



## METADATA

**Title:** Η επιστήμη των δεδομένων μέσα από τη γλώσσα R

**Other Titles:** ΒΑΣΙΚΕΣ ΚΑΙ ΠΡΟΧΩΡΗΜΕΝΕΣ ΕΝΝΟΙΕΣ

**Language:** Greek

**ISBN:** 978-960-603-394-0

**Subject:** MATHEMATICS AND COMPUTER SCIENCE,  
ENGINEERING AND TECHNOLOGY

**Keywords:** Data Mining / Statistics / Data Science / Big Data  
Analytics / Machine Learning

**Bibliographic Reference:** Verykios, V., Kagklis, V., & Stavropoulos, E. (2015). Η επιστήμη των δεδομένων μέσα από τη γλώσσα R [Undergraduate textbook]. Kallipos, Open Academic Editions. <http://dx.doi.org/10.57713/kallipos-734>

### Abstract

Η αυξανόμενη συσσώρευση μεγάλων ποσοτήτων δεδομένων δημιουργεί νέες ευκαιρίες στους τομείς: επιστήμη, οικονομία, εκπαίδευση, έρευνα, κ.λ.π. Παλαιότερα είτε η αποθήκευση τέτοιων δεδομένων δεν ήταν εφικτή, είτε η ανάλυσή τους ξεπερνούσε κατά πολύ τις υπολογιστικές δυνατότητες των σύγχρονων υπολογιστικών συστημάτων. Σήμερα που έχει επέλθει η τεχνολογική σύγκλιση στους παραπάνω τομείς υπάρχει ανάγκη για την εκπαίδευση επιστημόνων οι οποίοι θα μπορούν να ανταπεξέλθουν στις σύγχρονες ανάγκες, δίνοντας λύσεις σε σοβαρά ζητήματα μέσω της ευφυούς ανάλυσης δεδομένων.

Σκοπός του βιβλίου είναι να καταγράψει τις βασικές αρχές που διέπουν τη νέα αυτήν επιστήμη των δεδομένων, που είναι γνωστή και ως Εξόρυξη Δεδομένων. Το βιβλίο ξεκινάει με την περιγραφή της υπάρχουσας κατάστασης στον τομέα της εξόρυξης γνώσης από τα δεδομένα, τους τομείς στους οποίους αυτή βασίζεται και τις νέες προκλήσεις. Στη συνέχεια καλύπτει ζητήματα πολυδιάστατων δεδομένων και δημιουργίας αναφορών, όπως και ζητήματα νέων μορφών βάσεων

δεδομένων, (NoSQL συστήματα). Οι τύποι και η ποιότητα των δεδομένων αποτελούν αναπόσπαστο κομμάτι του βιβλίου που μαζί με την επεξεργασία και τα μέτρα ομοιότητας αποτελούν τη βάση ενός έργου εξόρυξης δεδομένων.

Η συνοπτική στατιστική και η οπτικοποίηση αποτελούν το πρώτο βήμα για την εξερεύνηση των δεδομένων. Το κύριο κομμάτι του βιβλίου αφορά τη λεπτομερή περιγραφή και ανάλυση εναλλακτικών αλγορίθμων για την εκπόνηση των βασικών λειτουργιών της Εξόρυξης Δεδομένων οι οποίες είναι η Κατηγοριοποίηση, η Ανάλυση Συσχετίσεων και η Ανάλυση Συστάδων. Πιο προχωρημένα ζητήματα που αφορούν την εκτίμηση των παραγόμενων μοντέλων, τη σύγκρισή τους και τη γρηγορότερη εκτέλεσή τους επίσης καλύπτονται στο τελευταίο κομμάτι του βιβλίου. Όλα τα θέματα αναλύονται σε βάθος μέσω προγραμμάτων στη γλώσσα R.

Ένας μεγάλος αριθμός από λυμένες ασκήσεις βοηθούν στην κατανόηση και στην εμπέδωση των εννοιών, ενώ άλυτα προβλήματα θα βοηθήσουν το φοιτητή να αναπτύξει μία ολοκληρωμένη άποψη για το θέμα αυτό.

