

Informe de Análisis de Tasa de Cancelación de Clientes (Churn)

Proyecto Challenge2 TelecomX LATAM

1. Introducción

El presente informe detalla los hallazgos y conclusiones de un análisis de datos enfocado en predecir y comprender los factores que influyen en la **tasa de cancelación de clientes (churn)** para una empresa de telecomunicaciones. El objetivo principal es identificar a los clientes con mayor riesgo de abandono y proporcionar recomendaciones estratégicas basadas en el análisis de los datos. La metodología incluyó la preparación de los datos, el desarrollo de modelos predictivos de clasificación y la interpretación de los resultados.

2. Metodología de Análisis

El proyecto se desarrolló en una serie de etapas secuenciales:

- **Preparación de Datos:** Se cargó el archivo `TelecomX_normalizado.csv` y se realizó una limpieza de datos, eliminando la columna de identificador único (`customerID`) y manejando los valores nulos. Las variables categóricas fueron codificadas a un formato numérico y se aplicó la técnica de sobremuestreo **SMOTE** para balancear la variable objetivo (`Churn`), corrigiendo el desbalance de clases detectado.
 - **Análisis Exploratorio y de Correlación:** Se visualizó la matriz de correlación para identificar las relaciones entre las variables, enfocándose en aquellas con una correlación significativa (valor absoluto > 0.1) con el `Churn`. Se investigaron variables clave como el tipo de contrato, la antigüedad (`tenure`) y el gasto total.
 - **Modelado Predictivo:** Se dividió el conjunto de datos en entrenamiento (70%) y prueba (30%). Se entrenaron dos modelos de clasificación para predecir el churn:
 - **Regresión Logística:** Un modelo lineal sensible a la escala, por lo que se aplicó normalización a las variables predictoras.
 - **Random Forest:** Un modelo basado en árboles de decisión, no sensible a la escala, que funciona bien con relaciones no lineales.
 - **Evaluación de Modelos:** El rendimiento de ambos modelos se evaluó utilizando métricas como **Accuracy, Precision, Recall, F1-score** y la **Matriz de Confusión** en el conjunto de datos de prueba.
-

3. Resultados y Hallazgos Clave

3.1 Análisis de Correlación

El análisis de correlación reveló las variables más influyentes en la tasa de cancelación:

- **Correlación positiva con Churn:** Las variables con mayor correlación positiva con la cancelación fueron `Contract_Month-to-month` (contrato mes a mes) e `InternetService_Fiber optic` (servicio de internet de fibra óptica). Esto indica que los clientes con estas características son significativamente más propensos a cancelar.
- **Correlación negativa con Churn:** Las variables con mayor correlación negativa fueron `tenure` (antigüedad del cliente), `Contract_Two year` (contrato de dos años) y `OnlineSecurity_Yes` (servicio de seguridad en línea). Esto sugiere que la lealtad y los servicios de valor agregado reducen la probabilidad de churn.

3.2 Desempeño de los Modelos Predictivos

Ambos modelos mostraron un buen desempeño, sin signos evidentes de overfitting o underfitting. Sin embargo, el modelo **Random Forest** superó ligeramente a la Regresión Logística en la mayoría de las métricas.

Análisis de la Matriz de Confusión: La matriz de confusión mostró que ambos modelos son competentes para identificar tanto a los clientes que no cancelan (True Negative) como a los que sí cancelan (True Positive). El modelo Random Forest fue particularmente efectivo para reducir los falsos negativos, lo que significa que es mejor para identificar a los clientes en riesgo de churn.

3.3 Importancia de las Variables

El análisis de la importancia de las variables en el modelo Random Forest y los coeficientes en la Regresión Logística confirmó que los principales impulsores del churn son:

1. **Tipo de Contrato:** El factor más relevante. Los clientes con **contratos mensuales** tienen un riesgo de churn mucho mayor que los que tienen contratos a largo plazo.
 2. **Antigüedad del Cliente (tenure):** La antigüedad es un fuerte predictor. Los clientes más nuevos son más propensos a cancelar.
 3. **Servicio de Internet:** El servicio de **fibra óptica** se asoció con un mayor riesgo de churn, lo que podría indicar problemas de calidad de servicio o insatisfacción con la relación costo-beneficio.
 4. **Gasto Total (Charges.Total):** El gasto total también influye, aunque en menor medida que las variables anteriores.
-

4. Conclusiones Estratégicas y Soluciones

El análisis ha permitido identificar con claridad los **factores de riesgo más importantes** que impulsan la cancelación de clientes. Basándose en estos hallazgos, se proponen las siguientes estrategias de retención:

1. Estrategia de Fidelización de Contratos:

- **Acción:** La empresa debe enfocar sus esfuerzos de marketing y ventas en incentivar a los clientes con contratos mensuales a migrar a planes de 1 o 2 años. Se pueden ofrecer descuentos, mejoras de servicio o beneficios adicionales para hacer esta transición atractiva.
- **Justificación:** El tipo de contrato es la variable más predictiva de churn. Al asegurar un compromiso a largo plazo, la empresa puede reducir drásticamente la tasa de cancelación.

2. Programa de Retención para Nuevos Clientes:

- **Acción:** Desarrollar un programa de "onboarding" proactivo para los clientes con baja antigüedad (primeros 3-6 meses). Esto puede incluir llamadas de seguimiento, encuestas de satisfacción o beneficios exclusivos para asegurar una experiencia positiva inicial.
- **Justificación:** Los clientes más nuevos son los más propensos a cancelar, por lo que es fundamental generar lealtad desde el inicio.

3. Análisis y Optimización del Servicio de Fibra Óptica:

- **Acción:** Se debe realizar un análisis más profundo sobre el servicio de fibra óptica para identificar las causas de la alta tasa de churn. Esto podría implicar revisar la calidad de la red, la atención al cliente o la competitividad de los precios. Es fundamental abordar cualquier punto de dolor en este servicio.
- **Justificación:** La correlación positiva con el churn sugiere que hay una insatisfacción latente en este segmento de clientes que debe ser resuelta.

4. Promoción de Servicios de Valor Agregado:

- **Acción:** Promocionar los servicios de seguridad en línea, soporte técnico y otros complementos a los clientes que aún no los tienen. Se pueden ofrecer paquetes combinados que incluyan estos servicios como parte de la oferta principal.
- **Justificación:** Las variables de servicios adicionales mostraron una correlación negativa con el churn, indicando que los clientes con estos servicios son más fieles y tienen una percepción de mayor valor.

Este informe proporciona una base sólida para que la empresa de telecomunicaciones tome decisiones informadas y estratégicas para combatir la pérdida de clientes y mejorar su rentabilidad a largo plazo.