

# Trabajo Final

Éste es el trabajo final del módulo Data Analyzing, donde se pide realizar lo siguiente con los datos de notas en matemática de estudiantes, obtenidos de la página:

<https://www.kaggle.com/janiobachmann/math-students>

Importar éstos datos en un notebook de Python. Los atributos para los conjuntos de datos de student-mat.csv (curso de matemáticas) son:

- 1 school - escuela del estudiante (binario: 'GP' - Gabriel Pereira o 'MS' - Mousinho da Silveira)
- 2 sex - sexo del estudiante (binario: 'F' - femenino o 'M' - masculino)
- 3 age - edad del estudiante (numérico: de 15 a 22)
- 4 address - tipo de dirección particular del estudiante (binario: 'U' - urbano o 'R' - rural)
- 5 famsize - tamaño de la familia (binario: 'LE3' - menor o igual a 3 o 'GT3' - mayor que 3)
- 6 Pstatus - estado de cohabitación de los padres (binario: 'T' - viviendo juntos o 'A' - separados)
- 7 Medu - educación de la madre (numérico: 0 - ninguno, 1 - educación primaria (4º grado), 2 - 5º a 9º grado, 3 - educación secundaria o 4 - educación superior)
- 8 Fedu - educación del padre (numérico: 0 - ninguno, 1 - educación primaria (4º grado), 2 - 5º a 9º grado, 3 - educación secundaria o 4 - educación superior)
- 9 Mjob - trabajo de la madre (nominal: 'maestra', relacionado con la atención de la 'salud', 'servicios' civiles (por ejemplo, administrativo o policial), 'en casa' u 'otro')
- 10 Fjob - trabajo del padre (nominal: 'maestro', relacionado con el cuidado de la salud, 'servicios' civiles (por ejemplo, administrativo o policial), 'en casa' u 'otro')

- 11 reason: razón para elegir esta escuela (nominal: cerca de 'hogar', escuela 'reputación', preferencia de 'curso' u 'otro')
- 12 guardian: tutor del estudiante (nominal: 'madre', 'padre' u 'otro')
- 13 traveltime: tiempo de viaje de la casa a la escuela (numérico: 1 - 1 hora)
- 14 studytime - tiempo de estudio semanal (numérico: 1 - 10 horas)
- 15 failures: número de fallos de clases anteriores (numérico: n si 1)
- 16 schoolup - apoyo educativo adicional (binario: sí o no)
- 17 famsup - apoyo educativo familiar (binario: sí o no)
- 18 paid - pagas extra dentro de la asignatura del curso (matemáticas o portugués) (binario: sí o no)
- 19 activities - actividades extracurriculares (binario: sí o no)
- 20 nursery - asistió a la guardería (binario: sí o no)
- 21 higher - quiere cursar estudios superiores (binario: sí o no)
- 22 internet - acceso a internet en casa (binario: sí o no)
- 23 romantic - con una relación romántica (binario: sí o no)
- 24 famrel - calidad de las relaciones familiares (numérico: de 1 - muy mala a 5 - excelente)
- 25 freetime - tiempo libre después de la escuela (numérico: de 1 - muy bajo a 5 - muy alto)
- 26 goout - salir con amigos (numérico: de 1 - muy bajo a 5 - muy alto)
- 27 Dalc - consumo de alcohol en la jornada laboral (numérico: de 1 - muy bajo a 5 - muy alto)
- 28 Walc - consumo de alcohol durante el fin de semana (numérico: de 1 - muy bajo a 5 - muy alto)

- 29 health - estado de salud actual (numérico: de 1 - muy malo a 5 - muy bueno)
- 30 absences - número de ausencias escolares (numérico: de 0 a 93)
- 31 G1 - nota del primer período (numérico: de 0 a 20)
- 32 G2 - nota del segundo período (numérico: de 0 a 20)
- 33 G3 - nota final (numérica: de 0 a 20)

La variable dependiente (objetivo de predicción) es la nota final, el resto son variables independientes o predictoras. Se desea contar con un apropiado modelo de predicción de las notas finales, tal que los errores de predicción sean mínimos. En ese sentido, se pide construir un algoritmo que estime varios modelos de regresión, de la nota G3 en función de un subconjunto de las variables independientes, y seleccionar aquél modelo cuyo error cuadrático medio sea el menor.

- Como variables predictoras puede utilizar las 32 variables independientes o un subconjunto menor, basado en criterios estadísticos para elegir qué variables podrían predecir la nota final.
- Los modelos que se pueden considerar son la regresión lineal, la regresión Ridge y/o la regresión Lasso.
- Puede incluir términos polinomiales de las variables numéricas.
- Si se va a considerar variables categóricas, tiene que utilizar necesariamente la transformación correspondiente en variables dicotómicas.

Enviar el notebook construido en Python y con su nombre completo al correo: [alcivale@gmail.com](mailto:alcivale@gmail.com), hasta el jueves 10 de septiembre a horas 23:59.