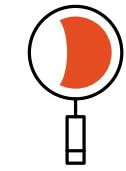


Caso Práctico

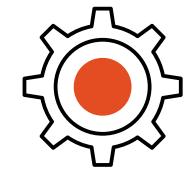
por David Clemente

Objetivo



- Desarrollar un pipeline reproducible para predecir **PAY_AMT4** mediante **Kedro**.
- Construir un modelo de clasificación para **default.payment.next.month**.
- **Garantizar buenas prácticas:** modularidad, no data leakage, validación robusta y reportes automáticos.

Metodología



- EDA y limpieza del dataset
- Feature engineering enfocado en información disponible (evitando leakage)
- Entrenamiento y validación con múltiples modelos
- Uso de pipelines Kedro:
 - Preprocessing → Modeling → Evaluating → Reporting



Resultados



PAY_AMT4 (Regresión)

Mejor modelo: Gradient Boosting / XGBoost

Métricas utilizadas: MSE, RMSE, MAE, R², Explained Variance

Resultado general:

- R² ≈ 0.80–0.83 en test
- Error moderado y sin overfitting significativo

Test de Diferencias Críticas (Nemenyi):

- XGBoost, Extra Trees y Random Forest fueron consistentemente top performers
- Modelos lineales (LR, Ridge, Lasso) quedaron en últimos lugares

Default Payment (Clasificación)

Mejor modelo: Gradient Boosting Classifier

Métricas utilizadas: AUC, Accuracy, Precision, Recall, F1

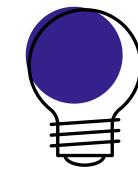
Resultados clave:

- AUC ≈ 0.78
- Accuracy ≈ 0.82
- Balance aceptable entre precision y recall
- Matriz de confusión y curva ROC generadas

Nemenyi:

- **Modelos con mejor ranking:** Decision Tree, KNN, Extra Trees
- GBC y XGBoost fueron estables y competitivos, pero no siempre top rankings

Conclusiones



- El pipeline desarrollado permiten un flujo **reproducible y escalable**.
- Para **PAY_AMT4**, los modelos de ensamble (**XGBoost, Extra Trees, Random Forest**) fueron los más robustos según R^2 y pruebas estadísticas.
- Para **default.payment.next.month**, el **Gradient Boosting Classifier** logró el mejor equilibrio entre AUC, precisión y recall.
- El análisis de **diferencias críticas (Nemenyi)** confirmó diferencias significativas entre familias de modelos.
- El proyecto demuestra un enfoque E2E, desde EDA hasta reportes listos para negocio.