

UNIVERSIDAD PRIVADA FRANZ TAMAYO

FACULTAD DE INGENIERÍA

INGENIERÍA DE SISTEMAS



**"PREDICCIÓN DE CHURN Y ESTRATEGIAS DE RETENCIÓN EN TELEFONÍA FIJA
DE COTEL R. L."**

CASO: "Cotel R.L."

AUTOR: Osco Valencia Juan Manuel

DOCENTE: Ing. Enrique Alejandro Laurel Cossío

LA PAZ – BOLIVIA

I – 2025

1. Introducción

La telefonía fija ha sido históricamente un pilar fundamental de las telecomunicaciones en Bolivia. Sin embargo, en las últimas dos décadas ha enfrentado una disrupción sin precedentes debido a la masificación de la telefonía móvil y servicios VoIP. A pesar de este contexto adverso, la Cooperativa de Telecomunicaciones La Paz (COTEL) mantiene una base sólida de clientes de telefonía fija que representa su principal fuente de ingresos y sustenta la viabilidad operativa de toda la organización.

Este proyecto desarrollará un sistema predictivo basado en análisis de Big Data y Machine Learning para identificar con 60 a 90 días de anticipación qué clientes de telefonía fija tienen mayor probabilidad de cancelar su servicio. Esta capacidad predictiva permitirá implementar estrategias de retención proactivas, personalizadas y costo-efectivas que protejan la base de clientes y aseguren la sostenibilidad financiera de la cooperativa.

El análisis utilizará datos históricos de múltiples años de operación, integrando información de contratos, facturación, comportamiento de pago, datos demográficos y servicios complementarios (internet y TV Cable).

2. Justificación

2.1 Relevancia Estratégica y Económica

La telefonía fija no es simplemente un servicio más; es el servicio ancla que sostiene financieramente a toda COTEL. Este servicio genera ingresos recurrentes predecibles que permiten proyecciones financieras estables y planificación de inversiones a largo plazo.

Impacto económico del churn:

Estudios del sector de telecomunicaciones indican que adquirir un nuevo cliente cuesta entre 5 y 7 veces más que retener uno existente. En telefonía fija, estos costos son aún mayores debido a la instalación física, gastos de marketing en un mercado en declive, y menor disponibilidad de nuevos clientes potenciales.

Un modelo predictivo efectivo que logre reducir la tasa de churn en un 25% podría generar:

- Retención de cientos de clientes adicionales anualmente
- Protección de ingresos por más de un millón de bolivianos al año

- ROI superior al 200% en el primer año

2.2 Relevancia Técnica y de Big Data

Este proyecto es ideal para Big Data debido a:

Volumen significativo: Décadas de operación generan decenas de miles de contratos con millones de transacciones que exceden las capacidades de herramientas tradicionales.

Complejidad del problema: La predicción de churn en telefonía es desafiante porque involucra múltiples causas (precio, calidad, competencia, cambios demográficos) y diferentes perfiles de clientes (residenciales vs comerciales, nuevos vs antiguos).

Necesidad de análisis temporal: El churn raramente es abrupto; está precedido por señales como disminución de uso, inicio de morosidad o cambios de plan. Identificar estos patrones requiere técnicas avanzadas de feature engineering y análisis de series temporales.

Integración de múltiples fuentes: El proyecto requiere fusionar datos de diferentes sistemas operacionales, un desafío típico de proyectos Big Data empresariales.

2.3 Relevancia Social e Institucional

Como cooperativa, COTEL tiene responsabilidades que trascienden la rentabilidad:

- **Sostenibilidad del empleo:** La pérdida masiva de clientes amenaza empleos en La Paz
- **Servicio a socios-clientes:** Retenerlos mantiene el espíritu cooperativista
- **Acceso a telecomunicaciones:** La telefonía fija sigue siendo esencial para adultos mayores y zonas con cobertura móvil limitada

2.4 Relación Preliminar con las 5V del Big Data

- **Volumen:** Registros históricos de múltiples años superarán ampliamente el requisito de 30,000 registros
- **Velocidad:** Datos mensuales con scoring periódico para intervenciones oportunas
- **Variedad:** Integración de datos transaccionales, maestros, temporales y categóricos

- Veracidad: Datos de sistemas operacionales auditados, con procesos de validación a implementar
- Valor: Reducción proyectada de churn con ROI superior al 200%

3. Planteamiento del Problema

3.1 Problema Central

COTEL enfrenta un desafío crítico: carece de capacidad predictiva para identificar anticipadamente qué clientes de telefonía fija están en riesgo de cancelar su servicio. Esta situación genera:

Gestión reactiva: COTEL solo detecta el riesgo cuando el cliente solicita cancelación o entra en mora prolongada, momento en que la decisión ya está tomada (solo 15-20% son revertibles en esta etapa).

Estrategias ineficientes: Sin segmentación predictiva, se implementan:

- Campañas masivas genéricas que desperdician recursos en clientes sin riesgo
- Ofertas irrelevantes que no abordan las causas específicas de insatisfacción
- Falta de priorización del equipo comercial

Desconocimiento de patrones: No hay respuestas a preguntas fundamentales:

- ¿Qué comportamientos preceden la cancelación?
- ¿Qué perfiles tienen mayor propensión al churn?
- ¿Cuáles son las mejores variables predictoras?
- ¿Cuánto tiempo antes se manifiestan las señales de riesgo?

Contexto agravante: La telefonía fija está en declive estructural por sustitución tecnológica (móviles, VoIP), lo que hace que cada cliente retenido sea una victoria estratégica en un mercado en contracción.

3.2 Preguntas de Investigación

1. **¿Qué variables tienen mayor poder predictivo del churn?** (morosidad, antigüedad, servicios complementarios, ubicación, tipo de cliente)

2. **¿Cuánto tiempo antes de la cancelación se manifiestan señales de riesgo?** (90, 60 o 30 días de anticipación)
3. **¿Existen patrones diferenciados por segmento?** (residenciales vs comerciales, nuevos vs antiguos, multi-servicio vs mono-servicio)
4. **¿Es posible construir un modelo operativamente útil?** (Meta: Recall $\geq 70\%$, Precision $\geq 50\%$, interpretable)
5. **¿Qué estrategias de retención serían más efectivas?** (descuentos para sensibles a precio, mejoras técnicas para insatisfechos con calidad, educación para usuarios con bajo engagement)

3.3 Decisiones que se Beneficiarán

Nivel Operacional:

- Call center: alertar cuando clientes de riesgo llaman
- Área técnica: priorizar mantenimiento para clientes en riesgo
- Cobranzas: diferenciar estrategias según riesgo de churn

Nivel Táctico:

- Comercial: campañas segmentadas, asignación óptima de presupuesto
- Marketing: contenido educativo para segmentos en riesgo
- Producto: rediseño de planes problemáticos, bundles atractivos

Nivel Estratégico:

- Gerencia: proyecciones de ingresos más precisas, evaluación de viabilidad del servicio
- Planificación: identificar zonas críticas, dimensionar equipos
- **Decisión crítica:** ¿Vale la pena seguir invirtiendo en infraestructura de cobre o migrar a VoIP?

4. Objetivos de Investigación

4.1 Objetivo General

Desarrollar un modelo predictivo de churn para telefonía fija de COTEL que identifique con 60-90 días de anticipación clientes con alta probabilidad de cancelación, proporcionando insights accionables sobre factores causales para implementar estrategias de retención proactivas y personalizadas que reduzcan la tasa de abandono en al menos 20% y protejan la sostenibilidad financiera de la cooperativa.

4.2 Objetivos Específicos

1. Integrar datos de sistemas operacionales de contratos, facturación, pagos, deudas y servicios complementarios en una estructura unificada que permita análisis longitudinal y trazabilidad completa del ciclo de vida del cliente.
2. Realizar análisis exploratorio exhaustivo identificando diferencias entre clientes que han cancelado vs activos, y crear variables compuestas de alto poder predictivo
3. Entrenar y comparar al menos tres algoritmos optimizando para Recall $\geq 70\%$ y Precision $\geq 50\%$, con énfasis en interpretabilidad.
4. Clasificar clientes en riesgo por nivel (Alto/Medio/Bajo) y causa probable, generando recomendaciones estratégicas personalizadas.
5. Crear visualizaciones ejecutivas, operativas y analíticas.

5. Fuentes de Datos

5.1 Origen de los Datos

Los datos para este proyecto provendrán de los registros oficiales de COTEL almacenados en bases de datos Oracle de los sistemas operacionales de la cooperativa. Estos sistemas gestionan la información transaccional y maestra de todos los servicios de telecomunicaciones que ofrece la organización.

5.2 Tipos de Información Requerida

El proyecto integrará información de múltiples módulos operacionales:

Información de Contratos:

- Datos maestros de contratos de telefonía fija
- Estados y transiciones de los servicios
- Fechas relevantes (contratación, activación, suspensiones, cancelaciones)
- Características del servicio contratado
- Información geográfica y de ubicación

Información de Clientes:

- Datos demográficos y de identificación
- Perfil del cliente (persona natural o jurídica)
- Información de contacto
- Historial de relación con la cooperativa

Información Transaccional:

- Registros de facturación mensual
- Histórico de consumo y cargos
- Estados de pago
- Tendencias de uso del servicio

Información Financiera:

- Estado de cuentas y saldos
- Historial de pagos
- Morosidad y deudas acumuladas
- Patrones de comportamiento crediticio

Información de Servicios Complementarios:

- Contratación de otros servicios (Internet, TV Cable)
- Paquetes y bundles
- Cross-selling histórico

5.3 Volumen Estimado

Se estima trabajar con un volumen de datos que supera ampliamente el requisito mínimo de 30,000 registros, considerando:

- Múltiples años de operación histórica de COTEL
- Base consolidada de clientes de telefonía fija (activos e históricos)
- Transacciones mensuales acumuladas a lo largo del tiempo
- El dataset final post-integración podría alcanzar cientos de miles de registros

5.4 Cronograma de Preparación de Datos

Fase	Actividad Principal	Duración	Entregables
Fase 1	Análisis y diseño	1 semana	Mapeo de tablas, queries SQL
Fase 2	Extracción	1 semana	Archivos CSV exportados
Fase 3	Integración y limpieza	2 semanas	Dataset integrado, reporte de calidad
Fase 4	Feature engineering	1 semana	Features creados, diccionario de variables
Fase 5	Preparación modelado	1 semana	Train/val/test splits, datasets procesados
Total		6 semanas	Base de datos analítica completa