

Ειδικά Θέματα Υπολογιστικής Στατιστικής: Μηχανική Δεδομένων (*Topics in Computational Statistics: Data Engineering*)

ΔΙΔΑΣΚΩΝ: Κ.Π.ΠΑΝΟΥΣΗΣ

Γενικά Στοιχεία Μαθήματος

Κωδικός: 61235

Τύπος: Επιλογής Ομάδας Μαθημάτων 2

Επίπεδο: Μεταπτυχιακό

Έτος σπουδών: Α'

Εξάμηνο σπουδών: 2^ο

ECTS: 3

Γλώσσα διδασκαλίας: Αγγλική

Περιεχόμενο Μαθήματος

Το μάθημα θέτει θεμέλια στη Μηχανική Δεδομένων δίνοντας έμφαση σε εφαρμογές Στατιστικής και Επιστήμης των Δεδομένων. Θα ξεκινήσουμε με βασικά στοιχεία της SQL όπως *queries*, *joins* και *aggregations*, καθώς και θεμελιώδεις έννοιες βάσεων δεδομένων και ανάλυσης δεδομένων στη μνήμη (*in-memory analytics*). Επιπλέον, θα κάνουμε εισαγωγή στην Python, μελετώντας βιβλιοθήκες, τύπους και δομές δεδομένων, ροές ελέγχου, λειτουργίες εισόδου/εξόδου και συνδέσεις με βάσεις δεδομένων. Στη συνέχεια, θα καλυφθούν θέματα από όλο τον κύκλο ζωής των δεδομένων, συμπεριλαμβανομένων της εισαγωγής, μετασχηματισμού, φόρτωσης, οπτικοποίησης, μοντελοποίησης, εφαρμογής, ενημέρωσης, παρακολούθησης, συντήρησης και συγγραφής εγχειριδίων. Τέλος θα επιλεχθούν προς έκθεση προχωρημένα θέματα συγγενή με την μηχανική λογισμικού που άπτονται στη σύγχρονη Στατιστική, όπως ο παραλληλισμός, η υπολογιστική νέφους, η διαχείριση σφαλμάτων, ο έλεγχος λογισμικού, ο έλεγχος εκδόσεων κ.α.

Προαπαιτούμενα

Δεν υπάρχουν

Επιδιωκόμενα Μαθησιακά Αποτελέσματα

Μετά την ολοκλήρωση του μαθήματος, οι φοιτητές θα

- Κατανοούν τις αρχές της Μηχανικής Δεδομένων και τη σημασία τους στη Στατιστική Ανάλυση.
- Έχουν στέρεες βάσεις στον προγραμματισμό με SQL και Python.
- Εκτελούν με άνεση τυπικές εργασίες μηχανικής δεδομένων με έμφαση στη στατιστική μοντελοποίηση.
- Θα έχουν πρακτική εμπειρία σε *in-memory computing* τεχνολογίες.

Συνιστώμενη Βιβλιογραφία

- McKinney, Wes. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.", 2012.
- Fangohr, Hans. Python for Computational Science and Engineering. Online Resource 2022
- Crickard, Paul. Data Engineering with Python: Work with massive datasets to design data models and automate data pipelines using Python. Packt Publishing Ltd, 2020.
- Beaulieu, Alan. Learning SQL: master SQL fundamentals. " O'Reilly Media, Inc.", 2009.
- Tanimura, Cathy. *SQL for Data Analysis*. " O'Reilly Media, Inc.", 2021.

Διδακτικές και Μαθησιακές Μέθοδοι

Δια ζώσης διδασκαλία σε 6 εβδομαδιαίες τρίωρες διαλέξεις. Κάθε εβδομάδα θα δίνεται μια εργασία (κάποιες προς βαθμολόγηση). Επιπλέον θα δοθεί και μία ομαδική εργασία.

Μέθοδοι Αξιολόγησης και Βαθμολόγησης

Ο τελικός βαθμός προκύπτει από το σταθμισμένο μέσο των βαθμών της γραπτής εξέτασης (50%) και των εργασιών (50%).