



**ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ**  
**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ – ΤΜΗΜΑ ΤΟΥΡΙΣΜΟΥ**  
**ΜΠΣ «ΨΗΦΙΑΚΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΣΤΗΝ ΔΙΟΙΚΗΣΗ ΤΗΣ**  
**ΦΙΛΟΞΕΝΙΑΣ ΚΑΙ ΤΟΝ ΤΟΥΡΙΣΜΟ»**  
**ΜΑΘΗΜΑ: ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΝΟΗΜΟΣΥΝΗ ΣΤΟΝ ΤΟΥΡΙΣΜΟ**

**«ΕΥΦΥΗΣ ΑΝΑΛΥΣΗ ΤΟΥΡΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ»**

**Στόχος της εργασίας**

Στόχος της εργασίας είναι η εφαρμογή ευφών τεχνικών ανάλυσης τουριστικών δεδομένων με τη χρήση του λογισμικού Weka (<https://www.cs.waikato.ac.nz/ml/weka/>).

**Δραστηριότητες (Tasks)**

**Δραστηριότητα 1: Βιβλιογραφική Επισκόπηση**

Χρησιμοποιώντας το [scholar.google.com](https://scholar.google.com), αναζητήστε σχετική βιβλιογραφία στον χώρο της μηχανικής μάθησης σε τουριστικά δεδομένα. Μια αρχή:

<https://link.springer.com/article/10.1007/s11135-021-01234-3#Tab2>

**Δραστηριότητα 2: Μελέτη των δεδομένων**

Φορτώστε στο Weka το αρχείο τουριστικών δεδομένων `booking.csv`.

Μελετήστε μέσω της καρτέλας `pre-process` του Weka τα δεδομένα σας.

Φτιάξτε πίνακες που να δείχνουν τις διαφορετικές τιμές που μπορεί να πάρει το κάθε διακριτό/ονοματικό χαρακτηριστικό εισόδου, και το εύρος τιμών του κάθε αριθμητικού χαρακτηριστικού εισόδου.

Φτιάξτε γραφήματα για να δείξετε πώς η τιμή του κάθε χαρακτηριστικού εισόδου κυμαίνεται βάσει των τιμών της εξόδου.

Φτιάξτε γράφημα που να δείχνει την κατανομή των τιμών της εξόδου στα δεδομένα. Είναι ισορροπημένη η κατανομή, ή υπάρχει ανισορροπία;

**Δραστηριότητα 3: Μηχανική Μάθηση -  $k$**

Πραγματοποιήστε πειράματα ταξινόμησης με τον αλγόριθμο των  $k$ -πλησιέστερων γειτόνων ( $k$ ) και 10-fold cross validation. Εφαρμόστε διάφορες τιμές του  $k$ , με στόχο να βρείτε εκείνη την τιμή που θα σας οδηγήσει στη βέλτιστη απόδοση ταξινόμησης. Παρουσιάστε τα αποτελέσματά σας για τις διάφορες τιμές του  $k$  σε ένα συγκριτικό γράφημα. Δείξτε σε πίνακα/γράφημα τα αποτελέσματά σας (precision και recall για την κάθε τιμή της εξόδου).

**Δραστηριότητα 4: Μηχανική Μάθηση – J48**

Πραγματοποιήστε πειράματα ταξινόμησης με δέντρα απόφασης (J48) και 10-fold cross validation. Τρέξτε πειράματα και με ακλάδευτο (unpruned TRUE) και με κλαδεμένο (Reduced Error Pruning) δέντρο. Δείξτε σε εικόνα το δέντρο που επάγεται. Δείξτε σε πίνακα/γράφημα τα αποτελέσματά σας (precision και recall για την κάθε τιμή της εξόδου).

**Δραστηριότητα 5: Μηχανική Μάθηση – Σύγκριση**

Συγκρίνετε τα αποτελέσματα των δυο ταξινομητών, φτιάχνοντας ένα συγκριτικό γράφημα με τις καλύτερες αποδόσεις του καθενός. Ποιος τα πάει καλύτερα και σε τι;

### **Δραστηριότητα 6: Ανάλυση και αξιολόγηση των αποτελεσμάτων**

Αναλύστε ποιοτικά τα αποτελέσματα που βγάλατε.

Ποια χαρακτηριστικά είναι σημαντικά για την πρόβλεψη της εξόδου;

Ποιά τιμή της εξόδου είναι πιο εύκολο/δύσκολο να προβλεφθούν σε σχέση με την άλλη; Πού πιστεύετε ότι οφείλεται αυτό;

Τι λάθη γίνονται στην ταξινόμηση; Παρατηρείστε τα παραδείγματα που ταξινομούνται λάθος.

Που πιστεύετε ότι οφείλεται το λάθος; Μπορείτε να αναγνωρίσετε κάτι που μπερδεύει τον ταξινομητή και τον οδηγεί σε λάθος απόφαση;

**ΠΑΡΑΔΟΤΕΟ (ΜΕΧΡΙ ΤΗΝ ΗΜΕΡΟΜΗΝΙΑ ΤΩΝ ΕΞΕΤΑΣΕΩΝ): Μια αναφορά μέχρι 10 σελίδες, που θα περιγράψει την δουλειά που κάνατε στα πλαίσια των Δραστηριοτήτων 1-6.**