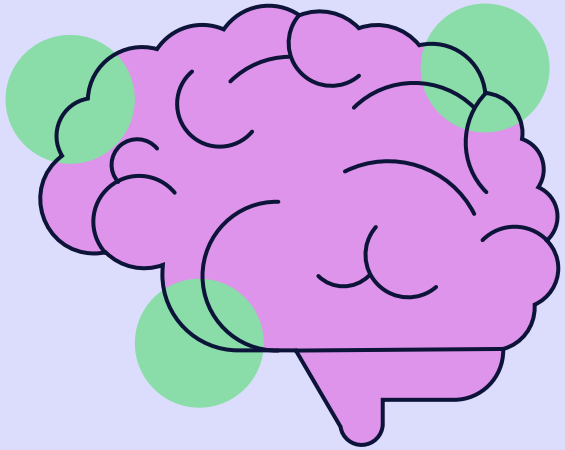
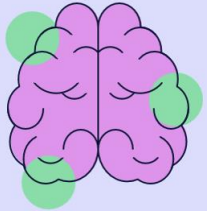


Proyecto ML:
Predicción del riesgo de sufrir un ictus

¿Qué es un ictus?

- Trastorno de la circulación cerebral que altera la función de una determinada región del cerebro
- 2ª causa de muerte en España, 1º en mujeres
- 120.000 casos al año



Proyecto de Machine Learning

Objetivo

- Predecir el riesgo de una persona de sufrir un ictus y prevenirlo

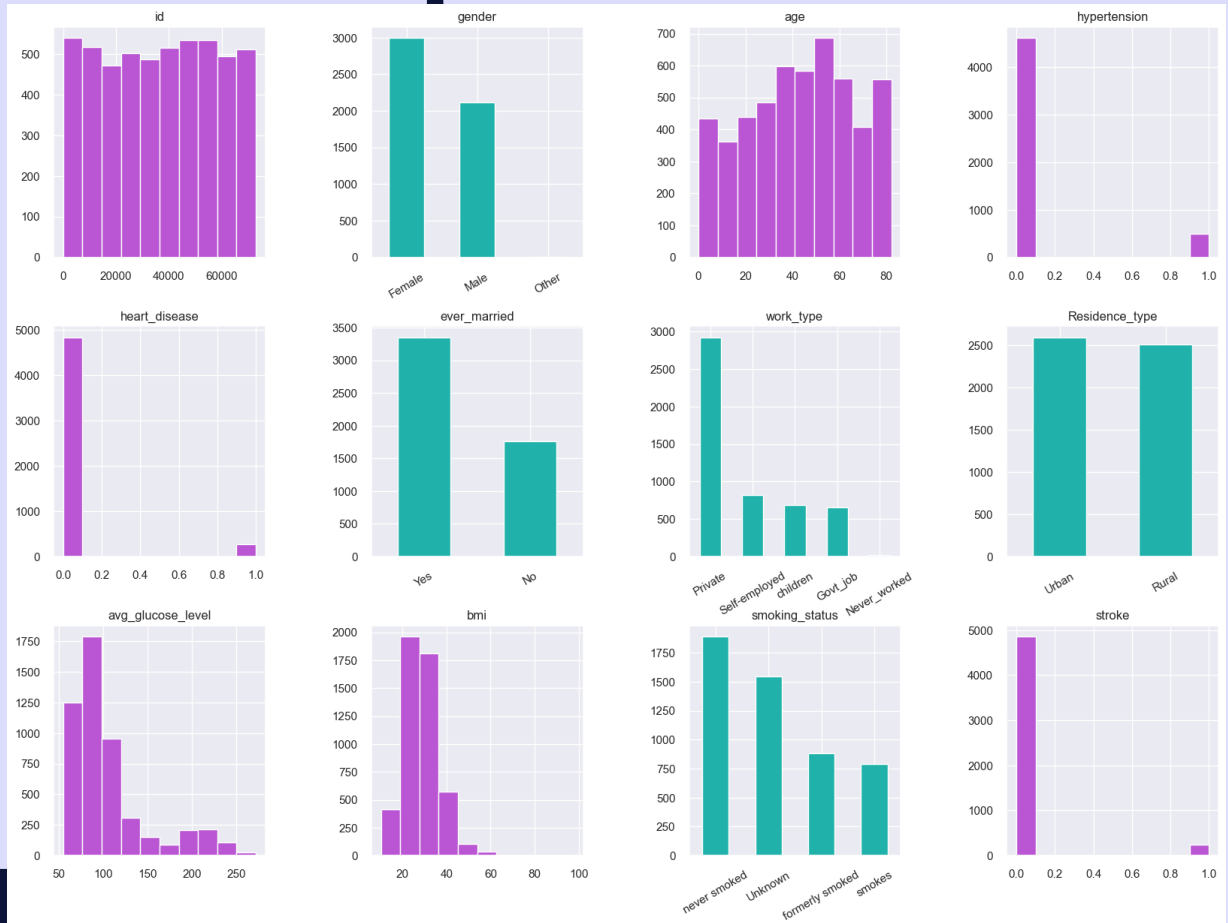
Dataset

- Datos personales: edad, gérenero...
- Datos clínicos: hipertensión, niveles de glucosa, si fuma...

Tipo de problema

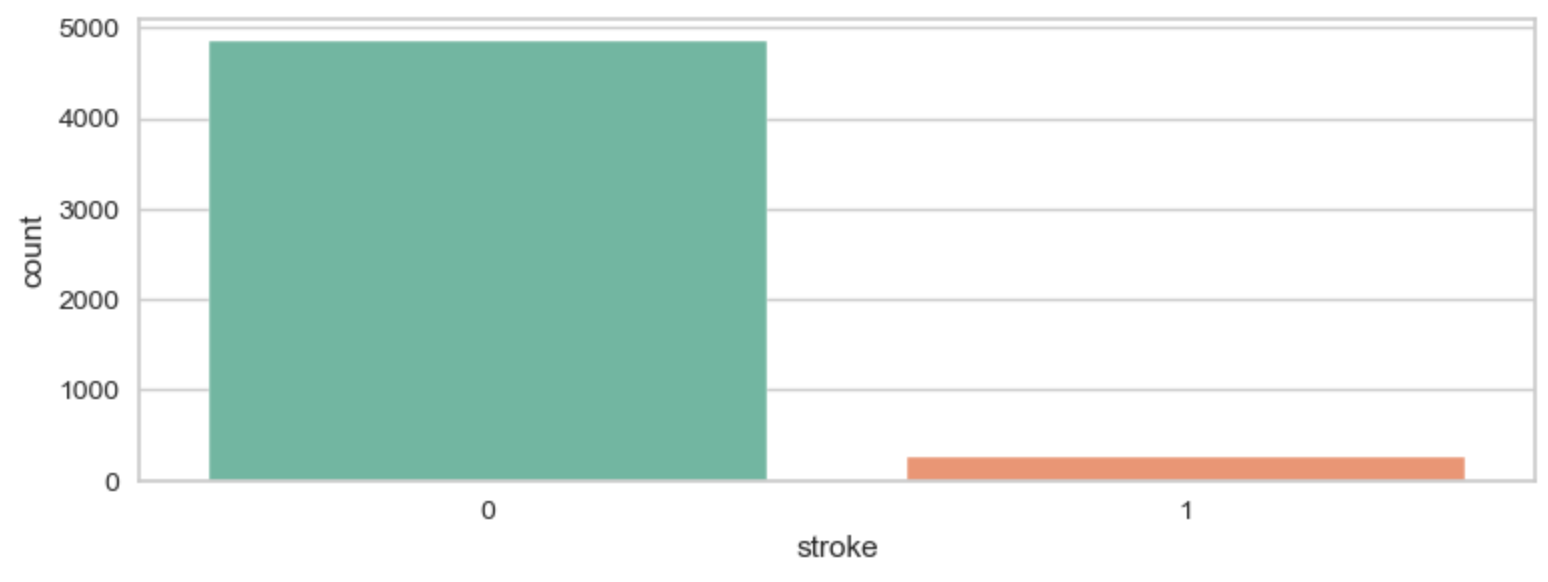
- Clasificación binario

Análisis exploratorio de datos

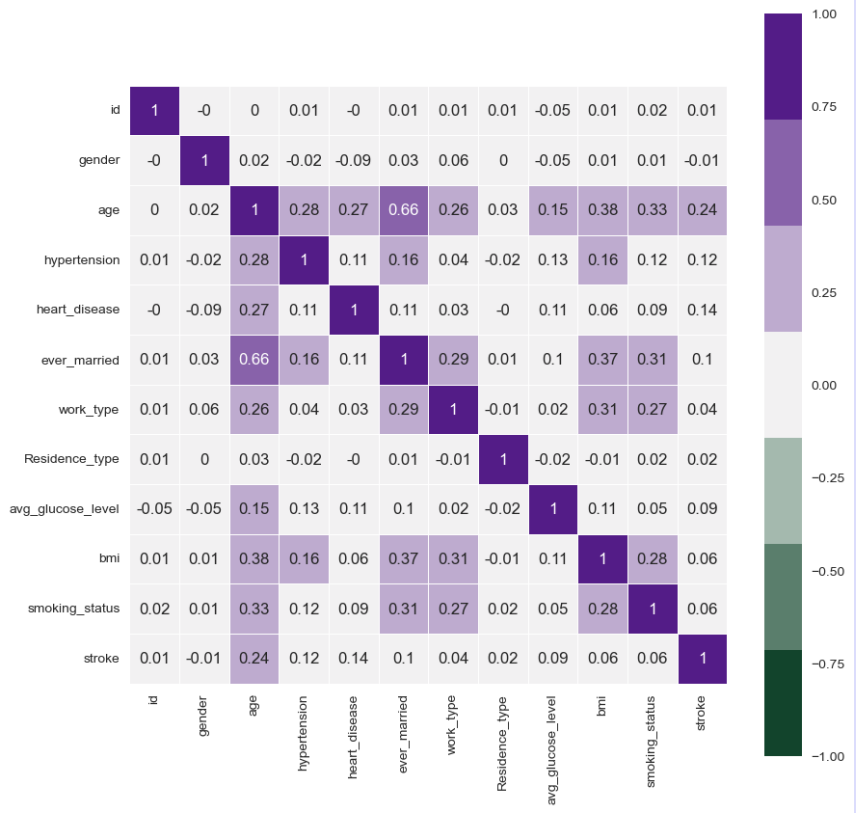


Análisis exploratorio de datos

Target



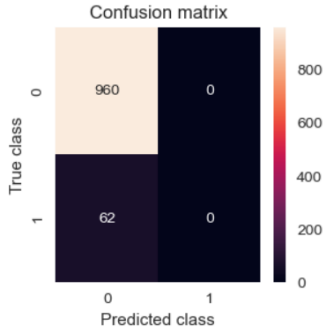
Análisis exploratorio de datos



Modelos

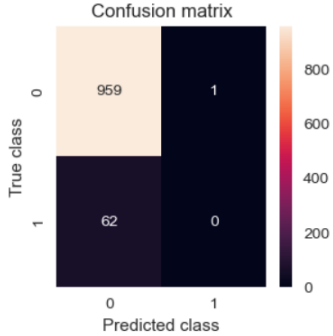
Regresión logística

	precision	recall	f1-score	support
0	0.94	1.00	0.97	960
1	0.00	0.00	0.00	62
accuracy			0.94	1022
macro avg	0.47	0.50	0.48	1022
weighted avg	0.88	0.94	0.91	1022



Random Forest

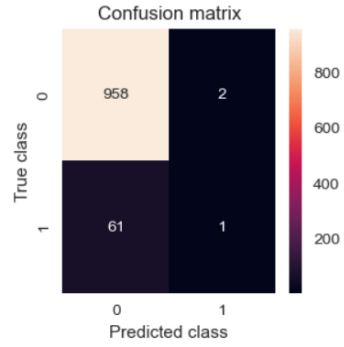
	precision	recall	f1-score	support
0	0.94	1.00	0.97	960
1	0.00	0.00	0.00	62
accuracy			0.94	1022
macro avg	0.47	0.50	0.48	1022
weighted avg	0.88	0.94	0.91	1022



Modelos

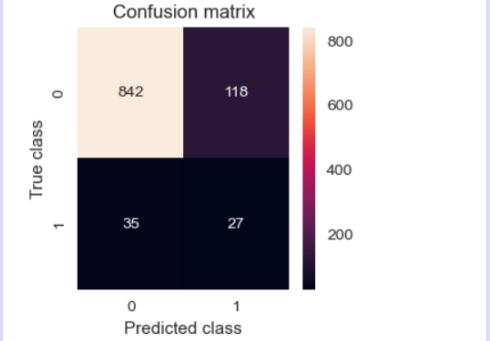
CatboostClassifier

	precision	recall	f1-score	support
0	0.94	1.00	0.97	960
1	0.33	0.02	0.03	62
accuracy			0.94	1022
macro avg	0.64	0.51	0.50	1022
weighted avg	0.90	0.94	0.91	1022



XGBoostClassifier

	precision	recall	f1-score	support
0	0.96	0.88	0.92	960
1	0.19	0.44	0.26	62
accuracy			0.85	1022
macro avg	0.57	0.66	0.59	1022
weighted avg	0.91	0.85	0.88	1022



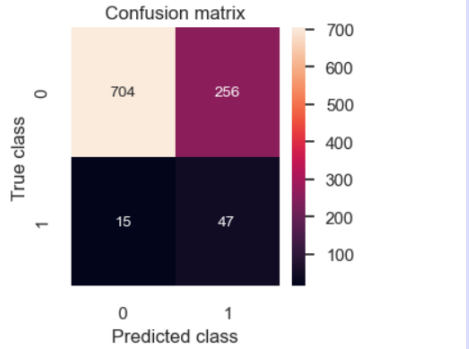
Modelos

Técnicas de balanceo de clases

01 Utilizar parámetros, como weight='balanced'

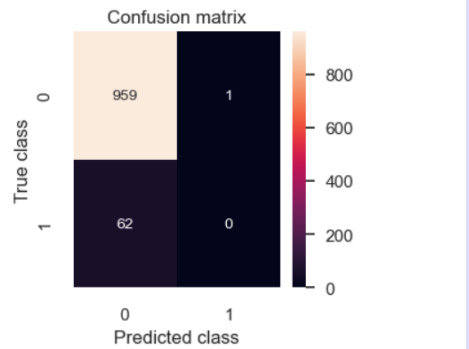
Regresión logística

	precision	recall	f1-score	support
0	0.98	0.73	0.84	960
1	0.16	0.76	0.26	62
accuracy			0.73	1022
macro avg	0.57	0.75	0.55	1022
weighted avg	0.93	0.73	0.80	1022



Random Forest

	precision	recall	f1-score	support
0	0.94	1.00	0.97	960
1	0.00	0.00	0.00	62
accuracy			0.94	1022
macro avg	0.47	0.50	0.48	1022
weighted avg	0.88	0.94	0.91	1022



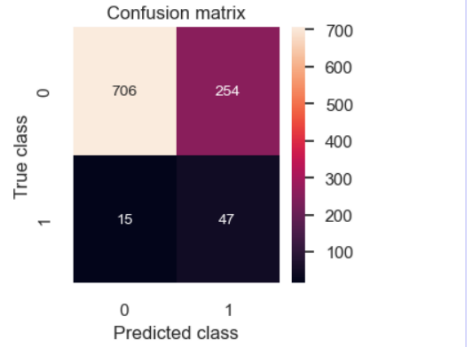
Modelos

Técnicas de balanceo de clases

02 Oversampling

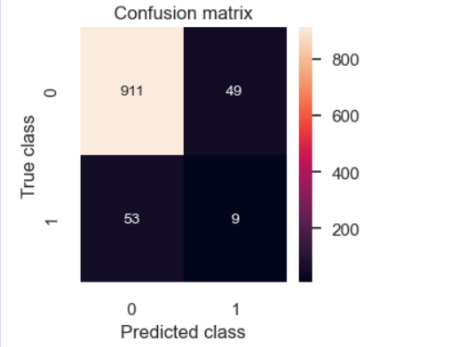
Regresión logística

	precision	recall	f1-score	support
0	0.98	0.74	0.84	960
1	0.16	0.76	0.26	62
accuracy			0.74	1022
macro avg	0.57	0.75	0.55	1022
weighted avg	0.93	0.74	0.80	1022



Random Forest

	precision	recall	f1-score	support
0	0.95	0.95	0.95	960
1	0.16	0.15	0.15	62
accuracy			0.90	1022
macro avg	0.55	0.55	0.55	1022
weighted avg	0.90	0.90	0.90	1022



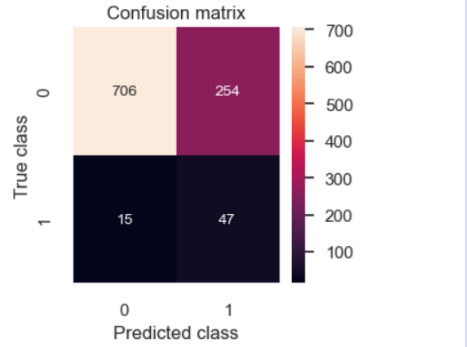
Modelos

Técnicas de balanceo de clases

03 Undersampling

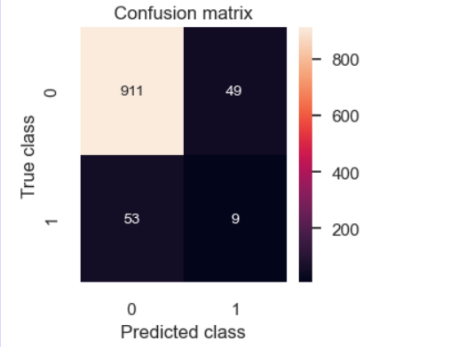
Regresión logística

	precision	recall	f1-score	support
0	0.98	0.74	0.84	960
1	0.16	0.76	0.26	62
accuracy			0.74	1022
macro avg	0.57	0.75	0.55	1022
weighted avg	0.93	0.74	0.80	1022



Random Forest

	precision	recall	f1-score	support
0	0.95	0.95	0.95	960
1	0.16	0.15	0.15	62
accuracy			0.90	1022
macro avg	0.55	0.55	0.55	1022
weighted avg	0.90	0.90	0.90	1022



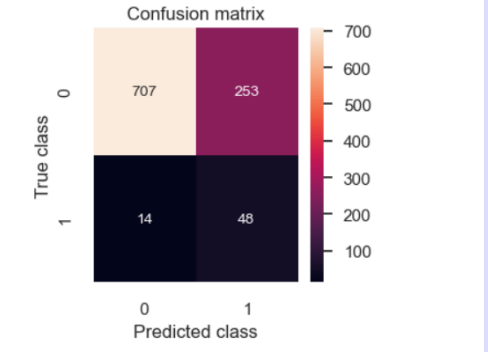
Modelos

Técnicas de balanceo de clases

04 Combinar Undersampling y OverSampling (SMOTETomek)

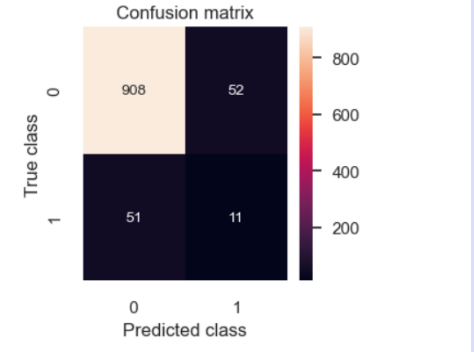
Regresión logística

	precision	recall	f1-score	support
0	0.98	0.74	0.84	960
1	0.16	0.77	0.26	62
accuracy			0.74	1022
macro avg	0.57	0.76	0.55	1022
weighted avg	0.93	0.74	0.81	1022



Random Forest

	precision	recall	f1-score	support
0	0.95	0.95	0.95	960
1	0.17	0.18	0.18	62
accuracy			0.90	1022
macro avg	0.56	0.56	0.56	1022
weighted avg	0.90	0.90	0.90	1022



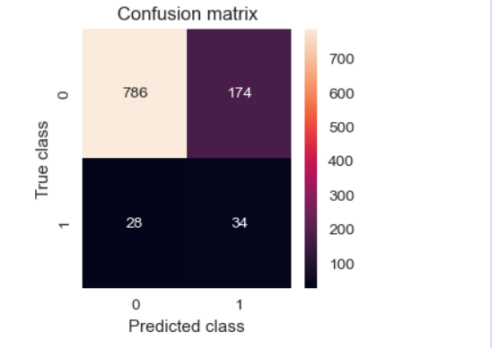
Modelos

Técnicas de balanceo de clases

05 Ensamble de modelos con balanceo

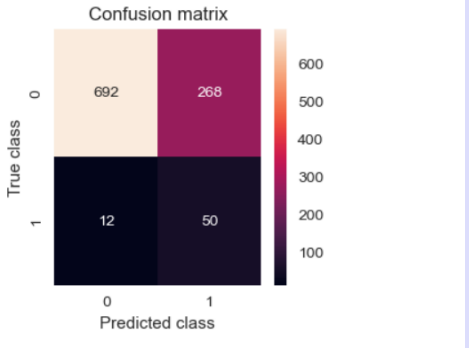
BalancedBaggingClassifier

	precision	recall	f1-score	support
0	0.97	0.82	0.89	960
1	0.16	0.55	0.25	62
accuracy			0.80	1022
macro avg	0.56	0.68	0.57	1022
weighted avg	0.92	0.80	0.85	1022



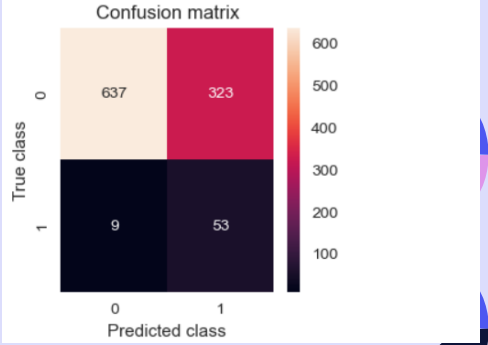
BalancedRandomForestClassifier

	precision	recall	f1-score	support
0	0.98	0.72	0.83	960
1	0.16	0.81	0.26	62
accuracy			0.73	1022
macro avg	0.57	0.76	0.55	1022
weighted avg	0.93	0.73	0.80	1022



EasyEsembleClassifier

	precision	recall	f1-score	support
0	0.99	0.66	0.79	960
1	0.14	0.85	0.24	62
accuracy			0.68	1022
macro avg	0.56	0.76	0.52	1022
weighted avg	0.93	0.68	0.76	1022



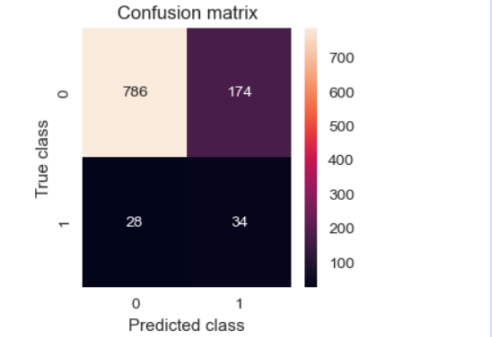
Modelos

Técnicas de balanceo de clases

05 Ensamble de modelos con balanceo

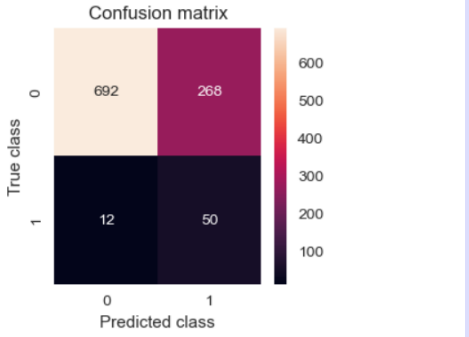
BalancedBaggingClassifier

	precision	recall	f1-score	support
0	0.97	0.82	0.89	960
1	0.16	0.55	0.25	62
accuracy			0.80	1022
macro avg	0.56	0.68	0.57	1022
weighted avg	0.92	0.80	0.85	1022



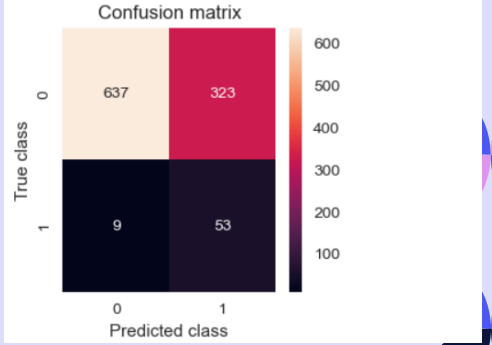
BalancedRandomForestClassifier

	precision	recall	f1-score	support
0	0.98	0.72	0.83	960
1	0.16	0.81	0.26	62
accuracy			0.73	1022
macro avg	0.57	0.76	0.55	1022
weighted avg	0.93	0.73	0.80	1022



EasyEnsembleClassifier

	precision	recall	f1-score	support
0	0.99	0.66	0.79	960
1	0.14	0.85	0.24	62
accuracy			0.68	1022
macro avg	0.56	0.76	0.52	1022
weighted avg	0.93	0.68	0.76	1022



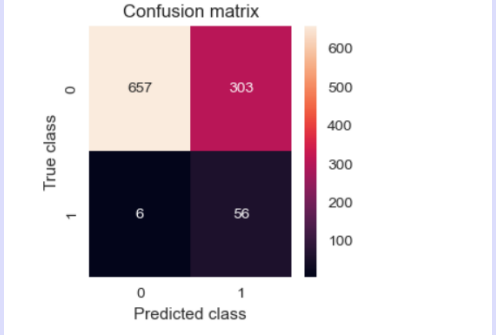
Mejor modelo y parámetros

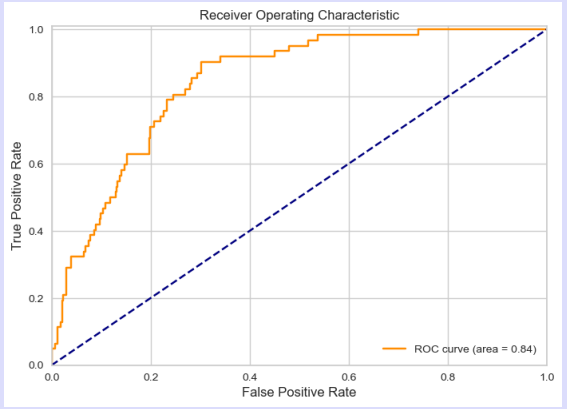
EasyEsembleClassifier

Con GridSearch y validación cruzada y los parámetros:

- base_estimator=CatboostClassifier (verbose=False, n_estimators=30)
- sampling_strategy=not minority
- replacement= True

	precision	recall	f1-score	support
0	0.99	0.68	0.81	960
1	0.16	0.90	0.27	62
accuracy			0.70	1022
macro avg	0.57	0.79	0.54	1022
weighted avg	0.94	0.70	0.78	1022



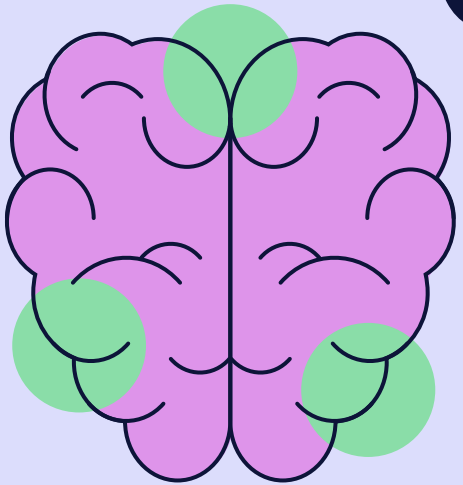


Mejoras del modelo

Qué hacer

- Actualizar periodicamente los datos de los pacientes que hay en la base de datos
- Ampliar la información de la base de datos con otras bases de datos de otros hospitales
- Ampliar la información recogida con nuevas variables: como si hacen ejercicio físico, medicación, problemas de coagulación, etc....

Gracias



María Mora López