

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΜΑΘΗΜΑ: ΕΠΙΧΕΙΡΗΜΑΤΙΚΗ ΕΥΦΥΙΑ ΚΑΙ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

ΕΡΓΑΣΙΑ ΕΞΑΜΗΝΟΥ-ΑΚΑΔ. ΕΤΟΣ 2026-2027

Στόχος της εργασίας

Στόχος της εργασίας είναι η εφαρμογή τεχνικών εξόρυξης δεδομένων σε δεδομένα, είτε με τη χρήση του λογισμικού Weka (https://www.cs.waikato.ac.nz/ml/weka/), είτε με τη χρήση της γλώσσας python.

Δραστηριότητες (Tasks)

Δραστηριότητα 1: Συλλογή Δεδομένων

Μπορείτε να επιλέξετε να χρησιμοποιήσετε για την εργασία σας ένα ήδη υπάρχον σετ δεδομένων ή να συλλέξετε δικά σας πρωτογενή δεδομένα. Ενδεικτικές πηγές για σετ δεδομένων θα βρείτε εδώ:

http://archive.ics.uci.edu/ml/datasets.html?sort=nameUp&view=list
http://www.cs.ubc.ca/labs/beta/Projects/autoweka/datasets/
https://www.kaggle.com/
https://datasetsearch.research.google.com/
https://www.stats.govt.nz/large-datasets/csv-files-for-download/
https://www.bl.uk/collection-metadata/downloads
https://openml.org/search?type=data&sort=runs&status=active
https://sigma.ai/open-datasets/

Φροντίστε το σετ δεδομένων σας να έχει τουλάχιστον 10 χαρακτηριστικά και τουλάχιστον τριψήφιο αριθμό παραδειγμάτων.

Επίσης, φροντίστε να καταλαβαίνετε τον μικρόκοσμο του σετ δεδομένων. Τι σημαίνει το κάθε χαρακτηριστικό εισόδου, οι τιμές του, η έξοδος. Διαφορετικά δεν θα μπορέσετε να κάνετε αξιολόγηση των αποτελεσμάτων σας.

Δραστηριότητα 2: Βιβλιογραφική Επισκόπηση

Μελετήστε την βιβλιογραφία σχετικά με

- την εφαρμογή που έχετε επιλέξει (πχ τι πρόσφατες καλές σχετικές εργασίες έχουν δημοσιευτεί για εξόρυξη δεδομένων πάνω στην ίδια εφαρμογή/θεματική περιοχή/domain που έχετε επιλέξει)
- το σετ δεδομένων που έχετε επιλέξει (ποιοι άλλοι το έχουν χρησιμοποιήσει σε έρευνά τους, και πώς).

Δραστηριότητα 3: Μετασχηματισμός των δεδομένων

Σε περίπτωση που συλλέγετε πρωτογενή δεδομένα, αυτά θα πρέπει

- να καθαριστούν από θόρυβο, διπλότυπα κλπ
- να μετασχηματιστούν σε διανύσματα χαρακτηριστικών-τιμών (feature value vectors, με την επιλογή κατάλληλων χαρακτηριστικών μάθησης).

Δραστηριότητα 4: Προεπεξεργασία των δεδομένων

Μελετήστε την προοπτική εφαρμογής διαφόρων τεχνικών προεπεξεργασίας των παραδειγμάτων μάθησης, εφόσον είναι εφαρμόσιμες και χρήσιμες στα δεδομένα σας:

- Διακριτοποίηση αριθμητικών μεταβλητών
- Κανονικοποίηση (Normalization of attribute values, …)
- Φιλτραρίσματα (Επιλογές σετ χαρακτηριστικών, επιλογές σετ παραδειγμάτων, feature selection, …)
- Εξισορρόπηση (Imbalanced datasets, SMOTE, resampling, one-sided sampling, Tomek links, …)
- Μετατροπή του αρχείου δεδομένων σε csv format

Δραστηριότητα 5: Μηχανική Μάθηση
Μελέτη τεχνικών μηχανικής μάθησης, και εκτενείς πειραματισμοί με εφαρμογή τους στα δεδομένα. Πειραματισμοί με διάφορα σετ παραμέτρων των αλγορίθμων, με διάφορα validation σχήματα κλπ.
Κανένας αλγόριθμος δεν εφαρμόζεται σαν μαύρο κουτί – πρέπει να είναι κατανοητή η λογική του, οι βασικές του παράμετροι και ο τρόπος λειτουργίας του.

Δραστηριότητα 6: Απεικόνιση, ανάλυση και αξιολόγηση των αποτελεσμάτων
Βρείτε τα κατάλληλα μέτρα αξιολόγησης για να εκτιμήσετε την απόδοση των πειραμάτων σας (πχ ποσοστό σωστών ταξινομήσεων, ακρίβεια, ανάκληση, f-score, Kappa-score, ROC, RMSE, RAE, SSE κλπ) και παρουσιάστε τα αποτελέσματά σας.
Απεικόνιση των αποτελεσμάτων με πίνακες, γραφήματα, ιστογράμματα κλπ.
Ανάλυση και ποιοτική αξιολόγηση των αποτελεσμάτων
Σύγκριση των διαφόρων μεθόδων προεπεξεργασίας
Σύγκριση των διαφόρων τεχνικών μηχανικής μάθησης

Άρθρο εργασίας (μέχρι την ημερομηνία των εξετάσεων)
Συγγραφή άρθρου τουλάχιστον 6 σελίδων, σε 12 pt γραμματοσειρά, single line-spacing, που θα περιλαμβάνει

Περίληψη (150 λέξεις)
Εισαγωγή
- Εισαγωγή στον ερευνητικό χώρο της εργασίας
- Βασικές προσεγγίσεις επίλυσής του
- Συνεισφορά της εργασίας
- Σχεδιάγραμμα του άρθρου

Ερευνητικός χώρος (με βιβλιογραφική επισκόπηση)
Αλγόριθμοι μηχανικής μάθησης
Μεθοδολογική διαδικασία
- Περιγραφή δεδομένων
- Περιγραφή προεπεξεργασίας
- Πειράματα μάθησης
- Ανάλυση και αξιολόγηση

Συμπεράσματα και προτάσεις για μελλοντικές βελτιώσεις
Βιβλιογραφία

Παραδοτέα
Ενδιάμεσες παρουσιάσεις προόδου: Θα ανακοινωθούν από τους διδάσκοντες

Το άρθρο της εργασίας θα πρέπει να έχει υποβληθεί στην αντίστοιχη εργασία που θα ανοίξει στο opencourses του μαθήματος (https://opencourses.ionio.gr/courses/DDI200/**) μέχρι την ημερομηνία των εξετάσεων του μαθήματος. Η εργασία θα μετρήσει 80% του τελικού βαθμού και η γραπτή εξέταση 20%.**

Την ημέρα των εξετάσεων του μαθήματος, κάθε φοιτητής
- **θα πρέπει να έχει ετοιμάσει 10λεπτη παρουσίαση της εργασίας του με διαφάνειες**
- **θα πρέπει να παρουσιαστεί την ώρα των εξετάσεων στην αίθουσα που αναγράφεται στο πρόγραμμα των εξετάσεων, προκειμένου να παρουσιάσει την εργασία του και να εξεταστεί πάνω σε αυτή. Όποιος φοιτητής δεν παρουσιαστεί στις εξετάσεις δεν θα βαθμολογηθεί για το μάθημα, ανεξάρτητα αν έχει υποβάλει ηλεκτρονικά το άρθρο της εργασίας του.**