

Proyecto Grupal 1

OBJETIVO

El objetivo del proyecto, el cual debe ser realizado por grupos formados de 5 personas máximo, es aplicar un algoritmo de árboles de decisión, clasificador u otros para construir un modelo predictivo de fuga de clientes en la industria de las telecomunicaciones. A partir de los resultados de este modelo y del análisis realizado, deberán proponer 4 acciones de retención de clientes.

ANTECEDENTES

Todos los grupos trabajarán sobre el mismo set de datos, con 6590 registros, los cuales corresponden al comportamiento de clientes de telefonía celular. Dentro del set encontrarán la variable dependiente “*Fugado*”, la cual corresponde a la observación realizada después de un mes respecto a la situación en la que se encuentra el cliente (*Fugado si o no*). Como se presenta en la Figura 1, el marco temporal de análisis tiene 1 mes como ventana de tiempo respecto a la predicción de fuga.

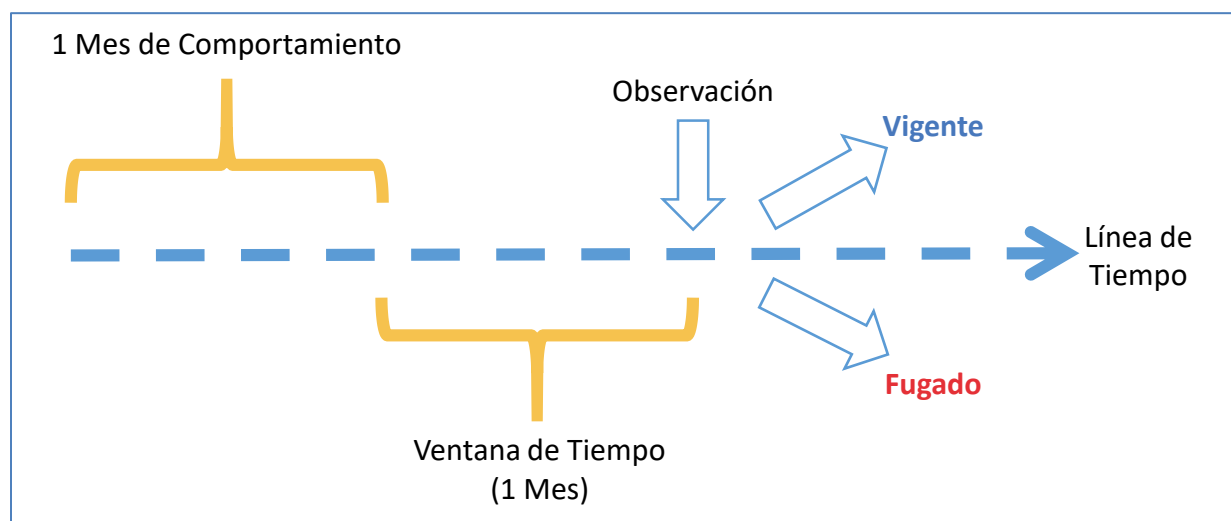


Figura 1: Marco temporal de análisis

SET DE DATOS

El set de datos, con cerca de 6.500 registros, contiene las siguientes variables:

- Sexo, Edad, Casado: Variables Demográficas
- Numero Dependientes: Personas dependientes que viven con el titular (hijos, abuelos, etc.)
- Recomendaciones realizadas: Numero de personas o familiares referidos por el cliente
- Meses como Cliente: Meses como cliente (se trunca en 72, que significa 72 o más meses)
- Plan: Tipo de plan (Ninguno, A,B,C,D,E) No se conoce la descripción de cada plan
- Cargo Mensual Llamadas: Cargo mensual por llamadas efectuadas (local+internacional)
- Múltiples líneas: ¿Plan contempla más de una línea?
- Servicio Internet: ¿Incluye plan de datos?
- GB mensuales consumidos: Consumo de datos en GB por mes
- Servicio Adicional Antivirus: Esta suscrito a un servicio de antivirus
- Servicio Respaldo en la Nube: Esta suscrito a espacio de respaldo y recuperación en la nube
- Seguro Protección Equipo: Seguro pérdida, robo, fallo equipo
- Servicio Soporte Premium: Esta suscrito a soporte premium 24/7
- Usa Streaming TV, Películas, Música: Suscrito a los diferentes servicios de streaming
- Plan Ilimitado Datos: Plan incluye datos ilimitados
- Tipo Contrato: Plan mensual, anual, 2 o más años(o empresas)
- Cobro Mensual: Boleta mensual (Negativo significa devolución/descuento)
- Histórico de Devoluciones: Total histórico de devoluciones (principalmente por reclamos/errores en facturación)
- Histórico Cargos extra datos: Total histórico de cobros por uso adicional de datos
- Histórico Cargos Llamadas: Total histórico de cobro por llamadas
- Histórico Cobro Acumulado: Total histórico cobro por cliente (*lifetime-value* actual)
- Fugado: Variable principal que indica si el cliente se ha fugado o no
- Causa Fuga: Causa de la Fuga

- Mejor la competencia: encontró un plan mejor en la competencia
- Disconforme: Alguna razón técnica o del servicio distinta a temas de precio o de servicio al cliente
- Por Precio: Muy alto el costo del servicio
- Servicio cliente: Mala atención o actitud frente a un contacto de soporte
- Otra: Otra razón distinta a las anteriores (no quiso responder en general)

DESARROLLO

Cada grupo deberá desarrollar las siguientes tareas:

1. Generar los pasos necesarios para el desarrollo de un modelo predictivo de fuga de clientes a partir del set de datos entregado. Lo cual incluye el contexto, preparación y limpieza de datos, análisis descriptivo/exploratorio, modelamiento, validación y robustez del modelo, presentación de resultados, etc.
2. Proponer 4 acciones de retención a partir de los patrones de fuga detectados. Explicitar el razonamiento utilizado para generar las acciones. Las acciones deberían ser acompañadas por una estimación del impacto o valor generado.
3. Preparar una presentación con el trabajo realizado. Consideren una presentación ejecutiva, que no exceda los 15 minutos y que exponga los puntos esenciales del desarrollo del modelo, sus resultados y sus propuestas.

CRITERIOS DE EVALUACIÓN

La nota de este proyecto se calculará de la siguiente manera:

Aspectos a ser Evaluados	Ponderación en la Nota Final
Análisis exploratorio, limpieza de datos, ingeniería de datos Entrenamiento de modelo usando SVM, árbol de decisión o random forest. Validación del modelo usando test	
Presentación: Estructura, claridad de contenido y atractivo visual de la ppt	30%
Exposición y contenido: Descripción del problema, metodología aplicada, resultados y propuesta de acciones de retención	70%

ENTREGA

Las presentaciones deben ser enviadas hasta las 23:59 horas del martes 3 de Octubre al correo regonzar@uc.cl

Las presentaciones se realizarán durante la clase del miércoles 4 de Octubre

- Máximo 15 slides
- 15 minutos presentación + 3 minutos preguntas.
- Puede presentar uno o varios miembros del equipo
- Repasar y practicar los tiempos, una diferencia de más de ± 1 minuto sobre lo estipulado podría ir en desmedro de la nota.
- No enviar la presentación en la fecha límite establecida también podría ir en desmedro de la nota