

Modelo de attrition para la campaña de retención proactiva de clientes

Objetivo

- Bajas del *Paquete Premium*, 154 k clientes
- Predicción dos meses previos (BAJA+2)

Resultados

- Ganancia = 125.4 M AR\$
- Predicción para 202109
- # Envíos de ofertas de retención = 9.5k

Insumos al proyecto

Datos crudos

- 152 atributos de 154k clientes
- fotos 201901 a 202109

Algoritmos disponibles

- LightGBM (lightgbm_3.3.5) sobre R (R6.2.5.1)
- Optimización Bayesiana de hiperparámetros
→ reutilización código z823
- Entrenamiento y ensayo contra foto
→ código z824

Hiperparámetros LightGBM

Hiperparámetros ensayados

- Learning rate
- Feature fraction
- Number of leaves
- Minimum data in leaf

Primer ensayo de optimización Bayesiana

- Entrenamiento sobre subset: 202010 a 202103
- Ensayo: 202105 / validación: 202104
- Ganancia: 75.8 M AR\$

Primeros ensayos de Feature Engineering

Nuevos features

- Indicadores de saldo negativo $\sum m^* < 0$
- Proxy actividad = $\sum |m^*|$
- ¿Más ATMs competencia? $\text{atm_oth} > \text{atm_oth}$
- Tarjetas: coalesce montos (m) y transacciones
 $\text{tarjeta}(m/c) = \text{visa}(m/c) + \text{mc}(m/c)$

zweite/sql_eng.ipynb

Evaluación features

- Ganancia = 81.8 M AR\$

zweite/823_lightgbm_binaria_BO_20231011.r

Insumo ensayos de Feature Engineering Histórico (FEH)

Baseline: 152 \Rightarrow 456 atributos

- lag 1, 3 y 6 meses sobre todos montos (m*) [flotantes]
transacciones (c*) [enteros]
- Extensión JupySQL operando sobre base SQL DuckDB (duckdb_0.2.6)

experimentos/sql_eng_baseline.ipynb

FEH1: Agregados al baseline \Rightarrow 912 atributos

- Atributos con datos reiterados \Rightarrow drop
- Lags 2, 4 y 5 meses \rightarrow completa semestre
- Todos: min, máx, media movil semestre

experimentos/sql_eng_all6.ipynb

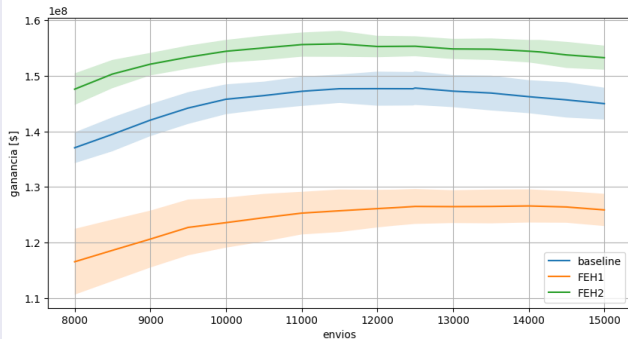
FEH 2: agregados al FEH 1 \Rightarrow 2128 atributos

- normalización semestral: $\text{atr} / \text{media semestral}$
- Δ normalizado $(\text{atr} - \text{lag\#}) / \text{media semestral}$

experimentos/sql_eng_all6.ipynb

Rendimientos FEH 1 y FEH 2

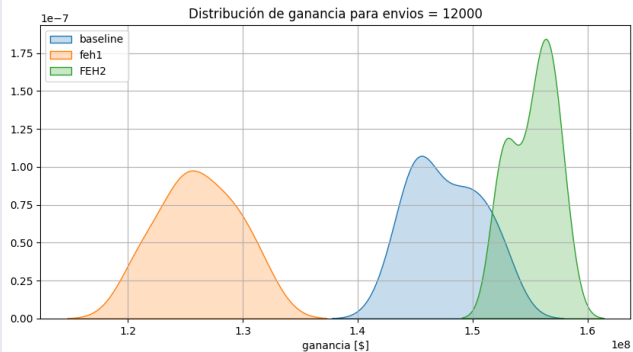
Ganancia vs # Envíos



Optimización Bayesiana: experimentos/823_(baseline/all6_a/all6).r
Ganancia vs semilla: experimentos/824_(baseline/feh1/all6_zweite).r
Gráficación: experimentos/graph.ipynb

Distribución de ganancias

Envíos con máxima ganancia en FEH2



Optimización Bayesiana: experimentos/823_(baseline/all6_a/all6).r
Ganancia vs semilla: experimentos/824_(baseline/feh1/all6_zweite).r
Gráficación: experimentos/graph.ipynb

Catástrofes

Análisis

- montos (m^*) == 0.0 en foto
- transacciones (c^*) == 0 ídem.
- >10 k registros \Rightarrow NULL todos en foto

dritte/sql_eng_all6cat.ipynb

Evaluación

- Foto objetivo predicción \rightarrow 202109
- Entrena: 2021901 a 202106 exceptúa confinamiento ASPO 202003-11

dritte/823_1126keinePanik.r

- Ganancia = 122.7 M AR\$

Promedio de predicciones de múltiples corridas es foto objetivo

Múltiples corridas LightGBM

- Foto objetivo predicción → 202109
- Distintas probabilidades por cliente: 18 semillas

`dritte/824_1129null.r`

- Promedio de predicciones
- Asignación de envíos

`dritte/1129_null_sem.r`

- Ganancia = 125.4 M AR\$