

# Next Educación Máster in Big Data & Business Intelligence 2023/2024

## Asignatura 3: Técnicas de análisis y explotación de datos

### Trabajo propuesto

Descargue el siguiente dataset adjunto, el cual esta basado en este [artículo](#). Dado este conjunto de datos, se pide lo siguiente:

1. Realizar un análisis exploratorio del conjunto de datos, investigando las variables que se crea que son más relevantes, justificando el procesado y las transformaciones que sean necesarias para llevarlo a cabo (20%)
2. Obtener dos modelos predictivos de clasificación binaria (30%). Evaluar los modelos y analizar las métricas pertinentes:
  - 2.1. Un modelo que detecte en base a los datos si un usuario es fumador o no.
  - 2.2. Un modelo que detecte en base a los datos si un usuario controla o no las calorías ingeridas.
3. Obtener un modelo predictivo de clasificación multiclase que en base a los datos obtenga el grado de obesidad de una persona (30%). Evaluar el modelo y analizar las métricas pertinentes.
4. Aplicar técnicas de clustering para segmentar a la población del dataset (20%). Evaluar los modelos y analizar las métricas pertinentes.

## Entregable

Se entregarán los siguientes archivos de resultados:

- Cuaderno de RStudio (archivo .Rmd y archivo .html) con la resolución del ejercicio
- Memoria de trabajo donde se detallen los pasos que se han dado, las decisiones tomadas, los resultados, la interpretación de los mismos y un último apartado de conclusiones

Otras consideraciones:

- El ejercicio podrá resolverse en grupos de 4 personas
- Se realizará un único entregable por grupo
- El documento de trabajo estará estructurado convenientemente y donde figure el nombre de todos los integrantes del grupo
- No se especifica límite de páginas
- No está permitido el plagio, entre compañeros u otras fuentes. Dicho plagio podrá ser penalizado con la anulación del trabajo para las partes involucradas.
- El código empleado tiene que ser 100% original del grupo.