

Técnicas de Machine Learning para Análisis de Datos

Obligatorio

13 de mayo de 2019

El objetivo del obligatorio es predecir si un cliente va a cancelar su abono en una empresa de telecomunicaciones. Este tipo de problemas de clasificación binario se conoce como “churn”.

Para ello, se dispone de un archivo `dataset.csv` provisto por la empresa de telecomunicaciones que contiene datos históricos etiquetados del comportamiento de un conjunto de clientes.

Además, la empresa definió la lista de clientes de los cuales está interesada en predecir si van a cancelar su suscripción. Los datos de estos clientes están en el archivo `test_sample.csv`.

La evaluación del obligatorio será de la siguiente manera:

1. Un máximo de 30 puntos se destinarán a evaluar la aplicación de los conceptos vistos en las clases teóricas y prácticas de la materia. Esto comprende:
 - a. Aplicación de todos los algoritmos de machine learning estudiados.
 - b. Comparación de las hipótesis usando métricas adecuadas de error.
 - c. Selección justificada de las hipótesis utilizando validación cruzada con 5 folds.
2. Un máximo de 10 puntos se destinarán a evaluar los resultados obtenidos en los clientes a predecir de acuerdo a la modalidad y al criterio que se definirán oportunamente.

La entrega consistirá de:

1. Los scripts R utilizados para la obtención de las hipótesis.
2. Un informe detallado de los resultados obtenidos y decisiones tomadas.
3. El archivo `test_sample_predict.csv` con la predicción en el formato que se definirá más adelante.