online Learning for Latent Dirichlet Allocation. *Advances in Neural Information Processing Systems*, 23. <https://proceedings.neurips.cc/paper/2010/hash/71f6278d140af599e06ad9bf1ba03cb0-Abstract.html>
23. Jannach, D., Karakaya, Z., & Gedikli, F. (2012). Accuracy improvements for multi-criteria recommender systems. *EC '12*.