

Hoja informaiva: Mejora del modelo

Práctica

```
# División de datos en conjuntos de entrenamiento y validación
# test_size - la fracción del número total de objetos que tiene que ser dividido como el conjunto de validación (un número de 0 a 1)
from sklearn.model_selection import train_test_split

df_train, df_valid = train_test_split(df, test_size=0.25, random_state=54321)
```

```
# Inicialización del modelo de árbol de decisión
# n_estimadores - número de árboles
from sklearn.ensemble import RandomForestClassifier

model = RandomForestClassifier(random_state=54321, n_estimators=3)
```

```
# Inicialización del modelo de regresión logística
from sklearn.linear_model import LogisticRegression

model = LogisticRegression(random_state=54321)
```

Teoría

Un **conjunto de validación** es un conjunto de datos que forma parte de un conjunto de datos fuente. Se utiliza para comprobar la calidad de un algoritmo durante el entrenamiento de un modelo.

Los **parámetros** son los ajustes del modelo obtenidos a partir de los datos de entrenamiento que determinan el trabajo del modelo.

Los **hiperparámetros** son los ajustes de los algoritmos de aprendizaje.