
Data Analytics με την χρήση R και KNIME

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 Analytics με την χρήση R και KNIME»**.

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 Analytics με την χρήση R και KNIME»**, 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 παρουσίαση βασικών τεχνικών ανάλυσης δεδομένων μέσα από την R και το KNIME.

Οι εκπαιδευόμενοι μετά το τέλος του προγράμματος θα κατέχουν τις απαραίτητες γνώσεις και εφόδια ώστε να μπορούν να ενταχθούν σε μια ομάδα που έργο της θα είναι η ανάλυση δεδομένων. Βασιζόμενοι στο πρόγραμμα αυτό θα μπορούν να συνεχίσουν την εκμάθηση αυτού του πεδίου και να εξελιχθούν μελλοντικά σε ανεξάρτητους αναλυτές δεδομένων.

Το πρόγραμμα απευθύνεται σε:

- ▶ **Επαγγελματίες που θέλουν να αναπτύξουν δεξιότητες ανάλυσης δεδομένων**
- ▶ **Κατόχους πτυχίου που θέλουν να συνδυάσουν το αντικείμενό τους με data analytics**

Ως προς την πρακτική χρησιμότητα του αντικειμένου, το πρόγραμμα αυτό έρχεται να βοηθήσει ανθρώπους με ποικίλα γνωστικά εφόδια και επαγγελματικές εμπειρίες να αναπτύξουν δεξιότητες ανάλυσης δεδομένων. Αυτήν τη στιγμή στην αγορά εργασίας υπάρχει μεγάλη ζήτηση για άτομα που γνωρίζουν καλά έναν επιχειρηματικό τομέα και συνδυάζουν τη γνώση αυτή με τεχνικές ανάλυσης δεδομένων.

Με την παρακολούθηση αυτού του προγράμματος ένας επαγγελματίας θα μπορεί να εκμεταλλευτεί την εμπειρία του ώστε να μεγιστοποιήσει τα οφέλη από τα δεδομένα που έχει στη διάθεσή του ένας οργανισμός ή επιχείρηση.

Το πρόγραμμα αποσκοπεί στην εξοικείωση με τη γλώσσα R, μία από τις πιο διαδεδομένες γλώσσες για την ανάλυση δεδομένων. Η R είναι ανοικτή γλώσσα προγραμματισμού που έχει μεγάλη βάση χρηστών και χιλιάδες (επίσης ανοικτές) βοηθητικές βιβλιοθήκες. Υποστηρίζεται από εταιρίες όπως η Microsoft, IBM, Oracle, και τη χρησιμοποιούν μεταξύ άλλων οι Google, Facebook, Twitter, Uber, Airbnb, HP, Ford, Accenture, Deloitte, EY, KPMG, McKinsey.

Οι εκπαιδευόμενοι θα γνωρίσουν και το KNIME, το οποίο είναι επίσης ανοικτό πρόγραμμα. Παρότι το χρησιμοποιούν εταιρίες όπως η Procter & Gamble, International Flavors & Fragrances, και Fractal Analytics, δεν είναι τόσο διαδεδομένο. Το περιβάλλον του όμως είναι όμοιο με δημοφιλή εργαλεία όπως τα IBM SPSS Modeler και Alteryx.

Μέσω του KNIME οι συμμετέχοντες στο πρόγραμμα θα έχουν τη δυνατότητα να εξοικειωθούν με τις βασικές αρχές των παραπάνω. Δεν απαιτείται κάποια προηγούμενη γνώση.

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 στο αντικείμενο εργασιακή εμπειρία

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 περισσότερων 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 τηλεφωνικώς 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 **4 θεματικές ενότητες (μαθήματα)**.

ΠΕΡΙΓΡΑΦΗ ΜΑΘΗΜΑΤΩΝ

Ακολουθεί η δόμηση των μαθημάτων και μία συνοπτική περιγραφή τι μελετά η κάθε διδακτική ενότητα.

Μάθημα- Εισαγωγή στην Ανάλυση Δεδομένων

Διδακτική Ενότητα 1: Εισαγωγή στην Ανάλυση Δεδομένων

Περιγραφή των συνθηκών που καθιστούν αναγκαία την ύπαρξη αναλυτών δεδομένων.

Εισαγωγή σε γενικές έννοιες του data analytics.

Διδακτική Ενότητα 2: Προετοιμασία δεδομένων

Data cleansing. Αναφορά σε regular expressions.

Μάθημα- Γλώσσα R

Διδακτική Ενότητα 1: Εισαγωγή στη γλώσσα R

Τι είναι η R; Ιστορία της γλώσσας. Εγκατάσταση και βασικές έννοιες.

Διδακτική Ενότητα 2: Μια πρώτη ματιά στα δεδομένα (Περιγραφική Στατιστική)

Αρχική ανάλυση των δεδομένων. Χρήση περιγραφικής στατιστικής.

Διδακτική Ενότητα 3: Dimension reduction - Feature selection

Επιλογή σημαντικών μεταβλητών. Τεχνικές μείωσης των διαστάσεων στα δεδομένα.

Διδακτική Ενότητα 4: Μελέτη περίπτωσης: marketing campaign (Λογιστική Παλινδρόμηση - Δένδρα Αποφάσεων)

Λογιστική παλινδρόμηση (logistic regression). Παλινδρόμηση με δένδρα αποφάσεων (decision trees).

Διδακτική Ενότητα 5: Μελέτη περίπτωσης: customer segmentation (Clustering)

Αλγόριθμοι συσταδοποίησης (clustering) όπως K-means και hierarchical trees.

Μάθημα- Πρόγραμμα KNIME

Διδακτική Ενότητα 1: Εισαγωγή στο KNIME

Τι είναι το KNIME; Ιστορία του προγράμματος. Εγκατάσταση και βασικές τεχνικές.

Διδακτική Ενότητα 2: Clustering με το KNIME

Αλγόριθμοι συσταδοποίησης (clustering) στο KNIME.

Διδακτική Ενότητα 3: Μελέτη περίπτωσης: Παλινδρόμηση (Regression)

Γραμμική παλινδρόμηση και δένδρα αποφάσεων.

Διδακτική Ενότητα 4: Προχωρημένες τεχνικές παλινδρόμησης (Regression - advanced techniques)

Παλινδρόμηση με support vector machines, neural nets, ensembles.

Μάθημα- Παρουσίαση αποτελεσμάτων και άλλες χρήσιμες οδηγίες

Διδακτική Ενότητα 1: Εισαγωγή στο Storytelling with data

Παρουσίαση των αποτελεσμάτων.

Διδακτική Ενότητα 2: Μελέτη περίπτωσης: customer segmentation

Παρουσίαση αντίστοιχης μελέτης περίπτωσης.

Διδακτική Ενότητα 3: Μελέτη περίπτωσης: movie ratings

Παρουσίαση μελέτης περίπτωσης που αφορά αξιολόγηση ταινιών κινηματογράφου.

Διδακτική Ενότητα 4: Μελλοντικές κατευθύνσεις - Εργαλεία

Επιπλέον δυνατότητες R και KNIME (π.χ. σύνδεση με Hadoop και Spark). Οδηγίες για περαιτέρω μελέτη.

Διδακτική Ενότητα 5: Πρακτική άσκηση στην Ανάλυση Δεδομένων

Πρακτική άσκηση με την ανάλυση δεδομένων που θα επιλέξει ο κάθε σπουδαστής.