

# Práctica de Aprendizaje Supervisado

## Introducción

El aprendizaje supervisado construye un modelo en base a ejemplos conocidos. El objetivo es ser poder realizar predicciones con datos futuros utilizando este modelo entrenado. Estas predicciones pueden ser categoricas (clasificación) o numericas (regresión).

En clase hemos visto diversos modelos: Naive Bayes, SVM, Redes de Neuronas, Decision Trees y Random Forests.

El objetivo de la práctica es obtener una mayor soltura en el uso de los mismos para problemas reales. Algunos modelos proporcionan una mayor precisión pero poca capacidad explicativa (caja negra) mientras otros proporcionan una mejor explicación y peor precisión (caja blanca).

Se propone la realización de una de las dos siguientes opciones utilizando al menos un modelo de caja negra y otro de caja blanca así como su aplicación a un problema de clasificación y de regresión.

### Opción 1: Bank Marketing

Se trata de un conjunto de datos que proporciona información de campañas de marketing de un banco Portugués, que se puede descargar de <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>.

El objetivo es construir un clasificador para predecir si el cliente va a contratar un depósito (variable y del conjunto de datos). Adicionalmente se solicita que se pruebe un modelo de regresión para predecir la edad (variable age) en función de las otras variables.

### Opción 2: Give Me Some Credit

Se trata del conjunto de datos de la competición de Kaggle “Give me some credit” que se puede descargar de <https://www.kaggle.com/c/GiveMeSomeCredit>.

El objetivo es extender el ejemplo inicial que se presento en clase y mejorar su capacidad predictiva. Para ello se pueden probar nuevas features, optimizar el clasificador elegido, probar con diferentes modelos, etc. Se debe evaluar el modelo haciendo submissions reales en kaggle en esta competición. Adicionalmente se solicita que se pruebe un modelo de regresión para predecir el sueldo mensual de la persona (variable MonthlyIncome).

## Criterios de evaluación

Se tendrán en cuenta los siguientes criterios de evaluación:

- Calidad de la memoria explicativa.
- El rigor de la experimentación (train/test/validation).
- El número de modelos utilizados y la forma de optimizar cada uno de ellos.
- La creatividad y la mejora predictiva en la creación de nuevas features.

## **Entrega**

Es necesario hacer entrega de los scripts necesarios para obtener los resultados así como de un documento (de preferencia en pdf) que resuma los resultados obtenidos.

Se puede utilizar indistintamente R o Python y cualquiera de los paquetes vistos en clase u otros.

La fecha límite de entrega para optar a la puntuación sobre 10 es el miércoles 8 de Abril a las 23:59. A partir de esa fecha las prácticas punturan sobre 8 restando 1 punto por cada semana de retraso acumulado. Por ejemplo si se entrega el martes 14 de Abril (sobre 8), el miércoles 15 de Abril (sobre 7), el miércoles 22 de Abril (sobre 6) y así sucesivamente....