

Curso Data Science Code Space Academy

Módulo 5. Algoritmos de Machine Learning

Challenge: Ajusta el mejor modelo predictivo que ayude a una entidad financiera de préstamos a clasificar las solicitudes de préstamos entre (1) tienen riesgo de impago y (0) son seguras.

Características del desafío:

1. Se compartirán dos *dataset* de la entidad Lending Club:
 - a. El primer *dataset* será el conjunto de datos que se utilizará para ajustar el modelo de predicción. Este *dataset* contiene un número elevado de variables independientes (X) las cuales se reducirán para quedarse con las más apropiadas (lo veremos en clase) y la variable dependiente (y) que se quiere aprender a predecir: *bad_loans*
 - b. El segundo *dataset* será el conjunto de datos que se utilizará para evaluar finalmente tu modelo predictivo (no confundir con la división entre train y test). Esta *dataset* estará contiene solamente un número elevado de variables independientes (X), pero no la variable independiente (y), de manera que no se pueda utilizar en el entrenamiento y sean datos que el modelo nunca haya visto.
2. Utiliza todas las técnicas que has aprendido en el módulo 4 para preparar los datos:
 - a. Análisis previo.
 - b. Detección y eliminación de outliers.
 - c. Imputación de datos.
 - d. Codificación.
 - e. Escalado.
3. Evalúa de forma preliminar todos los modelos que has visto en la primera clase del módulo 5 con técnicas como la validación cruzada.
4. Utiliza todas las técnicas de ajuste de hiperparámetros que has visto en clase:
 - a. Búsqueda exhaustiva.
 - b. Búsqueda random
5. (Para los más atrevidos) Organiza tu código en Pipelines.
6. Una vez has encontrado un modelo que se presente un buen comportamiento en el proceso de entrenamiento y test (*best_model*), prueba con el dataset de evaluación, guarda el resultado en formato .csv y envíalo a rubencpardo@gmail.com, con las siguientes indicaciones:
 - a. Características del modelo que has utilizado: tipo e hiperparámetros.
 - b. Features utilizadas.
 - c. Score (*accuracy_score*) obtenido durante el proceso de train y test para ese modelo.
7. Los resultados se podrán enviar hasta el **jueves 18 de mayo** coincidiendo con el comienzo del siguiente curso, y se publicará en el campus el ranking de resultados.

Enlace al colab de plantilla: <https://colab.research.google.com/drive/1IJrWei0E0c6-pjBQ8xd2AvG1EW2ZGXq?usp=sharing>