Blue text on a black background

Description automatically generated

***Σχολή Μηχανικών***

***Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών***

*Διαχείριση Δεδομένων Μεγάλης Κλίμακας*

***Αναφορά Ανάλυσης Δεδομένων***

*Εργασία: Best-value Smartphone Analysis*

*ΕΥΦΡΟΣΥΝΗ ΒΑΡΣΟΥ 21390021*

*ΑΓΓΕΛΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ ΜΕΝΤΖΕΛΟΣ 21390132*

# 1. Εισαγωγή

Αυτό το έγγραφο αποτελεί την συγγραφή της αναφοράς για την ανάλυση δεδομένων που πραγματοποιήσαμε σε ένα dataset που αφορά για τα διάφορα κινητά που υπάρχουν στην αγορά και τα αντίστοιχα χαρακτηριστικά τους. Θα επικεντρωθούμε στην ανάλυση σχετικά με την τιμή του κάθε κινητού σε αντιπαράθεση με τον επεξεργαστή που έχουν και την RAM.

Αρχικά, θα περιγράψουμε το πρόβλημα που επιχειρούμε να λύσουμε με την ανάλυση αυτή, έπειτα θα περιγράψουμε εν συντομία το dataset που χρησιμοποιήσαμε και τέλος θα επικεντρωθούμε στην περιγραφή του κώδικα που υλοποιήσαμε καθώς και την περιγραφή/αποτίμηση των αποτελεσμάτων που συμπεραίνουμε από αυτά.

Σε αυτήν την αναφορά δεν θα συμπεριλάβουμε όλους τους πίνακες/σχήματα καθώς είναι πολλά για τις δύο αναλύσεις που υλοποιήσαμε (Processor, RAM) αλλά υπάρχουν όλοι αναλυτικά στον κώδικα .ipynb. Ο κώδικας .py περιέχει όλη την ανάλυση αλλά συμπεριλήφθηκε για πρακτικούς λόγους συνεπώς η ανάλυση μαζί με τις πηγές που χρησιμοποιήσαμε βρίσκεται στον κώδικα .ipynb (θα ήταν προτιμότερο δηλαδή να προβληθεί από εκεί αφού είναι ενσωματωμένοι όλοι οι πίνακες και τα σχήματα καθώς και λεπτομερείς σχόλια).

# 2. Ορισμός Προβλήματος

*καλό είναι να συμπεριλάβουμε ένα παράδειγμα χρήσης των αποτελεσμάτων της, π.χ. για ποιόν είναι χρήσιμα ?*

# 3. Περιγραφή του συνόλου δεδομένων

ΑΓΓΕΛΟΣ <https://www.kaggle.com/datasets/nishantdeswal1810/smartphones/data> -> τα εξηγεί και εδώ, συνοπτικά κάθε πεδίο και κυρίως τα πεδία που χρησιμοποιήσαμε εμεις. Οπότε εξηγησε γενικά και μετά για τα πεδία που χρησιμοποιήσαμε εμεις και ΠΟΛΥ συνοπτικα τα αλλά (σαν απλή αναφοράά τους)

# 4. Περιγραφή των μεθόδων ανάλυσης δεδομένων

*Η υποενότητα της ανάλυσης των δεδομένων θα πρέπει να περιγράφει τις τεχνικές που χρησιμοποιήσατε και μια εξήγηση γιατί! Πολύ σημαντικό είναι να προσπαθήσετε να πείσετε τον αναγνώστη ότι μια συγκεκριμένη τεχνική που χρησιμοποιείται είναι αυτή που ταιριάζει στο πρόβλημα. Να είστε σαφείς και περιεκτικοί*

ΑΓΓΕΛΟΣ: Association Rules, Clustering

ΦΡΟΣΩ: One Hot Encoding

# 5. Πειραματικά Αποτελέσματα

*Η ενότητα των πειραματικών αποτελεσμάτων θα πρέπει να περιλαμβάνει όλα τα πειράματα που χρησιμοποιήσατε. Συζητήστε τις παραμέτρους και ιδιαίτερα τους χρόνους εκτέλεσης, καθώς και τυχόν μέτρα αξιολόγησης που χρησιμοποιήθηκαν. Συμπεριλάβετε πίνακες/σχήματα όπως κρίνετε απαραίτητο (τα περισσότερα έγγραφα ανάλυσης δεδομένων τα διαθέτουν).*

*MIN SUPPORT, MIN CONFIDENCE -> Discord τι σημαίνει έχουμε μεγαλύτερο ή μικρότερο min\_sup κτλ.*

*ΧΡΟΝΟΙ ΕΚΤΕΛΕΣΕΙΣ ΕΙΔΙΚΑ ΓΙΑ CLUSTERING (το γράφει το .ipynb) -> πολυπλοκότητα clustering?*

*ΣΧΗΜΑΤΑ CLUSTERING -> 4 clusters (δοκιμή αρχικά με 3, αλλά υπήρχαν κάποια πολύ εξω οπότε 4)*

# 6. Κριτική αποτίμηση αποτελεσμάτων

ΦΡΟΣΩ για price processor (συνοπτικά για κάθε cluster, ποια είναι τα frequent items και οι συσχετίσεις του και αν βγάζει νόημα να έχουν αυτά απο την λογική)

ΑΓΓΕΛΟΣ: για price ram (συνοπτικά για κάθε cluster, ποια είναι τα frequent items και οι συσχετίσεις του και αν βγάζει νόημα να έχουν αυτά απο την λογική)

# 7. Συμπεράσματα

Τι συμπεράσματα βγάζουμε για τα διαφορετικά είδους κινητά (low-end, middle-end, higher-end, luxury). Απο τα αποτελέσματα ποιο είναι το καλύτερο? Σε σύγκριση με την τιμή δηλαδή πάντα. Γενικά αν εμπίπτουν τα αποτελέσματα με την πραγματικότητα,