



## Inteligencia y Analítica de Negocios

Profesor: Felipe Ortega.

### Práctica final T2 (Parte I)

- **Datos:** Fichero <https://www.kaggle.com/fivethirtyeight/the-ultimate-halloween-candy-power-ranking/data> (origen: <http://walthickey.com/2017/10/18/whats-the-best-halloween-candy/> ).
- **Tecnologías:** Python y Jupyter notebook o R/RStudio y RMarkdown.
- **Fecha de entrega:** Domingo, 6 de mayo de 2018, a las 23:55.
- En esta primera práctica vamos a afianzar el uso de herramientas de evaluación de modelos y validación de resultados.

### Evaluación de modelos

1. Cargar el fichero de datos "candy-data.csv", definiendo para ello la información de tipos de datos necesaria para interpretar las columnas adecuadamente en el lenguaje de programación seleccionado.
2. El modelo que debemos ajustar es una regresión múltiple, utilizando como variable de salida los valores de la columna `winpercent` y como variables explicativas (entrada) todas las demás columnas. El objetivo del análisis es identificar qué factores están más relacionados con la elección de un tipo de caramelo como ganador sobre los restantes.

Se pueden obtener ideas adicionales sobre cómo se han utilizado los datos para un análisis real en este [post de 538](#) sobre el mismo dataset.

- a) Primero divide la muestra total de datos en un 75% para training y un 25% para testing. Ajusta un modelo de regresión múltiple sobre los datos de training y luego comprueba su efectividad para predecir los valores de los datos de testing. **(3 puntos)**.

- b) Utilizando *bootstrap*, calcula nuevos intervalos de confianza (95%) para los estimadores de todos los regresores del modelo. ¿Son consistentes todos los valores obtenidos con los resultados del apartado anterior?. ¿Qué variables seleccionaría para mantenerlas en el modelo? **(3 puntos)**.
  
- c) Ahora realiza una validación cruzada (*k-fold validation*) con  $K=5$ . Compara la nueva estimación del MSE con la obtenida inicialmente ¿Qué podemos concluir acerca de la robustez del modelo?. **(4 puntos)**.