

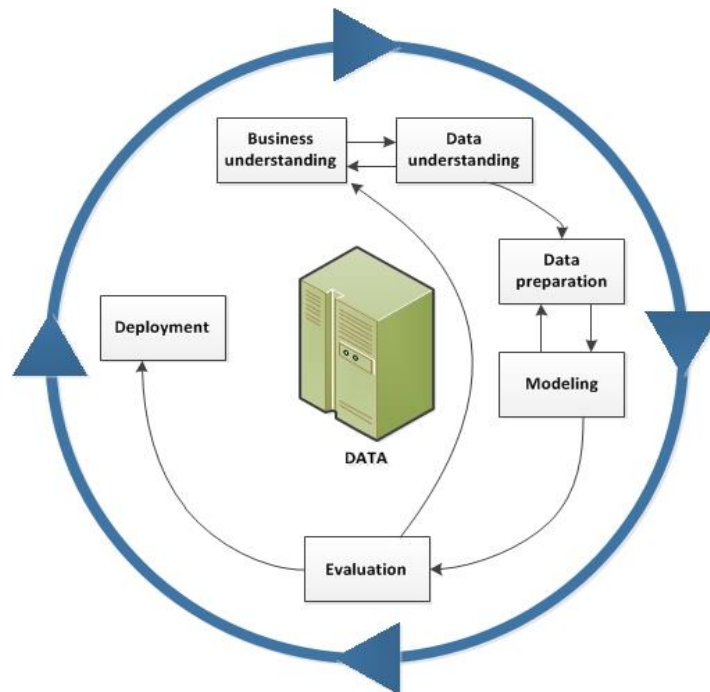
Obligatorio de Analítica de Negocios 2

11 de mayo de 2021

Vamos a trabajar con la base de datos **Hitters** del paquete **ISLR**, la cual contiene 322 observaciones y 20 variables de jugadores de la MLB (Major League Baseball) para los años 1986 y 1987.

El objetivo es crear un modelo que permita, en base a las estadísticas históricas y del año anterior, estimar el salario para el próximo año.

Para elaborar este modelo hay que seguir las fases de un proyecto analítico, es decir, definir el problema, entender los datos (análisis descriptivo), modelar distintos modelos y evaluarlos, luego se determina el modelo a implementar y se comunica la decisión.



En base a la base de datos anterior, se pide:

- 1) Realizar un análisis descriptivo de la base de datos (10 puntos).
- 2) Del punto anterior, se puede apreciar que hay 59 datos perdidos para la variable Salary la cual queremos estimar. Dado este problema se sugieren dos alternativas, se deberá elegir una de las dos justificando la respuesta (10 puntos).
 - a. Eliminar los 59 datos correspondientes a los datos perdidos.
 - b. Realizar un análisis de cluster con todas las variables salvo Salary, y estimar los datos perdidos como el Salary promedio según el grupo.
- 3) Aplicar una técnica de Análisis de Componentes Principales, mejora los resultados de los modelos, fundamente la respuesta (5 puntos).
- 4) Elaborar un cuadro que compare los errores de los modelos en train y test de al menos tres modelos posibles (10 puntos).
- 5) Elegir el modelo para implementar, justifique su respuesta (5 puntos).