
Data Science και
Μεγάλα Δεδομένα με
τη Γλώσσα R και
Rstudio

1. ΕΙΣΑΓΩΓΗ

Το Κέντρο Επιμόρφωσης και Δια Βίου Μάθησης (Κ.Ε.ΔΙ.ΒΙ.Μ.) του **Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών (Ε.Κ.Π.Α.)** σας καλωσορίζει στο Πρόγραμμα Συμπληρωματικής εξ Αποστάσεως Εκπαίδευσης και συγκεκριμένα στο πρόγραμμα επαγγελματικής επιμόρφωσης και κατάρτισης με τίτλο **«Data Science και Μεγάλα Δεδομένα με τη Γλώσσα R και Rstudio»**.

Η ανάγκη συνεχούς επιμόρφωσης και πιστοποίησης επαγγελματικών δεξιοτήτων οδήγησε το **Πρόγραμμα Συμπληρωματικής εξ Αποστάσεως Εκπαίδευσης (E-Learning)** του Ε.Κ.Π.Α. στο σχεδιασμό των πρωτοποριακών αυτών Προγραμμάτων Επαγγελματικής Επιμόρφωσης και Κατάρτισης, με γνώμονα τη **διασύνδεση της θεωρητικής με την πρακτική γνώση**, αναπτύσσοντας κυρίως, την εφαρμοσμένη διάσταση των επιστημών στα αντίστοιχα επαγγελματικά πεδία.

Στη συνέχεια, σας παρουσιάζουμε αναλυτικά το πρόγραμμα σπουδών για το πρόγραμμα επαγγελματικής επιμόρφωσης και κατάρτισης: **«Data Science και Μεγάλα Δεδομένα με τη Γλώσσα R και Rstudio»**, τις προϋποθέσεις συμμετοχής σας σε αυτό, καθώς και όλες τις λεπτομέρειες που πιστεύουμε ότι είναι χρήσιμες, για να έχετε μια ολοκληρωμένη εικόνα του προγράμματος.

2. ΣΚΟΠΟΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ

Σκοπός του προγράμματος είναι η εκπαίδευση των συμμετεχόντων/ουσών στην επεξεργασία και ανάλυση δεδομένων, με έμφαση σε μεγάλα σύνολα δεδομένων (large datasets), μέσω της γλώσσας R και του περιβάλλοντος RStudio.

Στο πλαίσιο αυτό, θα πραγματοποιηθεί, η βήμα προς βήμα εκμάθηση της χρήσης της γλώσσας R στην Επιστήμη των Δεδομένων, συμπεριλαμβανομένων των εργαλείων, τεχνικών και των βιβλιοθηκών για Data Science, καθώς και του ολοκληρωμένου περιβάλλοντος ανάπτυξης Rstudio. Ειδικότερα, οι εκπαιδευόμενοι/ες θα μάθουν να αξιοποιούν τις υπολογιστικές μεθόδους και τεχνολογίες για την ανάλυση των Μεγάλων Δεδομένων (Big Data) που θα τους επιτρέψουν να πραγματοποιούν data web scraping, να κάνουν προεπεξεργασία των δεδομένων (Data Munging), να υπολογίζουν διάφορα στατιστικά μέτρα και να κάνουν οπτικοποίηση των δεδομένων (Data Visualization). Επιπρόσθετα, κατά τη διάρκεια του προγράμματος, οι συμμετέχοντες/ουσες θα εξοικειωθούν με την στατιστική και ανάλυση δεδομένων, καθώς και με τη χρήση των αλγορίθμων μηχανικής μάθησης. Πιο συγκεκριμένα, θα μάθουν τις βασικές αρχές δειγματοληψίας (Sampling), να κάνουν Προγνωστική Ανάλυση (Predictive Analytics) και να επιλέγουν το κατάλληλο μοντέλο (Model selection), ανάλογα με τη φύση του προβλήματος. Ακόμη, θα μάθουν πώς πραγματοποιείται η ανίχνευση των ακραίων τιμών (Outlier detection) και η βηματική παλινδρόμηση (Stepwise regression). Τέλος, μέσα από μελέτες περίπτωσης

(Case studies) με Πραγματικά Δεδομένα, οι εκπαιδευόμενοι/ες θα εξοικειωθούν με τη γλώσσα R και θα είναι σε θέση να επιλύουν πρακτικά προβλήματα με πραγματικά δεδομένα.

3. ΚΑΤΗΓΟΡΙΕΣ ΥΠΟΨΗΦΙΩΝ ΠΟΥ ΓΙΝΟΝΤΑΙ ΔΕΚΤΟΙ ΣΤΟ ΠΡΟΓΡΑΜΜΑ - ΤΡΟΠΟΣ ΕΝΤΑΞΗΣ

Αίτηση συμμετοχής μπορούν να υποβάλλουν:

- ▶ **απόφοιτοι Πανεπιστημίου/ΤΕΙ της ημεδαπής και της αλλοδαπής**
- ▶ **απόφοιτοι Δευτεροβάθμιας Εκπαίδευσης με συναφή εργασιακή ενασχόληση.**

Αναλυτικότερα, το πρόγραμμα απευθύνεται σε απόφοιτους σχολών πληροφορικής και τηλεπικοινωνιών, μηχανικούς, μηχανολόγους ηλεκτρονικούς, ηλεκτρολόγους και απόφοιτους σχολών θετικών επιστημών, καθώς και σε όσους έχουν ένα σχετικό τεχνολογικό υπόβαθρο γνώσεων και θέλουν να επιμορφωθούν στο αντικείμενο της Επιστήμης των Δεδομένων και στα Μεγάλα Δεδομένα με τη Γλώσσα R και το Rstudio.

Η αίτηση συμμετοχής υποβάλλεται ηλεκτρονικά, μέσω της ιστοσελίδας:

<https://elearningekpa.gr/>

4. ΠΡΟΑΠΑΙΤΟΥΜΕΝΑ

Τα προαπαιτούμενα για την παρακολούθηση του Προγράμματος από τους εκπαιδευόμενους είναι:

- ▶ Πρόσβαση στο Διαδίκτυο
- ▶ Κατοχή προσωπικού e-mail
- ▶ Βασικές γνώσεις χειρισμού ηλεκτρονικών υπολογιστών

5. ΤΡΟΠΟΣ ΔΙΕΞΑΓΩΓΗΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ

Η διδασκαλία στα προγράμματα εξ αποστάσεως επαγγελματικής επιμόρφωσης και κατάρτισης του Κέντρου Επιμόρφωσης και Δια Βίου Μάθησης του ΕΚΠΑ διεξάγεται μέσω του διαδικτύου, προσφέροντας στον εκπαιδευόμενο «αυτονομία», δηλαδή δυνατότητα μελέτης ανεξαρτήτως περιοριστικών παραγόντων, όπως η υποχρέωση της φυσικής του παρουσίας σε συγκεκριμένο χώρο και χρόνο.

Το εκπαιδευτικό υλικό του προγράμματος διατίθεται σταδιακά, ανά διδακτική ενότητα, μέσω ειδικά διαμορφωμένων ηλεκτρονικών τάξεων. Κατά την εξέλιξη κάθε θεματικής ενότητας αναρτώνται σε σχετικό link οι απαραίτητες για την ομαλή διεξαγωγή της εκπαιδευτικής διαδικασίας ανακοινώσεις.

Ο εκπαιδευόμενος, αφού ολοκληρώσει τη μελέτη της εκάστοτε διδακτικής ενότητας, καλείται να υποβάλει ηλεκτρονικά, το αντίστοιχο τεστ αξιολόγησης. Τα τεστ μπορεί να περιλαμβάνουν ερωτήσεις αντιστοίχισης ορθών απαντήσεων, πολλαπλής επιλογής, αληθούς/ψευδούς δήλωσης, ή upload, όπου ο εκπαιδευόμενος θα πρέπει να διατυπώσει και να επισυνάψει την απάντησή του. Η θεματική ενότητα μπορεί να συνοδεύεται από τελική εργασία, η οποία διατίθεται κατά την ολοκλήρωση της θεματικής ενότητας (εφόσον το απαιτεί η φύση της θεματικής ενότητας) και αφορά το σύνολο της διδακτέας ύλης.

Παράλληλα, παρέχεται **πλήρης εκπαιδευτική υποστήριξη** δεδομένου ότι ο εκπαιδευόμενος μπορεί να απευθύνεται ηλεκτρονικά (για το διάστημα που διαρκεί το εκάστοτε μάθημα) στον ορισμένο εκπαιδευτή του, μέσω ενσωματωμένου στην πλατφόρμα ηλεκτρονικού συστήματος επικοινωνίας, για την άμεση επίλυση αποριών σχετιζόμενων με τις θεματικές ενότητες και τις ασκήσεις αξιολόγησης ή την τελική εργασία.

6. ΤΡΟΠΟΣ ΕΞΕΤΑΣΗΣ ΚΑΙ ΒΑΘΜΟΛΟΓΗΣΗΣ

Σε κάθε διδακτική ενότητα ο εκπαιδευόμενος θα πρέπει να επιλύει και να υποβάλλει ηλεκτρονικά το αντίστοιχο τεστ, τηρώντας το χρονοδιάγραμμα που έχει δοθεί από τον εκπαιδευτή του. Η κλίμακα βαθμολογίας κυμαίνεται από 0 έως 100%. Συνολικά, η βαθμολογία κάθε θεματικής ενότητας προκύπτει κατά το 60% από τις ασκήσεις αξιολόγησης και κατά το υπόλοιπο 40% από την τελική εργασία, η οποία εκπονείται στο τέλος του συγκεκριμένου μαθήματος και εφόσον το απαιτεί η φύση αυτού.

Η χορήγηση του **Πιστοποιητικού Επιμόρφωσης** πραγματοποιείται, όταν ο εκπαιδευόμενος λάβει σε όλα τα μαθήματα βαθμό μεγαλύτερο ή ίσο του 50%. Σε περίπτωση που η συνολική βαθμολογία ενός ή περισσοτέρων μαθημάτων δεν ξεπερνά το 50%, ο εκπαιδευόμενος έχει τη δυνατότητα επανεξέτασης των μαθημάτων αυτών μετά την ολοκλήρωση της εκπαιδευτικής διαδικασίας του προγράμματος. Η βαθμολογία που θα συγκεντρώσει κατά τη διαδικασία επανεξέτασής του είναι και η οριστική για τα εν λόγω μαθήματα, με την προϋπόθεση ότι ξεπερνά εκείνη που συγκέντρωσε κατά την κανονική διάρκεια της εκπαιδευτικής διαδικασίας. Σε διαφορετική περίπτωση διατηρείται η αρχική βαθμολογία.

7. ΛΟΙΠΕΣ ΥΠΟΧΡΕΩΣΕΙΣ ΕΚΠΑΙΔΕΥΟΜΕΝΩΝ - ΠΡΟΫΠΟΘΕΣΕΙΣ ΧΟΡΗΓΗΣΗΣ ΠΙΣΤΟΠΟΙΗΤΙΚΟΥ

Πέρα από την **επιτυχή ολοκλήρωση** του προγράμματος για τη χορήγηση του Πιστοποιητικού απαιτούνται τα εξής:

► **Συμμετοχή του εκπαιδευόμενου στη διαδικασία Δειγματοληπτικού Ελέγχου Ταυτοποίησης**

Η διαδικασία Δειγματοληπτικού Ελέγχου Ταυτοποίησης Εκπαιδευόμενου στοχεύει στη διασφάλιση της ποιότητας των παρεχομένων εκπαιδευτικών υπηρεσιών. Συγκεκριμένα, εξουσιοδοτημένο στέλεχος του Κέντρου Επιμόρφωσης και Δια Βίου Μάθησης του ΕΚΠΑ, επικοινωνεί τηλεφωνικά με ένα τυχαίο δείγμα εκπαιδευόμενων, προκειμένου να διαπιστωθεί εάν συμμετείχαν στις εκπαιδευτικές διαδικασίες του προγράμματος, εάν αντιμετώπισαν προβλήματα σε σχέση με το εκπαιδευτικό υλικό, την επικοινωνία με τον ορισμένο εκπαιδευτή τους, καθώς και με τη γενικότερη μαθησιακή διαδικασία. Η τηλεφωνική επικοινωνία διεξάγεται με την ολοκλήρωση του εκάστοτε προγράμματος, ενώ η μέση χρονική διάρκειά της συγκεκριμένης διαδικασίας είναι περίπου 2-3 λεπτά.

Σε περίπτωση μη συμμετοχής του εκπαιδευόμενου στη διαδικασία Δειγματοληπτικού Ελέγχου Ταυτοποίησης, εφόσον κληθεί, ή μη ταυτοποίησής του κατά τη διεξαγωγή της, δεν χορηγείται το πιστοποιητικό σπουδών, ακόμα και αν έχει ολοκληρώσει επιτυχώς την εξ αποστάσεως εκπαιδευτική διαδικασία.

► **Συμμετοχή του εκπαιδευόμενου στη διαδικασία Δειγματοληπτικού Ελέγχου Εγγράφων**

Ο δειγματοληπτικός έλεγχος εγγράφων διασφαλίζει την εγκυρότητα των στοιχείων που έχει δηλώσει ο εκπαιδευόμενος στην αίτηση συμμετοχής του στο Πρόγραμμα και βάσει των οποίων έχει αξιολογηθεί και εγκριθεί η αίτηση συμμετοχής του σε αυτό.

Κατά τη διάρκεια ή μετά το πέρας του προγράμματος, πραγματοποιείται δειγματοληπτικός έλεγχος εγγράφων από τη Γραμματεία. Ο εκπαιδευόμενος θα πρέπει να είναι σε θέση να προσκομίσει τα απαραίτητα δικαιολογητικά τα οποία πιστοποιούν τα στοιχεία που έχει δηλώσει στην αίτηση συμμετοχής (Αντίγραφο Πτυχίου, Αντίγραφο Απολυτήριου Λυκείου, Βεβαίωση Εργασιακής Εμπειρίας, Γνώση Ξένων Γλωσσών κ.τ.λ.).

Σε περίπτωση μη συμμετοχής του εκπαιδευόμενου στη διαδικασία Δειγματοληπτικού Ελέγχου Εγγράφων, εφόσον κληθεί, ή μη ύπαρξης των δικαιολογητικών αυτών, δεν χορηγείται το πιστοποιητικό σπουδών, ακόμα και αν έχει ολοκληρώσει επιτυχώς την εξ αποστάσεως εκπαιδευτική διαδικασία.

► **Αποπληρωμή του συνόλου των διδάκτρων**

Ο εκπαιδευόμενος θα πρέπει να μην έχει οικονομικής φύσεως εκκρεμότητες. Σε περίπτωση που υπάρχουν τέτοιες, το πιστοποιητικό σπουδών διατηρείται στο αρχείο της Γραμματείας, μέχρι την ενημέρωση της για τη διευθέτηση της εκκρεμότητας.

Αναλυτική περιγραφή των παραπάνω υπάρχει στον Κανονισμό Σπουδών:

<https://elearningekpa.gr/regulation>

8. ΣΥΓΓΡΑΦΕΙΣ ΤΟΥ ΕΚΠΑΙΔΕΥΤΙΚΟΥ ΥΛΙΚΟΥ

Οι συγγραφείς του εκπαιδευτικού υλικού είναι μέλη ΔΕΠ του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών ή και ειδικοί εμπειρογνώμονες με ιδιαίτερη συγγραφική καταξίωση, οι οποίοι κατέχουν πολύ βασικό ρόλο στην υλοποίηση του προγράμματος.

9. ΠΩΣ ΔΙΑΜΟΡΦΩΝΕΤΑΙ Η ΥΛΗ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ

Το πρόγραμμα επαγγελματικής επιμόρφωσης και κατάρτισης περιλαμβάνει **3 θεματικές ενότητες (μαθήματα)**.

ΠΕΡΙΓΡΑΦΗ ΜΑΘΗΜΑΤΩΝ

Μάθημα - Εισαγωγή στην Επιστήμη των Δεδομένων και στην R/
Rstudio

Διδακτική Ενότητα 1: Η Επιστήμη των Δεδομένων

Διδακτική Ενότητα 2: Η Χρήση της Γλώσσας R/Rstudio

Διδακτική Ενότητα 3: Εφαρμογές Επιστήμης Δεδομένων

Μάθημα - Ανάλυση Μεγάλου Όγκου Δεδομένων (Big Data)

Διδακτική Ενότητα 1: Τεχνολογίες για Ανάλυση Μεγάλων Δεδομένων (Big Data)

Διδακτική Ενότητα 2: Data Web Scraping

Διδακτική Ενότητα 3: Data Handling - Large Datasets

Διδακτική Ενότητα 4: Προεπεξεργασία Δεδομένων (Missing Values and Imputation)

Διδακτική Ενότητα 5: Data Presentation / Graphic Overview

Μάθημα - Στατιστική και Ανάλυση Δεδομένων και Μηχανική
Μάθηση

Διδακτική Ενότητα 1: Βασικές Αρχές Δειγματοληψίας (Sampling)

Διδακτική Ενότητα 2: Προγνωστικά Μοντέλα και Model Selection

Διδακτική Ενότητα 3: Outlier Detection και Stepwise Regression

Διδακτική Ενότητα 4: Case Studies με Πραγματικά Δεδομένα