

PRÀCTICA 3: PREDICCIÓ DE RESULTATS ACADÈMICS



L'any 2008 **Paulo Cortez** and **Alice Silva** van presentar a una conferència a Porto l'estudi [*Using Data Mining to Predict Secondary School Student Performance*](#).

L'estudi recull dades del curs 2005 - 2006 de dos instituts públics de la regió portuguesa d'Alentejo: *Gabriel Pereira* i *Mousinho da Silveira*.

Els investigadors van crear el set de dades amb informació pròpia de les escoles (al que van afegir camps com les notes finals i trimestrals a més de les absències) i dels resultats d'un qüestionari que ajudaven a identificar dades demogràfiques (educació dels pares o ingressos familiars), socials (consum d'alcohol) i escolars (nombre d'assignatures suspeses en cursos anteriors).

El qüestionari va ser validat per un grup de docents i provat inicialment amb un conjunt de 15 alumnes.

La versió final del qüestionari va tenir 37 preguntes que van respondre 788 alumnes. Per problemes de qualitat de les dades acadèmiques es van haver de descartar 111 qüestionaris al no poder vincular-los amb exactitud amb el set acadèmic.

Les dades es van dividir en dos grups per analitzar separatament el rendiment acadèmic de les assignatures Matemàtiques (395 observacions) i Portuguès (649 observacions).

Alguns dels atributs plantejats al qüestionari també es van desestimar de l'anàlisi per manca de resposta, com els ingressos familiars. Els atributs finals són:

- **Sex:** female (F)/male (M)
- **Age:** 15 - 22
- **School:** Gabriel Pereira/Mousinho da Silveira

- **Address:** urban/rural
- **Pstatus** (Parent's cohabitation status): living together (T)/living apart (A)
- **Medu** (Mother's education): 0 (none)/1 (primary education - 4th grade)/2 (5th to 9th grade)/3 (secondary education)/4 (higher)
- **Mjob** (Mother's job): Nominal description
- **Fedu** (Father's education): mateix rang de valors que Medu
- **Fjob** (Father's job): Nominal description
- **Guardian:** mother, father or other (nominal)
- **Famsize** (Family size): <=3 o >= 3
- **Famrel** (Quality of family relationship): from 1 - very bad - 5 - excellent
- **Reason** (Reason to choose this school): nominal
- **Traveltime** (Home to school travel time): 1: < 15 minutes, 2: 15 - 30 minutes, 3: 30 - 60 minutes, 4: >60 minutes
- **Studytime** (Weekly study time): 1: < 2 hours, 2: 2 - 5 hours, 3: 5 - 10 hours, 4: >10 hours
- **Failures** (number of past class failures): 1 to 4
- **Schoolsup** (extra educational school support): Yes/No
- **Famsup** (family educational support): Yes/No
- **Activities** (extra curricular activities): Yes/No
- **Paidclass** (extra paid class): Yes/No
- **Internet** (Internet access at home): Yes/No
- **Nursery** (attended nursery school): Yes/No
- **Higher** (wants to take higher education): Yes/No
- **Romantic** (with a romantic relationship): Yes/No
- **Freetime** (free time after school): 1: very low - 5: very high
- **Goout** (going out with friends): 1: very low - 5: very high
- **Walk** (weekend alcohol consumption): 1: very low - 5: very high
- **Dalc** (workday alcohol consumption): 1: very low - 5: very high
- **Health** (current health status): 1: very low - 5: very high
- **Absences:** number of school absences (0 to 93)
- **G1** (first period grade): 0 - 20
- **G2** (second period grade): 0 - 20
- **G3** (final grade): 0 - 20

NOTA: Les qualificacions en el sistema portuguès s'assignen en una escala entre 0 i 20. S'aconsegueix aprovar a partir de 10.

A partir del dataset centrat en avaluar el rendiment acadèmic dels alumnes matriculats a portuguès (*portuguese_hs_students.csv*) elabora la resposta dels següents apartats:

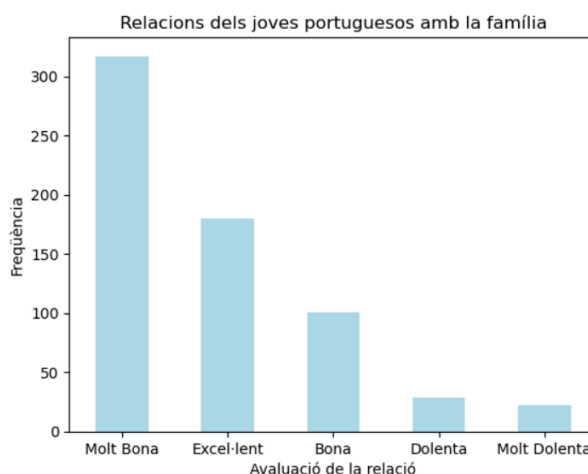
TREBALL FINAL

- Explica a partir de les dades com és l'alumnat d'aquesta assignatura (10%)

Exemples:

- Com són les relacions familiars?

```
def catFamRel (valor):  
    if valor == 1:  
        return "Molt Dolenta"  
    elif valor == 2:  
        return "Dolenta"  
    elif valor == 3:  
        return "Bona"  
    elif valor == 4:  
        return "Molt Bona"  
    else:  
        return "Excel·lent"
```



- Com són les relacions familiars en funció de l'edat?
- Com són les relacions familiars en funció dels resultats acadèmics finals?
- ...
- Implementa un algorisme que intenti predir les absències de l'alumnat. Avalua el resultat del model (30%)
- Implementa un algorisme que intenti predir si l'alumnat aprovarà o no ($G3 [Nota final] \geq 10$). Avalua el resultat del model (30%)
- Implementa un algorisme que ajudi a predir si l'alumnat serà Excepcional ($G3 [Nota final] \geq 19$). El resultat d'aquest model haurà d'ajudar al Ministeri d'Ensenyament en la presa de decisió d'assignació de beques. Avalua el resultat del model (30%)