

Organización de datos, Curso Rodriguez 1er Cuatrimestre 2023



Docente: Rodriguez, Juan Manuel.

Estudiantes:

- Mendoza Hernandez, Sabrina Scarlet (108524)
- Valeriani, Matias Gabriel (108570)
- Jang, Lucas (109151)

Checkpoint 3: Ensamble de modelos

En esta tercera parte del trabajo práctico se construyeron una serie de clasificadores como K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), XGBoost (Extreme Gradient Boosting) , y dos ensambles híbridos como Voting y Stacking. Analizando los diferentes score, obtuvimos como mejor modelo (mejores métricas): **Stacking**. Este modelo combina varios modelos de aprendizaje automático de nivel inferior en un modelo de nivel superior que utiliza las predicciones de los modelos de nivel inferior como entradas. El modelo de nivel superior se entrena para predecir el resultado final en función de las salidas de los modelos de nivel inferior. Luego de realizar un estudio exhaustivo, obtuvimos como modelos base: Random Forest, XGBoost y Regresión logística; como modelo meta: Regresión logística, y como resultados de la predicción: una precisión de 0.87 para los que no fueron cancelados y 0.89 para los cancelados, un recall de 0,88 para los no cancelados y 0.87 para los cancelados, por un F1-score de 0,88 en nuestro dataframe, y pudiendo obtener así 0,8737 en la competencia de Kaggle.

- Matriz de confusión y métricas obtenidas con el mejor predictor Stacking:

