

Trabajo Práctico 1: Reservas de Hotel

Checkpoint 3: Ensambles

75.06 / 95.58 Organización de Datos - FIUBA

Ing. Rodríguez - 1°C 2023



Grupo 31 - "Datazo":

- 106203 - Kisinovsky, Diego Andrés
- 102685 - Mena Giraldo, Michael Gustavo
- 104256 - Brocca, Pablo Martín

Introducción

En este checkpoint armamos y entrenamos distintos modelos predictivos para usar como base de modelos de ensamble de tipo Stacking y Voting, con el objetivo de realizar predicciones más precisas sobre nuestro dataset de test.

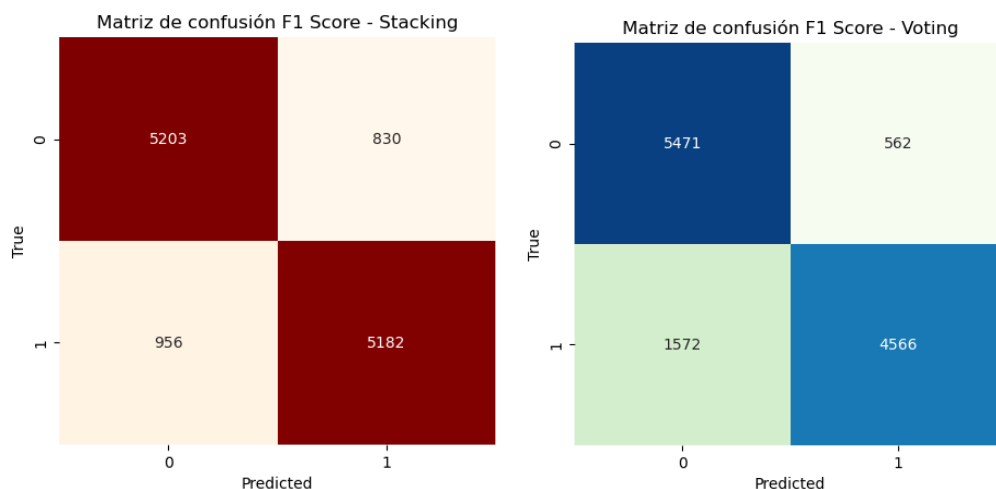
Observaciones del análisis

Los modelos entrenados son K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest (RF) y XGBoost (XGB). Con la excepción de SVM, para el cual elegimos SVM normalizado mediante Standard Scaling porque produjo los mejores scores, para el resto de los modelos optimizamos los hiperparámetros utilizando K-Fold Cross-Validation de 10 folds con Randomized Search. Luego de entrenar los modelos de ensamble, obtuvimos los siguientes scores de accuracy en comparación con los modelos base:



Podemos observar que Stacking obtuvo la mejor accuracy entre todos los modelos.

Graficamos también las matrices de confusión para los dos modelos de ensamble:



Vemos que, si bien Voting tuvo mejores cantidades de TN y FP (mejor precisión), Stacking en general presenta un mejor F1 Score, que tiene en cuenta tanto precisión como recall.