

Diplomado en ML Cloud - UCB CBBA

Módulo 4: Machine Learning Cloud MLOps

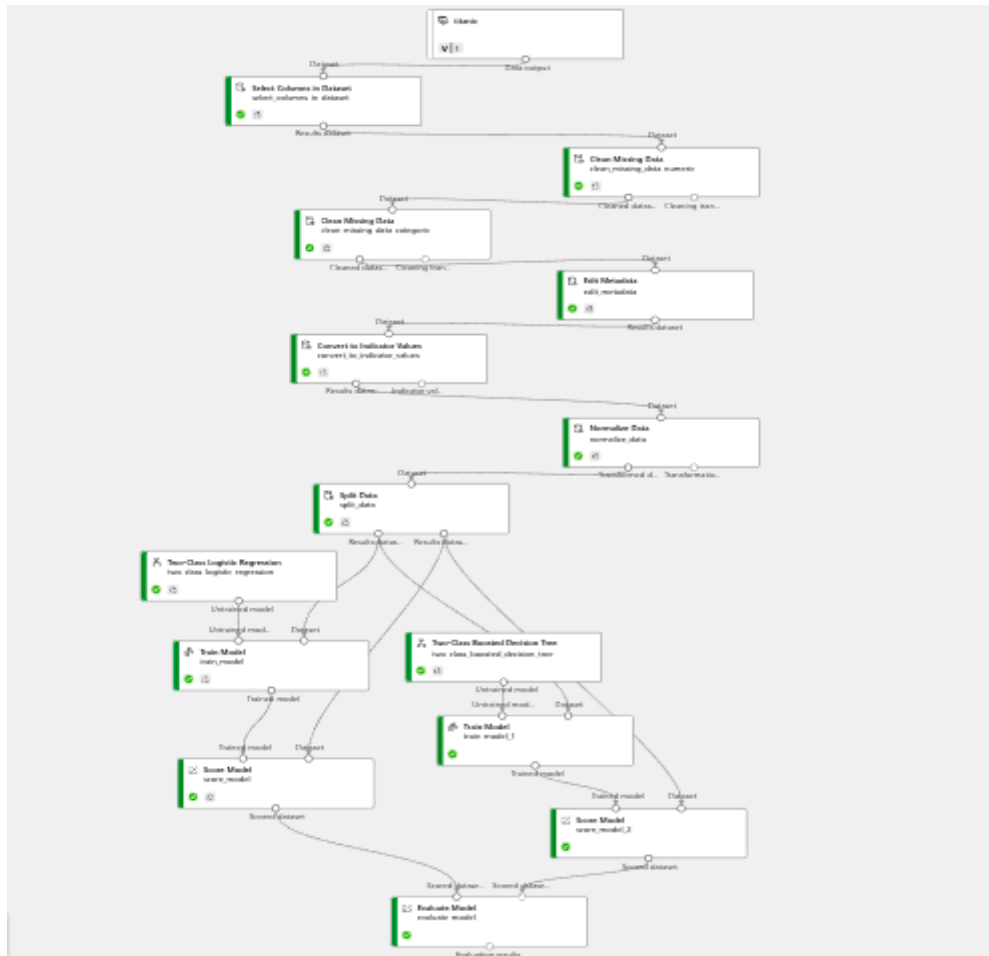
Docente: Ing. Mauricio Alejandro Quezada

Estudiante: Jose Carlos Iriarte

Fecha : Agosto del 2025

Laboratorio 3

Captura del DAG completo del pipeline



Captura de configuración de Clean Missing Data

Practica-3-Titanic-010 Completed Share Job overview

Clean Missing Data

Overview **Parameters** Outputs + logs Metrics Child jobs Images

Refresh Register model Debug and monitor

Columns to be cleaned Edit column

Column types: Numeric
Exclude column names: survived

Minimum missing value ratio 0.0

Maximum missing value ratio 1.0

Cleaning mode Replace with median

Generate missing value indicator column False

Cols with all missing values Remove

Practica-3-Titanic-010 Completed Share Job overview

Clean Missing Data

Overview Parameters Outputs + logs Metrics Child jobs Images

Refresh Register model Debug and monitor

Columns to be cleaned Edit column

Column names: embarked

Minimum missing value ratio Edit

0.0

Maximum missing value ratio Edit

1.0

Cleaning mode Edit

Replace with mode

Generate missing value indicator column Edit

False

Cols with all missing values Edit

Remove

Tabla comparativa de métricas (Logistic vs Boosted)

Practica-3-Titanic-010 Completed Share Job overview

Evaluate Model

Overview Parameters Outputs + logs Metrics Child jobs Images

Refresh Create custom chart View as...

Select metrics

Accuracy (left port)	Accuracy (right port)	AUC (left port)
0.8091603	0.8129771	0.8546605
AUC (right port)	F1 Score (left port)	F1 Score (right port)
0.8506790	0.7340426	0.7322404
Precision (left port)	Precision (right port)	Recall (left port)
0.7840909	0.8072289	0.69

Practica-3-Titanic-010 Completed Share Job overview

Evaluate Model

Overview Parameters Outputs + logs Metrics Child jobs Images

Refresh Create custom chart View as...

Select metrics

Lift curve (left port) Edit

Lift curve (left port) Number of true positive

Lift curve (left port) Positive rate

Practica-3-Titanic-010 Completed Share Job overview

Evaluate Model

Overview Parameters Outputs + logs Metrics Child jobs Images

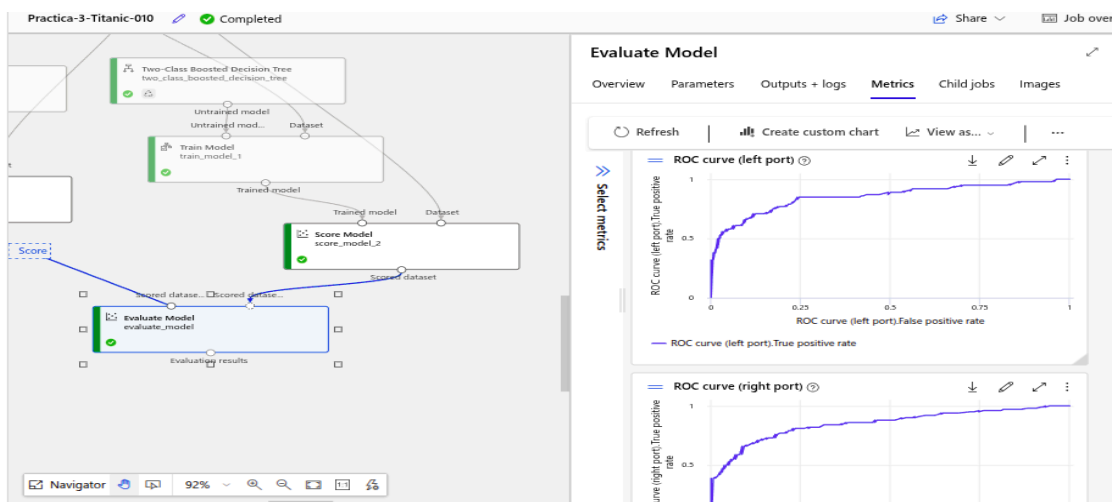
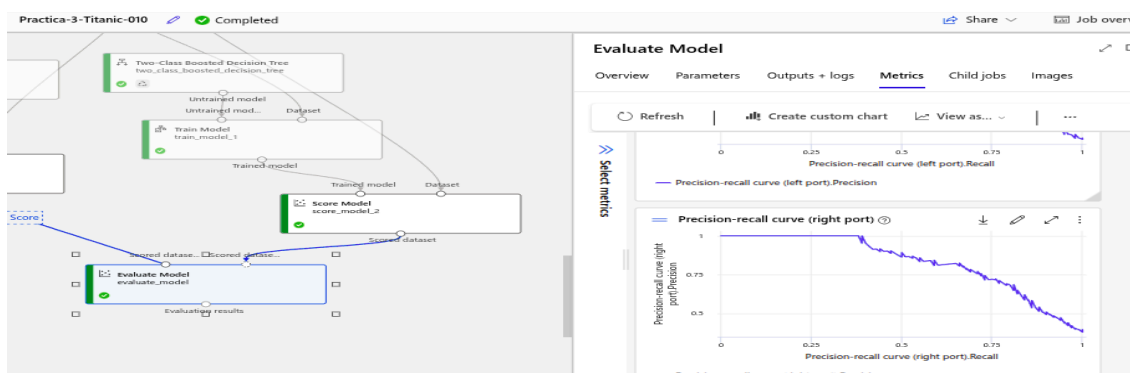
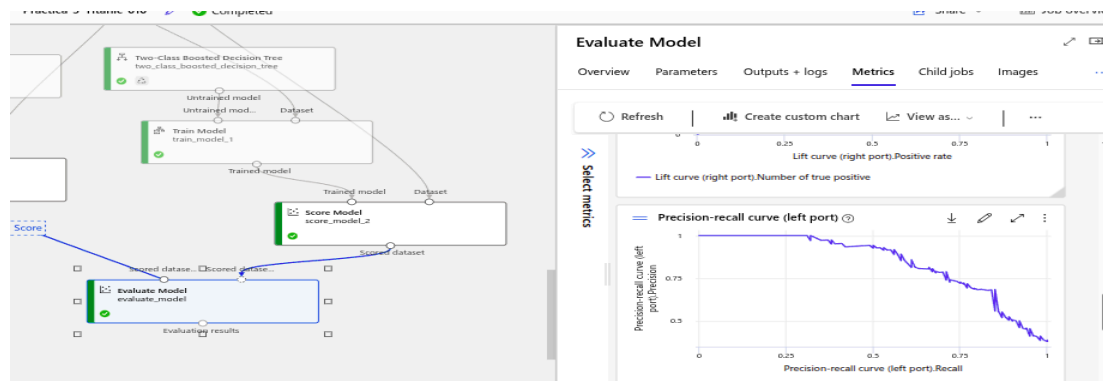
Refresh Create custom chart View as...

Select metrics

Lift curve (right port) Edit

Lift curve (right port) Number of true positive

Lift curve (right port) Positive rate



Conclusiones breves

- Ambos modelos (Logistic Regression y Boosted Decision Tree) alcanzaron un rendimiento aceptable, con accuracies alrededor de 0.81.
- El Boosted Decision Tree mostró mejor desempeño general, destacando en AUC (0.85) y precisión (≈ 0.80), lo que indica mayor capacidad discriminativa.
- La Logistic Regression tuvo métricas más equilibradas, pero quedó ligeramente por debajo en F1 y Recall.
- En problemas como Titanic, donde el costo de falsos negativos es relevante (no predecir sobrevivientes), un modelo con mayor Recall podría ser más útil, aunque implique sacrificar algo de precisión.