

Informe — Predicción de Recompra tras Promoción

Proyecto: Modelizado de Minería de Datos

Archivo: `Mini_Proyecto_Clientes_Promociones_from_Python.xlsx` (20 filas, 8 columnas)

1. Resumen ejecutivo

Con un conjunto pequeño de 20 clientes, entrenamos un árbol de decisión para predecir si un cliente realizará una **recompra** después de recibir (o no) una promoción.

Resultados clave:

- **Accuracy (prueba):** 0.75 (3/4 predicciones correctas en el set de prueba).
- **Importancia de variables:** `Total_Compras` explica el 100% de la decisión del modelo (importancia = 1.0); las demás variables (Género, Edad, Recibió_Promo, Monto_Promo, Ingreso_mensual) tienen importancia 0.
- **Profundidad del árbol:** 1 (modelo muy simple: una sola regla).
- **Score de entrenamiento:** 1.00 (perfecto en entrenamiento — señal de posible sobreajuste debido a muestra muy pequeña).
- **Matrices y métricas:** reporte de clasificación con precisión/recall F1 razonables pero calculadas sobre un test muy pequeño (soportes por clase = 2).

Conclusión principal: En este dataset, el historial de compras (`Total_Compras`) es el predictor dominante de recompra. Esto sugiere que **clientes con mayor número de compras previas** tienen mayor probabilidad de volver a comprar tras (o independientemente de) una promoción.

2. Descripción del dataset (salida del script)

- Filas: 20
- Columnas: 8 (`Cliente_ID`, `Genero`, `Edad`, `Recibio_Promo`, `Monto_Promo`, `Recompra`, `Total_Compras`, `Ingreso_mensual`)
- No hay valores nulos.
- Estadísticos (resumen):
 - Edad: media 35.8 (sd 7.76), rango 23–50.
 - Monto_Promo: media 321.5, sd 317.6, min 0, max 800 (alta varianza).
 - Total_Compras: media 2.2 (min 1, max 5).
 - Ingreso_mensual: media 37.750 (moneda local), sd ~9.497.

3. Preprocesamiento realizado

- Mapeos categóricos:
 - `Genero`: 'F' → 0, 'M' → 1.
 - `Recibio_Promo`: 'Si' / 'Sí' → 1, 'No' → 0.
 - `Recompra`: 'Si' / 'Sí' → 1, 'No' → 0.
- División de datos: **train/test 80/20**, estratificando por la variable objetivo (`Recompra`).
- Modelo: `DecisionTreeClassifier(random_state=42, max_depth=3)` — aunque el árbol final quedó con profundidad 1.

4. Resultados del modelo

Matriz de confusión (test)

```
[[1 1]
 [0 2]]
```

- Clase 0 (No Recompra): 2 muestras en test; 1 correctamente clasificadas, 1 mal.
- Clase 1 (Recompra): 2 muestras en test; 2 correctamente clasificadas.

Reporte de clasificación (test; soporte=2 por clase)

- Clase 0: precision 1.00, recall 0.50, f1 0.67
- Clase 1: precision 0.67, recall 1.00, f1 0.80
- Accuracy global: **0.75**

Importancia de variables

Variable	Importancia
Total_Compras	1.0
Genero	0.0
Edad	0.0
Recibio_Promo	0.0
Monto_Promo	0.0
Ingreso_mensual	0.0

Métricas adicionales

- Profundidad del árbol: **1**
- Número de hojas: **2**
Score entrenamiento: **1.0000** (100%) — posible sobreajuste por pequeño tamaño de datos.
- Score prueba: **0.7500**

5. Interpretación y hallazgos

1. **Total_Compras domina la predicción.** El árbol con profundidad 1 significa que una *única* condición sobre **Total_Compras** decide la mayoría de las predicciones. En términos prácticos: los clientes con más compras previas tienen mayor probabilidad de recompra.
2. **Las promociones y monto promocional no aparecieron como decisivas** en este modelo — **no implica que las promociones no funcionen**, sino que, en este conjunto reducido de 20 ejemplos, **Total_Compras** captura la señal más fuerte. Posibles razones:
 - Muestra pequeña (20 filas) y división test muy pequeña (4 filas) — inseguridad estadística.
 - Poca variabilidad o correlación entre **Recibio_Promo/Monto_Promo** y **Recompra** en este sample.
3. **Sobreajuste potencial:** score de entrenamiento perfecto y test menor; con tan pocos datos un árbol puede memorizar.
4. **Estadística de soporte:** las métricas por clase calculadas sobre 2 observaciones por clase en test son poco robustas; cambiaría con más datos y validación cruzada.

6. Recomendaciones para marketing (acción)

Con cautela (por tamaño de muestra):

- **Priorizar clientes con historial de compras alto** (mayor **Total_Compras**) para campañas que busquen recompra — parece ser el perfil con más probabilidad de volver.
- **Diseñar A/B tests** de promoción focalizados en diferentes segmentos (novatos vs. frecuentes) para cuantificar efecto real de la promoción y del monto.
- **No descartar las promociones a nuevos clientes**, pero medir su ROI: quizá requieren incentivos distintos (descuento en la 2ª compra, envío gratis, etc.).
- **Recolectar más datos** (más clientes, más campañas, fecha de última compra, canal, tipo de producto) antes de tomar decisiones definitivas.

7. Dashboard sugerido para el equipo de marketing

Elementos visuales e indicadores clave (KPIs):

- KPI: Tasa de recompra global y por segmento (**Total_Compras** buckets: 1, 2–3, 4+).
- Gráficos:
 - Boxplot: **Monto_Promocion** por **Recompra** (ver si existe diferencia de medianas).
 - Scatter/line: **Ingreso_mensual** vs probabilidad de recompra.
 - Bar charts: distribución de **Recompra** por **Genero**, por **Recibio_Promo**.
 - Visualización del árbol de decisión (regla simple encontrada).
- Tabla: perfiles con mayor probabilidad (ej.: **Cliente_ID**, **Total_Compras**, **Recompra_prob_estimada**).

8. Limitaciones del análisis

- **Tamaño de la muestra (n=20)**: resultados poco estables; métricas de test calculadas sobre 4 observaciones.
- **Variables faltantes importantes**: recencia (días desde última compra), canal de promoción, tipo de producto, respuesta temporal, cohortes.
- **Modelo simple** y sensible al muestreo; falta validación cruzada y ajuste de hiperparámetros en conjunto más grande.

9. Próximos pasos recomendados (técnicos)

1. **Aumentar la muestra**: recolectar más registros y más campañas.
2. **Agregar variables** clics, apertura de mail, visita web.

10. Comunicación para no técnicos (ejemplo)

“Al analizar 20 clientes encontramos que el mejor indicador para saber si alguien va a volver a comprar es si ya compró varias veces antes. Esto quiere decir que, por ahora, tiene más sentido invertir promociones en clientes con historial de compras. Sin embargo, nuestra muestra es pequeña y recomendamos hacer pruebas controladas y recopilar más datos antes de cambiar la estrategia general.”

11. Conclusión

El análisis realizado demuestra que el **historial de compras del cliente** (`Total_Compras`) es el factor más determinante para predecir la recompra, superando al género, la edad, el ingreso mensual o incluso el hecho de haber recibido una promoción.

El modelo —un árbol de decisión sencillo— alcanzó una **precisión del 75%** en el conjunto de prueba, mostrando una buena capacidad predictiva pese al **tamaño reducido del dataset (solo 20 registros)**.

Sin embargo, debido a la poca cantidad de datos y a la simplicidad del modelo (profundidad 1), los resultados deben interpretarse con cautela. Se recomienda **ampliar la base de clientes, incluir más variables relevantes y realizar validaciones más robustas** antes de extraer conclusiones definitivas.

En términos de negocio, los resultados sugieren que **los clientes con mayor número de compras previas son los más propensos a recomprar**, por lo que el área de marketing podría enfocar sus promociones en estos perfiles para maximizar el retorno de inversión. A futuro, combinar este enfoque con análisis más amplios permitirá optimizar las estrategias de fidelización y segmentación de clientes.