



ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΤΕΧΝΟΛΟΓΙΕΣ ΕΥΦΥΟΥΣ ΔΙΑΧΕΙΡΙΣΗΣ ΑΝΘΡΩΠΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ
ΕΡΓΑΣΤΗΡΙΟ

ΑΣΚΗΣΗ: WEKA: DECISION TREES
([HTTP://WWW.CS.WAIKATO.AC.NZ/ML/WEKA](http://www.cs.waikato.ac.nz/ml/weka))

1. Ξεκινήστε το Weka και φορτώστε το αρχείο *student-mat.v2* . Το αρχείο αυτό δείχνει την επίδοση μαθητών δευτεροβάθμιας εκπαίδευσης στην Πορτογαλία στα Μαθηματικά.

```
# Attributes for both student-mat.csv (Math course) and student-por.csv (Portuguese language course) datasets:
1 school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
2 sex - student's sex (binary: 'F' - female or 'M' - male)
3 age - student's age (numeric: from 15 to 22)
4 address - student's home address type (binary: 'U' - urban or 'R' - rural)
5 famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6 Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
7 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 for 5th to 9th grade, 3 for
secondary education or 4 for higher education)
8 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 for
secondary education or 4 for higher education)
9 Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police),
'at_home' or 'other')
10 Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police),
'at_home' or 'other')
11 reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12 guardian - student's guardian (nominal: 'mother', 'father' or 'other')
13 traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1
hour)
14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15 failures - number of past class failures (numeric: n if 1<=n<3, else 4)
16 schoolsup - extra educational support (binary: yes or no)
17 famsup - family educational support (binary: yes or no)
18 paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19 activities - extra-curricular activities (binary: yes or no)
20 nursery - attended nursery school (binary: yes or no)
21 higher - wants to take higher education (binary: yes or no)
22 internet - Internet access at home (binary: yes or no)
23 romantic - with a romantic relationship (binary: yes or no)
24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29 health - current health status (numeric: from 1 - very bad to 5 - very good)
30 absences - number of school absences (numeric: from 0 to 93)

# these grades are related with the course subject, Math or Portuguese:
31 G1 - first period grade (numeric: from 0 to 20)
31 G2 - second period grade (numeric: from 0 to 20)
32 G3 - final grade (numeric: from 0 to 20, output target) ΔΙΑΚΡΙΤΟΠΟΙΗΜΕΝΟ
```

2. Στην καρτέλα Classify, στο Classifier και επιλέξτε στον φάκελο Trees τον αλγόριθμο *j48*.
3. Η επιλογή 'minNumObj' καθορίζει τον ελάχιστο αριθμό αντικειμένων στα φύλλα του δέντρου.
4. Επιλέξτε Unpruned=True, για να φτιάξετε ένα δέντρο χωρίς κλάδεμα.
5. Τρέξτε τον αλγόριθμο. Πώς είναι τα αποτελέσματα;
6. Παρατηρήστε το δέντρο σας. Ποιο είναι το πιο κατάλληλο χαρακτηριστικό για την ταξινόμησή σας;
7. Επιλέξτε Reduced Error Pruning. Ξανατρέξτε τον αλγόριθμο. Πώς είναι τα αποτελέσματά σας;
8. Αφαιρέστε τα χαρακτηριστικά των βαθμών των προηγούμενων τριμήνων. Ξανατρέξτε τον αλγόριθμο. Τι παρατηρείτε;
9. Φορτώστε το αρχείο *student-mat.v3*. Τρέξτε τον *j48*. Πώς τα πάει σε σχέση με πριν; Παρατηρείστε το καινούριο δέντρο. Έχει αλλάξει;
10. Φορτώστε το αρχείο *student-mat.v4*. Τρέξτε τον *j48*. Πώς τα πάει σε σχέση με πριν; Παρατηρείστε το καινούριο δέντρο. Έχει αλλάξει;