

certamen II - Data Mining

Sebastian Niklitschek

Parte práctica

Problema 1: 30 puntos.

El objetivo de este problema es establecer un modelo que apoye una campaña de promoción vía telefónica de un producto en una determinada institución financiera brasileña. Se le solicita construir un modelo que pueda predecir con anticipación, si un cliente aceptará o no contratar un producto ofertado por la vía antes mencionada. Para ello, la institución financiera cuenta con data histórica de campañas similares realizadas en el pasado, las que contienen información respecto al cliente contactado y una variable que indica si el cliente decidió contratar la oferta o no. Encontrará el conjunto de datos en la dirección:

<https://archive.ics.uci.edu/ml/datasets/bank+marketing>

Se le solicita lo siguiente:

- Describa las variables que componen el conjunto de datos bank-additional-full.csv y genere estadísticos de resumen apropiados para ellas. Describa los resultados (5 puntos);
- Genere modelos utilizando los métodos de Bosques aleatorios, Redes neuronales, Máquinas de soporte vectorial y potenciación. Evalúe cada modelo utilizando validación cruzada con 10 grupos en paralelo. ¿Qué método entrega los mejores resultados? (Utilice la regla de Condorcet para realizar su elección en caso de dudas).

Nota: El problema tiene una bonificación adicional de 5 puntos si descarga directamente el archivo (.zip) con R y extrae de él sólo lo necesario creando un archivo temporal.