Εργασία 3 – Classification and Clustering (με το WEKA)

Το αρχείο δεδομένων **car.arff** περιλαμβάνει την αξιολόγηση 1728 αυτοκινήτων και την κατάταξή τους σε τέσσερεις κατηγορίες (στήλη class): μη αποδεκτό (unacc), αποδεκτό (acc), καλό (good) και πολύ καλό (vgood). Επίσης, περιλαμβάνει τα ακόλουθα χαρακτηριστικά των αυτοκινήτων:

- buying: τιμή αγοράς (χιλιάδες ευρώ)
- maint: έξοδα συντήρησης (χιλιάδες ευρώ)
- doors: πλήθος θυρών
- persons: πλήθος ατόμων
- lug_boot: μέγεθος αποθηκευτικού χώρου (λίτρα)
- safety: επίπεδο ασφάλειας (1-χαμηλό, 2-μέτριο, 3-υψηλό)

Κατηγοριοποίηση

(1) Χρησιμοποιείστε τον WEKA Experimenter και πειραματιστείτε με τους αλγόριθμους IBk και J48 χρησιμοποιώντας διάφορες παραμέτρους τους και κατασκευάζοντας εναλλακτικά μοντέλα πάνω στο car.arff. Δώστε μια αναφορά σχετικά με τα αποτελέσματα των πειραμάτων σας. Βρήκατε κάποιο μοντέλο που να είναι καλύτερο από τα άλλα (με επίπεδο σημαντικότητας 5% στο t-test) ;

(2) Τρέξτε νέα πειράματα χρησιμοποιώντας τον παρακάτω πίνακα κόστους:

```
a    b    c    d   <-- classified as
0    3    9   27 |    a = unacc
0    0    3    9 |    b = acc
0    0    0    3 |    c = good
0    0    0    0 |    d = vgood
```

Η κατανομή των εμφανίσεων των τιμών (unacc, acc, good, vgood) της class variable στα δεδομένα είναι (1210, 384, 69, 65) αντίστοιχα. Στην παρατήρηση αυτή και στο γεγονός ότι οι τιμές της κλάσης είναι διαβαθμισμένες (ranked) βασίζεται η φιλοσοφία της εκθετικής κατανομής των βαρών του πίνακα κόστους που σας δίνεται. Με άλλα λόγια μας ενδιαφέρει ένα μοντέλο να μην προβλέπει ως αποδεκτά μη-αποδεκτά αυτοκίνητα. Σχολιάστε τα αποτελέσματα των πειραμάτων σας.

Συσταδοποίηση

(3) Χρησιμοποιείστε τους αλγόριθμους συσταδοποίησης Hierarchical και KMeans και προσπαθήστε να βρείτε την καλύτερη συσταδοποίηση σε 4 συστάδες. Ως καλύτερη εννοούμε αυτή που πετυχαίνει τη μεγαλύτερη ακρίβεια με την μέθοδο αξιολόγησης classes to clusters evaluation.

(4) Χρησιμοποιείστε τον πίνακα κόστους του ερωτήματος (2) για να αξιολογήσετε ξανά τις 3 καλύτερες συσταδοποιήσεις που βρήκατε στο ερώτημα (3).