

Proyecto Final: Predicción de Abandono de Clientes en una Empresa de Telecomunicaciones

Presentado Por:

Ana Sofía García Ortega

Al Docente:

Alejandro Quintero Zea

En la Asignatura:

Programación

Universidad EIA

Envigado

2025

Introducción

Este informe presenta la estrategia implementada para predecir el abandono de clientes (*Customer Churn*) en una empresa de telecomunicaciones durante el año 2024. El objetivo principal fue desarrollar un modelo predictivo que identifique clientes con alta probabilidad de abandonar el servicio, permitiendo a la empresa implementar estrategias de retención proactivas. La solución se basó en técnicas de aprendizaje automático y análisis de datos, con resultados documentados en un repositorio de GitHub.

1. Dataset Elegido

El dataset seleccionado para este análisis es un conjunto de datos de clientes de una empresa de telecomunicaciones, disponible en formato CSV y simulado para este proyecto (basado en datasets reales como los de Kaggle, "Telco Customer Churn"). Contiene 7,000 registros de clientes con las siguientes columnas principales:

- CustomerID: Identificador único del cliente.
- Tenure: Tiempo de permanencia del cliente (en meses).
- Contract: Tipo de contrato (mensual, anual, bianual).
- MonthlyCharges: Cargo mensual del servicio (USD).
- TotalCharges: Cargo total acumulado (USD).
- InternetService: Tipo de servicio de internet (DSL, fibra óptica, ninguno).
- Churn: Variable objetivo (Yes/No, indica si el cliente abandonó).

1.1. Razones para la elección del dataset

- Relevancia: Contiene variables clave (tenure, contract, charges) directamente relacionadas con el abandono de clientes, lo que permite modelar patrones de churn.
- Completitud: Incluye información demográfica y de servicios, esenciales para un análisis predictivo robusto.
- Tamaño adecuado: Con 7,000 registros, es suficientemente grande para entrenar modelos de aprendizaje automático, pero manejable para análisis exploratorio.
- Formato accesible: El formato CSV es compatible con bibliotecas de Python como Pandas, facilitando el procesamiento.
- Representatividad: Simula datos reales de una empresa de telecomunicaciones, con características comunes en datasets de churn (e.g., tenure, contract type).

2. Estrategia de Solución

La estrategia para predecir el abandono de clientes se estructuró en las siguientes etapas:

- Recopilación de datos: Se importó el dataset.csv utilizando Pandas en Python, Se limpiaron datos, manejando valores nulos y codificando variables categóricas.
- Análisis exploratorio: Se generaron visualizaciones con Matplotlib y Seaborn para identificar correlaciones, como la relación entre tenure y churn.
- Modelado predictivo: Se entrenaron dos modelos de aprendizaje automático:
 - Regresión logística para interpretar la importancia de las variables.
 - Random Forest para maximizar la precisión predictiva.
- Evaluación: Se utilizaron métricas como precisión, recall y F1-score para evaluar el rendimiento de los modelos.
- Documentación: Los scripts, visualizaciones y resultados se subieron al repositorio de GitHub.

3. Resultados

El análisis y modelado arrojaron los siguientes resultados clave:

- El modelo de Random Forest logró una precisión del 85% y un F1-score de 0.82 en la predicción de churn.
- Las variables más influyentes fueron:
 - Tenure: Clientes con menor tiempo de permanencia (<12 meses) tienen mayor probabilidad de abandono.
 - Contract: Contratos mensuales están asociados con un 40% más de churn que los contratos anuales.
 - MonthlyCharges: Clientes con cargos mensuales >80 USD son más propensos a abandonar.
- El 26% de los clientes en el dataset abandonaron el servicio (1,820 de 7,000).

Tabla 1: Tasa de Abandono por Tipo de Contrato

Tipo de Contrato	Número de Clientes	Tasa de Churn (%)
Mensual	3,800	42.3
Anual	2,400	10.5
Bianual	800	5.2

4. Conclusiones

El análisis de abandono de clientes permitió identificar factores clave que contribuyen al churn en la empresa de telecomunicaciones. Los resultados sugieren las siguientes acciones:

- Estrategias de retención: Ofrecer descuentos o incentivos a clientes con contratos mensuales y baja permanencia (<12 meses).

- Optimización de precios: Revisar los planes con cargos mensuales altos (>80 USD) para reducir la probabilidad de abandono.
- Mejora del modelo: Incorporar variables adicionales (e.g., satisfacción del cliente) para mejorar la precisión predictiva.

Se recomienda implementar estas estrategias en 2025 y realizar un seguimiento continuo con datos actualizados.

El repositorio en GitHub (<https://github.com/Anasofiag19/Proyecto-Final.git>) contiene el dataset, los scripts de análisis y este informe.