

UNIVERSIDAD PRIVADA FRANZ TAMAYO
FACULTAD DE INGENIERÍA INGENIERÍA DE SISTEMAS



**Predicción de Churn y Análisis de Retención de Clientes
de Internet en COTEL**

CASO: COTEL R.L.

ESTUDIANTE: JUAN DAVID QUISPE INCA

MATERIA: BIG DATA

LA PAZ – BOLIVIA

I - 2025

1. Introducción

La industria de telecomunicaciones en Bolivia ha experimentado una transformación significativa en las últimas dos décadas, caracterizada por una creciente competencia entre operadoras y una demanda cada vez más exigente de servicios de conectividad de alta calidad. En este contexto, la Cooperativa de Telecomunicaciones La Paz Ltda. (COTEL), una de las principales proveedoras de servicios de Internet, telefonía y televisión por cable en el departamento de La Paz, enfrenta desafíos constantes relacionados con la retención de sus clientes.

El abandono de clientes, conocido en la industria como "churn", representa uno de los problemas más críticos para la sostenibilidad financiera de empresas de telecomunicaciones. Cada cliente que cancela su servicio no solo significa una pérdida de ingresos recurrentes, sino también el desperdicio de los recursos invertidos en su adquisición inicial y la potencial transferencia de valor hacia la competencia.

El análisis de datos masivos (Big Data) ofrece una oportunidad sin precedentes para abordar este desafío de manera proactiva. A través del procesamiento y análisis de grandes volúmenes de datos históricos sobre contratos, facturación, patrones de pago y comportamiento de clientes, es posible identificar señales tempranas de insatisfacción o riesgo de cancelación. Este proyecto busca aplicar técnicas avanzadas de ciencia de datos y aprendizaje automático para desarrollar un sistema predictivo que permita a COTEL anticiparse al churn y diseñar estrategias de retención personalizadas y efectivas.

2. Justificación

2.1 Relevancia Económica

Múltiples estudios en la industria de telecomunicaciones han demostrado que **adquirir un nuevo cliente puede costar entre 5 y 7 veces más que retener uno existente**. En el contexto de COTEL, donde la competencia incluye operadoras nacionales e internacionales, cada cliente perdido representa no solo la pérdida de ingresos mensuales recurrentes (aproximadamente Bs. 100-250 según el plan contratado), sino también:

- Costos de marketing y publicidad desperdiciados
- Inversión en infraestructura de instalación no recuperada
- Pérdida de valor de vida del cliente (Customer Lifetime Value - CLV)
- Riesgo de reputación negativa por recomendaciones adversas

Un modelo predictivo efectivo que identifique con 60-90 días de anticipación a clientes en riesgo puede reducir la tasa de churn en un 15-25%, lo que para una base de clientes de 50,000+ suscriptores de internet podría significar **ahorros anuales de varios millones de bolivianos**.

2.2 Relevancia Técnica y de Big Data

Este proyecto es particularmente adecuado para un enfoque de Big Data debido a:

- **Volumen significativo de datos:** Se trabajará con más de 30,000 registros de contratos históricos, incluyendo datos de facturación mensual que pueden generar cientos de miles de transacciones.
- **Complejidad analítica:** La predicción de churn requiere el análisis de múltiples variables interrelacionadas (comportamiento de pago, cambios de plan, suspensiones, deudas acumuladas) y la aplicación de algoritmos de machine learning supervisado.
- **Procesamiento de series temporales:** Los patrones de churn solo pueden identificarse mediante el análisis longitudinal de datos históricos.

- **Generación de valor accionable:** Los resultados del modelo pueden integrarse directamente en sistemas CRM para automatizar alertas y campañas de retención.

2.3 Relevancia Institucional y Social

Para COTEL como cooperativa, la retención de socios-clientes va más allá del aspecto económico:

- **Fortalecimiento del modelo cooperativo:** Retener socios contribuye a la estabilidad y crecimiento sostenible de la organización.
- **Mejora de la calidad del servicio:** Identificar causas de insatisfacción permite implementar mejoras operativas.
- **Responsabilidad social:** Mantener empleos y servicios de conectividad estables contribuye al desarrollo económico regional.

2.4 Relación con las 5V del Big Data

Este proyecto aborda de manera integral las cinco características fundamentales del Big Data:

Volumen: El proyecto trabajará con más de 30,000 registros de contratos de internet, junto con sus respectivas facturas mensuales históricas (estimadas en 200,000+ transacciones), datos de clientes y consultas de deudas, superando ampliamente el umbral de Big Data para análisis cooperativo.

Velocidad: Los datos de COTEL se actualizan mensualmente con cada ciclo de facturación. El modelo predictivo está diseñado para procesar estos datos de forma periódica (mensual o quincenal) y generar alertas en tiempo casi real sobre clientes en riesgo.

Variedad: El proyecto integra múltiples tipos de datos:

- Datos estructurados transaccionales (facturas, pagos, deudas)
- Datos maestros (información de clientes, planes comerciales)
- Datos categóricos (estados de contrato, formas de pago)
- Datos temporales (fechas de contrato, rescisión, instalación)

Veracidad: Los datos provienen directamente de los sistemas operacionales de COTEL, garantizando su confiabilidad. Sin embargo, el proyecto incluirá procesos de limpieza y validación para manejar inconsistencias, valores nulos y registros duplicados comunes en sistemas legacy.

Valor: El resultado final del proyecto genera valor tangible y medible:

- Reducción de la tasa de churn entre 15-25%
- ROI positivo en campañas de retención (menor inversión por cliente retenido)
- Mejora en la segmentación de clientes para estrategias comerciales
- Base de conocimiento para decisiones estratégicas de la gerencia

3. Planteamiento del Problema

3.1 Descripción del Problema

Actualmente, la Cooperativa de Telecomunicaciones La Paz (COTEL) **no cuenta con un sistema predictivo que identifique con anticipación qué clientes de internet tienen alta probabilidad de cancelar su servicio**. Las acciones de retención se llevan a cabo de manera reactiva, es decir, después de que el cliente ya ha manifestado su intención de cancelar o directamente ha procedido con la rescisión del contrato.

Esta situación genera las siguientes consecuencias negativas:

1. **Pérdida inevitable de clientes:** Cuando un cliente expresa su deseo de cancelar, generalmente ya ha tomado la decisión y evaluado alternativas con la competencia, reduciendo significativamente las posibilidades de retención.
2. **Ineficiencia en recursos de retención:** Las campañas de retención masivas (sin segmentación) resultan en:
 - Inversión desperdiciada en clientes con bajo riesgo de churn
 - Falta de recursos para atender adecuadamente a clientes de alto riesgo
 - Ofertas genéricas que no abordan las causas específicas de insatisfacción

3. **Desconocimiento de patrones de churn:** La empresa carece de información estructurada sobre:
 - Qué comportamientos preceden a la cancelación
 - Qué perfiles de clientes son más propensos al churn
 - Qué factores (plan, morosidad, suspensiones) tienen mayor impacto
4. **Pérdida de oportunidades de mejora operativa:** Sin análisis sistemático, no se identifican problemas recurrentes en la calidad del servicio, atención al cliente o pricing que podrían estar causando insatisfacción generalizada.

3.2 Preguntas de Investigación

Este proyecto busca responder las siguientes interrogantes:

1. **¿Qué variables tienen mayor poder predictivo sobre el churn?**
 - ¿Es la morosidad el factor principal? ¿Los cambios de plan indican insatisfacción? ¿La antigüedad del cliente es un factor protector?
2. **¿Existen patrones temporales identificables?**
 - ¿Cuántos meses antes de la cancelación comienzan a manifestarse señales de riesgo? ¿Hay estacionalidad en el churn?
3. **¿Se pueden segmentar los clientes en riesgo por causa probable de churn?**
 - Clientes en riesgo por: precio (buscan planes más baratos), calidad del servicio, morosidad crónica, competencia.
4. **¿Qué perfil de clientes tiene mayor tasa de churn?**
 - ¿Clientes nuevos vs antiguos? ¿Planes básicos vs premium? ¿Modalidad de pago prepago vs postpago?
5. **¿Es posible construir un modelo con precisión suficiente para ser operacionalmente útil?**
 - ¿Se puede alcanzar un recall $\geq 70\%$ para identificar correctamente a clientes en riesgo sin generar demasiadas falsas alarmas?

3.3 Decisiones que se Beneficiarían de los Resultados

El modelo predictivo y los insights generados permitirán a COTEL tomar decisiones informadas en múltiples áreas:

Área Comercial:

- Diseñar campañas de retención proactivas dirigidas a segmentos específicos
- Crear ofertas personalizadas según la causa probable de churn
- Priorizar esfuerzos de retención en clientes de alto valor (high CLV)

Área de Atención al Cliente:

- Alertar a agentes cuando un cliente en riesgo contacta al call center
- Priorizar tickets de soporte de clientes con alto riesgo de churn
- Implementar programas de seguimiento post-instalación para clientes nuevos

Gerencia Estratégica:

- Identificar productos o zonas geográficas con mayor churn
- Evaluar el impacto de cambios de precio o políticas en la retención
- Proyectar ingresos con mayor precisión considerando churn esperado

Área Técnica:

- Detectar zonas con problemas de conectividad que generen insatisfacción
- Priorizar mantenimiento preventivo en áreas críticas

4. Objetivos de Investigación

4.1 Objetivo General

Desarrollar un **modelo predictivo de churn** para clientes de servicio de internet de COTEL que permita identificar con **60 a 90 días de anticipación** a clientes con alta probabilidad de cancelación del servicio, facilitando la implementación de **estrategias de retención proactivas y segmentadas** que reduzcan la tasa de abandono y maximicen el retorno de inversión en acciones comerciales.

4.2 Objetivos Específicos

1. Construir una base de datos analítica integrada

Integrar información de contratos de internet, facturación, historial de pagos y datos de clientes en una estructura unificada que permita análisis longitudinal y la creación de variables derivadas relevantes para la predicción de churn.

2. Identificar variables predictoras mediante análisis exploratorio de datos (EDA)

Realizar un análisis exploratorio exhaustivo para:

- Caracterizar la distribución de variables clave (planes, deudas, antigüedad)
- Identificar patrones de comportamiento diferencial entre clientes que cancelaron vs. clientes activos
- Detectar correlaciones y multicolinealidad entre variables

3. Crear dashboards interactivos para visualización y monitoreo

Desarrollar visualizaciones de datos que incluyan:

- Dashboard ejecutivo (tasa mensual, proyecciones)
- Panel de monitoreo de clientes en riesgo

5. Fuentes de Datos

5.1 Descripción de Bases de Datos Disponibles

El proyecto utilizará datos operacionales reales de la Cooperativa de Telecomunicaciones La Paz (COTEL), específicamente del servicio de Internet. Las fuentes principales son:

5.1.1 Tabla de Contratos de Internet

Origen: Sistema de gestión de contratos de COTEL

Formato: CSV exportado desde base de datos relacional (PostgreSQL/Oracle)

Volumen estimado: 50,000+ contratos (históricos y actuales)

Periodo temporal: 2006 - 2025 (19 años de datos)

Variables clave:

- **contrato**: ID único del contrato
- **cod_cliente**: ID del cliente (puede tener múltiples contratos)
- **plan_comercial**: Plan contratado (CTL 5MB, CTL 7MB, etc.)
- **forma_pago**: POST PAGO / PRE PAGO
- **cod_estado_contrato**: Estado actual (1=ACTIVO, 2=INACTIVO, 3=SUSPENDIDO, etc.)
- **anulado**: Descripción del estado (EN SERVICIO, RETIRADO, INACTIVO(RESCINDIDO))
- **f_contrato**: Fecha de inicio del contrato
- **f_rescision**: Fecha de cancelación (si aplica)
- **f_instalacion**: Fecha de instalación del servicio
- **direccion**: Ubicación del servicio
- **cod_servicio**: Tipo de servicio (3 = Internet)

5.1.2 Tabla de Clientes

Origen: Base de datos de clientes COTEL

Formato: CSV

Volumen estimado: 30,000+ registros únicos

Variables clave:

- **cod_cliente**: ID único del cliente
- **nombres, ape_paterno, ape_materno**: Datos demográficos
- **nro_documento**: CI del cliente
- **direccion**: Dirección del cliente
- **telefono_ref**: Teléfono de contacto
- **f_nacimiento**: Fecha de nacimiento (para calcular edad)
- **sexo**: Género del cliente

5.1.3 Tabla de Facturación

Origen: Sistema de facturación mensual COTEL

Formato: CSV

Volumen estimado: 200,000+ facturas (histórico de varios años)

Frecuencia: Datos mensuales

Variables clave:

- **contrato:** ID del contrato facturado
- **cod_cliente:** ID del cliente
- **periodo:** Periodo de facturación (YYYYMM)
- **fecha_emision:** Fecha de emisión de factura
- **monto_total:** Monto total facturado
- **monto_cotel:** Monto correspondiente al servicio
- **estado:** Estado de la factura (P=Pagado, G=Generado, NP=No Pagado)
- **telefono:** Línea asociada (si aplica)

5.1.4 Query de Deudas

Origen: Consulta SQL al sistema de cobranzas

Formato: Tabla dinámica consultable

Contenido: Deudas acumuladas por cliente en los tres servicios (Internet, Telefonía, TV Cable)

Variables extraíbles:

- Total deuda por servicio
- Antigüedad de la deuda
- Número de facturas vencidas

