

**Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων**  
**“Εισαγωγή στην Αναλυτική των Δεδομένων”, Χειμερινό Εξάμηνο 2022-2023**

**Παραγωγή Συστάσεων στο Περιβάλλον R/RStudio**

Η Επεξεργασία σε περιβάλλον του MS-Excel/Google Sheets, όπως παρουσιάστηκε/εξετάστηκε στις διαλέξεις:

(α) [Η άσκηση](#)

(β) [Η λύση](#)

**Ζητείται** η διεκπεραίωση της επεξεργασίας με πιο ευέλικτο τρόπο σε περιβάλλον R/RStudio, σύμφωνα με το παρακάτω πλαίσιο το οποίο **πρέπει να τηρηθεί επακριβώς**:

**(α) Μονάδες: 0.5**

Ανάγνωση των αρχείων [train\\_data.xls](#) και [test\\_data.xls](#) στα εξής δύο dataframes: [train\\_data](#) και [test\\_data](#), αντίστοιχα. Στη συνέχεια, να δημιουργηθούν και να αποδοθούν σε αυτές οι κατάλληλες τιμές (αρχικές ή μη) οι εξής δομές:

**readers\_training**: το πλήθος των αναγνωστών ( $U_1, \dots$ ) στο train\_data (δυναμικά η τιμή, κατά την εκτέλεση)

**readers\_test**: το πλήθος των αναγνωστών ( $NU_1, \dots$ ) στο test\_data (δυναμικά η τιμή, κατά την εκτέλεση)

**columns\_training**: το πλήθος των στηλών του train\_data (δυναμικά η τιμή, κατά την εκτέλεση)

**k\_NN**: ο αριθμός των πλησιέστερων γειτόνων που θα καθορίσουν τις συστάσεις (αρχική τιμή: 3)

**books**: διάνυσμα με τους τίτλους των βιβλίων ([δημιουργείται δυναμικά](#) από τα δεδομένα)

**recommendations**: dataframe που προορίζεται να καταχωρήσει τις συστάσεις που θα παράξει ο αλγόριθμος. [Αρχική δομή και περιεχόμενο](#): reader= character(0), book= character(0), rank = numeric(0)

**MAE**: διάνυσμα που προορίζεται να καταχωρήσει τις τιμές Mean Absolute Error που θα παράξει ο αλγόριθμος. Αρχικοποιείται στο κενό.

**(β) Μονάδες: 5.0**

Δημιουργία συνάρτησης **calculate\_similarities()** η κλήση της οποίας δημιουργεί το dataframe similarities που καταχωρεί τις τιμές ομοιότητας των αναγνωστών του test\_data προς κάθε έναν από τους αναγνώστες του train\_data. Παράδειγμα: [για τα δεδομένα της άσκησης \(οι πρώτοι U1.....U10\)](#).

(γ) Δίνεται ο “κορμός”/main() του αλγόριθμου:

```
for (i in 1:readers_test) {  
  k_nearest <- get_k_nearest(i, k_NN)  
  predictions <- calculate_predictions(i, k_nearest)  
  need_recommendation <- spot_the_NAs(i)  
  recommendations <- calculate_recommendations(i, need_recommendation, predictions, recommendations)  
  MAE <- mean_absolute_error(i, predictions, MAE)  
}
```

**(δ) Μονάδες: 1.0**

Να συνταχθεί ο κώδικας της συνάρτησης **get\_k\_nearest()** η οποία θεωρεί τον τρέχοντα αναγνώστη (i) του `test_data` και υπολογίζει για αυτόν τους `k_NN` πλησιέστερους γείτονές του, βάσει των τιμών ομοιότητας που καταχωρεί το `similarities`. Το αποτέλεσμα της εκτέλεσης της `get_k_nearest()` καταχωρείται στο `dataframe k_nearest`. Παράδειγμα: [για τα δεδομένα της άσκησης και για τον πρώτο \(NU1\) αναγνώστη του test\\_data](#).

**(ε) Μονάδες: 2.0**

Να συνταχθεί ο κώδικας της συνάρτησης **get\_predictions()** η οποία δημιουργεί διάνυσμα μήκους ίσου προς το μήκος του διανύσματος `books` όπου τοποθετούνται (σε σχήμα αντιστοίχισης ένα-προς-ένα) οι τιμές των προβλεπόμενων τιμών διαβάθμισης που αναμένεται να δώσει ο τρέχων (i) αναγνώστης NU στο κάθε ένα βιβλίο, με οδηγό την ομοιότητά του προς τους `k_NN` εγγύτερους γειτόνους του που καταχωρούνται στο `k_nearest`. Παράδειγμα: [για τα δεδομένα της άσκησης και για τον πρώτο \(NU1\) αναγνώστη του test\\_data](#).

**(στ) Μονάδες: 0.5**

Να συνταχθεί ο κώδικας της συνάρτησης **spot\_the\_NAs()** που εντοπίζει τα κενά (NAs) στη διαβάθμιση βιβλίων για τον τρέχοντα αναγνώστη (i) και 'περνά' τις θέσεις τους στο διάνυσμα `need_recommendation`. Πρόκειται για τις αύξουσες θέσεις των βιβλίων για τα οποία καλείται να παράξει συστάσεις ο αλγόριθμος στην περίπτωση του συγκεκριμένου αναγνώστη NU. Παράδειγμα: [για τα δεδομένα της άσκησης και για τον πρώτο \(NU1\) αναγνώστη του test\\_data](#).

**(ζ) Μονάδες: 0.5**

Να συνταχθεί ο κώδικας της συνάρτησης **calculate\_recommendations()** που χρησιμοποιεί το περιεχόμενο των `predictions` και `need_recommendation` για να επιπροσθέσει τις συστάσεις που έχουν παραχθεί για τον τρέχοντα αναγνώστη (i) στο περιεχόμενο του `dataframe recommendations`. Παράδειγμα: [για τα δεδομένα της άσκησης και για τον πρώτο \(NU1\) αναγνώστη του test\\_data](#).

**(η) Μονάδες: 0.5**

Να συνταχθεί ο κώδικας της συνάρτησης **mean\_absolute\_error()** που υπολογίζει το μέσο απόλυτο σφάλμα απόκλισης που αφορά στις συστάσεις που έχουν παραχθεί για τον τρέχοντα αναγνώστη (i) στο περιεχόμενο του διανύσματος MAE. Παράδειγμα: [για τα δεδομένα της άσκησης και για τον πρώτο \(NU1\) αναγνώστη του test\\_data](#).

Με την **ολοκλήρωση** της εκτέλεσης του αλγόριθμου για τα δεδομένα της άσκησης: το περιεχόμενο που **πρέπει να καταχωρούν** τα [recommendations](#) και [MAE](#).

=====

**Καταληκτική ημερομηνία παράδοσης: 15/01/2023**

Το πως και το που της κατάθεσης της εργασίας θα ανακοινωθούν έγκαιρα.

~~~~~