

# Inteligencia de Negocio: Práctica 3

## Competición de Kaggle

David Cabezas Berrido

Grupo 2: Viernes

[dxabezas@correo.ugr.es](mailto:dxabezas@correo.ugr.es)

1 de enero de 2021

# Índice general

1. Pruebas realizadas . . . . .	2
---------------------------------	---

## 1. Pruebas realizadas

Intento	Fecha-hora	Posición	Validación	Test	Preprocesado	Modelo
1	19/12/2020 14:35:08	5	0.8185	0.7420	Preprocesado 1	RandomForest por defecto
2	20/12/2020 10:51:19	4-5	0.8900	0.7580	Preprocesado 2	RandomForest por defecto
3	20/12/2020 12:34:18	4-5	0.9040	0.7498	Preprocesado 2 eliminando de train los ejemplos que inicialmente tenían descuento	RandomForest con 350 estimadores de profundidad máxima 20
4	20/12/2020 17:05:16	3-4	0.9179	0.7645	Preprocesado 3	MLPClassifier con capas ocultas de tamaño (200,200)
5	21/12/2020 09:47:11	3-4	0.8750	0.6877	Preprocesado 3	2-NN con distancia Manhattan y pesos inversamente proporcionales a la distancia
6	21/12/2020 13:30:36	3	0.9143	0.7739	Preprocesado 3	C-SVM con C=65 y kernel RBF
7	21/12/2020 15:32:20	3	0.9144	0.7645	Preprocesado 3	RandomForest por defecto
8	22/12/2020 09:49:44	2	0.9143	0.8059	Preprocesado 3	GradientBoosting con 500 estimadores
9	22/12/2020 09:50:03	2	0.9304	0.8016	Preprocesado 3	Stacking: - RandomForest por defecto - MLPClassifier con capas ocultas (200,200) - C-SVM con C=65 y kernel RBF - GradientBoosting con 500 estimadores
10	22/12/2020 14:40:39	2	0.9312	0.7636	Preprocesado 3	AdaBoost con 500 árboles de profundidad 12 y learning rate de 1.1
11	23/12/2020 10:24:20	2	0.9163	0.7990	Preprocesado 3 sustituyendo los valores perdidos por 0 en la columna descuento en lugar de eliminar la columna	GradientBoosting con 500 estimadores
12	23/12/2020 12:12:38	2		0.7998	Preprocesado 3	Moda (predicción más frecuente) de los intentos 4, 6, 7, 8, 9, 10 y 11
13	24/12/2020 19:54:10	2	0.9205	0.7886	Preprocesado 3	GradientBoosting con 500 árboles de profundidad 6, learning rate de 1.175 y submuestras del 70 %
14	25/12/2020 09:59:17	4	0.8535	0.7239	Preprocesado 3 + PCA con 0.95 de varianza explicada	GradientBoosting con 500 estimadores
15	25/12/2020 10:25:21	4	0.9159	0.8007	Preprocesado 3 tras corregir error en LabelEncoder	GradientBoosting con 500 estimadores

16	25/12/2020 11:33:44	4		0.8093	Preprocesado 3	Stacking de cuatro GradientBoosting con número de estimadores 450, 500, 550 y 600 respectivamente; tasas de aprendizaje 0.14, 0.12, 0.1, 0.08 respectivamente; todos con submuestras del 90 %
17	26/12/2020 10:58:03	4		0.8085	Preprocesado 3	Stacking anterior con la opción passthrough
18	26/12/2020 11:12:21	4	0.9268	0.7886	Preprocesado 3	HistGradientBoosting por defecto
19	26/12/2020 11:31:13	4		0.8024	Preprocesado 3	Stacking de tres GradientBoosting con 500, 550 y 600 estimadores respectivamente; tasas de aprendizaje 0.12, 0.1 y 0.08 respectivamente; todos con submuestras del 90 %. También tres HistGradientBoosting con 100, 150 y 200 iteraciones máximas respectivamente
20	27/12/2020 11:00:35	4		0.8016	Preprocesado 3	Stacking de cuatro GradientBoosting con número de estimadores 450, 500, 550 y 600 respectivamente; tasas de aprendizaje 0.14, 0.12, 0.1, 0.08 respectivamente; todos con submuestras del 90 %. También dos HistGradientBoosting con 100 y 200 iteraciones máximas respectivamente
21	27/12/2020 13:54:7	4	(0.8392)	0.7886	Preprocesado 3	GradientBoosting con 550 árboles de profundidad 2, tasa de aprendizaje de 0.15 y submuestras del 90 %
22	27/12/2020 14:30:41	4	(0.8462)	0.7790	Preprocesado 3	LightGBM con 125 árboles de profundidad máxima 8 y 27 nodos hoja como máximo; tasa de aprendizaje del 0.08
23	28/12/2020 09:21:46	4	0.9290	0.7808	Preprocesado 3	LightGBM con 200 árboles con profundiad máxima 14
24	28/12/2020 09:22:13	4	0.9291	0.7843	Preprocesado 3	LightGBM con 125 árboles con 29 nodos hoja como máximo; tasa de aprendizaje del 0.11
25	28/12/2020 09:16:14	4	0.9266	0.7817	Preprocesado 3	LightGBM por defecto
26	29/12/2020 10:42:25	4	(0.8151) 0.8990	0.7964	Preprocesado 3	MLP con early stopping
27	29/12/2020 10:52:45	4	(0.7920) 0.9135	0.778	Preprocesado 3	SVM con C=40
28	29/12/2020 13:27:51	4	(0.8352) 0.9207	0.8016	Preprocesado 3	XGBoost con 200 árboles de profundidad 3
29	30/12/2020 09:45:53	4	(0.8370) 0.9262	0.7929	Preprocesado 3	HistGradientBoosting con 75 iteraciones máximas

30	30/12/2020 09:43:11	4	0.9300	0.7774	Preprocesado 3	HistGradientBoosting con 200 iteraciones máximas, tasa de aprendizaje del 0.08 y árboles con 29 nodos hoja como máximo
31	30/12/2020 10:08:35	4	0.9304	0.8110	Preprocesado 3	Stacking de GradientBoosting con 500 estimadores, MLP con early stopping, XGBoost con 200 árboles de profundidad 3 y HistGradientBoosting con 75 iteraciones máximas
32	31/12/2020 10:14:38	5	0.9268	0.8162	Preprocesado 3	Stacking de GradientBoosting con 500 estimadores, MLP con early stopping y XGBoost con 200 árboles de profundidad 3
33	31/12/2020 11:29:18	5	0.9225	0.8067	Preprocesado 3	Stacking de GradientBoosting con 500 estimadores y MLP con early stopping

Tabla 1: Pruebas realizadas