

## Checkpoint 3 - Grupo 23

### Introducción

Las técnicas más utilizadas durante esta etapa fueron la de One Hot Encoding para que nuestros modelos puedan utilizar la información de variables categóricas. Otra técnica utilizada fue el escalador de datos para que nuestro modelo SVM pueda trabajar correctamente y de mejores resultados.

Cabe recalcar que los modelos fueron entrenados con los mismos datos y técnicas para que luego puedan ser utilizados, los que mejor resultado dan, en los ensambles híbridos.

En el dataset no realizamos nuevas modificaciones, mantuvimos el dataset del capítulo anterior.

### Construcción del modelo

Detallar como mínimo los siguientes puntos:

- Hiperparámetros optimizados para KNN: algoritmo, peso, distancia y cantidad de vecinos
- Hiperparámetros optimizados para SVN: kernel, gradoy gamma
- Hiperparámetros optimizados para RF: max depth, min simples leaf y cantidad de estimadores
- Mencionar hiperparámetros optimizados para XGBoost
- Modelos utilizados para el ensamble tipo voting: random forest y xg boost
- Modelos utilizados para el ensamble tipo stacking: random forest y xg boost  
Meta modelo ensamble stacking: regresión logística

### Cuadro de Resultados

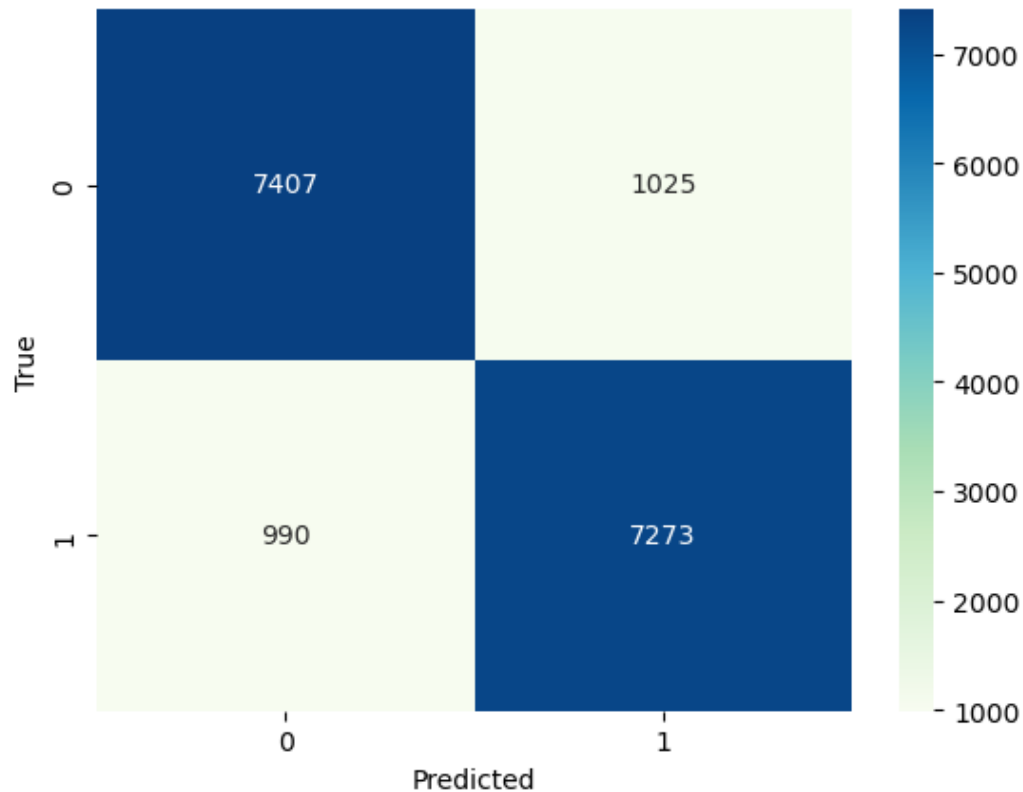
Medidas de rendimiento en el conjunto de TEST:

- F1
- Precision
- Recall
- Métrica XXX (Cualquier otra que consideren relevante)
- Resultado obtenido en Kaggle.

La métrica X a utilizar va a ser el f1 score sobre los datos de entrenamiento para determinar y evaluar si hubo overfitting o underfitting.

Modelo	F1-Test	Precision Test	Recall Test	Metrica X	Kaggle
KNN	0,7787	0,7455	0,8152	0,9959	
SVM	0,8550	0,8742	0,8367	0,8770	
Random Forest	0,8800	0,8651	0,8953	0,9551	0,8692
XGBoost	0,8792	0,8663	0,8925	0,9550	
Voting	0,8669	0,8820	0,8581		
Stacking	0,8778	0,8772	0,8785	0,9615	0,8674

### Matriz de Confusion



Este es el heatmap de las predicciones de nuestro modelo RandomForest optimizado con sus hiperparametros. Como se puede observar, nuestro modelo predice correctamente en la gran mayoría de los casos. Sin embargo, notamos que el modelo tiene ciertos inconvenientes a la hora de predecir que fue cancelado.

## **Tareas Realizadas**

Como somos solo 2 personas, ambos trabajamos todos los modelos en conjunto.