

# Εργασία 1

---

## Εισαγωγή

Η εργασία αυτή έχει ως στόχο την εξοικείωση με τις τεχνικές πολλαπλών μοντέλων πρόβλεψης (Μάθημα 2), την εμπέδωση των προβλημάτων του διαφορετικού κόστους εσφαλμένης ταξινόμησης (Μάθημα 3) και της ασυμμετρίας κλάσεων (Μάθημα 4) καθώς και των λύσεων που υπάρχουν σε αυτά τα προβλήματα, και τέλος την απόκτηση πρακτικής εμπειρίας στο σωστό τρόπο εκτέλεσης πειραμάτων συγκριτικής αξιολόγησης μεταξύ αλγορίθμων (Μάθημα 5).

## Μέρος Α

Στο πρώτο μέρος της εργασίας θα δουλέψετε στο αντικείμενο των τεχνικών πολλαπλών μοντέλων πρόβλεψης σε συνδυασμό με το αντικείμενο της συγκριτικής αξιολόγησης μεταξύ αλγορίθμων. Συγκεκριμένα, θα πρέπει να επιλέξετε 10 dataset από κάποιο repository (π.χ. [UCI](#)) και να παρουσιάσετε μια συγκριτική αξιολόγηση τεσσάρων από τις τεχνικές πολλαπλών μοντέλων που είδαμε στο Μάθημα 2, ακολουθώντας τις πρακτικές που είδαμε στο Μάθημα 5.

## Μέρος Β

Στο δεύτερο μέρος της εργασίας θα δουλέψετε στο πρόβλημα του διαφορετικού κόστους ταξινόμησης. Χρησιμοποιείτε το σύνολο Statlog (Heart) από το [UCI repository](#) και τον πίνακα κόστους που το συνοδεύει. Εφαρμόστε τρεις από τις τεχνικές που συζητήσαμε στο μάθημα 3, συνδυάζοντας τες με τους αλγορίθμους μάθησης Random Forest, Linear SVM και Naive Bayes.

## Μέρος Γ

Στο τρίτο μέρος της εργασίας θα δουλέψετε στο πρόβλημα της ασυμμετρίας κλάσεων. Χρησιμοποιείτε το παρακάτω dataset: <https://www.kaggle.com/mlg-ulb/creditcardfraud>. Εφαρμόστε τρεις από τις τεχνικές που συζητήσαμε στο μάθημα 4, συνδυάζοντας τες με τους αλγορίθμους μάθησης Random Forest, Linear SVM και Naive Bayes.

## Λογιστικά

Οι εργασίες προτείνεται να γίνουν σε ομάδες 2-3 ατόμων. Θα πρέπει να παραδώσετε από ένα notebook (<https://jupyter.org/>) για κάθε μέρος της εργασίας. Κάθε notebook θα περιέχει τον κώδικα και συνδέσμους προς τα δεδομένα που χρησιμοποιήσατε. **Ακόμη, θα παρουσιάζει και θα σχολιάζει τα αποτελέσματα των πειραμάτων σας.** Προτεινόμενη γλώσσα εργασίας, αλλά όχι περιοριστική, είναι η γλώσσα Python (3.6 +). Θα πρέπει να παραδώσετε μέσω του elearning ένα αρχείο .zip με όλα τα notebooks. Προθεσμία υποβολής: 3/5/2019. Εκπρόθεσμη υποβολή μέχρι και 2 μέρες μετά την προθεσμία θα έχει ως αποτέλεσμα τη μείωση του βαθμού κατά 10% την ημέρα.