

Análisis de la influencia del consumo de alcohol en los resultados académicos utilizando aprendizaje automático

José Antonio Yébenes Montenegro
Máster en Inteligencia de Negocio y Big Data
Directora: Dra. Laia Subirats Maté

Índice

1. Contexto y justificación
2. Objetivos
3. Metodología
4. Conclusiones
5. Líneas de trabajo futuro

Contexto y justificación del trabajo

El consumo de alcohol en exceso es perjudicial a cualquier edad pero en menores es un problema que nos afecta a todos.

El objetivo de este trabajo es estudiar si existe relación entre el consumo de alcohol y los resultados académicos en estudiantes de educación secundaria.

Más de 480.000 menores se han emborrachado en el último mes. ¿Hacemos algo? Campaña de FAD

**EL PROBLEMA CON EL ALCOHOL EN MENORES
ES QUE NO VEMOS QUE SEA NUESTRO PROBLEMA**

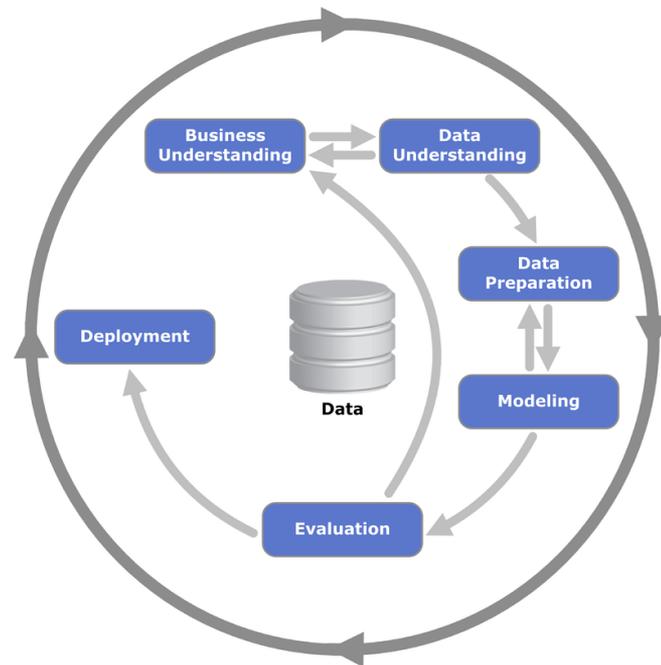
Objetivos

Los datos utilizados corresponden a información social y académica de un grupo de estudiantes de educación secundaria de edades comprendidas entre los 15 y los 22 años de dos escuelas Portuguesas.

- Analizar la calidad de los datos
- Entender los datos mediante visualizaciones sencillas
- Entender los datos a través del análisis estadístico univariante y bivariante
- Segmentar los datos
- Detectar reglas de asociación
- Detectar las variables más influyentes en el hecho de superar o no la asignatura
- Construir un modelo predictivo sobre la variable objetivo (superar o no la asignatura)

Metodología CRISP-DM

1. Comprensión del negocio
2. Comprensión de los datos
3. Preparación de datos
4. Modelado
5. Evaluación
6. Despliegue



Conclusiones

En el grupo que suspende la proporción de estudiantes que no quiere realizar estudios superiores es mayor (20 %) que en general (8,52 %). En el grupo que suspende la proporción de estudiantes del colegio Mousinho da Silveira es mayor (36,95 %) que en general (26,95 %). El número de suspensos previos es una de las variables más influyente a la hora de predecir si se aprueba o no, la nota media final no supera el aprobado cuando hay suspensos previos.

En el grupo que suspende la proporción de estudiantes que no quiere realizar estudios superiores es mayor (20 %) que en general (8,52 %).

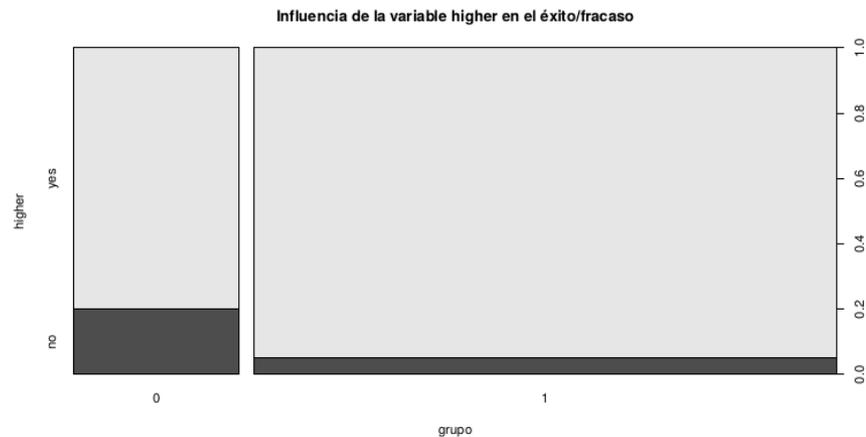


Figura 3.10: Influencia de la intención de continuar con estudios superiores

Conclusiones

El consumo de alcohol influye en las notas finales pero existen otras variables que influyen más como el colegio al que va, si quiere seguir estudiando, los suspensos previos, si hace refuerzo escolar, los estudios de los padres o el número de ausencias.



Attribute usage:

100.00%	school	Escuela del alumno
100.00%	failures	Número de suspensos previos
100.00%	higher	Quiere realizar estudios superiores
99.04%	traveltime	Tiempo de casa a la escuela
97.53%	schoolsup	Clases extras
95.47%	fedu	Formación del padre
94.10%	absences	Número de ausencias
93.96%	paid	Clases extras de pago
91.77%	famrel	Relación familiar
84.22%	pstatus	Situación de los padres
83.26%	fjob	Trabajo del padre
83.26%	guardian	Persona responsable del alumno
82.44%	goout	Sale con los amigos
78.88%	studytime	Tiempo de estudio semanal
78.88%	freetime	Tiempo libre después de la escuela
73.25%	mjob	Trabajo de la madre
69.68%	dalc	Consumo de alcohol entre semana
69.14%	health	Estado de salud
64.75%	sex	Género del alumno
62.55%	nursery	Atención en la enfermería
62.00%	reason	Motivo elección de la escuela
58.02%	age	Edad
51.17%	activities	Actividades extraescolares
46.91%	alcohol	Consumo global de alcohol
41.98%	address	Tipo de domicilio del alumno

Líneas de trabajo futuro

Algunas tareas que se podrían hacer para profundizar en el estudio y que no han dado tiempo de abortar serían;

- Mejorar el modelo predictivo usando técnicas de extracción de características PCA o SVD
- Selección de variables, incluso añadir variables derivadas de las existentes
- El conjunto de datos tiene más 1 que 0, seguramente por eso el modelo predice mucho mejor los 1 (Sensitivity = 0.9593) y en cambio predice correctamente muy pocos 0 (Specificity = 0.2464). Seguramente balanceando el conjunto de datos seguramente se conseguiría mejorar el modelo.

Muchas gracias