



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA INDUSTRIAL**

**MODELO DE PREDICCIÓN DE FUGA DE CLIENTES DE
TELEFONÍA MÓVIL POST PAGO**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

PATRICIO ALFREDO PÉREZ VILLANUEVA

PROFESOR GUÍA:
MARCEL GOIC FIGUEROA

MIEMBROS DE LA COMISIÓN:
IGNACIO CALISTO LEIVA
ALEJANDRA PUENTE CHANDÍA

**SANTIAGO DE CHILE
2014**

RESUMEN DE LA MEMORIA PARA OPTAR AL
TÍTULO DE INGENIERA CIVIL INDUSTRIAL
POR: PATRICIO PÉREZ VILLANUEVA
FECHA: 23/12/2013
PROFESOR GUÍA: MARCEL GOIC F.

MODELO DE PREDICCIÓN DE FUGA DE CLIENTES DE TELEFONÍA MÓVIL POST PAGO

El mercado de las telecomunicaciones presenta dinamismo y flexibilidad tanto en los servicios que entregan como en las estrategias para aumentar la rentabilidad de sus clientes. Sin embargo, existen comportamientos de uso de los usuarios que los direcciona en el camino de la fuga (o *churn* en inglés) el cual es un tema de interés para las compañías de telefonía y que acapara grandes cantidades de recursos para hacerle frente, pasando desde aspectos comerciales hasta estratégicos.

Diversos factores pueden afectar la decisión de cancelar un contrato por parte de un cliente. Estos aspectos pueden tener relación no tan solo con el servicio, sino que también con su red de contactos la que puede influenciar su comportamiento de fuga mediante el efecto "boca a boca". Así, el objetivo general del trabajo es construir un modelo de comportamiento de fuga para clientes de telefonía móvil post-pago que considere el efecto de la actividad de la red de contactos.

Para cumplir con el objetivo del trabajo, se necesita determinar qué variables afectan el comportamiento de fuga, para posteriormente utilizar estas variables como entrada en modelos predictivos de Regresión Logística y Árboles de Decisión. Lo anterior no sólo busca predecir la fuga, sino medir además el impacto económico de anticipar dicho comportamiento en la empresa, con el fin de proponer acciones comerciales que permitan retener a los clientes.

Los resultados muestran que el mejor modelo es un Árbol de Decisión, el cual logra tener un 41% de exactitud en la predicción y un *sensitivity* del 73%, lo cual indica que la proporción verdaderos positivos (fugados) que el modelo clasifica es buena. Por otro lado la información de la red es poco

concluyente respecto a la predicción, pues estos datos tienen bajo poder predictivo, por lo tanto se postulan nuevas variables y datos que serían importantes considerar en un trabajo futuro como lo es la información del detalle de llamadas por ejemplo.

AGRADECIMIENTOS

Hoy cierro una etapa de mi vida y lo primero que se me viene a la cabeza es mi madre. Chari, sin ti nada de lo que estoy alcanzando ahora hubiese sido posible sin tu entrega como madre y como la bella persona que eres. Desde tu simple sonrisa hasta el más cálido aliento desde la distancia me sirvieron para sacar adelante esta carrera. Te lo agradezco enormemente y espero algún día poder devolverte todo el amor que me entregaste y me sigues entregando.

Kafe, gracias a los astros que te tengo como hermano. Eres el mejor compañero de historias y risas que pude haber recibido. Me cuesta decírtelo en la cara, pero sabes que te quiero un kilo. Sigue siendo el mismo hombre de pelos gruñón de siempre.

Muchos momentos uno vive en la universidad; momentos buenos como también malos, pero sin duda un pilar fundamental son los amigos y sin ellos, mi vida en la "ciudad" hubiese sido compleja. Hugo, Valita, Pepe, Dani, Pasti, Coni, Lilí y Chávez, mis eternas gracias a ustedes por compartir su amistad y cariño conmigo. Espero hayan podido sentir lo mismo de parte mía. Son el recuerdo más bello que me llevaré y guardaré de estos 6 años.

Pero no puedo dejar de lado a mis amigos de la vida: Vidal, Tarro y Ángel. Ustedes siempre tuvieron un momento para subirme el ánimo o para olvidarme de lo que era estar en la U y recordar el pasado, reírnos del resto y disfrutar de lo simple. Jamás los olvidaré cabros.

Gracias a los "marketeros": Dani, Benja, Lety, Pasita y Lucía. Ustedes siempre estuvieron ahí con un "vamos si saldrá todo bien", "te va a ir la raja". Sentí su apoyo y preocupación. Me hicieron sentir querido. Se los agradezco en el corazón y estoy seguro que serán unos grandes profesionales. Sigán siendo las bellas personas que son.

Agradecer a la gente de Claro. En especial a Carlos y Yean por la paciencia y la buena onda entregada en el tiempo que estuve con ustedes.

Gracias Marcel, tu ayuda como docente y profesional lograron impulsar el término de este trabajo. Muchas gracias.

Finalmente quisiera decirles que se merecen lo mejor y busquen siempre ser felices en esta vida. Un abrazo apretado a todos. Hasta siempre...

TABLA DE CONTENIDO

1. INTRODUCCIÓN.....	1
2. JUSTIFICACIÓN Y DESCRIPCIÓN DEL PROYECTO	5
3. OBJETIVOS.....	10
4. ALCANCES.....	10
5. RESULTADOS ESPERADOS	11
6. MARCO CONCEPTUAL	11
7. MARCO METODOLÓGICO.....	12
8. ANÁLISIS DESCRIPTIVO DE LOS DATOS.....	14
10. EVALUACIÓN ECONÓMICA.....	34
11. SUGERENCIA DE ACCIONES PARA RETENCIÓN.....	39
12. TRABAJO FUTURO	39
13. BIBLIOGRAFÍA	41
14. ANEXOS	44

ÍNDICE DE TABLAS

Tabla 1: Cantidad de Fugados y No fugados presentes en la Base de datos.....	17
Tabla 2: Ejemplo de la estructura de la base de datos.....	18
Tabla 3: Evolución ARPU en 6 meses para Fugados y No fugados.....	19
Tabla 4: Evolución MOU salida en 6 meses para Fugados y No fugados.....	21
Tabla 5: Porcentaje de clientes Fugados y No fugados que cambió su rol en el último mes.....	22
Tabla 6: Porcentaje de clientes Fugados y No fugados que aumentaron su nº de seguidores, marginales, % de Externos y disminuyeron su nº de nodos y % de Claros en sus redes.....	22
Tabla 7: Porcentaje de clientes Fugados y No fugados que disminuyeron su nº de llamadas y mensajes en sus redes.....	23
Tabla 8: Matriz de confusión de Árbol de Decisión CHAID con proporción 60/40 en conjunto de entrenamiento.....	25
Tabla 9: Métricas del árbol elegido en el conjunto de entrenamiento.....	25
Tabla 10: Matriz de confusión de Árbol de Decisión CHAID con proporción 60/40 en conjunto de testeo.....	26
Tabla 11: Métricas del árbol elegido en el conjunto de testeo.....	27
Tabla 12: Variables relevantes asociadas a la fuga en el árbol elegido.....	28

Tabla 13: Matriz de confusión de Regresión con proporción 90/10 y punto de corte 0,7 en conjunto de entrenamiento.....	29
Tabla 14: Métricas de la regresión elegida en el conjunto de entrenamiento.....	30
Tabla 15: Variables relevantes asociadas a la fuga en la regresión elegida.....	30
Tabla 16: Matriz de confusión de Regresión con proporción 90/10 y punto de corte 0,7 en conjunto de testeo.....	32
Tabla 17: Métricas de la regresión elegida en el conjunto de testeo.....	32
Tabla 18: Comparación de modelos mediante las métricas más relevantes: <i>lift</i> , <i>accuracy</i> y <i>sensitivity</i>	33
Tabla 19: Matriz de confusión de Árbol de Decisión CHAID con proporción 60/40 en conjunto de entrenamiento con información de redes de contacto.....	34
Tabla 20: Matriz de confusión de Regresión con proporción 90/10 y punto de corte 0,7 en conjunto de entrenamiento con información de redes de contacto.....	34
Tabla 21: Valores de betas asociados a la información de red.....	34

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Evolución del PIB en el Sector Comunicaciones.....	5
Ilustración 2: Crecimiento anual de planes de telefonía móvil post pago. Elaboración propia.....	6
Ilustración 3: Total tráfico de planes Post pago en los últimos 6 años. Elaboración Propia.....	7
Ilustración 4: Resultado Neto de Portabilidad Móvil de Post pago.....	8
Ilustración 5: <i>Churn</i> histórico (últimos 4 años) por Empresa. Elaboración Propia.....	10
Ilustración 6: Ejemplo de una red social.....	11
Ilustración 7: Betas asociados a las variables significativas de la Regresión seleccionada.....	31

1. INTRODUCCIÓN

Hoy en día la telefonía móvil a nivel mundial es un medio de comunicación masivo. Según datos de la OCDE¹, para el año 2009 existían 1.256,8 millones de suscripciones móviles en el mundo, de los cuales aproximadamente el 57% son asociados a telefonía móvil post pago. Para el caso de Chile, y según datos de la Subsecretaría de Telecomunicaciones (SUBTEL) [1], el total de abonados llega a 24.130.754, de los cuales sólo el 28,4% pertenecen a telefonía móvil con contrato. Este porcentaje del mercado de telefonía móvil hoy en día es abastecido por 9 operadores, de los cuales Entel, Movistar y Claro se adjudican aproximadamente el 97% del mercado².

La industria de las telecomunicaciones a nivel nacional se ha caracterizado por tener un crecimiento constante en los últimos años como se puede apreciar en la Ilustración 1. Según el informe sectorial de la SUBTEL [1] el tercer trimestre de 2012 el PIB del sector mostró un aumento de un 5,6% con respecto a igual período de 2011, lo que muestra un crecimiento similar a la economía nacional representada por el PIB (5,7%). Tanto el crecimiento en telefonía como en internet móvil son las principales fuentes de crecimiento del sector de telecomunicaciones, lo que corrobora que este mercado aún continúa en crecimiento.

¹ OECD Communications Outlook 2011 [www.oecd.org/sti/telecom/outlook]

² Para más información ver Anexos 14.1.

Diferencia entre PIB Total y PIB Telecomunicaciones

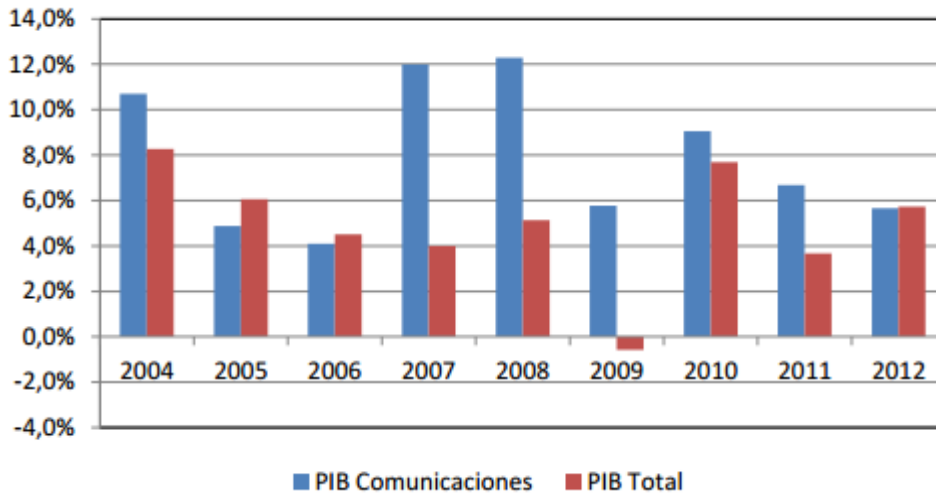


Ilustración 1: Evolución del PIB en el Sector Comunicaciones.

A lo expuesto se agrega el crecimiento en términos económicos de la industria el número de abonados a planes post pago ha presentado un crecimiento durante los últimos 8 años. Si bien en los últimos dos años (2011 y 2012) se ven crecimientos a menores tasas estos siguen siendo positivos lo que denota que el mercado de la telefonía móvil post pago sigue en alza. Lo anterior se puede apreciar en la Ilustración 2.

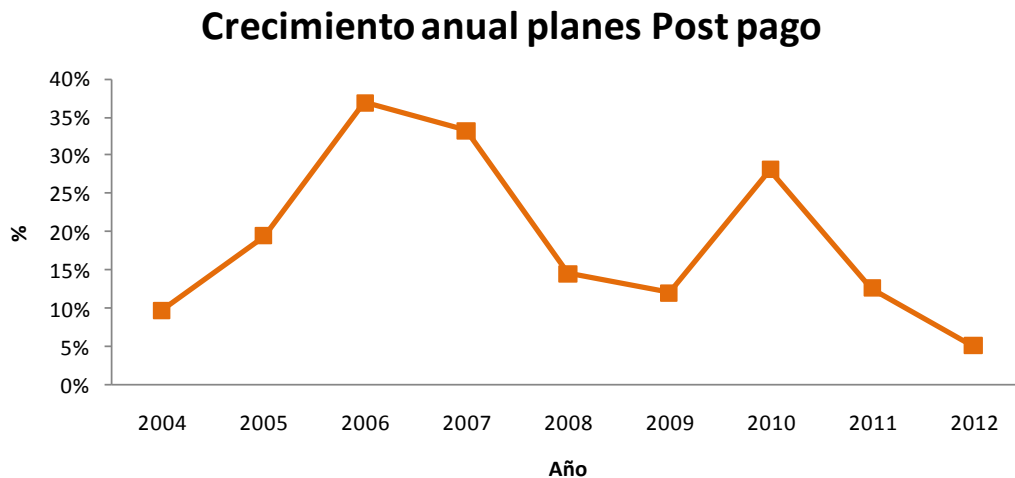


Ilustración 2: Crecimiento anual de planes de telefonía móvil post pago. Elaboración propia.

Ahora, considerando el crecimiento en el número de abonados en planes con contrato de los primeros 6 meses del año 2013 este ha disminuido (en

promedio) comparado con el año 2012 para el mismo periodo de tiempo. El año 2012 tuvo un crecimiento promedio del 0,6% mientras que para el año 2013 fue de 0,43%. Esto último puede estar asociado a factores macroeconómicos que afectan la decisión de contratar un plan o al hecho de que existe una mayor oferta comercial para telefonía prepago.

Por otro lado, durante los últimos 6 años se ha producido un aumento en el tráfico entre móviles medido en minutos. La Ilustración 3 da cuenta de lo anterior, destacando el hecho de que los clientes con planes post pago están consumiendo más minutos impulsado en parte por el hecho de que hoy en día el cobro por minuto es más barato que años anteriores.

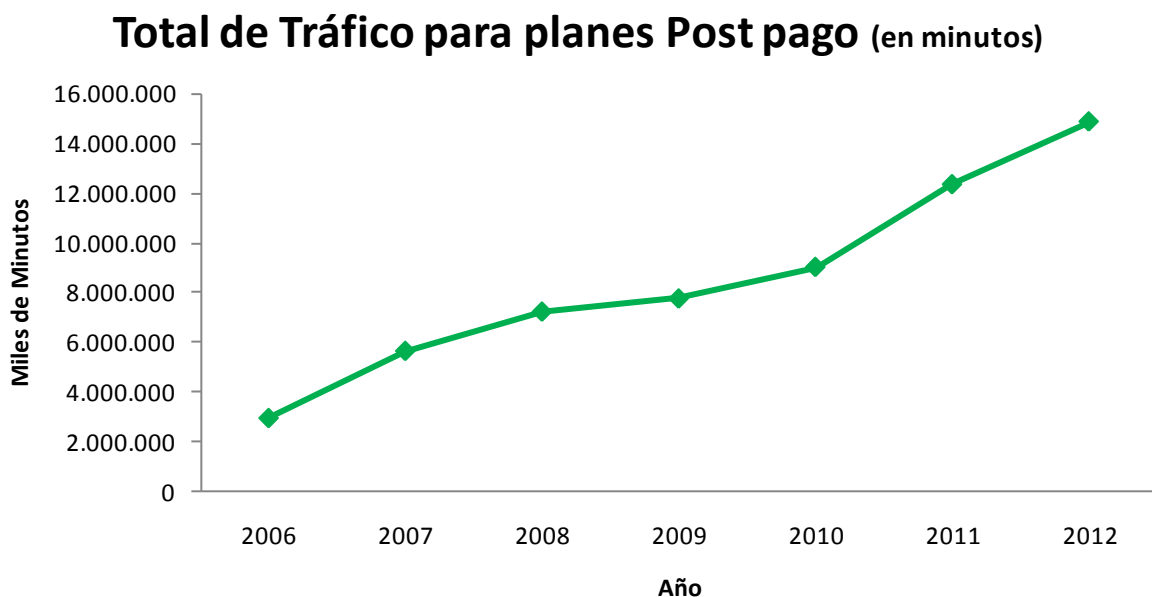


Ilustración 3: Total tráfico de planes Post pago en los últimos 6 años. Elaboración Propia.

El gobierno actual ha desarrollado y propuesto iniciativas que buscan incentivar la competencia en el sector. Una de estas propuestas es la portabilidad numérica, la cual comenzó a regir el año 2012. Dicha iniciativa derriba la barrera de movilidad entre compañía por parte de los usuarios, aportando mayor flexibilidad a este mercado. A fines del año 2012 habían hecho uso de la portabilidad casi un 4% del total de conexiones de telefonía en uso en el país (843.422 usuarios), de los cuales un 92,1% corresponden a usuarios de telefonía móvil y de estos el 44,3% corresponde a telefonía móvil post pago. La empresa que se ha visto más afectada con esta iniciativa es Movistar, seguido de Claro, los cuales han presentado una alta variación

respecto a la portabilidad de sus clientes. En la siguiente ilustración (Ilustración 4) se aprecia lo anteriormente dicho [2]:

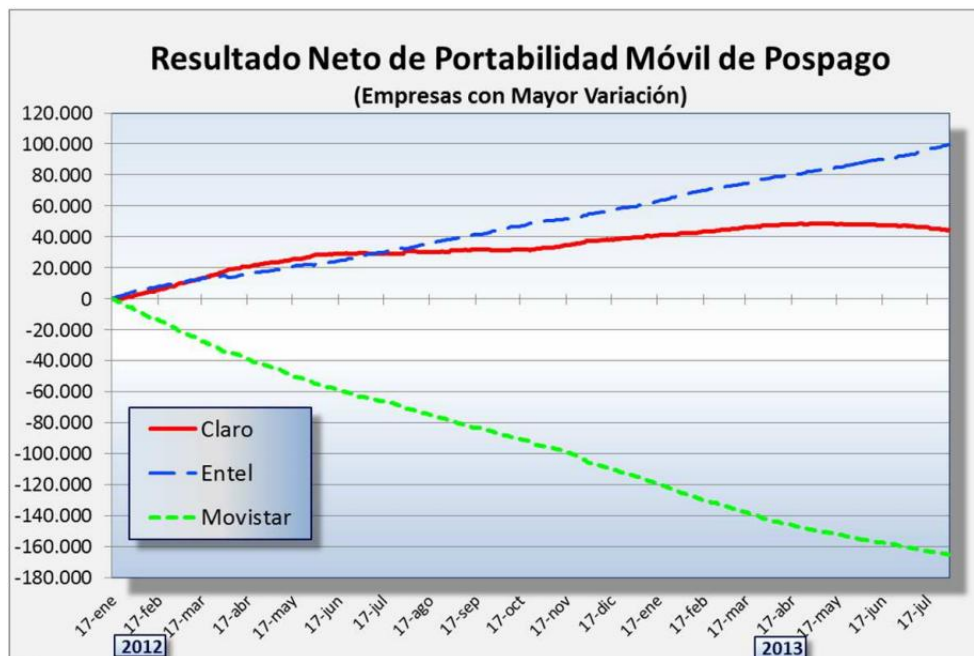


Ilustración 4: Resultado Neto de Portabilidad Móvil de Post pago.

Otra iniciativa que es parte del plan del gobierno es el proceso de homologación de la numeración de la telefonía local a 9 dígitos, la que tiene por objetivo que todas las redes telefónicas (fijas, móviles o voz por internet), tengan el mismo largo de numeración, lo que apunta a que todas las comunicaciones en Chile no tengan diferenciación de precios por distancias o tecnologías de acceso. Esto último podría tener un impacto en cómo se van modificando las redes de contacto de los usuarios, pues el aspecto económico (es decir, mantener contacto con clientes de la misma compañía para no incurrir en costos mayores) pasaría a ser un lazo menos fuerte de unión de redes.

Un aspecto relevante de mencionar es el aumento de operadores que ofrecen servicios de telecomunicación, pasando de ser 3 a 9 en la actualidad, lo cual aumenta la competitividad dentro del mercado y la oferta de productos y servicios. A la vez, los avances tecnológicos como internet y las redes sociales permiten que los clientes puedan acceder a mayor información y así comparar entre las empresas.

Bajo este contexto de competitividad, sumado al mayor acceso a internet y redes sociales por parte de los clientes, que les permite obtener mayor información y conocimiento de las empresas, es que se genera un marco de flexibilidad y dinamismo respecto a la movilidad de los clientes entre una compañía y otra.

Un aspecto relevante para las empresas relacionado a dicha movilidad, es la fuga (o *churn* en su terminología en inglés) que según [3] es "*la acción de cancelar el servicio prestado por la compañía*".

Diversos factores contribuyen al comportamiento de fuga en telecomunicaciones. Por un lado está el arista del servicio entregado por la empresa (calidad de la señal, cobertura, servicio al cliente, precios, etc.) y por otro lado está la red de contactos en la cual se desenvuelve el cliente que impacta mediante el efecto "boca a boca" y que según [4], es un factor determinante en las decisiones de compra de un consumidor. El cómo anticiparse a esta problemática, es decir, lograr identificar qué determina el término de contrato, es atractivo para las empresas de telecomunicaciones, pues entrega diversos beneficios como por ejemplo lo es la menor inversión en retener un cliente (gracias a la recomendación de la red de contactos) [5], el hecho de que adquirir a un nuevo cliente cuesta seis veces más que retener a uno, además de que clientes que se mantienen más tiempo en la empresa generan mayores ingresos y son menos sensibles a las acciones de marketing de la competencia convirtiéndose en consumidores menos costosos de servir [6].

En la presente memoria se buscará predecir el comportamiento de fuga en clientes de telefonía móvil con contrato y evaluar el impacto de la red de contacto en dicho fenómeno para el caso de la empresa Claro Chile S.A.

2. JUSTIFICACIÓN Y DESCRIPCIÓN DEL PROYECTO

Claro para fines del año 2012 capturaba el 17% del total de clientes con planes post pago (1.164.074). De este total de clientes de la empresa, para

ese mismo año el 5,1% se había fugado³ (59.367 aproximadamente), tasa bastante más alta que el año 2008 (3,8%⁴). Sumado a lo anterior en la Ilustración 5 se aprecia la evolución del *churn* para las principales empresas de la industria en los últimos 4 años. Se puede ver que el *churn* de Claro sobrepasa en casi 2 puntos porcentuales al resto de las compañías a lo largo de los años y si bien ha tenido una baja en el año 2012, sigue siendo el más alto del mercado. Si bien la fuga de clientes está presente en todas las empresas de la industria, el hecho de que este comportamiento se haga más notorio en Claro es un aspecto preocupante el cual la empresa está buscando hacerse cargo.

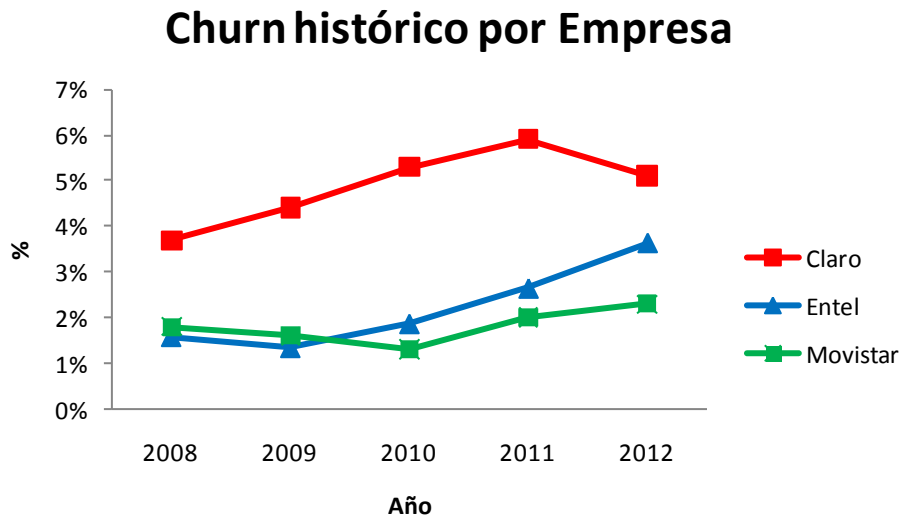


Ilustración 5: *Churn* histórico (últimos 4 años) por Empresa. Elaboración Propia.

Actualmente la compañía invierte aproximadamente 1,8 ARPU⁴ (*average revenue per user*) por cliente por año en temas de retención, lo que da una suma de \$1.638.469.833 que se invierten y que no están teniendo el efecto que se espera. Por lo tanto, si se logra mejorar la retención de clientes en un 1%, es decir, retener a 11.641 clientes (si se compara con los clientes que se fueron el 2012) se lograría aumentar la rentabilidad de la empresa en \$270.852.214⁵. Así descubrir qué está incidiendo en el comportamiento de fuga de los clientes es un asunto de relevante hoy en día en la empresa no solo en temas estratégicos sino también económicos.

³ Fuente: <http://www.duna.cl/noticias/2013/02/26/clientes-menos-fieles-tasa-de-cambio-entre-empresas-de-celulares-se-duplica-en-cinco-anos/>

⁴ Fuente: Reportes internos de Claro S.A. Se utiliza esta medida por temas de confidencialidad.

⁵ Considerando un ARPU promedio de cliente fugado de \$23266 y que el cliente se queda un año más.

Existe evidencia bibliográfica de que las redes sociales afectan las decisiones de los usuarios. Según [18] el 75% de los clientes fugados contó sus experiencias negativas con la compañía a al menos una persona. Siguiendo la misma línea [26] muestra que la recomendación es la principal razón para cambiar de servicios, por lo que la inclusión de la red de contactos en la predicción de fuga puede significar un aporte relevante en el entendimiento de la decisión de término de contrato.

Para descubrir qué influencia en mayor medida la decisión de término de contrato se analizarán datos transaccionales de los clientes, como también el impacto que tiene la red de contactos en este fenómeno.

El problema de negocio que persigue este trabajo, es que actualmente Claro realiza grandes esfuerzos por mejorar la imagen corporativa de la empresa, pero si esto no va de la mano con una buena política de servicios y retención de clientes, no será posible alcanzar la meta que la misma empresa se ha dispuesto: ser la primera compañía de telecomunicaciones del país para 2020. Por ello la predicción de fuga de clientes, aporta una componente importante al mejoramiento de la gestión de clientes de la empresa.

2.1 Redes Sociales y Fuga.

Las redes sociales son estructuras de relaciones entre agentes que pueden representarse mediante grafos que unen a dos elementos dentro de un sistema que están vinculados mediante algún tipo de criterio (relación de parentesco por ejemplo) y que pueden ser estudiadas mediante *Social Network Analysis* (SNA). En la Ilustración 6 se aprecia un ejemplo de un grafo de una red social, el cual representa una gran red de personas (por ejemplo la red de telefonía móvil de Claro). Dentro de dicha red los nodos son las personas y los arcos que los unen representan el vínculo entre estos agentes. Además se puede apreciar que dentro de esta macro red se pueden formar sub redes, dado que las personas no se vinculan con todos los participantes de la red, formando así aglomeraciones con distintos tipos de comportamientos y patrones de comunicación.

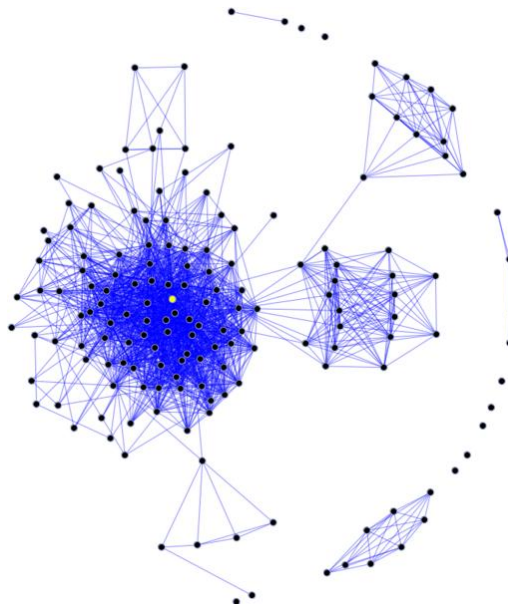


Ilustración 6: Ejemplo de una red social.

Según [31] existen algunos indicadores que permiten caracterizar una red en particular como por ejemplo la densidad, que hace referencia a la cohesión que puede tener la red. Además está la centralidad el cual destaca los actores más importantes de la red. De aquí nacen dos definiciones que permiten identificar una red. Por un lado está la proximidad central que está basado en el largo promedio de aristas de la red que unen a un nodo a los demás nodos y que revela la capacidad de un nodo de ser alcanzado. Por el otro lado se tiene la centralidad de intermediación la cual se centra en la capacidad de un nodo para ser intermediario entre otros dos nodos. Estos elementos permiten mediante SNA entender de mejor manera cómo es el dinamismo de estas redes.

Las redes sociales pueden influenciar distintos aspectos de una persona como por ejemplo contratar un servicio, comprar un producto o el hecho de abandonar una compañía, todo esto mediante las referencias o el conocido efecto "boca a boca". Existen estudios que estudian el efecto de estos lazos, por ejemplo [28] analiza el efecto de programas de marketing viral y cómo la recomendación mediante e-mails puede llevar a clientes a obtener un beneficio. Por otro lado [18] analiza cómo la influencia de pares cercanos puede llevar a tomar la decisión de descargar un videojuego basado en que otros ya lo han hecho. Sin embargo, también puede ocurrir que el rol de las redes sociales pueda ser causante de abandono de un servicio como lo

demuestra [30] donde determina que la recomendación es la principal razón para cambiar de servicios por parte de un cliente.

Dentro de una red existen lazos que unen a dos o más actores en la red y estos pueden ser débiles o fuertes basados en las interacciones que tiene dichos agentes. Según el trabajo de [32] la influencia de lazos débiles es al menos tan fuerte como los lazos fuertes. Además concluye que si bien los esfuerzos de marketing de las empresas son efectivos, la eficacia de éstos disminuye rápidamente con el tiempo y tanto los lazos fuertes como débiles se convierten en la principal fuerza de atracción para clientes. Finalmente menciona que el efecto de lazos fuertes disminuye a medida que el tamaño de la red de cada cliente decrece, lo cual va en relación con lo que se aprecia en la realidad. Todo lo anterior bajo el contexto del lanzamiento de un nuevo producto.

Para el caso de las telecomunicaciones la red social de un individuo se entiende como aquella red cercana de números telefónicos con los cuáles mantiene contacto frecuentemente. Esta red social puede ser medida mediante la cantidad de llamadas, mensajes de texto y/o tráfico de información entre los distintos clientes, también se pueden establecer relaciones entre distintos usuarios mediante la duración de cada llamada, de dónde se origina la llamada o mensaje, es decir, quién se comunica con quién, etc. Dentro de cada red existen clientes que ejecutan una mayor cantidad de llamadas, realizan una mayor cantidad de conexiones con otras personas y pasan a ser los líderes de una o más redes, pero por otro lado existen clientes que son más "pasivos" en términos de tráfico de datos y son conocidos como seguidores en la red. Por lo general realizan llamadas de manera menos frecuente y pueden ser influenciados por el resto de los miembros de la red (ya sean líderes o no). En efecto, en redes donde el líder no pertenece a la compañía, los miembros de estos grupos son 1,6 veces más propensos a fugarse [21], lo que da cuenta de que la fuerza del líder tiene una importante influencia en la probabilidad de fuga de un suscriptor. Existe evidencia bibliográfica que da cuenta del efecto que tiene la red de contacto de un cliente en la decisión de fuga de este. En [29] se estudia la evolución de fugados utilizando información de clientes con planes de prepago y da cuenta que la propensión de un consumidor a fugarse va a depender del número de lazos o nodos que ya se han fugado. De hecho en [21] se demuestra que la probabilidad de fuga de un cliente que es miembro

de grupos pequeños (menor o igual a 20 personas) es 2,7 mayor que aquellos suscriptores que pertenecen a grupos más grandes. Finalmente según [18] la información de redes sociales tiene un impacto positivo en la precisión de predicción de los modelos (incremento del 19,5%), mostrando que el efecto "boca a boca" efectivamente influencia la decisión de fuga de un suscriptor.

3. OBJETIVOS

3.1 Objetivo General

Construir un modelo de predicción de comportamiento de término de contrato que considere el efecto de la actividad de la red de contactos.

3.2 Objetivos Específicos

- Determinar las variables de mayor incidencia en la decisión de fuga.
- Evaluar el impacto de la red de contactos en la decisión de terminar el contrato.
- Evaluar económicamente el impacto de mejorar la predicción.
- Sugerir acciones que permitan reducir el comportamiento de fuga.

4. ALCANCES

Dentro de los alcances de este trabajo se pueden mencionar:

- Se predecirá la fuga de clientes de servicio de telefonía móvil post pago y no clientes de otro servicio.
- Los clientes serán personas naturales, por lo que se deja de lado a empresas y/o instituciones.
- Se utilizarán datos proporcionados por la empresa, tanto transaccional como de la red de contacto. No se levantará nueva información. La data de la red de contacto es parcial, es decir, no se cuenta con la información del detalle de las llamadas, sólo ciertos indicadores.

- La evaluación económica se realizará con los precios dispuestos para el cliente final en su contrato, sin considerar costos.

5. RESULTADOS ESPERADOS

En base a lo que se desarrollará en esta memoria, se espera alcanzar los siguientes resultados:

- Listado de las variables claves que influyen el comportamiento de fuga de los clientes.
- La probabilidad de fuga con y sin datos de la red de contactos.
- El efecto de la red de contactos en la decisión de término de contrato.
- El valor económico de la mejora en la predicción.

6. MARCO CONCEPTUAL

6.1 *Churn*

Según lo definido en [3], *churn* es "la acción de cancelar el servicio prestado por la compañía". Esta cancelación puede ser por voluntad propia del cliente, o bien, la empresa puede cancelar el contrato. Para efectos de esta memoria se considerará el *churn* en términos del abandono voluntario por parte del cliente. Con esta definición entonces la gestión del *churn* "consiste en desarrollar técnicas que permitan a las firmas mantener a sus clientes rentables y apuntar al incremento de la lealtad del cliente" [23].

La clasificación del tipo de *churn* puede ser vista desde el sentido del grado de fuga del cliente. Con esto, el tipo de fuga que se considerará será el de Absoluto, es decir, suscriptores que se han desligado sobre la base total en un periodo [5].

También existe otra visión del *churn* que está asociada al negocio de una empresa. Aquí la categorización relevante para el tema a tratar es [9]:

Churn de la compañía: Es la más costosa, pues el cliente emigra hacia la competencia, por lo que afecta no solo el ámbito económico, sino que además la imagen de la compañía que se ve reflejado en el *market share* de la competencia.

7. MARCO METODOLÓGICO

7.1 Definición de las variables claves

Parte importante de la modelación del problema es determinar cuáles son las variables importantes que influyen la decisión de terminar el contrato. Para determinar esto, primero se hará una selección de variables considerando criterios como varianza y test Chi-Cuadrado para la medición de correlación entre las variables. Al tener seleccionadas las variables, se procederán a realizar transformaciones (como el cálculo de diferencias entre meses) de estas para así capturar ciertas tendencias que se puedan presentar en los datos.

7.2 Desarrollo y Evaluación de Modelos

Como lo que se busca es poder predecir la fuga, es necesario aplicar distintos modelos predictivos para así evaluar qué metodología tiene mejor desempeño.

Los modelos a utilizar serán Árboles de Decisión y Regresión Logística. Esto basado en que:

- Existen estudios que demuestran efectividad en su uso para predicción de fuga en la industria de telecomunicaciones [8] [9] [10] [11] [12] [13] [14] [15] [17].
- Existen aplicaciones en Chile [4].
- Dentro de un torneo en que compitieron 33 modelos de predicción de fuga, estos dos fueron los con mejor desempeño [7].
- Se ajustan y se desempeñan bien en diversos conjuntos de datos [16].

Respecto a las ventajas y desventajas de cada modelo estas se explican brevemente a continuación [4] [24] [25]:

Árboles de Decisión: Respecto a las reglas de asignación éstas son sensibles a pequeñas perturbaciones en los datos. Por otro lado, presenta inconvenientes cuando la cantidad de alternativas es grande dentro de una variable, lo que dificulta su partición.

En relación a sus ventajas es que los árboles de decisión son tolerantes al ruido y a valores faltantes. Por otro lado reduce el número de variables independientes. Así el entendimiento de los resultados que se obtienen con este modelo es claro respecto a la predicción pudiendo así conocer y perfilar tanto al cliente como su comportamiento previo a la fuga.

Regresión Logística: La principal ventaja de este modelo es que se pueden incluir términos que hacen del modelo uno no lineal, lo que entrega flexibilidad a este.

En cambio, la desventaja es que esta misma flexibilidad, si es alta, puede conllevar a un alto grado de sobreajuste del modelo afectando la precisión de este.

Para la evaluación de los modelos se utilizarán *Lift*, *F-Measure* y aspectos relacionados a la exactitud (*accuracy*), sensibilidad, especificidad y precisión. Dichos criterios de evaluación son los utilizados en [4], [7], [10], [12], [18], [19], [20] y [21] para comparar modelos.

7.3 Evaluación del impacto de la red de contacto en la predicción.

Debido a que no solo existen factores relacionados a aspectos transaccionales que logran explicar la fuga, es que se quiere medir el impacto que tiene la red de contactos del cliente en la decisión de fuga de este. Para evaluar este efecto, se realizará el mismo procedimiento que el

desarrollado en el primer paso, pero ahora considerando como *input* adicional indicadores de la red de contacto de los clientes. Para modelar el efecto, se utilizarán los mismos métodos y medidas de evaluación mencionados en la sección anterior.

7.4 Evaluación económica de la predicción

Teniendo los resultados de ambos modelos (con y sin efecto de la red de contactos) se evaluará económicamente cuánto le reportaría a la empresa retener a los clientes que fueron predichos como propensos a fuga. Para esto se tendrá en consideración los precios vigentes de los contratos que cada cliente posee y se proyectarán los meses que se espera que el consumidor se quede en la compañía y así traer a valor presente dichos flujos.

8. ANÁLISIS DESCRIPTIVO DE LOS DATOS

8.1 Datos

La base de datos con la que se cuenta contempla una temporalidad de 6 meses, la cual abarca desde Octubre de 2012 hasta Marzo de 2013. Se filtró la base con la finalidad de obtener los datos de clientes tipo personas, descartando el tipo empresa u otra categoría. Con ello se llegó a la siguiente distribución entre Fugados y No fugados, destacando el hecho que se tienen una sobre representación de Fugados en la base [Tabla 1]:

Tipo	Cantidad de clientes únicos
No fugados	13.208
Fugados	36.891

Tabla 1: Cantidad de Fugados y No fugados presentes en la Base de datos.

Esto da un total de 50.099 clientes únicos para los 6 meses. Cada usuario está descrito por 281 atributos. Cabe mencionar que se cuenta con la historia transaccional de los clientes fugados también de un largo de 6 meses, es

decir, si por ejemplo un cliente se fugó en Noviembre de 2012, se cuenta con la data desde Mayo de 2012 a Octubre de 2012.

En una primera limpieza de datos se eliminaron atributos que poseían menos del 5% de registros válidos, como también aquellas de varianza cero y que no representaban aporte al objetivo del trabajo por tener información repetida o similar como lo fueron variables asociadas a la región, modelo del teléfono que el cliente usa, etc. Con esta primera limpieza 26 variables fueron eliminadas. Posterior a ello se utiliza test Chi-cuadrado para obtener variables correlacionadas. Finalmente cada cliente es definido por 62 variables las cuales se pueden clasificar en tres categorías:

- Información de comportamiento: minutos de uso (*MOU*), duración de las llamadas (últimos 3 y 6 meses), número de llamadas/mensajes, ARPU (*average revenue per user*), saldo a pagar, días de deuda, etc.
- Información de interacción con la compañía: tipo de reclamo, cantidad de kilobytes usados, morosidad (medido en días), cantidad de reclamos.
- Aspectos demográficos: ciudad, rut.

Para estructurar mejor la base de datos, cada cliente tiene asociado sus 62 variables divididas por temporalidad, es decir, cada variable se separa por mes. Por ejemplo, un cliente cualquiera estaría descrito en la base de datos como aparece en la [Tabla 2]. El término "t" denota el mes de Marzo por lo que "t-1" es Febrero, "t-2" es Enero y así sucesivamente. Esto se repite para el resto de las variables y el resto de los clientes.

ID	ARPU t	ARPU t-1	ARPU t-2	ARPU t-3
Cliente 1	13.208	14.898	15.546	15.222

Tabla 2: Ejemplo de la estructura de la base de datos.

Para capturar de mejor manera la información de los datos, es que se crearon nuevas variables que miden la diferencia entre un mes y otro. Por ejemplo, para el caso de la variable "ARPU t" y "ARPU t-1", se genera la variable "delta ARPU t" que es la diferencia entre los dos atributos anteriores.

Esto se replica para demás variables que sean continuas. Para el caso de variables ordinales, como por ejemplo "Nota de pago", se crean variables dicotómicas que captura el hecho de si un cliente bajó su nota de pago o no para cada mes.

Según lo realizado por [27] se dividirá la base muestral en proporción 70/30, tanto para el conjunto de entrenamiento y de testeo. Dentro del primer conjunto se probarán distintas proporciones entre No fugados y Fugados, con el fin de probar qué distribución se desempeña mejor en el entrenamiento de los datos. Dichas proporciones serán: 50/50, 60/40, 70/30, 80/20, 90/10 y 95/5 respectivamente. Lo anterior se basa en el trabajo realizado por J. Burez y D. Van den Poel [20]. Para el conjunto de testeo se utilizará una proporción 95/5. Esto se basa en lo que ocurre en la realidad dentro de la compañía con la proporción entre No fugados y Fugados para el periodo de análisis.

Para modelar la red de contactos se cuenta con una base de datos que comprende los mismos clientes que la base anteriormente descrita y tres meses de datos. Para los No fugados la temporalidad es de Enero – Marzo 2013 y para los Fugados 3 meses hacia tras desde su mes de fuga. Al igual que con los datos transaccionales se crean nuevas variables que capturan las diferencias de los atributos mes a mes. La base cuenta con las siguientes categorías de variables:

- Rol de cada usuario: este puede ser Seguidor, Líder o Marginal (el cual fue calculado por la empresa en base a la cantidad de llamadas, mensajes y transferencia de datos con otros usuarios de la misma compañía como de otras empresas).
- Tráfico de la red: cantidad de llamadas y mensajes de texto que cada cliente realiza dentro de su red.
- Conformación de la red: cantidad de seguidores y/o marginales que esta posee, cantidad de nodos que cada red tiene y qué porcentaje de estos nodos son de la empresa Claro y qué porcentaje no. Esto último determina el Tipo de Comunidad a la cual pertenece el cliente, que es una clasificación interna de la empresa para identificar las distintas redes respecto a su composición entre clientes de la empresa y clientes externos. Así, los tipos de comunidad pueden ser: Externa de grado 1

(0-34% de Claro), Externa de grado 2 (34,1-49,9% de Claro), Neutral (50% Claro), Fiel de grado 2 (50,1-66% de Claro) y Fiel de grado 1 (66,1-100% de Claro).

Se entenderá nodo como un número telefónico. Por ejemplo, si la red de un cliente tiene 5 nodos, es que dicho usuario se comunica con 4 personas.

8.2 Análisis Descriptivo

Considerando la data transaccional tanto de los clientes No fugados como Fugados existen algunas consideraciones. Por un lado aquellos suscriptores que se fueron de la compañía presentan, en promedio, un valor comercial (medido en ARPU) 0,7% mayor que los No fugados durante los 6 meses de estudio. Sin embargo, al mirar la evolución de esta medida mes a mes ocurre que los No fugados presentan una tendencia a la baja en esta medida, mientras que los Fugados si bien mantienen en promedio este valor, es más alto comparado con los No fugados. Este aspecto puede deberse en parte a que los clientes Fugados tienen asociado planes de mayor valor comercial o al hecho de que gastan más minutos en llamadas a clientes de otras compañías (que tienen una tarifa más elevada). Esto da cuenta que la compañía está perdiendo clientes valiosos en términos de valor comercial. Lo anterior se puede apreciar en la Tabla 3:

Tipo Cliente	ARPU t (\$)	ARPU t-1 (\$)	ARPU t-2 (\$)	ARPU t-3 (\$)	ARPU t-4 (\$)	ARPU t-5 (\$)
No fugado	17.711,3	18.216,5	18.743,5	18.289,9	27.773,3	32.926,6
Fugado	23.268,9	23.773,4	23.830,3	23.256,4	23.298,0	17.260,9

Tabla 3: Evolución ARPU en 6 meses para Fugados y No fugados.

Si bien en términos de ARPU los Fugados son más valiosos, al mirar el margen que dejan para la compañía, éste es 5% mayor comparada con los No fugados, lo que podría dar cuenta que es levemente más costoso para la empresa mantener a estos clientes, lo cual es lógico puesto que la empresa tuvo que invertir dinero para retener a estos clientes durante un tiempo más.

Otra característica que diferencia a los Fugados de los No fugados está relacionada a la conectividad. Se vislumbra que aquellos que se fugan tienen una menor propensión a conectarse y transmitir datos. Esto se ve reflejado en que los No fugados consumen en promedio 122% más kilobytes de información que lo Fugados, además la cantidad de conexiones que realizan durante el mes es 45% menor que la cantidad de conexiones de los No fugados. Si bien tanto Fugados como No fugados tienen asociados planes multimedia o planes de cuenta exacta (que les permiten conectividad y transferencia de datos) son estos últimos los que más aprovechan este servicio.

Respecto a la cantidad de minutos que un cliente usa en un mes, se da el hecho que los clientes Fugados usan en promedio 4% menos de minutos en promedio para los 6 meses que los No fugados para llamadas salientes, pero viendo la evolución mes a mes, se aprecia que éstos comienzan a aumentar sus minutos gastados en llamadas salientes comparado con los No fugados [Tabla 4], lo cual se vincula con un punto relevante que es el hecho que los clientes Fugados comenzaron a gastar más minutos extras de su plan para realizar llamadas *offnet*, es decir, a clientes de otras compañías. En promedio para los 6 meses, los Fugados gastaron 94 minutos extras de sus planes para realizar este tipo de llamadas, mientras que los No fugados gastaron en promedio 40 minutos. Siguiendo en esta misma línea la duración de llamadas *offnet* (medido en minutos) de los Fugados es en promedio 20% mayor que la de los No fugados. Es más, durante los últimos 4 meses, los clientes que se habían ido comenzaron a aumentar paulatinamente la duración de sus llamadas a clientes de otras compañías, no así los No fugados. Caso contrario ocurre con la duración de las llamadas *onnet*, donde los No fugados gastan en promedio 24% más de minutos que los Fugados, además se da el hecho que los clientes que se van comienzan a disminuir la duración de sus llamadas a clientes de la misma compañía.

Referente al número de llamadas que realizan los clientes está el caso de los Fugados, quienes realizaron un 27% más de llamadas (en promedio) a clientes fuera de la compañía comparado con los No fugados, mientras que para llamadas *onnet* los Fugados realizaron un 4% más de llamadas que los No fugados. Esto sugiere que los clientes que se fueron de la empresa son

suscriptores que tiene mayor frecuencia de llamada que los que clientes que se quedan. Todo lo anterior da pie para pensar que los Fugados se ven incentivados a irse de la empresa dado el hecho de que realizan más llamadas y gastan más minutos en comunicarse con clientes externos a Claro.

Tipo Cliente	MOU salida t	MOU salida t-1	MOU salida t-2	MOU salida t-3	MOU salida t-4	MOU salida t-5
No fugado	195,1	209,8	225,1	224,0	233,3	235,8
Fugado	207,3	216,2	223,9	225,0	226,1	190,9

Tabla 4: Evolución MOU salida en 6 meses para Fugados y No fugados.

El hecho de manifestar a la compañía lo que está haciendo mal es un punto que los No fugados realizan más seguido que los Fugados, ya que en promedio los primeros realizaron 2 reclamos en los últimos tres meses, mientras que los Fugados no realizaron ninguno. Lo que da cuenta de dos posibles aristas: por un lado los clientes Fugados no confían en el sistema de atención al cliente de la empresa y/o no presentan problemas que requieran este tipo de atención.

Un aspecto relevante para evaluar a un cliente es el referente a su comportamiento de pago. Para esto se analiza la variable Nota de Pago que toma un valor de 1 a 7 y que se calcula mes a mes observando cuánto se demora en pagar la(s) cuenta(s) que tiene asociada el cliente Para el caso de los Fugados estos tienen asociados una Nota de Pago de 2 en promedio valor que es menor comparada con la nota de los No fugados (que es de 4 en promedio). Es decir, los que se han ido tienen peor comportamiento de pago que los que se quedan. Esto se ve reforzado por la variable Morosidad (que mide el número de días que se atrasa un cliente en pagar) el cual para los Fugados es en promedio 10 días de atraso, versus los 5 días de atraso de los No fugados.

Considerando ahora la información de la redes de contactos se analiza la información considerando las 3 categorías de variables:

Rol de cada usuario:

La Tabla 5 muestra el porcentaje de clientes que cambiaron su rol en el último mes:

Tipo	%
No fugados	23%
Fugados	28%

Tabla 5: Porcentaje de clientes Fugados y No fugados que cambió su rol en el último mes.

De lo anterior se puede apreciar que existe poca movilidad respecto al rol que cumple un cliente dentro de su red. Cabe mencionar que sólo el 2% de los clientes cumple un rol de Líder, mientras que casi el 50% cumple un rol de Seguidor durante el último mes.

Conformación de la Red:

La Tabla 6 muestra los porcentajes de clientes para los cuales aumentó su número de Seguidores, Marginales y de Porcentaje de Externos en sus redes. Esto se puede apreciar en las primeras tres columnas de la tabla. En las redes de los fugados existe una mayor presencia de contactos externos, lo que podría influenciar en una posible decisión de fuga dado el hecho de que se están comunicando con usuarios que no son de la compañía. Además de la columna 5, se puede ver que en casi la mitad de las redes de los fugados, existió una disminución (de casi un 7%) en la cantidad de clientes que son de la compañía, lo que refuerza lo dicho anteriormente. Caso contrario ocurre para el caso de los No fugados, los cuales presentan un cambio menor en sus contactos externos.

VARIABLE					
Tipo	Nº Seguidores	Nº Marginales	% Externos	Nº Nodos	% Claro
No fugados	54.4%	58%	49.2%	48.3%	46.6%

Fugados	57%	59.1%	55.1%	45.3%	48.4%
---------	-----	-------	-------	-------	-------

Tabla 6: Porcentaje de clientes Fugados y No fugados que aumentaron su n° de seguidores, marginales, % de Externos y disminuyeron su n° de nodos y % de Claros en sus redes.

Tráfico de la red

La Tabla 7 muestra los porcentajes de clientes para los cuales disminuyó su cantidad de llamadas y mensajes en sus redes para el último mes tanto para Fugados como No fugados.

VARIABLES		
Tipo	N° Llamadas	N° Mensajes
No fugados	57.2%	46.4%
Fugados	44.5%	43.7%

Tabla 7: Porcentaje de clientes Fugados y No fugados que disminuyeron su n° de llamadas y mensajes en sus redes.

El número promedio de llamadas que disminuye dentro de las redes de los No fugados es de 143, mientras que para los Fugados es de 139. Si bien es mayor el porcentaje de redes de los No fugados en donde existe disminución en el número de llamadas, en promedio la cantidad de llamadas menos es similar para ambos grupos, lo que indicaría que los clientes Fugados realizan menos llamadas dentro de sus redes.

Lo anteriormente descrito permite formular algunas hipótesis respecto a la información transaccional y de red. Por un lado se tiene que los clientes Fugados se ven influenciados por el hecho de que sus comunicaciones (minutos de extras de llamadas, duración de llamadas, cantidad de llamadas, etc.) paulatinamente van dirigidas a clientes que no pertenecen a la compañía, por lo que sus incentivos a permanecer en Claro van cambiando en el tiempo, prefiriendo a la larga cambiar de compañía. Por otro lado, se puede decir que un mal comportamiento de pago por parte de los clientes

está relacionado con que un suscriptor tome la decisión de irse y se hace notorio con un par de meses de anticipación. Con los resultados de los modelos, se espera corroborar lo anteriormente expuesto.

9. RESULTADOS

9.1 Información Transaccional

Se reportan los resultados obtenidos para los modelos Árbol de Decisión y Regresión utilizando sólo la información transaccional de los clientes.

9.1.1 Árbol de Decisión

Como se mencionó en la sección 8.1 se probaron distintas proporciones para entrenar el modelo, el cual considera 4 meses de datos (Octubre 2012 a Enero 2013). La validación en el conjunto de testeo contempla el uso de los otros dos meses restantes (Febrero – Marzo 2013).

Se usaron distintos tipos de crecimiento de Árboles (CHAID, CRT y QUEST) para todas las proporciones de entrenamiento. Posteriormente para el elegir el árbol que mejor desempeño tenía se utilizaron las métricas mencionadas en 7.2, las cuales se calculan de la siguiente manera:

$$Lift = \frac{Precisión}{\frac{P}{P+N}}$$

$$F - Measure = \frac{2}{\frac{1}{Sensitivity} + \frac{1}{Presición}}$$

$$Accuracy = \frac{TP + TN}{P + N}$$

$$\text{Sensitivity} = \frac{TP}{P}$$

$$\text{Specificity} = \frac{TN}{N}$$

Donde,

P = Número de *churns* en el conjunto de datos (Positivos).

N = Número de *non-churns* en el conjunto de datos (Negativos).

TP = Número de Verdaderos Positivos.

TN = Número de Verdaderos Negativos.

FP = Número de Falsos Positivos.

Y,

$$\text{Presición} = \frac{TP}{TP + FP}$$

El primer indicador muestra qué tan mejor se logra desempeñar un modelo comparándolo con una muestra aleatoria de clientes. Por otro lado *F-Measure* mide el desempeño de la clasificación considerando tanto la *sensitibity* y la precisión, los cuales miden la proporción de clientes fugados bien clasificados versus el total de fugados en base y la proporción de clientes fugados bien clasificados versus el total de clientes que fue predicho como fugado, respetivamente.

En Anexos 14.2 se muestran las matrices de confusión para las distintas proporciones y los tres tipos de crecimiento de los árboles. Al compararlos la proporción que tiene mejor desempeño basado en las métricas anteriores es un QUEST con proporción 60/40, el cual posee la siguiente matriz de confusión en el conjunto de entrenamiento [Tabla 8]:

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9202	45
Fugado	67	6097

Tabla 8: Matriz de confusión de Árbol de Decisión CHAID con proporción 60/40 en conjunto de entrenamiento.

Las métricas que arrojó este árbol son las siguientes [Tabla 9]:

Medida	Valor
Lift	2,48
F – Measure	0,99
Accuracy	0,9927
Sensitivity	0,9891
Specificity	0,9951

Tabla 9: Métricas del árbol elegido en el conjunto de entrenamiento.

Al aplicar las reglas de clasificación que este árbol arroja en el conjunto de testeo, se tiene la siguiente matriz de confusión [Tabla 10]:

Observado	Pronosticado	
	No fugado	Fugado
No fugado	1545	2417
Fugado	57	152

Tabla 10: Matriz de confusión de Árbol de Decisión CHAID con proporción 60/40 en conjunto de testeo.

A pesar de que la cantidad de Falsos Positivos es grande (2417) comparado con la cantidad de Falsos Negativos (57), es preferible esto pues mal clasificar a un Fugado genera mayor perjuicio económico que mal clasificar a un No fugado, ya que si se clasifica a un cliente con la etiqueta de no fuga, siendo que en realidad se fuga, no se estarán percibiendo los ingresos que ese cliente reporta a la empresa, en cambio si se etiqueta como fugado a un cliente que no lo está la situación cambia y hasta se puede entregar un mejor servicio (mediante estrategias de retención) a este suscriptor.

Las métricas que arroja el árbol en el conjunto de testeo son las siguientes [Tabla 11]:

Medida	Valor
Lift	1,48
F – Measure	0,12
Accuracy	0,41
Sensitivity	0,73
Specificity	0,39

Tabla 11: Métricas del árbol elegido en el conjunto de testeo.

La métrica *Lift* indica “cuán mejor un clasificador predice comparado a una selección aleatoria. Compara la precisión con la tasa de churn total en la base de prueba” [20] por lo que un predictor con un valor mayor a 1, indica que el modelo predice bien. Para este caso el valor es aceptable y bueno. Para el caso de la medida *Accuracy* el valor es de un 41%, el cual se ve afectado principalmente por el hecho de que se tiene una mala clasificación de No fugados alta, sin embargo, de igual manera se considera que el modelo tiene un ajuste aceptable en esta medida. Esto además se ve reflejado en la métrica *Sensitivity*, el cual indica la proporción de Fugados bien clasificados por sobre el total de fugados en el conjunto de datos. Esta medida tiene un valor de 0,73 (con un máximo de 1) lo cual es positivo dado que la clase de interés es la de Fugados y por lo explicado en el párrafo anterior.

El esquema del árbol de decisión se encuentra en Anexos 14.3. Dicho árbol considera las siguientes variables como las más relevantes para explicar la fuga de un cliente [Tabla 12]:

VARIABLES
- Bajó Nota de pago en periodo t-4
- Variación factura total en periodo t-4
- Variación número de cargas en periodo t-3
- Variación monto total de carga en periodo t-3
- Variación saldo final vencido en periodo t-3
- Variación morosidad (en días) en periodo t-2

Tabla 12: Variables relevantes asociadas a la fuga en el árbol elegido.

Dada las variables, la fuga exhibe un patrón temporal de ocurrencia y es algo que se espera puesto que este fenómeno no es algo que ocurra de manera espontánea. La primera señal que se envía está relacionada con si un cliente bajó su nota de pago en el periodo t-4 (considerando como t=0 Enero de 2013), que es la nota que cada suscriptor tiene asociado basado en su patrón de pago de la o las cuentas que tenga asociado a su rut. Para aquellos clientes que efectivamente bajaron su nota de pago, es decir, empeoraron en este aspecto, ocurre que por lo demás tienen una diferencia en su factura total de su cuenta de -\$46.000 aproximadamente para el mismo periodo (t-4). Dicha factura se ve influenciada por los montos no pagados que el cliente puede haber tenido, aspecto que no se puede saber dado que no se cuenta con más historia pasada de los Fugados. De igual manera estas dos variables dan las primeras luces respecto a la fuga de un cliente asociándolo a un comportamiento de pago. Posterior a esto, ya en el periodo t-3, ocurre que el cliente disminuye el número de cargas que realiza a su celular en 6 unidades. En varios planes, salvo aquellos de Cuenta Exacta, es posible recargar el celular con más monto del que el plan trae, para así seguir ocupando los servicios de telefonía. De hecho, cerca 60% de los fugados tiene asociado un plan que les permite realizar recargas (como planes Multimedia, Blackberry, etc.). Esto indicaría que el cliente ya no estaría interesado en tomar esta opción para seguir utilizando los servicios, sino que en caso de que se le acaben los recursos de su plan opta por no recargar como antes.

Ahora para aquellos clientes que no presentaron una baja en su nota de pago (que son aproximadamente el 10%), se tiene otro tipo de comportamiento el cual también se relaciona con las recargas, pero esta vez asociado al monto de éstas. Cuando estos montos disminuyen (ya en el periodo $t-3$) en aproximadamente \$3.000 se establece la regla que el cliente es propenso a fuga. Esto puede estar determinando que un cliente no está interesado en seguir cargando con más montos su plan al momento de pasarse. En el mismo periodo ocurre además que el cliente aumenta su saldo final que está vencido, es decir, se sigue asociando a un comportamiento de pago. Ya en el periodo siguiente ($t-2$) ocurre que la morosidad del cliente (medido en días) aumenta en 3 días, es decir, que ahora atrasa el pago de su cuenta en 3 días.

Lo anteriormente descrito indica que el comportamiento de fuga de un suscriptor se relaciona directamente con el patrón de pago que este tiene.

9.1.2 Regresión Logística

Al igual que el modelo anterior, se probaron distintas proporciones de entrenamiento, además de utilizar distintos puntos de corte, 0,5, 0,7 y 0,9 para definir como fugado a un cliente. Para evaluar los modelos se usaron las mismas métricas que el modelo anterior.

En Anexos 12.4 se muestran las matrices de confusión para las distintas proporciones y los tres puntos de cortes. Al compararlos la proporción que tiene mejor desempeño es la regresión con punto de corte 0,7 con proporción 90/10, el cual posee la siguiente matriz de confusión en el conjunto de entrenamiento [Tabla 13]:

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9143	103
Fugado	162	865

Tabla 13: Matriz de confusión de Regresión con proporción 90/10 y punto de corte 0,7 en conjunto de entrenamiento.

En la Tabla 14 se aprecian las métricas que arroja el modelo en el conjunto de entrenamiento.

Medida	Valor
Lift	8,93
F – Measure	0,86
Accuracy	0,97
Sensitivity	0,84
Specificity	0,98

Tabla 14: Métricas de la regresión elegida en el conjunto de entrenamiento.

Las variables más relevantes para explicar la fuga de un cliente según este modelo se presentan en la Tabla 15:

VARIABLES
- Bajó Nota de pago
- Variación factura total
- Variación número de cargas
- Variación monto total de carga
- Variación morosidad (en días)
- Variación saldo final vencido

Tabla 15: Variables relevantes asociadas a la fuga en la regresión elegida.

Dichas variables tiene los siguientes valores de beta asociados [Ilustración 7]:

Valores de Betas para las Variables significativas del modelo de Regresión

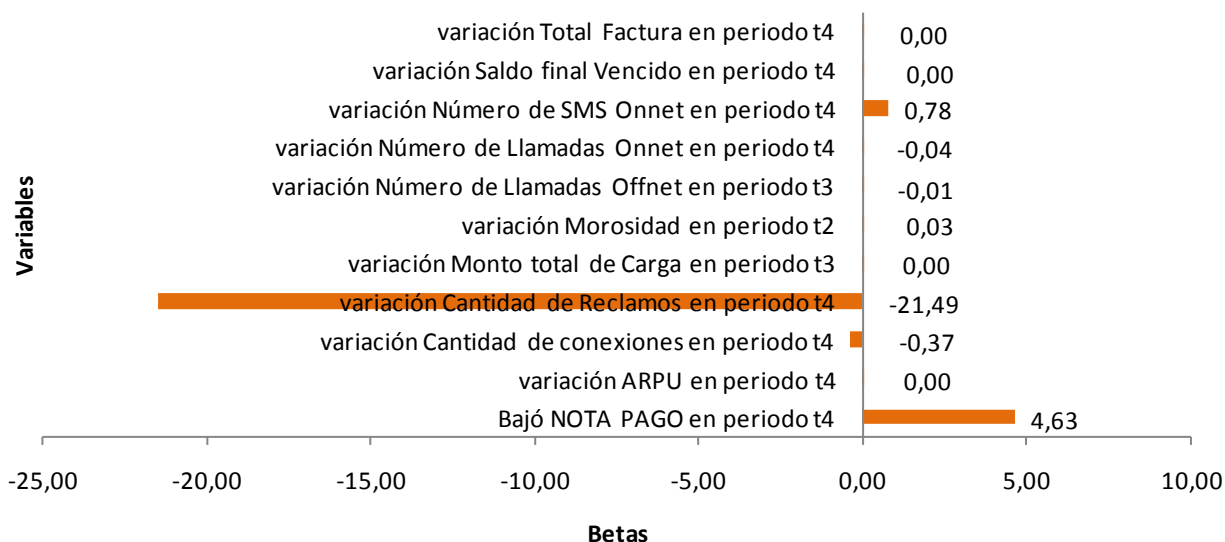


Ilustración 7: Betas asociados a las variables significativas de la Regresión seleccionada.

La variable que más peso tiene en la explicación de la fuga es la variación en la Cantidad de Reclamos que realiza un cliente ($\beta = -21,49$). Cuando esta variación es negativa, es decir, disminuye la cantidad de reclamos que el cliente hace respecto de un mes a otro, estaría indicando que el cliente o está más satisfecho con el servicio o no ha presentado problemas con su plan/equipo, por lo que disminuiría la probabilidad que un cliente decida fugarse. Por otro lado el hecho de que baje su nota de pago (que es la segunda variable más relevante) afecta positivamente la fuga de un suscriptor. Esto está asociado directamente con un comportamiento de pago el cual aparece de nuevo (al igual que el modelo anterior) como una componente importante en la decisión de fuga. Y en relación a este punto es que aparece otra variable vinculada al comportamiento de pago como lo es la Morosidad, que si bien tiene menor incidencia (0,03) afecta positivamente la fuga. Si un cliente aumenta su Morosidad (número de días en que atrasa el pago de su cuenta) es más probable que en los próximos meses incurra en fuga. Un aspecto llamativo ocurre con la variación de Mensajes *onnet* que el cliente gasta, el cual afecta de manera positiva la fuga ($\beta = 0,78$). Esto se

puede explicar por el hecho de que si un suscriptor disminuye (variación negativa) la cantidad de mensajes que envía a otros clientes de la misma compañía puede estar ocurriendo que o bien comienza a establecer comunicaciones con clientes de otras compañías o está paulatinamente dejando de ocupar los servicios de mensajería de su plan. Finalmente otras de las variables que influyen la fuga, pero de manera negativa, es la cantidad de conexiones ($\beta = -0,37$). Si un suscriptor aumenta la cantidad de conexiones que realiza con su equipo, da a entender que está ocupando de manera más constante los servicios de conexión que su plan le ofrece, por lo que tendría menos incentivos a querer dejar la compañía.

Al probar este modelo en el conjunto de testeo, se obtiene la siguiente matriz de confusión [Tabla 16]:

Observado	Pronosticado	
	No fugado	Fugado
No fugado	2360	1602
Fugado	111	98

Tabla 16: Matriz de confusión de Regresión con proporción 90/10 y punto de corte 0,7 en conjunto de testeo.

Es posible ver que el modelo mal clasifica una gran cantidad No fugados, pero es menor comparado con el modelo anterior. Sin embargo, ocurre lo contrario con la cantidad de Fugados mal clasificados siendo el caso que la Regresión tiene más Fugados mal etiquetados que el Árbol de Decisión. Las métricas que arroja este modelo en el conjunto de testeo se aprecian en la Tabla 17:

Medida	Valor
Lift	1,21
F – Measure	0,11
Accuracy	0,59
Sensitivity	0,47
Specificity	0,60

Tabla 17: Métricas de la regresión elegida en el conjunto de testeo.

Al observar la métrica *lift* se aprecia que está por sobre 1, lo cual indica que el modelo se desempeña mejor que una selección aleatoria, pero tampoco de una manera destacada. Al evaluar la exactitud del modelo este alcanza el 59%, lo cual es considerado bueno, pero esto básicamente se debe al hecho de que la cantidad de No fugados mal clasificados es menor comparada con el árbol. No obstante, al ver el parámetro *sensitivity* este tiene un valor de 0,47, lo cual indica que el modelo no logra capturar ni al 50% de los fugados de la base, lo que sería un aspecto negativo del modelo ya que es la clase de interés.

9.1.3 Comparación entre Modelos

Si bien el Árbol de Decisión posee un 18% menos de exactitud que la Regresión, este último se desempeña peor tanto en *lift* como en *sensitivity*. Los datos de los fugados se ajustan mejor en el Árbol de Decisión, por lo que éste puede aprender de mejor manera el fenómeno de fuga que la Regresión, además dado el hecho que se prefiere un modelo con ajuste aceptable, que se desempeñe bien en la clasificación de Fugados, pues es la clase de interés, y que se pueda lograr un mejor entendimiento de la fuga en términos de la sucesión de hechos, es que se prefiere el Árbol de Decisión como el modelo a utilizar.

La Tabla 18 compara ambos modelos utilizando las métricas más relevantes:

Medida	Árbol	Regresión
	Quest - 60/40	0,7 - 90/10
Lift	1,48	1,21
Accuracy	0,41	0,59
Sensitivity	0,73	0,47
TP	152	98
TN	1545	2360
FP	2417	1602
FN	57	111

Tabla 18: Comparación de modelos mediante las métricas más relevantes: *lift*, *accuracy* y *sensitivity*.

9.2 Información Transaccional y de Redes de Contacto

Al agregar la información de las redes de contacto de los clientes ocurre que para ambos modelos el poder predictivo de esta información es insignificante. Las matrices de confusión en el conjunto de entrenamiento para ambos modelos no varía [Tablas 19 – 20]. Por otro lado, las variables más relevantes que arrojan ambos modelos no consideran los datos de red, por lo que la historia detrás de la fuga no cambia al considerar las redes de contacto.

	Pronosticado	
	No fugado	Fugado
Observado		
No fugado	9202	45
Fugado	67	6097

Tabla 19: Matriz de confusión de Árbol de Decisión CHAID con proporción 60/40 en conjunto de entrenamiento con información de redes de contacto.

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9143	103
Fugado	162	865

Tabla 20: Matriz de confusión de Regresión con proporción 90/10 y punto de corte 0,7 en conjunto de entrenamiento con información de redes de contacto.

De todas maneras, con el modelo de Regresión es posible obtener los coeficientes asociados a las variables de red, los que puede dar ciertas directrices respecto a cómo esta información guía la fuga o impacta en esta.

La Tabla 21 muestra los betas asociados a las variables de la red de contactos.

Variable	Valor
- Cambió de rol en periodo t	-0,757
- Cambió de rol en periodo t-1	-0,903
- Variación cantidad de llamadas en t	0,001
- Variación cantidad de llamadas en t-1	0,001
- Variación cantidad de nodos en t	-0,004
- Variación cantidad de nodos en t-1	0,055
- Variación porcentaje de Claros	-1,515
- Variación cantidad de seguidores	-0,31

Tabla 21: Valores de betas asociados a la información de red.

Se ve que la variable con mayor peso es la variación en el porcentaje de clientes de Claro en la red ($\beta = -1,515$), es decir, que si aumenta esta porción, un cliente está menos incentivado a querer fugarse, lo cual va en la lógica de lo que ocurre en la realidad, pues un usuario está más propenso a mantener contacto con aquellas personas de su red y si estas son de la misma compañía que la del cliente, el efecto fuga es menos predominante. Por otro lado los efectos de variación de nodos, son bajos ($\beta = 0,001$) lo cual es esperable, ya que si bien esta variable indica el número de personas con

la que se relaciona, esta no dice mucho respecto a la fuga, pues si aumenta este valor este puede ser tanto de clientes de la misma compañía o de clientes externos. Lo mismo ocurre con la cantidad de llamadas. Al no tener el detalle de estas, ya sea la dirección de la llamada (quién llama a quién), minutos hablados, etc., poco se puede inferir sobre la respuesta que esta variable pueda tener en la fuga. De igual manera si este número de llamadas se asocia a clientes de la misma compañía se podría tener un indicio más claro sobre la influencia esta variable, es decir, que si un cliente (por ejemplo) disminuye su cantidad de llamadas a clientes claro, es más esperable que dicho suscriptor se vea propenso a fuga. Por otro lado, el hecho de que un cliente cambie su rol dentro de la red, por ejemplo pasando de ser Seguidor a ser Líder, afecta negativamente la fuga, pues si esto ocurre es menos probable que el cliente decida dejar la empresa dado que para haber llegado a tal posición dentro de la red, debe estar realizando y recibiendo llamadas/mensajes de manera más frecuente que en periodos pasados.

10. EVALUACIÓN ECONÓMICA

Para la evaluación económica se considerará el supuesto que al momento de lograr retener a un cliente este puede fugarse de la compañía con la misma probabilidad que un cliente al cual no se le ha realizado una acción de retención. Es decir, si mediante algún tipo de acción de retención un cliente propenso a fuga se queda en la compañía, en los periodos siguientes éste puede irse de la compañía con igual probabilidad que el resto de los clientes.

El beneficio que se obtiene al retener a un grupo de clientes con propensión a fuga viene dado por la siguiente fórmula:

$$\text{Beneficio (B)} = \text{N}^\circ \text{ de Clientes} * \text{ARPU} * \frac{1}{\% \text{ Churn}}$$

El primer término se refiere a la cantidad de clientes que efectivamente se logran retener y alcanzar con las acciones que la compañía ejecuta. El segundo término se utiliza como medida de valorización de un cliente y éste

valor corresponde al ARPU promedio de un cliente que se ha fugado de Claro (considerando los 6 meses de datos) el cual es de \$22.500 aproximadamente. El último término da cuenta de la temporalidad con la que un cliente se quedaría en la empresa después de retenido, el cual está basado en la tasa del *churn* presente en la compañía.

Pero para poder retener a un cliente es necesario realizar una inversión, la cual corresponde a $1,8 * ARPU^6$ por cliente. Para el término "N° de Clientes" se considerará todo aquel suscriptor que esté propenso a fuga. Con ello la inversión sigue la siguiente fórmula:

$$\boxed{\text{Inversión (Inv)} = N^{\circ} \text{ de Clientes} * 1,8 * ARPU}$$

Por lo tanto el Ingreso Neto vendrá dado por la resta de las dos fórmulas:

$$\boxed{\text{Ingreso Neto (IN)} = B - Inv}$$

Primero se considerará el caso base en donde no se ha logrado retener a ningún cliente y la tasa de fuga es del 5% (tasa de fuga para el año 2012). Con esto la cantidad de dinero que la empresa ha dejado de percibir es la siguiente:

$$\text{Beneficio no percibido} = 59.367 * \$22.500 * \frac{1}{5\%} = \text{MM\$26.715,15}$$

Ahora se considerará un escenario positivo en donde la tasa de fuga disminuye al 3% (promedio entre las tasas de fuga de los competidores directos de la compañía). Con esta tasa, 34.922 clientes aproximadamente se estarían yendo de la compañía (considerando como la base de clientes postpago de Claro 1.164.074). Dentro de este escenario positivo respecto a la predicción de fuga se considerarán 3 escenarios más respecto a la totalidad de clientes que efectivamente se logran alcanzar y retener un

⁶ Por temas de confidencialidad se utiliza esta medida.

tiempo más en la compañía: escenario positivo, intermedio y negativo. Para el primero se considerará que se logra retener el 80% de los fugados, el intermedio un 50% y el malo un 20%.

La inversión bajo este escenario es de:

$$Inv = 34.922 * 1,8 * ARPU = MM\$977$$

Escenario Positivo (80%):

$$B = 27.938 * \$22.500 * \frac{1}{3\%} = MM\$20.954$$

Por lo tanto,

$$IN = MM\$19.977$$

Escenario Intermedio (50%):

$$B = 17.461 * \$22.500 * \frac{1}{3\%} = MM\$13.096$$

Por lo tanto,

$$IN = MM\$12.119$$

Escenario Negativo (20%):

$$B = 6.984 * \$22.500 * \frac{1}{3\%} = MM\$5.238$$

Por lo tanto,

$$IN = MM\$4.261$$

Ahora se considerará un escenario intermedio en donde la tasa de fuga es del 3,8% (promedio entre las tasas de fuga de los competidores directos de

la compañía). Con esta tasa, 44.235 clientes aproximadamente se estarían yendo de la compañía. Se considerarán los mismos 3 escenarios anteriores respecto a la retención.

La inversión bajo este escenario es de:

$$Inv = 44.235 * 1,8 * ARPU = MM\$1.238$$

Escenario Positivo (80%):

$$B = 35.388 * \$22.500 * \frac{1}{3,8\%} = MM\$20.953$$

Por lo tanto,

$$IN = MM\$19.715$$

Escenario Intermedio (50%):

$$B = 22.118 * \$22.500 * \frac{1}{3,8\%} = MM\$13.096$$

Por lo tanto,

$$IN = MM\$11.858$$

Escenario Negativo (20%):

$$B = 8.847 * \$22.500 * \frac{1}{3,8\%} = MM\$5.238$$

Por lo tanto,

$$IN = MM\$4.000$$

Finalmente se considerará un escenario negativo en donde la tasa de fuga alcanza el 4,5% (promedio entre las tasas de fuga de los competidores directos de la compañía). Con esta tasa, 52.383 clientes aproximadamente

se estarían yendo de la compañía. Teniendo en cuenta los mismos 3 escenarios anteriores se tienen los siguientes resultados:

La inversión bajo este escenario es de:

$$Inv = 52.383 * 1,8 * ARPU = MM\$1.466$$

Escenario Positivo (80%):

$$B = 41.906 * \$22.500 * \frac{1}{4,5\%} = MM\$20.953$$

Por lo tanto,

$$IN = MM\$19.487$$

Escenario Intermedio (50%):

$$B = 26.192 * \$22.500 * \frac{1}{4,5\%} = MM\$13.096$$

Por lo tanto,

$$IN = MM\$12.363$$

Escenario Negativo (20%):

$$B = 10.477 * \$22.500 * \frac{1}{4,5\%} = MM\$5.238$$

Por lo tanto,

$$IN = MM\$4.945$$

Con lo anterior es posible apreciar que ya en el caso más negativo, es decir, bajando la tasa de fuga a un 4,5% y reteniendo un 20% de los clientes, el ingreso neto que recibe la empresa es altamente positivo, por lo tanto la

mejora en la predicción de fuga junto con las acciones de retención presentan un escenario favorable para Claro.

11. SUGERENCIA DE ACCIONES PARA RETENCIÓN

Actualmente Claro utiliza la siguiente secuencia de acciones para la retención de un cliente cuando éste da señales de fuga. Primero consiste en enviarle un mensaje de texto el cual contiene algún tipo de beneficio (entrada al cine, descuento en algún lugar de comida rápida, etc.). Cuando esta acción no tiene la respuesta esperada, se le envía un mail al cliente ofreciéndole ahora otro tipo de beneficios, como bolsas de minutos o de navegación, etc. Si con esta última acción el cliente no cambia su comportamiento de fuga, se le realiza un llamado telefónico en donde la oferta es más directa y la respuesta es inmediata. En esta acción la oferta puede ser algún descuento en el plan o aumento de minutos pagando el mismo monto de su plan, etc.

Considerando el hecho de que la fuga de un suscriptor tiene directa relación con su comportamiento de pago, se le sugiere a la empresa considerar distintas opciones de pago por parte de los clientes más morosos. Se podría contactar al 20% de los clientes con más morosidad en términos monetarios y ofrecer un descuento en el total de la deuda o bien pagar la deuda total en distintas cuotas. Otra opción a tener en cuenta sería el considerar beneficios por pagar la cuenta a tiempo o dentro de ciertos días antes del plazo de vencimiento. Dichos beneficios podrían ser los mismos que actualmente la empresa maneja (descuentos en cine, bolsas de minutos, etc.). Esto podría incentivar por parte del cliente el hecho de pagar su cuenta con antelación y se estaría premiando un buen comportamiento de pago, cosa que hoy en día no se hace.

12. TRABAJO FUTURO

Considerando el hecho de que existe evidencia bibliográfica de que las redes de contacto sí tienen un impacto en el comportamiento de fuga de un cliente, es que se cree que esta componente podría tener una alta influencia la decisión de fuga de los clientes de Claro. Sin embargo, dada la estructura de los datos con los que se contaba y la poca información que se pudo obtener

de estos, es que se postulan otras métricas de red que pueden ayudar a capturar de mejor manera la información que esta data contiene. Por un lado sería valioso contar con la información del detalle de las llamadas entre clientes, es decir, aparte de la información de cantidad de llamadas y mensajes que realizan dos suscriptores, agregar la información de frecuencia de llamadas, el origen de la llamada (quién llama a quién) y minutos de la llamada. Con esto se podría generar un grafo que permita visualizar de mejor manera los flujos de tráfico entre clientes. Otra componente relevante es el hecho de poder contar con la información del porcentaje de clientes que se han fugado de las redes de cada suscriptor, esto ayudaría para determinar qué tanto influye en la fuga de un suscriptor el hecho de que un alto (o bajo) porcentaje de los clientes Claro de la red se hayan fugado.

Uno de los indicadores actuales de la red hace referencia al número de nodos (cantidad de clientes con las que se comunica un suscriptor) de la red, la cual puede ir variando mes a mes. Sin embargo, si ocurre un aumento (o disminución) de esta variable, no se sabe a qué porcentaje de ese aumento (o disminución) corresponde a clientes de Claro o de una compañía externa. Esta información sería útil, pues se podría testear el hecho de que si en la red de un cliente aumenta el número de nodos y de este aumento un gran porcentaje es de clientes externos a la compañía, podría existir mayor probabilidad de fuga.

Un cliente puede pertenecer a distintas redes de contacto, cumpliendo distintos roles dentro de cada una de ellas. Además dentro de cada red existen distintas configuraciones de llamadas, tráficos de datos, etc. Por lo tanto sería interesante contar con un seguimiento de cada red durante varios meses y evaluar el impacto relativo de cada una de ellas en la decisión de fuga de un cliente.

Finalmente, dado que esta memoria está enfocada en clientes del tipo persona, sería valioso para la compañía evaluar la fuga de los clientes tipo empresa, los cuales tienen otro tipo de uso de los servicios, comportamientos de pagos distintos y un tratamiento por parte de Claro distintos, ya que representan clientes de mayor valor comercial, esto con la finalidad de seguir mejorando el servicio entregado por la empresa para este tipo de clientes.

13. BIBLIOGRAFÍA

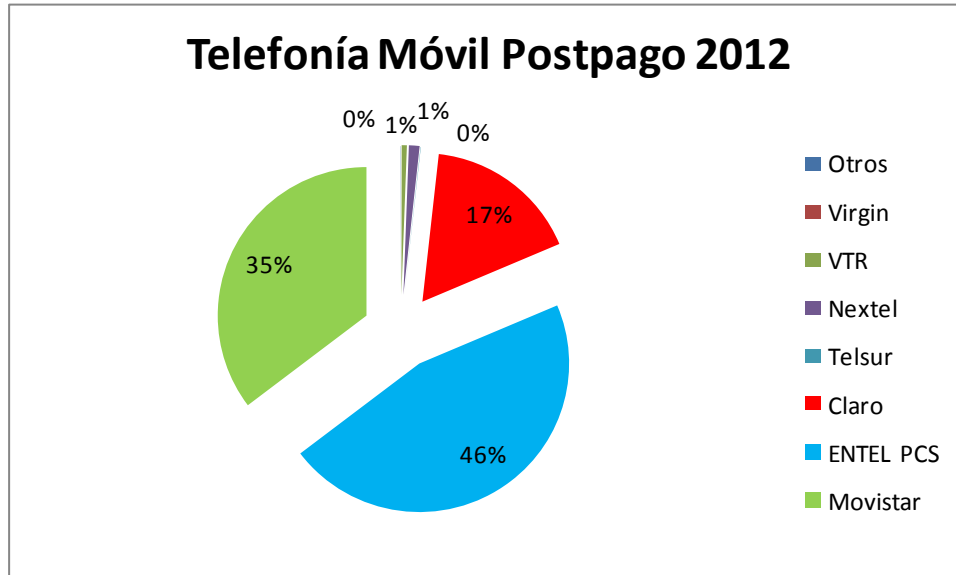
- [1] Subsecretaria de Telecomunicaciones, "Telecomunicaciones en Chile", Santiago, 2013.
- [2] Subsecretaria de Telecomunicaciones, "Información de Portabilidad", Santiago, 2013.
- [3] J. Lu, "Predicting customer churn in the telecommunications industry an application of survival analysis modeling using sas", *SAS User Group International (SUGI27) Online Proceedings*, pp. 114-27, 2001.
- [4] F. Barrientos, "Diseño e implementación de una metodología de predicción de fuga de clientes en una compañía de telecomunicaciones", 2011.
- [5] M. Richeldi and A. Perrucci, "Analyzing churn of customers".
- [6] W. Verbeke, D. Martens, C. Mues and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques", *Expert Systems with Applications*, 38, pp. 2354-2364, 2010.
- [7] S. Neslin, S. Gupta, W. Kamakura, J. Lu, C. Mason, "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models", *Journal of Marketing Research*, vol XLIII, pp. 204-211, 2006.
- [8] J. Hadden, A. Tiwari, R. Roy, D. Ruta, "Churn Prediction using Complaints Data", *World Academy of Science, Engineering and Technology*, 2006.
- [9] B. Huang, T. Kechadi, B. Buckley, G. Kiernan, E. Keogh, T. Rashid, "A new feature set with new windows techniques for customer churn prediction in land-line telecommunications", *Expert Systems with Applications*, 37, pp. 3657-3665, 2010.
- [10] S. Hung, D. Yen, H. Wang, "Applying data mining to telecom churn management", *Expert Systems with Applications*, 31, pp. 515-524, 2006.
- [11] J. Ahn, S. Han, Y. Lee, "Customer Churn Analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry", *Telecommunications Policy*, 30, pp. 552-568, 2006.

- [12] L. Bin, S. Peiji, L. Juan, "Customer Churn Prediction based on the Decision Tree in Personal Handyphone System Service", *School of Management, University of Electronic Science and Technology of China*, 2007.
- [13] M. Mozer, R. Wolniewicz, E. Johnson, H. Kaushansky, "Predicting Subscriber Dissatisfaction and Improving Retention in the wireless Telecommunications Industry", *Transactions on Neural Networks, Special issue on Data Mining and Knowledge Representation*, 2000.
- [14] S. Chandrasekhar, "Predicting the Churn in Telecom Industry", 2006.
- [15] C. Wei, I. Chiv, "Turning telecommunications call details to churn prediction: a data mining approach", *Expert Systems with Applications*, 23, pp. 103-112, 2002
- [16] D. Michie, D.J. Spiegelhalter, C.C. Taylor, "Machine learning, Neural and Statistical Classification", 1994.
- [17] J. Hadden, A. Tiwan, R. Roy, D. Ruta, "Computer assisted customer churn management: State of the art and future trends", *Computers & Operations Research*, 34, pp. 2902-2917, 2005.
- [18] T. Dierkes, M. Bichler, R. Krishnan, "Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic network", *Decision Support Systems*, 51, pp. 361-371, 2011.
- [19] V. Lazarov, M. Capota, "Churn Prediction", *Database Applications*, 2005.
- [20] J. Burez, D. Van den Poel, "Handling class imbalance in customer churn prediction", *Expert Systems with Applications*, 36, pp. 4626-4636, 2009.
- [21] Y. Richter, E. Yom-Tov, N. Slonim, "Predicting customer churn in mobile networks through analysis of social groups", 2007.
- [22] T. Fawcett "Roc graphs: Notes and practical considerations for researchers", *Technical Report, HP Laboratories*, 2004.
- [23] M. Lejenue, "Measuring the impact of data mining on churn management", *Internet Research*, 11, pp. 375-387, 2001.
- [24] S. Dreiseitl, L. Ohno-Machado, "Logistic regression and artificial neural network classification models: a methodology review", *J. of Biomedical Informatics*, 35, pp. 352-359, 2002.

- [25] C. Pérez. Técnicas de Segmentación: Conceptos, herramientas y aplicaciones. Madrid: Gaceta Grupo Editorial. 2011.
- [26] R. East, W. Lomax, R. Narain, "Customer tenure, recommendation and switching, Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior", *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 14, pp. 46-54, 2001.
- [27] J. Ferreira, M. Vellasco, M. Pacheco, C. Borbosa, "Data mining techniques on the evaluation of wireless churn", *European Symposium on Artificial Neural Networks*, pp. 483-488, 2004.
- [28] J. Leskovec, L. Adamic, B. Huberman, "The dynamics of viral marketing", *7th ACM Conference on Electronic Commerce*, pp. 228-237, 2006.
- [29] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjee, A. Nanavti, A. Joshi, Social ties and their relevance to churn in mobile telecom networks, EDBT (Nantes, France), 2008.
- [30] R. East, W. Lomax, R. Narain, Customer tenure, recommendation and switching, Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, 14, pp. 46-54, 2001.
- [31] G. Erétéo, M. Buffa, F. Gandon, P. Grohan, M. Leitzelman, P. Sander, "A State of the Art on Social Network Analysis and its Applications on a Semantic Web", SDOW2008 (Germany), 2008.
- [32] J. Goldenberg, B. Libai, E. Muller, "Talk of the Network: A Complex Systems Look at the Underlying Process of Word of Mouth", *Marketing Letters*, pp. 211-223, 2001.

14. ANEXOS

14.1 Gráfico de *Market share* en telefonía móvil Post pago año 2012.



Fuente: Elaboración Propia.

14.2 Matrices de confusión para el modelo Árbol de Decisión considerando las distintas proporciones y métodos de crecimiento en el conjunto de entrenamiento.

Proporción 50/50:

CHAID

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9143	104
Fugado	68	9179

CRT

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9156	91
Fugado	97	9150

QUEST

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9148	99
Fugado	69	9176

Proporción 60/40:

CHAID

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9162	85
Fugado	102	6062

CRT

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9202	45
Fugado	67	6097

Proporción 70/30:

CHAID

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9180	67
Fugado	102	3860

CRT

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9097	150
Fugado	89	3873

QUEST

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9126	121
Fugado	105	3857

Proporción 80/20:

CHAID

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9151	96
Fugado	108	2204

CRT

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9169	78
Fugado	145	2167

QUEST

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9169	78
Fugado	115	2197

Proporción 90/10:

CHAID

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9169	78
Fugado	105	922

CRT

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9150	97
Fugado	203	824

QUEST

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9145	102
Fugado	100	927

Proporción 95/05:

CHAID

Observado	Pronosticado	
	No fugado	Fugado
No fugado	8980	267
Fugado	64	423

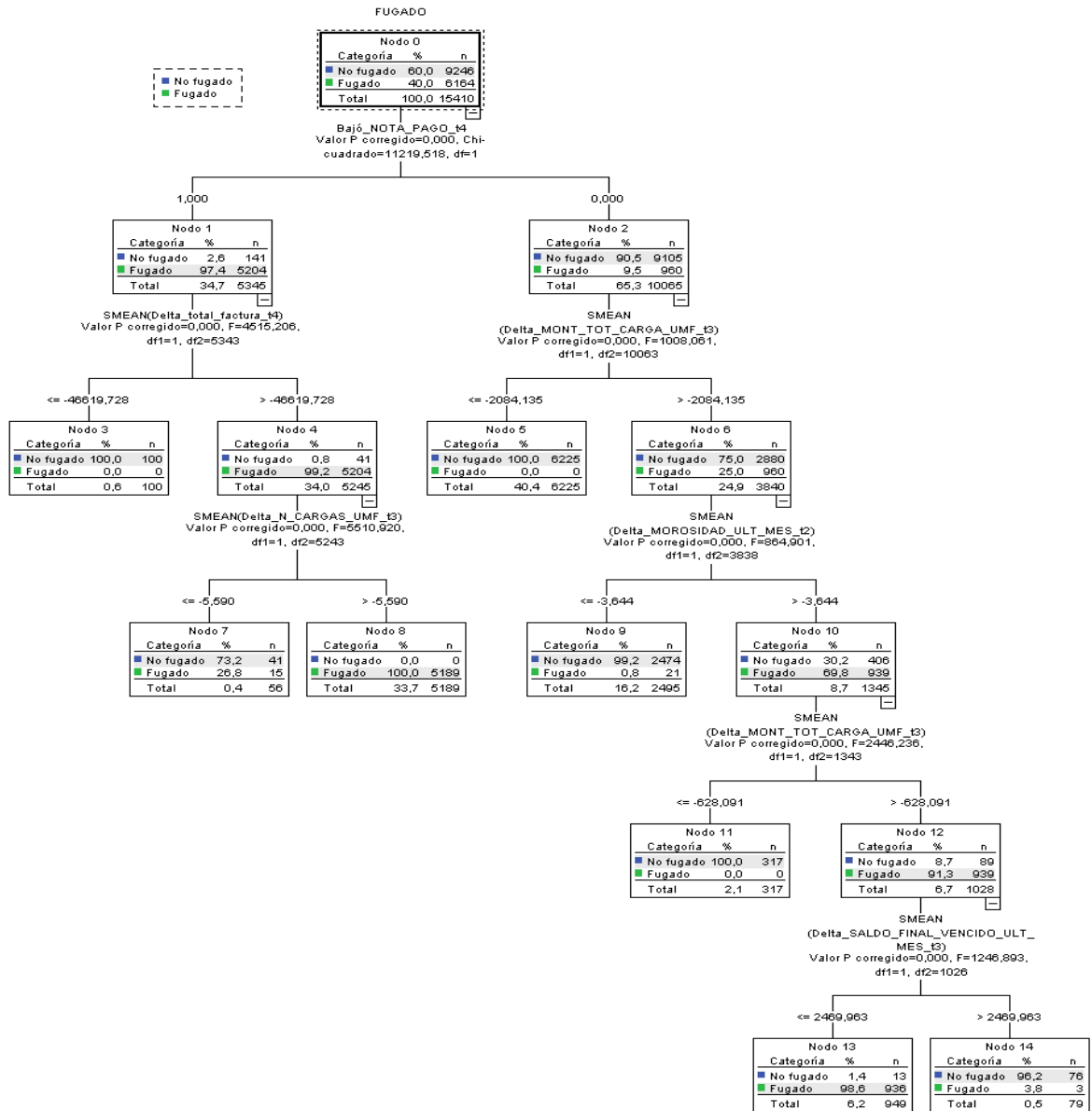
CRT

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9124	123
Fugado	99	388

QUEST

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9149	98
Fugado	75	412

14.3 Árbol de Decisión QUEST en conjunto de Testeo con proporción 60/40



14.4 Matrices de confusión para el modelo de Regresión considerando las distintas proporciones y puntos de corte en el conjunto de entrenamiento.

Proporción 50/50:

Punto de corte 0,5

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9105	141
Fugado	1482	7763

Punto de corte 0,7

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9105	141
Fugado	1514	7731

Punto de corte 0,9

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9105	141
Fugado	1587	7658

Proporción 60/40:

Punto de corte 0,5

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9105	141
Fugado	960	5204

Punto de corte 0,7

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9127	119
Fugado	831	5333

Punto de corte 0,9

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9109	137
Fugado	945	5219

Proporción 70/30:

Punto de corte 0,5

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9089	157
Fugado	643	3319

Punto de corte 0,7

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9101	145
Fugado	602	3360

Punto de corte 0,9

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9015	231
Fugado	634	3328

Proporción 80/20:

Punto de corte 0,5

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9109	137
Fugado	354	1958

Punto de corte 0,7

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9105	141
Fugado	344	1968

Punto de corte 0,9

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9069	177
Fugado	370	1942

Proporción 90/10:

Punto de corte 0,5

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9046	200
Fugado	245	782

Punto de corte 0,9

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9079	167
Fugado	192	835

Proporción 95/05:

Punto de corte 0,5

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9090	156
Fugado	189	298

Punto de corte 0,7

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9079	167
Fugado	178	309

Punto de corte 0,9

Observado	Pronosticado	
	No fugado	Fugado
No fugado	9001	245
Fugado	134	353