

Diplomado en ML Cloud - UCB CBBA

Módulo 4: Machine Learning Cloud MLOps

Docente: Ing. Mauricio Alejandro Quezada

Estudiante: Jose Carlos Iriarte

Fecha : Agosto del 2025

Laboratorio 4

Captura del DAG completo del pipeline

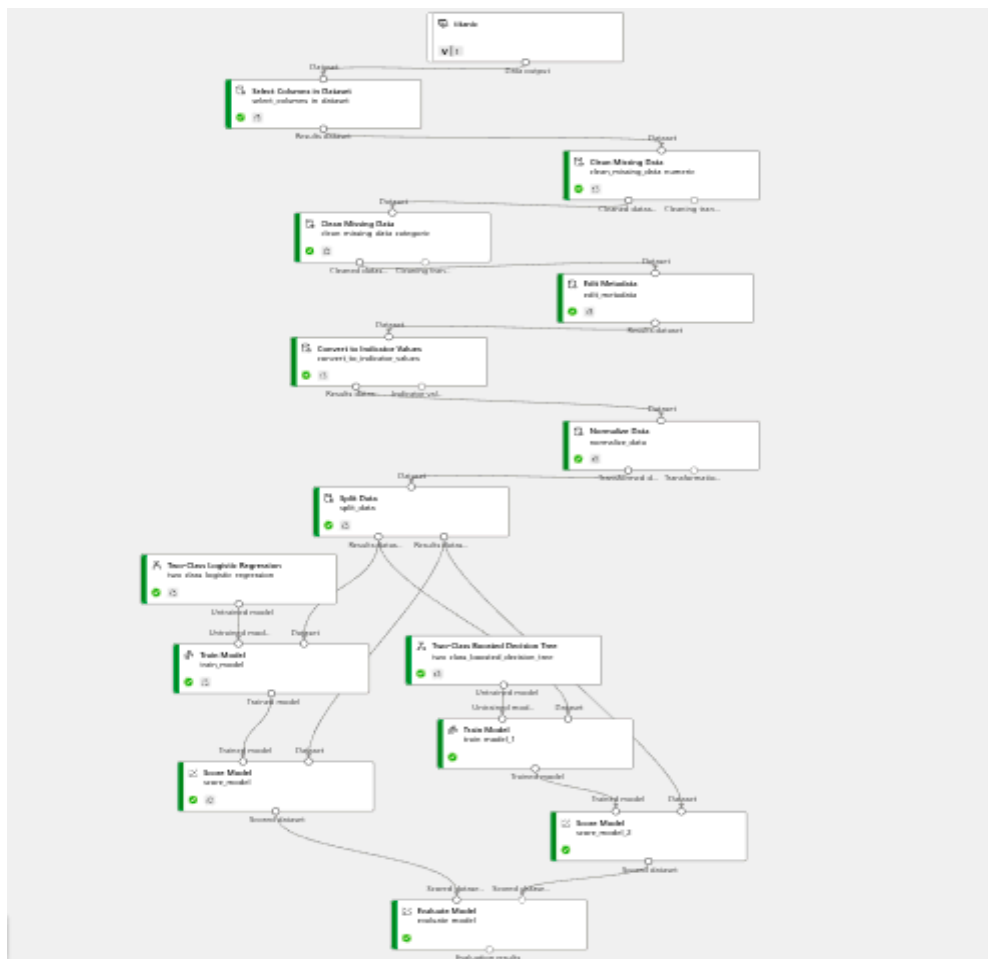
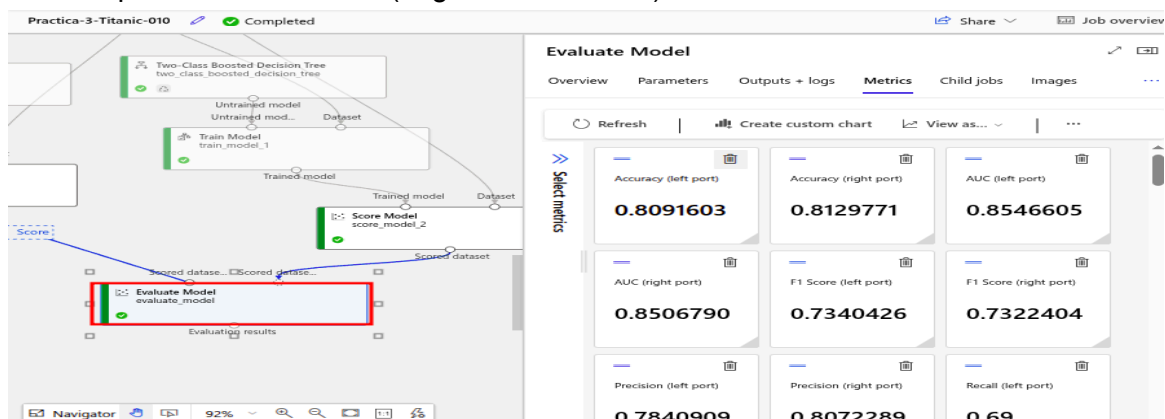
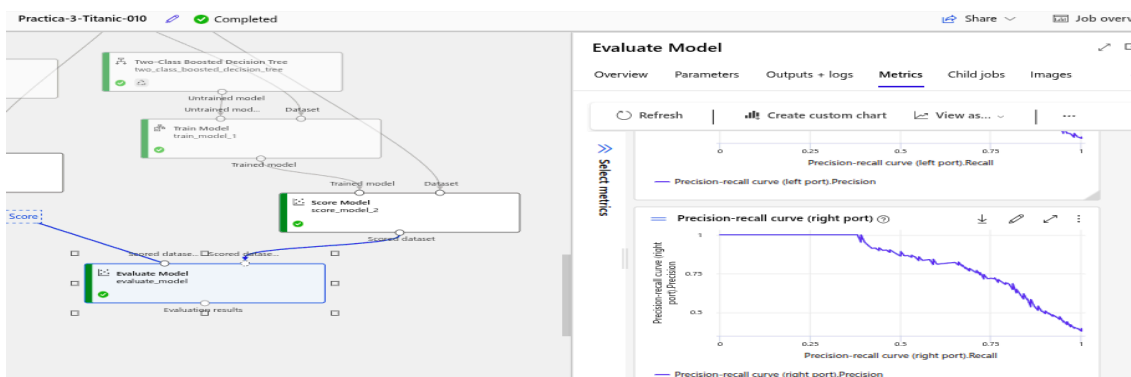
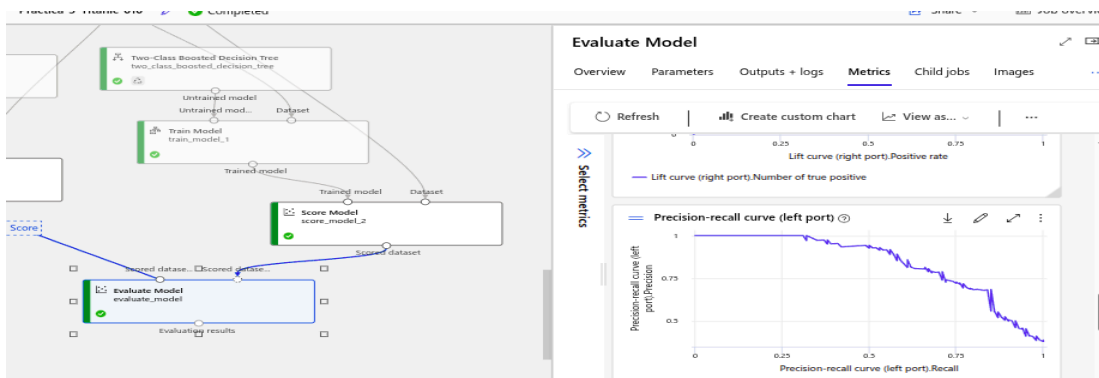
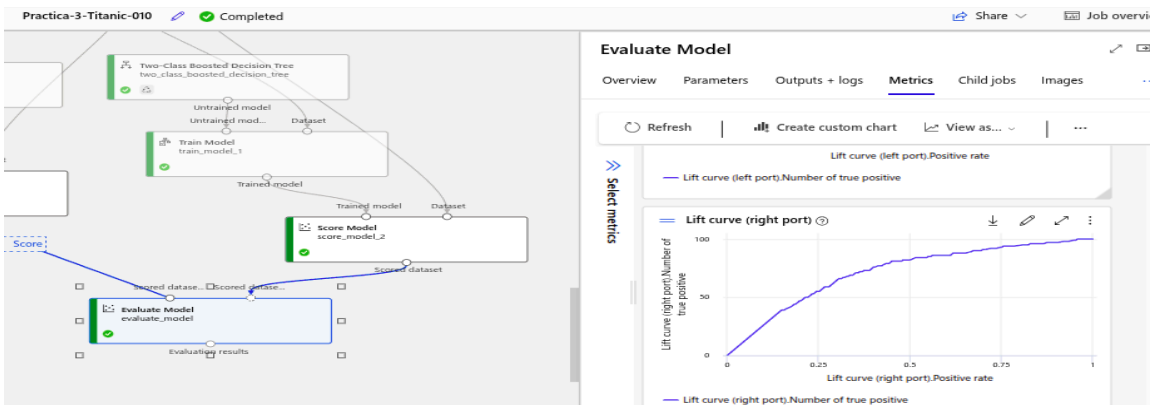
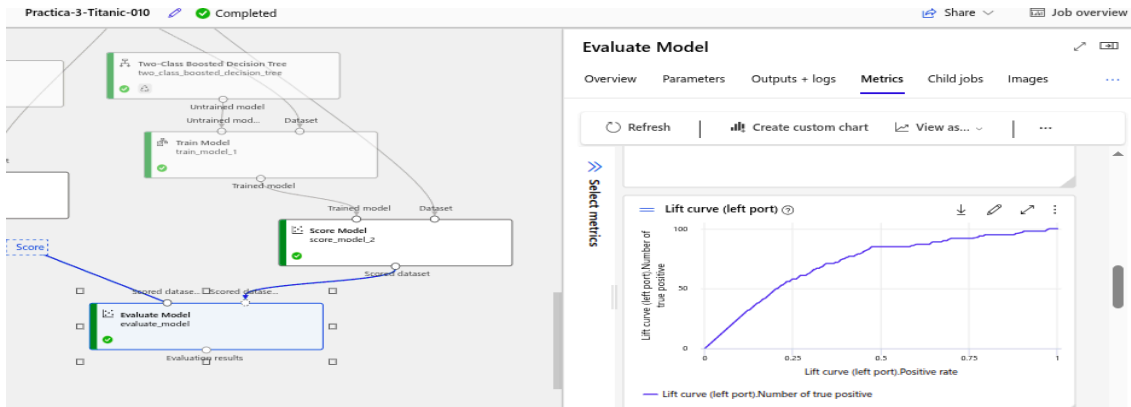


Tabla comparativa de métricas (Logistic vs Boosted)





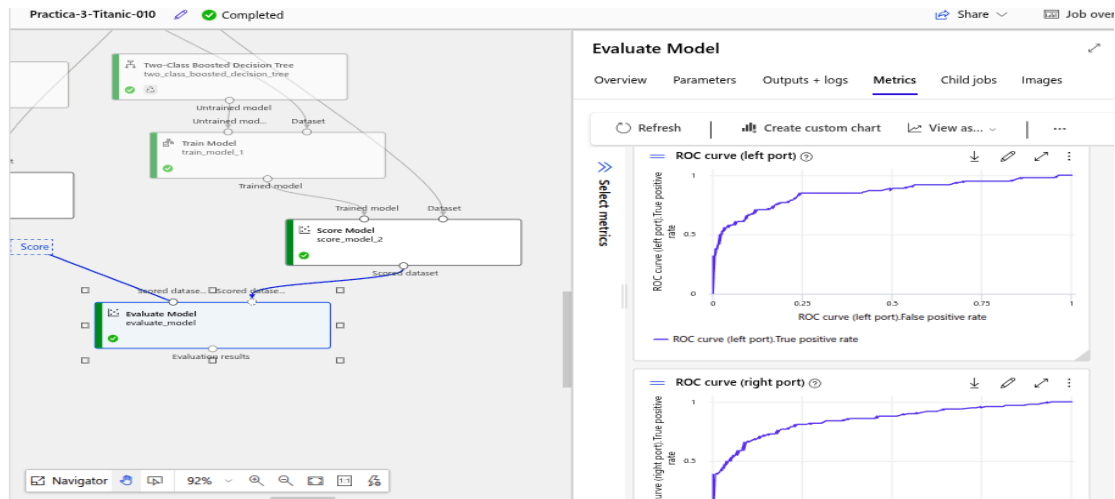


Tabla comparativa

Métrica	Modelo 1 (Logistic Regression)	Modelo 2 (Boosted Decision Tree)
Accuracy	0.8091603	0.8129771
AUC	0.8546605	0.850679
F1 Score	0.7340426	0.7322404
Precisión	0.7840909	0.8072289
Recall	0.69	0.67

¿Cuál fue el mejor modelo?

- No hay un ganador absoluto, porque ambos tienen rendimientos muy similares.
- **Logistic Regression** tuvo mejor **AUC (0.8547)** y **Recall (0.69)**, lo que indica que identifica mejor a los casos positivos (menos falsos negativos).
- **Boosted Decision Tree** tuvo mejor **Accuracy (0.813)** y **Precisión (0.807)**, lo que significa que clasifica mejor los casos negativos y reduce falsos positivos.

¿Por qué?

- **Logistic Regression** es lineal, por lo tanto más simple y menos propenso a sobreajuste. Su mejor Recall lo hace útil si lo importante es **no perder casos positivos**.
- **Boosted Decision Tree** aprovecha la combinación de árboles, lo que le da más poder predictivo en términos de **exactitud global** y **precisión**, aunque sacrificó algo de Recall. Es mejor si lo importante es que **las predicciones positivas sean más confiables**.

¿Cómo influyó el ajuste de hiperparámetros?

- En **Logistic Regression**, los hiperparámetros (regularización, tasa de aprendizaje en optimización, etc.) ayudan a controlar el balance entre sobreajuste y generalización. Un buen ajuste mejoró Recall y AUC.

- En **Boosted Decision Tree**, parámetros como **número** de árboles, profundidad máxima, tasa de aprendizaje influyen mucho. Un ajuste adecuado permitió mayor precisión y exactitud, aunque el modelo se volvió un poco menos sensible (menor Recall).

Conclusiones:

- Si tu objetivo es detectar la mayor cantidad de positivos (maximizar Recall y AUC), conviene Logistic Regression.
- Si tu objetivo es predecir con más exactitud y evitar falsos positivos, conviene el Boosted Decision Tree.