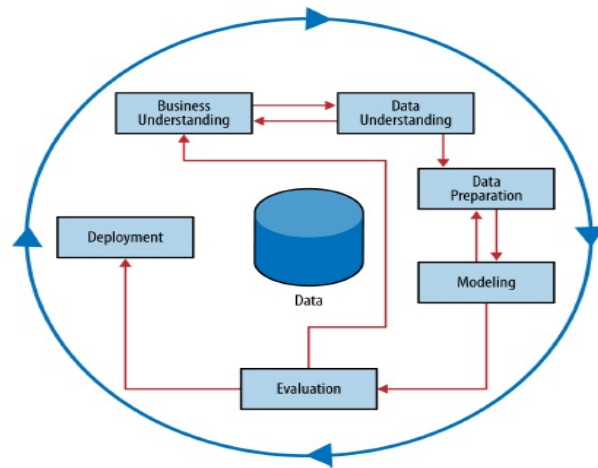


Examen de Mayo

Prof. Eduardo M. Sánchez Vila

1. **[2 puntos]** Recomendación de tapas. **El ejercicio consiste en predecir la valoración que el usuario “u31” realizará sobre la tapa “t1” que ofrece un determinado restaurante.** Para ello dispones de dos bloques de datos: (1) las valoraciones que ha dado un conjunto de usuarios a la tapa “t1”, 2) las valoraciones del usuario “u31” sobre las tapas que ha consumido previamente. Los datos se encuentran en el fichero **acue_examen_dataset.csv**, y los atributos de cada registro se describen en la primera fila del fichero (“characterTapa” indica el tipo de tapa y tiene dos posibles valores: daring (atrevida) y traditional (tradicional)). Se pide lo siguiente:
 - **[0,5 puntos]** Propón **1 método de predicción** que haga uso de los datos disponibles. Explica los cálculos que hay que realizar
 - **[1 punto]** Implementa el método, realiza los cálculos, y calcula la predicción.
 - **[0,5 puntos]** Por último, si pudieras acceder a más datos, indica qué variables o factores adicionales, tanto de usuario como de producto, necesitarías para poder utilizarlo en la práctica. Con esos nuevos datos, ¿qué estrategia o modelo de recomendación aplicarías para resolver este problema?
2. **[1 punto]** Ejercicio del accidente del taxi. Suponed que ahora tenemos una tercera compañía de taxis que son de color rojo, y que la proporción de taxis es: Verde (70%), Azul (20%) y Rojo (10%) . El testigo dice que vio el taxi de color Azul e identifica correctamente un determinado color en un 80% de los casos. ¿Cuál es la probabilidad de que el taxi responsable del accidente sea Azul?
3. **[1 punto]** ¿Cómo se estiman las preferencias del usuario en un sistema de recomendación basado en contenido? Calcula las preferencias de un usuario cuyas compras en un supermercado han sido recogidas en el fichero TT_dataset_compras.xls.
4. **[1 punto]** En un proyecto de *Data Science* se realizan diferentes fases, comenzando por el entendimiento del problema (*business understanding*) y concluyendo con el despliegue final de la solución (*deployment*).



- **[0,5 puntos]** Selecciona y justifica qué fase del proyecto : (1) es la más importante, y (2) es la que requiere más tiempo de trabajo.
- **[0.5 puntos]** Si consideramos los perfiles más populares del científico de datos (ingeniero informático, estadístico, matemático, físico, economista, etc.), selecciona y justifica qué perfil o perfiles serían los más idóneos para trabajar en cada fase del proyecto.