

Análisis Predictivo de Churn: Un Escudo de Datos (CRISP-DM)

Fase #1: Comprensión del Negocio (*Business Understanding*)

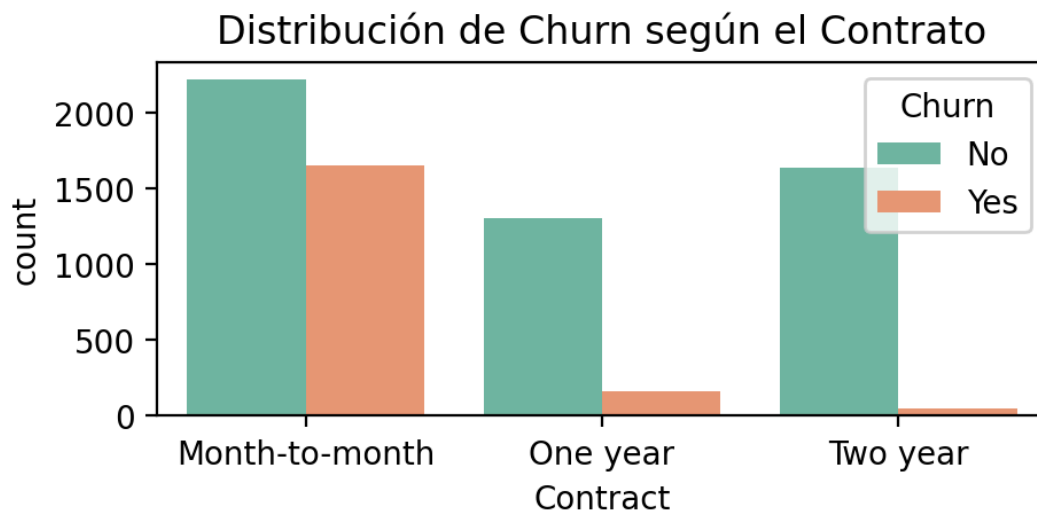
El Problema: Nuestro desafío es detener la **pérdida constante de clientes (Churn)**, que golpea los ingresos recurrentes y nos obliga a gastar de más en adquisición de nuevos usuarios.

- **Objetivo General: Dejar de reaccionar y empezar a predecir.** Queremos saber, con **alta certeza**, quién se irá antes de que lo haga.
 - **Impacto Clave: Optimización y Foco.** Al predecir con un **Recall del 73.3%**, garantizamos que el 73% de las campañas de retención van dirigidas a personas que *realmente* iban a abandonar. Esto minimiza el costo y maximiza el LTV (Lifetime Value).
 - **Stakeholders:** La solución empodera a **Marketing/Comercial** (usan la lista de riesgo para el rescate) y a **Gerencia/Dirección** (monitorean el impacto financiero).
-

Fase #2: Comprensión de los Datos (*Data Understanding*)

La Materia Prima: Datos de Servicio. La adquisición se centró en datos históricos (CRM y facturación). Nuestro principal reto fue identificar los problemas para que el modelo no se "engañara."

- **Factores Clave (EDA):** Identificamos los "**villanos**" del negocio: el **contrato Mes a Mes**, el servicio de **Fibra Óptica** (propenso al abandono), y la **Baja Antigüedad**.



El contrato **Mes a Mes** tiene la tasa de abandono más alta, demostrando el riesgo.

- **Problemas Críticos:** El problema técnico más grande fue el **desbalance de clases** (más clientes se quedan que los que se van). También tuvimos que limpiar datos faltantes en **TotalCharges**.

Fase #3: Preparación de los Datos (*Data Preparation*)

El Refinado de la Materia Prima. El objetivo fue hacer los datos digeribles para el algoritmo y corregir su sesgo natural.

- **Tratamiento de Variables:** Limpiamos y estandarizamos variables numéricas (como el gasto mensual) para ponerlas en la misma escala. También codificamos las variables de texto (One-Hot Encoding) para que el modelo pudiera "entender" los datos.
- **La Solución Avanzada (SMOTE):** Para vencer el sesgo del desbalance, aplicamos **SMOTE**. Esta técnica actuó como un "maquillador de datos", creando **muestras sintéticas** para la clase minoritaria ("Churn: Sí"). Esto forzó a nuestro modelo a aprender los patrones reales.

Fase #4: Modelado (*Modeling*)

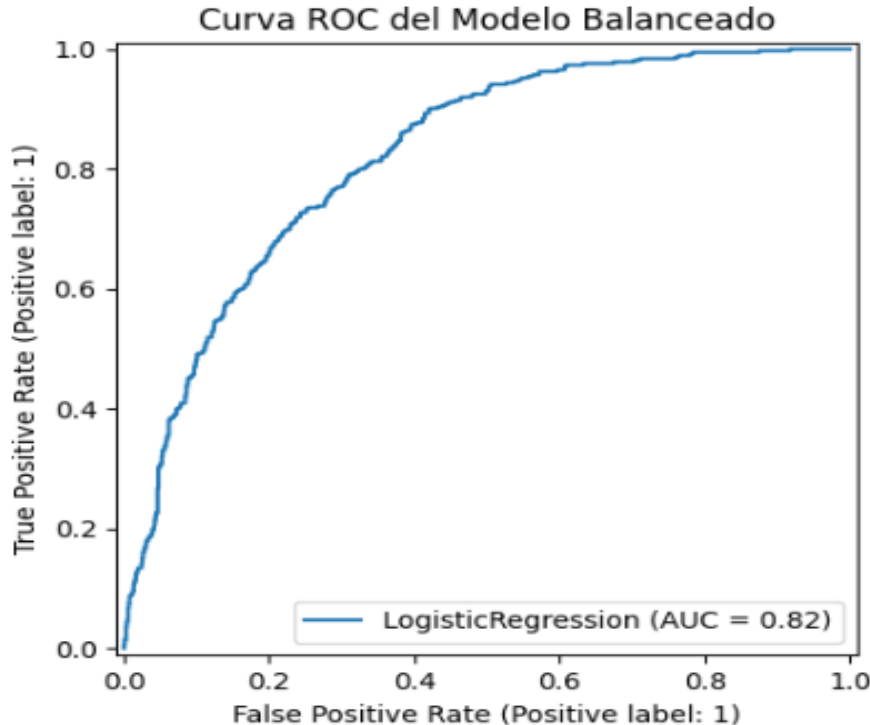
El Corazón de la Solución: Transparencia. Elegimos la **Regresión Logística** no por ser el modelo más complejo, sino por ser el más **transparente** y útil para la acción de negocio.

- **Algoritmo Elegido: Regresión Logística.**
 - **Justificación:** Le permite al equipo de Marketing entender: "**¿Cuánto riesgo me aumenta cada factor?**" La interpretabilidad es la clave del valor de negocio para diseñar ofertas de rescate.
 - **Proceso:** El entrenamiento se realizó **después** de aplicar la solución **SMOTE**.
-

Fase #5: Evaluación (*Evaluation*)

El Juicio Final: Éxito en el Equilibrio. El balance logrado después de aplicar SMOTE nos permitió obtener métricas robustas.

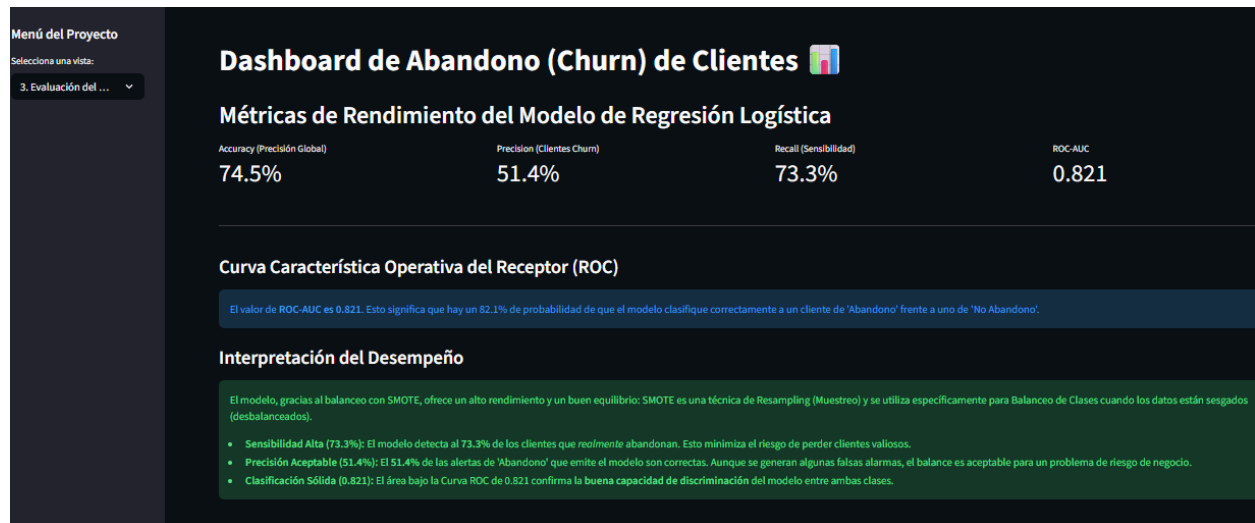
- **Métrica Clave (Recall): 73.3% Nuestra Sensibilidad:** El modelo levanta la mano por 7 de cada 10 clientes que nos abandonarían. Esto es crucial para minimizar la pérdida de ingresos.
- **Métrica de Eficiencia (Precision): 51.4% Nuestra Eficiencia:** Más de la mitad de nuestras alertas son correctas, lo que garantiza que la inversión en retención sea efectiva y controlada.
- **Conclusión:** El modelo es robusto (ROC-AUC 0.821) y logra un **equilibrio óptimo** entre la capacidad de detectar el riesgo y la eficiencia del gasto de marketing.



Recall (sensibilidad) y la precisión de la Regresión Logística, validando los números del 73.3% y 51.4%.

Fase #6: Despliegue (*Deployment*)

La Solución en Acción: Del Código al Escritorio. La solución fue empaquetada en un **Dashboard de Streamlit** y desplegada en la nube, poniendo el poder de la predicción directamente en manos del equipo de Marketing.



- **Plan de Mantenimiento:** La solución está viva. Para que el modelo no envejezca y pierda precisión (*model drift*), definimos un ciclo de **re-entrenamiento trimestral**.

El reentrenamiento del modelo lo realiza principalmente el **Científico de Datos** (*Data Scientist*) o el **Ingeniero de Machine Learning** (*ML Engineer*).

1. **Científico de Datos:** Generalmente es la persona que **analiza la causa de la pérdida de precisión** (identifica por qué el *drift* ocurrió) y **ejecuta el reentrenamiento** para asegurar que el modelo vuelva a ser preciso.
 2. **Ingeniero de ML:** Se encarga de **automatizar** este proceso, creando un *pipeline* que puede reentrenar el modelo automáticamente (o con un solo clic) cada trimestre o cada vez que la precisión cae por debajo de un umbral aceptable.
-
- **Proceso de Actualización:** El equipo de Data adquiere nuevos datos, vuelve a entrenar el modelo (regenerando `.pkl`), y lo sube a GitHub. Esto activa la **actualización automática** en Streamlit Cloud.
 - **Uso Operacional:** La solución genera una **"Puntuación de Riesgo"** diaria, que se integra en el flujo de trabajo del equipo Comercial para realizar las acciones de rescate de manera inmediata.