Prediciendo la duración de un cliente en una compañía de telecomunicaciones

Carlos Alberto Trujillo Moreno
Maestría en inteligencia analítica para toma de decisiones, Universidad de los Andes, Bogotá, Colombia.

Con el presente trabajo se busca aplicar el análisis de supervivencia dentro del contexto del calculo del valor de un cliente para una compañía de telecomunicaciones. El objetivo principal es predecir la duración de un cliente por medio de varias metodologías tradicionales como la metodología de Kaplan-Meier y Metodologías de Machine learning, en estas ultimas haciendo enfoque en los métodos basados en árboles como Survival Trees y Random Survival Forest. Se busca entender como la duración se ve afectada por variables relacionadas con el cliente y su experiencia en el marco de un contrato de suscripción pospago. Las metodologías serán implementadas con el software R.

Keywords: Survival Analysis, Customer Lifetime Value, Churn, Duración, Random Survival Forest, Survival Trees, COX Regression

INTRODUCCIÓN

En la actualidad la gestión de los clientes se ha convertido en uno de los principales enfoques de las compañías que manejan contratos de suscripción, sobretodo porque dichos contratos son los generadores de ingreso de las firmas y como es usual, el objetivo es mantener una cartera de clientes rentables que al final del ejercicio maximicen el valor de la empresa.

Dentro de las medidas de rentabilidad que implementan las empresas encontramos el Customer Lifetime Value, el cual evalúa cliente a cliente los ingresos y egresos que se perciben por cada uno y al final determina que tan rentable es el cliente en el tiempo. Esto ultimo es esencial en la medición de la rentabilidad del cliente ya que es imprescindible conocer cuanto va a durar el cliente para así establecer su valor. El objetivo de este trabajo es emplear varias metodologías analíticas para estimar dicha duración de la forma mas acertada posible.

En la primera parte del presente documento se abarcan conceptos esenciales para entender el contexto general del problema, se define el Customer Lifetime Value, el Churn (o fuga de clientes) y como se relaciona este con la supervivencia o duración de los suscriptores. En la segunda parte se desarrolla el concepto de análisis de supervivencia y como este es aplicable al contexto de una empresa de telecomunicaciones. En la tercera parte se profundiza en las metodologías disponibles para estimar la supervivencia.

En la cuarta parte se muestran los datos y las variables disponibles para la aplicación de metodologías. Seguido, en la quinta parte se discuten los resultados de los modelos aplicados, cual fue el modelo con mejores pronósticos y como se puede implementar en la compañía que proporcionó los datos. Por ultimo se muestran las conclusiones finales y próximos análisis a realizar.

MARCO TEÓRICO

Customer Lifetime Value

Dentro de las diferentes estrategias de mercadeo y experiencia de cliente en una compañía de telecomunicaciones, en las que se cuenta con un presupuesto limitado y en las que se enfrentan decisiones estratégicas que pueden generar repercusiones en la rentabilidad futura, se hace necesario establecer nuevas formas de evaluar la base de suscriptores; si bien es importante desarrollar estrategias de segmentación, se hace necesario entender si un cliente es rentable para la firma y de esta forma establecer las estrategias para gestionarlo, es por esto que se hace necesario calcular un indicador conocido como “Customer Lifetime Value (CLTV)”.

Una definición formal de CLTV es el valor presente neto de los réditos vinculados a un cliente en específico una vez ha sido adquirido, después de restar los costos asociados al mercadeo, venta, producción y servicio durante la vida del cliente [Blattberg et al., 2008b]. En otras palabras, es cuan rentable es un cliente considerando los costos que conlleva obtener y mantener dicho cliente durante un tiempo determinado. Es una visión financiera de un cliente, en la que se analiza a cada cliente como una inversión por aparte.

Cuando se analiza el CLTV, se analiza si la inversión realizada en el mismo va a ser recuperada en el futuro. Es por esto que CLTV es una metodología que permite determinar la rentabilidad de una compañía desde un punto de vista diferente al tradicional, en el que no se evalúan los flujos de fondos de la compañía en general, sino que se evalúan los flujos de fondo generados por cada uno de los clientes.

El cálculo de CLTV para una compañía de telecomunicaciones en las que es tradicional mantener una relación contractual con el cliente en la cual este periódicamente generará unos ingresos a la compañía por la suscripción que maneja permite establecer una metodología de cálculo específica. Según [Rosset et al., 2003] para calcular el CLTV de un cliente se deben tener en cuenta tres factores. El primero el valor del cliente, los réditos que se generaran de su relación con la firma. El segundo es la duración del modelo de servicio, la cual describe la duración de la relación con el cliente, comúnmente expresada a través de una función de supervivencia. En tercer y último lugar una tasa de descuento, la cual permite expresar una unidad de dinero que se percibirá en el futuro en términos de valor actual, según esta aproximación y si se tratara de términos continuos se puede definir el CLTV como:

$$CLTV = \int S(t)V(t)D(t)dt \tag{1}$$

En donde:
$S(t)$ =duración del cliente
$V(t)$ =réditos generados
$D(t)$ =tasa de descuento
En otras palabras y teniendo en cuenta el que los flujos de fondo percibidos por cada cliente se perciben durante periodos cortos o discretos en [Blattberg et al., 2008b] se plantea que la ecuación (1) se puede re-expresar como:

$$CLTV = \sum_{t=1}^{\infty} \frac{V_t}{(1 + D_t)^{S-1}} \tag{2}$$

Una vez se tiene calculado el CLTV para cada cliente, se puede utilizar esta medida para establecer varias estrategias dentro del ciclo de vida del cliente, la asignación de recursos para optimizar el CLTV y adicionalmente se pueden tomar decisiones relacionadas con los productos que se otorgan a los clientes, como el precio o canales de ventas para optimizar el CLTV.

Churn

Como es común en una empresa que presta servicios de suscripción, las compañías de telecomunicaciones se enfrentan a un problema conocido como fuga de clientes o más comúnmente conocido como Churn, es decir, la cantidad de clientes de la cartera que dan por finalizado el contrato por diferentes razones. El Churn también se puede definir como la probabilidad de que un cliente deje la firma dado un periodo de tiempo, pero para efectos de indicadores de compañía es el porcentaje de clientes de la cartera que se fueron en un periodo de tiempo [Blattberg et al., 2008a].

$$Churn = \frac{\text{Clientes que cancelaron}}{\text{Total de Clientes al inicio del periodo}} \tag{3}$$

En el contexto de las compañías de telecomunicaciones y específicamente en las compañías que prestan servicios bajo la modalidad de contrato, se suele hablar de dos tipos de Churn - que deben ser medidos para entender el impacto que generan - el primero, el Churn voluntario, se refiere a aquellos clientes que dan por cancelado el contrato por su propia voluntad, usualmente este tipo de Churn se gestiona desde las estrategias de retención de las que se comentará más adelante. Por otro lado, tenemos el Churn involuntario, es decir, aquellos clientes a los que la empresa les da por finalizado el contrato por no cumplir con los pagos mensuales que se establecen en el contrato, al igual que el Churn voluntario, el involuntario debe ser gestionado a través de cobranzas por diferentes métodos.

El Churn puede ser generado por diferentes razones y dependiendo de las mismas se suele clasificarlo para luego gestionarlo. Existen diferentes causas tanto exógenas como endógenas en el contexto en que la firma se desenvuelve; por ejemplo, dentro de las razones exógenas en las que el Churn voluntario se produce están las diferentes acciones de la competencia en el mercado y en especial en el mercado de las telecomunicaciones colombiano en donde su crecimiento se ha venido estancando durante los últimos 5 años como se puede observar en la Tabla 1; esto ha generado que las empresas de telecomunicaciones desarrollen estrategias en las que compiten con precios bajos entrando así en una lucha de suscriptores constante.

Año	Millones de Usuarios	Crecimiento
2012	58.1	
2013	56.6	-2.7 %
2014	61.4	8.5 %
2015	62.7	2.2 %
2016	63.8	1.8 %

Tabla 1
Usuarios del sector Telecomunicaciones en Colombia. Tomado de [MarketLine, 2017]

Como es claro también el Churn voluntario se genera por diferentes procesos que afectan la relación con el cliente, si estos procesos no satisfacen al mismo, se producirá en el tiempo un deseo de terminar la relación con la firma. Como lo menciona Khan [Khan, 2012] en mercados emergentes la lealtad y satisfacción juegan un papel crucial en el Churn.

Por otro lado, dentro de las razones exógenas que generan el Churn involuntario se encuentran la inflación y el desempeño de la economía, en Colombia, un país en el que según datos del banco de la república de Colombia el salario mínimo crece a un nivel más lento que los precios, se genera una necesidad del consumidor a ahorrar y priorizar gastos.

CLTV, Supervivencia y Churn

Para este momento nos encontramos en una problemática común en el cálculo del CLTV y es responder a la pregunta ¿Cuánto tiempo va a durar un cliente?

La duración de un cliente es de por si la única variable que no es conocida al momento de calcular el valor del mismo; los márgenes (los cuales ya fueron observados) y la tasa de descuento (la cual se puede calcular con metodologías como el CAPM [1]) son medidas conocidas por la empresa, pero saber de antemano cuánto tiempo el cliente va a generar dichos márgenes no es trivial. Estimar correctamente la duración de un cliente teniendo en cuenta el churn que se ha presentado y entendiendo que el tiempo o duración del cliente se convierte en la piedra angular del cálculo del CLTV y por ende debe ser estimada mediante procesos analíticos.

Según [Blattberg et al., 2008a] el Porcentaje de Churn es un indicador de la duración esperada o supervivencia de un cliente y nos permite establecer cuantos periodos de tiempo va a durar el cliente en la compañía generando réditos:

$$\text{Duración esperada del Cliente} = \frac{1}{Churn} \tag{4}$$

Esta medida es la forma más simple y sencilla de estimar la duración de un cliente, especialmente porque en la mayoría de las compañías de comunicaciones se cuenta con una medición juiciosa del churn mes a mes, pero a su vez presenta varias desventajas por los siguientes motivos:

La tasa de Churn es una medida calculable solo para grupos de clientes y no calculable para un solo cliente, además, en niveles estrictos y teniendo en cuenta que en una compañía de telecomunicaciones se tienen varios tipos de clientes, se debería calcular una tasa de Churn para un grupo homogéneo de clientes en sus características lo cual puede ser complejo de hacer en la mayoría de los casos de compañías de telecomunicaciones dada la poca información demográfica con la que cuentan y que en varias de ellas no se desarrollan procesos de segmentación juiciosos.

Otra desventaja de asumir que la Duración estimada de un cliente está en razón de la tasa de churn es que esta última no es constante en el tiempo para un grupo de clientes, ya que debido a las diferentes variables tanto exógenas como endógenas esta medida tiene a acelerarse o a disminuirse durante el tiempo.

Por otro lado, hay que tener en cuenta que esta medida es aplicable en un cliente nuevo, pero si se quiere calcular la duración estimada de un cliente que ya tiene una antigüedad con la empresa no es correcto utilizar esta medida ya que en esta no está involucrada la duración actual del cliente.

Es por lo anterior que se hace necesario utilizar una metodología más robusta para el cálculo de la duración estimada del cliente teniendo en cuenta las variables y condiciones mencionadas anteriormente. Para ello se plantea usar el Análisis de Supervivencia.

ANÁLISIS DE SUPERVIVENCIA

En [Kleinbaum and Klein, 2012] encontramos una definición clara del análisis de supervivencia: "análisis de supervivencia es un conjunto de métodos estadísticos para el análisis de datos en los que la variable de interés a ser estudiada es el tiempo hasta que ocurre un evento". Dado lo anterior, lo que se busca al desarrollar una metodología de análisis de supervivencia es entender cómo se comporta un evento determinado a través del tiempo y como se ve dicho tiempo afectado por diferentes variables. En esta definición se engloban dos conceptos interesantes para hacer profundidad; el tiempo, que para este caso se puede expresar de acuerdo a la medición que se esté realizando, se puede hablar de años, meses, semanas, etc. Por otro lado se habla de un "evento", en general en los análisis de supervivencia el evento que se busca observar es la muerte y de ahí el nombre de "supervivencia" pero el análisis puede ser desarrollado para cualquier tipo de eventos, que para el caso del presente documento es que el cliente deje de ser cliente (haga churn) de una compañía de telecomunicaciones, el cual para este análisis es un único evento por cliente y se asume que no es un evento que se pueda presentar varias veces para el mismo individuo.

Censura

Como el objetivo del análisis de supervivencia es estimar el tiempo entre el inicio de un estudio y la ocurrencia de un evento determinado [Wang et al., 2017], se debe abarcar un concepto muy importante en este tipo de análisis: La censura.

En la mayoría de los análisis de supervivencia, la ventana de observación y de medición del tiempo que pasa hasta que ocurre el evento es limitada, es decir, tiene un inicio y un fin, este hecho genera la presencia de un problema conocido como Censura, en otras palabras, por el hecho de ser una ventana de estudio limitada "no conocemos con exactitud el tiempo de supervivencia de todos los individuos que hacen parte del estudio" [Kleinbaum and Klein, 2012]. En general, la censura puede presentarse por tres motivos:

- La ventana de observación termina: Como se mencionó anteriormente, el tiempo durante el cual se estudia el evento no es ilimitado, por ende, cuando se termina dicha ventana de observación varios de los individuos aún siguen vivos. En este caso, el tiempo de supervivencia se asume como el tiempo que paso desde que se inició la observación del individuo y el fin de la ventana de observación. Para la aplicación del caso en una compañía de telecomunicaciones se va a presentar esta situación ya que a la fecha en que se generan los datos (fin de la ventana de observación) se tienen aún clientes que siguen siendo clientes.

[1] Capital Assets Pricing Model. Sharpe, William F. (1964)

- Se pierde el seguimiento de un individuo: Cuando por alguna razón no se puede seguir haciendo seguimiento al individuo se genera censura, ya que aunque la ventana de observación siga vigente, no se conoce si el individuo presentó o no el evento. En el caso de telecomunicaciones no se presenta este tipo de problemas ya que no se presentará el caso en que no se sepa si un cliente sigue siéndolo o no.
- Retiro: Cuando el individuo se retira del experimento y no se conoce si posteriormente presento el evento o no. Para nuestro caso este motivo no aplica ya que nuestro evento es cuando deje de ser cliente, el cliente no presenta "retiros".

A su vez, cuando se habla de Censura, se puede hablar de 2 tipos de censura:

Censura a la derecha, la cual es la más común y se presenta cuanto el tiempo de supervivencia real es mayor al tiempo de supervivencia observado dentro de la venta de observación. Esto ocurre siempre que el individuo continúe vivo una vez ha terminado la ventana de observación del experimento o cuando se pierde el seguimiento del mismo durante el experimento.

Por otro lado, la Censura a la izquierda se da cuando el tiempo de supervivencia real del individuo es menor al observado en el experimento. Este tipo de censura es muy inusual e implicaría que el experimento se inicie con individuos que ya presentan el evento de interés (sobre todo en experimentos relacionados con la adquisición de algún tipo de enfermedad).

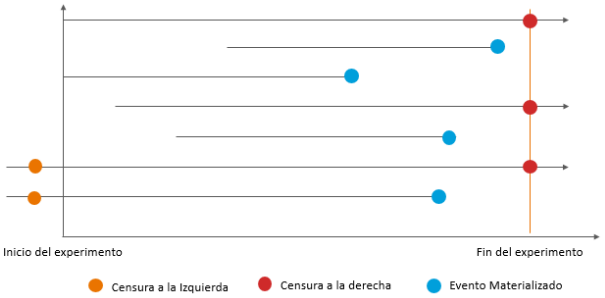

Figura 1. Tipos de Censura. Elaboración Propia.

En [Wang et al., 2017] se hace un planteamiento formal del problema dentro del contexto del análisis de supervivencia:

Para determinado individuo i, el cual se puede caracterizar por (X_i, y_i, δ_i) donde $X_i \in \mathbb{R}^{1\times P}$ es el vector de características del individuo i; δ_i es la variable binaria que indica si el evento es censurado ($\delta_i = 1$) o no censurado ($\delta_i = 0$); y y_i es el tiempo observado desde el inicio de la ventana de observación hasta el fin de la misma y es igual al tiempo de supervivencia T_i para un individuo sin censura y C_i para un individuo con censura.

$$y_i = \begin{bmatrix} T_i & si & \delta_i = 1 \\ C_i & si & \delta_i = 0 \end{bmatrix} \tag{5}$$

Función de Supervivencia

La función de supervivencia es aquella que sirve para representar la probabilidad de que el tiempo en que se presenta un evento no es antes de un tiempo especificado [Lee and Wang, 2013] y está definida por la siguiente expresión:

$$S(t) = Pr(T \geq t) \tag{6}$$

La función de supervivencia decrece a medida que pasa el tiempo t y siempre $S_{(t=0)} = 1$ dado que al inicio de la ventana de observación el sujeto está vivo, como se puede ver en la Figura 2 en la curva teórica. En la práctica dicha curva no es continua sino escalonada debido a que el tiempo se mide en unidades discretas (Años, meses, días, etc.).

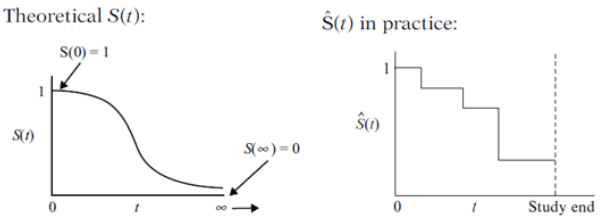

Figura 2. Forma de la curva de supervivencia. Tomada de [Kleinbaum and Klein, 2012]

Por otro lado, la función acumulada de muerte $F(t)$ representa la probabilidad de que el evento se presente antes del tiempo t y está definida como:

$$F(t) = 1 - S(t) \tag{7}$$

Por ende, la función de densidad puede ser obtenida aplicando la derivada de la función anterior (para casos continuos) o calculando la variación de $F(t)$ en los diferentes intervalos (para casos discretos, es decir, cuando Δt es pequeño)

Función Hazard o Función de Riesgo

Otra de las funciones necesarias para el análisis de supervivencia es la Función Hazard $h(t)$ la cual es útil para determinar la tasa de ocurrencia del evento de interés en un momento determinado dado que antes no se había presentado dicho evento. También conocida como tasa de sobrevivencia condicional [Kleinbaum and Klein, 2012] debido a que se trata de una probabilidad condicional:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \tag{8}$$

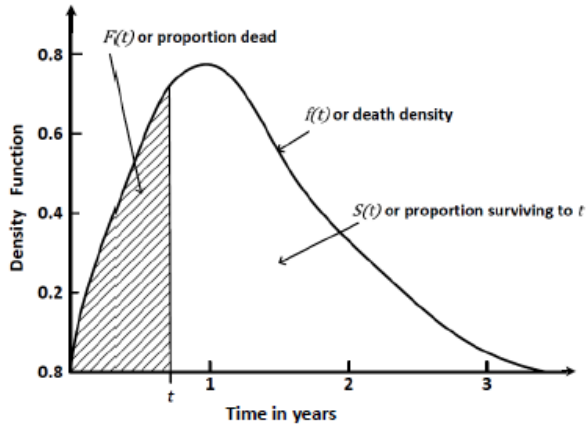

Figura 3. Funciones de Densidad de Supervivencia. Tomada de [Wang et al., 2017]

En la ecuación anterior se muestra como el tiempo de supervivencia de un sujeto T se va a ubicar en un intervalo de tiempo entre t y $t + \delta t$ dado que el tiempo de supervivencia T es mayor o igual a t

En [Godoy, 2009] se define a esta función como una medida de propensión a falla como una función de la edad del individuo relacionándola como la probabilidad de que la edad sea interrumpida por el evento de interés. Esta función describe la forma en que cambia la tasa de ocurrencia del evento de interés y por ende su única restricción es la no negatividad $h(t) >= 0$ y a su vez puede crecer o decrecer permanecer constante, ver Figura 4.

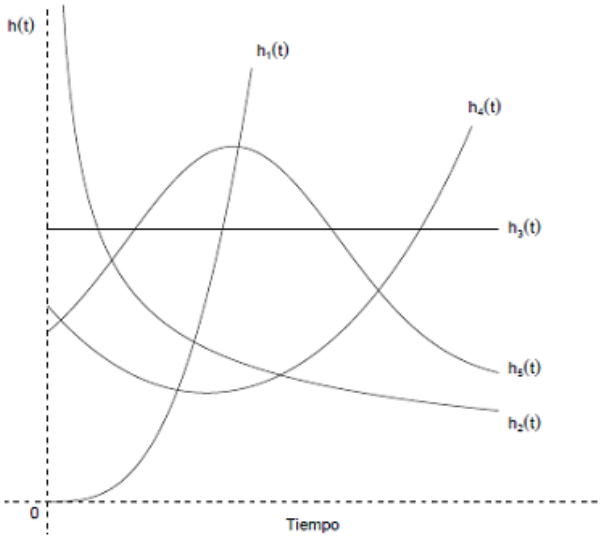

Figura 4. Posibles formas de la Función Hazard. Tomada de [Godoy, 2009]

Es importante tener en cuenta que la función Hazard se debe interpretar mejor como una tasa y no como una probabilidad ya que como se vio en la expresión anterior, se divide la probabilidad por un δt el cual expresa un intervalo de tiempo, por ende, la función Hazard expresa una probabilidad por unidad de tiempo.

La función Hazard sirve para determinar también si la función de supervivencia se ajusta a un modelo paramétrico ya que con la forma de la función Hazard se pueden prever algunas distribuciones presentes en la supervivencia del sujeto, en la Figura 5 se pueden notar algunos ejemplos:

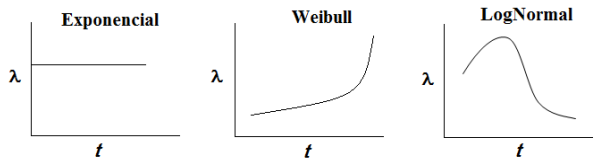

Figura 5. Algunas distribuciones presentes en la función Hazard. Tomada de [Kleinbaum and Klein, 2012]

La función Hazard también es muy útil ya que se puede utilizar para calcular la función de supervivencia S(t) pero para ello se debe calcular primero la función Hazard acumulada $H(t)$:

$$H(t) = \int_0^t h(u)du = -ln(S(t)) \tag{9}$$

Como se puede observar, la función de supervivencia y la función Hazard están relacionadas. La función de supervivencia se puede obtener si se calcula la función Hazard acumulada.

$$S(t) = exp(-H(t)) \tag{10}$$

METODOLOGÍAS PARA EL ANÁLISIS DE SUPERVIVENCIA

En [Wang et al., 2017] se hace una revisión de las diferentes metodologías estadísticas tradicionales. Según los autores existen dos grandes ramas de estudio para el análisis de supervivencia: Los modelos estadísticos tradicionales y los modelos de Machine Learning. En el presente trabajo, se busca utilizar dos métodos estadísticos, uno no paramétrico y uno semi-paramétrico (Kaplan-Meier y Regresión de Cox) y utilizar dos métodos de Machine Learning (Survival trees, Random Survival Forest) para estimar la duración de los clientes de una compañía de telecomunicaciones.

En el esquema representado en la Figura 6 se hace un resumen de las metodologías aplicables en análisis de supervivencia.

Los modelos estadísticos tradicionales

Se refieren a aquellos procedimientos que buscan estimar el las funciones de supervivencia y la función Hazard, por lo general estos modelos están diseñados para datos de pequeñas dimensiones y en general son utilizados en los campos

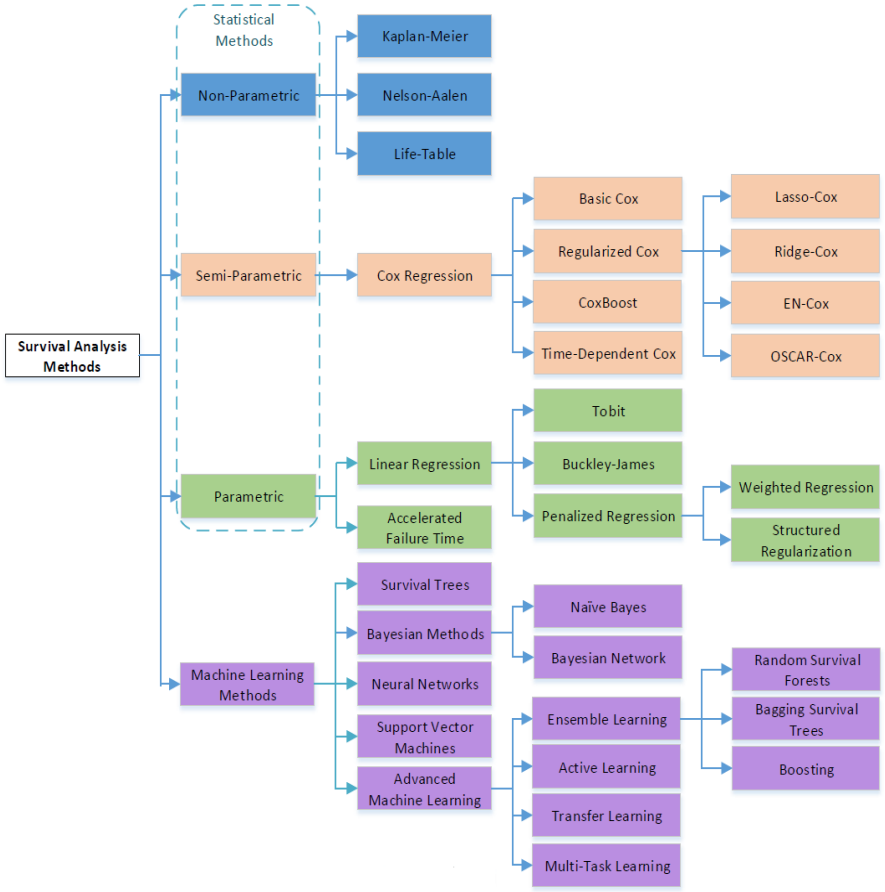

Figura 6. Metodologías de Supervivencia. Tomada de [Wang et al., 2017]

de la salud y análisis médicos. Dentro de los modelos tradicionales que se plantean en dicha investigación, se muestran tres tipos de modelos: Los no paramétricos, los semi-paramétricos y los paramétricos.

- Métodos Tradicionales no paramétricos: Se realiza una estimación empírica de la función de supervivencia y no se conoce bajo que distribución se desenvuelve la variable objetivo. La metodología más utilizada es la de Kaplan-Meier en la que se estima la función de supervivencia como una probabilidad condicional dado que el individuo ha sobrevivido hasta determinado tiempo. En estas metodologías también se estima la función de Riesgo (Hazard) con el fin de establecer en que momentos el individuo tiene mayores probabilidades a morir.

- Métodos tradicionales Semi-paramétricos: La función de distribución de los datos sigue siendo desconocida. La metodología más común es la de Regresión de riesgos proporcionales de Cox, en la que se construye función de riesgo y a esta se asocia a una estimación de parámetros que la afectan (regresión) para estimar cuales covariables afectan en mayor o menor medida el cambio de la función de riesgo.

- Métodos Tradicionales paramétricos: En estas metodologías la función de distribución de la variable objetivo se conoce a priori, lo cual los hace no tan comunes ya que en la mayoría de ocasiones es difícil establecer dicha función. En los casos en los que se conoce, se hacen los métodos ideales para estimar la supervivencia.

Metodología 1: Estimador de Kaplan-Meier. Este es uno de los métodos no paramétricos de estimación de supervivencia y es también conocido como el Método de Producto Limite; fue propuesto por [Kaplan and Meier, 1958] en su investigación titulada “Nonparametric estimation from incomplete observations”. Este método calcula las curvas de supervivencia basándose en la cantidad de eventos de interés que se presentan en un intervalo de tiempo observado y calculando las probabilidades condicionales entre dichos intervalos. Según la interpretación que realiza [Godoy, 2009] para determinar el estimador Kaplan-Meier de una función de supervivencia de una muestra que contiene datos censurados

por la derecha, se forma una serie de intervalos de tiempo, donde cada intervalo contiene un tiempo de falla (se presenta el evento de interés) y se considera que esta falla sucede al inicio del intervalo.

Entonces y basándonos en la notación de la investigación de [Wang et al., 2017] Se tiene $T_1 < T_2 < T_3 < \cdots < T_K$ como un conjunto de momentos en los que se presentó un evento para $N(K < N)$ sujetos. Dichos momentos, que se suponen independientes, pueden presentar censura ya que para algunos sujetos no se presentó el evento durante el periodo de observación. Para un momento especifico $T_{j(j=1,2,3,\cdots,K)}$ se tendrá que d_j es el numero de sujetos que presentaron el evento entre T_{j-1} y T_j, todos aquellos sujetos r_j para los que el momento del observación del evento o de censura sea mayor o igual a T_j serán considerados en riesgo a partir de T_j, es decir, $r_j = r_{j-1} - d_{j-1} - c_{j-1}$ (c_{j-1} son todos los sujetos que presentan censura durante T_{j-1} y T_j). Por consiguiente, la probabilidad condicional de sobrevivir mas allá del periodo T_j se puede definir como:

$$p(T_j) = \frac{(r_j - d_j)}{r_j} \tag{11}$$

En donde r_j son todos los sujetos vivos en el momento T_j, y d_j es la cantidad de sujetos que presento el evento de interés entre $T_(j-1)$ y T_j. Dada esta probabilidad condicional, el producto-limite que estima la función de supervivencia $S(t)$ esta definida por:

$$S(t) = \prod_{T_j<t}^{j} p(T_j) = \prod_{T_j<t}^{j} (1 - \frac{d_j}{r_j}) \tag{12}$$

Con esta estimación se pueden construir una tabla de supervivencia que muestre para cada intervalo de tiempo la probabilidad condicional de supervivencia, en la Tabla 2 se puede ver la generalización. Con la columna $S(t)$ de dicha tabla se procede a realizar un gráfico de la curva de supervivencia, un ejemplo gráfico se puede ver en la Figura 7.

T_j	r_j	d_j	c_j	$S(t)$
T_0	r_0	0	c_0	$S(t_0) = \frac{r-0}{r} = 1$
T_1	$r - c_0$	d_1	c_1	$S(t_1) = S(t_0) * \frac{r-c_0-d_1}{r-c_0}$
T_2	$r - c_1$	d_2	c_2	$S(t_2) = S(t_1) * \frac{r-c_1-d_2}{r-c_1}$
T_3	$r - c_2$	d_3	c_3	$S(t_3) = S(t_2) * \frac{r-c_2-d_3}{r-c_2}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
T_k	$r - c_{k-1}$	d_k	$_k$	$S(t_k) = S(t_{K-1}) * \frac{r-c_{k-1}-d_{k-1}}{r-c_{k-2}}$

Tabla 2
Ejemplo tabla supervivencia estimador Kaplan-Meier

Con esta función obtenida se pueden estimar las diferentes probabilidades de supervivencia de un sujeto en un momento determinado.

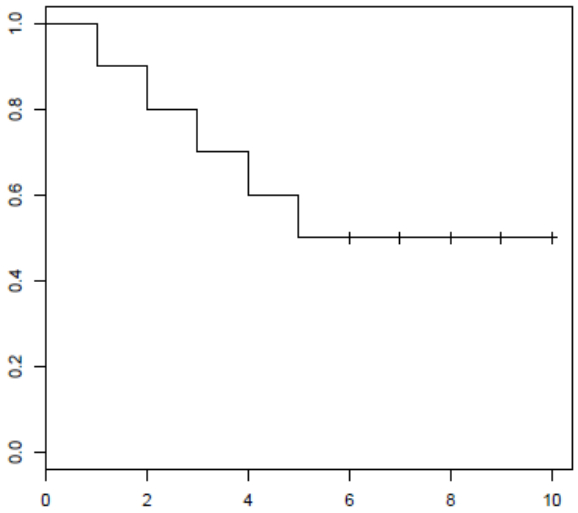

Figura 7. Ejemplo curva de supervivencia obtenida con Kaplan-Meier. Tomada de [Godoy, 2009]

Metodología 2: Riesgos proporcionales de COX. Este modelo propuesto en 1972 por David R. Cox es un modelo del tipo semi-paramétrico y es uno de los mas comunes a la hora de estimar el tiempo hasta que suceda un evento [Wang et al., 2017]. En este modelo no es requerido el conocimiento a priori de un a distribución de probabilidad que describa la variable objetivo (duración) y tiene la ventaja de que establece el impacto de covariables en el comportamiento del resultado. Su principal supuesto radica en que el riesgo de presentar el evento de interés para un individuo de un grupo es proporcional al riesgo de un individuo de otro grupo en el mismo lapso de tiempo [Godoy, 2009].

Para un sujeto i al que se le asocia una serie de covariables X_i un tiempo de supervivencia y_i y una descripción de censura δ_i, la función Hazard $h(t, X_i)$ el modelo sigue el supuesto de proporcionalidad de riesgo dado por:

$$h(t, X_i) = h_0(t) exp(X_i \beta) \tag{13}$$

para $i = 1, 2, 3, \cdots, N$ en donde $h_o(t)$ es la función Hazard base y puede ser una función de tiempo arbitraria, $X_i = (x_{i1}, x_{i2}, x_{i3}, \cdots, x_{iP})$ es el vector de covariables correspondiente al individuo i y $\beta^T = (\beta_1, \beta_2, \beta_3, \cdots, \beta_P)$ es el vector de coeficientes que determinan el impacto de las covariables en la función de riesgo base. El modelo de riesgos proporcionales se define como semi-paramétrico ya que la función Hazard base no tiene que ser definida ya que si se comparan dos instancias esta se hará irrelevante:

$$\frac{h(t, X_1)}{h(t, X_1)} = \frac{h_0(t) exp(X_1 \beta)}{h_0(t) exp(X_2 \beta)} = exp[(X_1 - X_2)\beta] \tag{14}$$

Lo anterior significa que el Hazard Ratio (Impacto de las covariables en el riesgo) no dependerá de la función Hazard

base. Si este Hazard Ratio es < 1 significa que el riesgo de que el individuo del grupo 1 experimente el evento de interés es menor comparado con el individuo del grupo 2 y si el Hazard Ratio es > 1 se concluye que el individuo del grupo 1 presenta más riesgo que el individuo del grupo 2.

El modelo de riesgos proporcionales lo es porque asume que el Hazard Ratio es constante y los individuos comparten la misma función Hazard base, por lo que se puede expresar la función de supervivencia $S(t)$ de la siguiente forma:

$$S(t) = exp(-H_0(t)exp(X\beta)) \tag{15}$$

$$S(t) = S_0(t)^{exp(X\beta)} \tag{16}$$

En don de $H_o(t)$ es la función Hazard base acumulada y $S_0(t) = exp(-H_o(t))$ representa la función de supervivencia base.

Los modelos de Machine Learning

Hacen referencia a aquellos procedimientos avanzados para predecir o clasificar la variable de estudio (en este caso supervivencia) estableciendo relaciones lineales y/o no lineales entre las diferentes variables que la explican. Estas metodologías son útiles además cuando se presentan conjuntos grandes de información tanto a nivel de individuos como a nivel de covariables.

Entre las metodologías más conocidas se encuentran las basadas en árboles: Survival Trees y Random Survival Forest, pero también se han implementado metodologías como Redes neuronales, Métodos Bayesianos y Support Vector Machines.

Metodología 3: Survival Trees. Los modelos basados en árboles fueron propuestos en 1963 por [Morgan and Sonquist, 1963] y se hicieron populares con el desarrollo del paradigma CART (Árbol de clasificación y regresión por sus siglas en ingles) de [Breiman et al., 1984]. Solo hasta que en [Gordon and Olshen, 1985] se planteó la primera aplicación de esta metodología en los problemas de supervivencia.

Los árboles de supervivencia o "Survival Trees" son una de las metodologías de machine learning que pueden ser aplicadas en problemas de estimación de duración de suceso de un evento, si bien es cierto que las metodologías semi-paramétricas como el modelo de riesgos proporcionales de COX son una de las más comunes en el desarrollo de estos problemas, en términos prácticos se enfrentan varias situaciones que conllevan al uso de metodologías más robustas; por ejemplo, el hecho de que en los métodos semiparamétricos en ocasiones el analista debe definir a priori algunas relaciones entre las covariables y la variable objetivo [Bou-Hamad et al., 2011], también, en ocasiones las metodologías de machine learning están diseñadas para el uso de grandes grupos de datos y ofrecen desempeños mejores bajo estas condiciones [Wang et al., 2017].

Los árboles de supervivencia determinan las relaciones existentes entre las covariables y por otro lado determinan los grupos de sujetos que tienen comportamientos similares en la duración de suceso del evento de interés. Adicionalmente, los árboles pueden ser fusionados para obtener modelos más robustos (por ejemplo: Random Survival Forest).

La metodología desarrollada en CART consiste en encontrar divisiones al espacio de covariables creando "nodos" o ramas del árbol con sujetos similares con respecto a la variable objetivo. Esta división se realiza minimizando una medida de "impureza" del modelo (Indicador de Gini o Entropía para variables categóricas y SSE para variables continuas). Las divisiones se realizan siempre con una sola de las covariables y en la primera división se realiza una búsqueda exhaustiva para determinar cuál de las covariables muestra mejor reducción de medida de impureza, luego se continúan realizando divisiones hasta encontrar el parámetro de parada el cual generalmente es una cantidad mínima de sujetos que debe tener un nodo terminal. Luego de esto se procede a realizar la "poda" de las ramas que no generan una separación significativa, este paso es clave ya que si no se realiza el modelo estará presentando overfitting.

La principal diferencia de los árboles de supervivencia y los árboles de clasificación tradicionales es la medida de impureza [Wang et al., 2017]. Los árboles de supervivencia no tienen una medida natural de impureza debido a la presencia de censura [Zhou and McArdle, 2015]. En Gordon y Olsen [Gordon and Olshen, 1985] se propone utilizar la medida Wasserstein, la cual mide la distancia entre dos funciones que para nuestro caso serán dos funciones de supervivencia obtenidas con el estimador Kaplan-Meier; con ello se busca minimizar la homogeneidad de las funciones de un mismo nodo y maximizar la heterogeneidad de funciones entre dos nodos diferentes. En [Ciampi et al., 1986] se propone usar el estadístico "Log-rank" utilizado para la comparación de dos funciones de supervivencia obtenidas en Kaplan-Meier, la partición será realizada cuando se maximice el valor de dicho estadístico. En [LeBlanc and Crowley, 1992] se propone usar una medida que reduzca al "desviación de un paso" en la generación de cada uno de los nodos, (según el autor, esta es una medida muy similar en comportamiento al uso del estadístico Log-rank). Este último estadístico de Split es el implementado en el paquete *rpart* [Therneau et al., 2017] el cual será utilizado en el presente trabajo.

Metodología 4: Random Survival Forest. Los árboles de decisión son conocidos por su facilidad de interpretación, pero a su vez por su inestabilidad a la hora de predecir sobre todo porque un pequeño cambio en los datos de entrenamiento puede conducir a un cambio en su poder de predicción [Bou-Hamad et al., 2011] generando altos niveles de varianza. [Breiman, 2001] propuso metodologías para solventar es-

te problema, entre ellas, Random Forest.

Esta metodología consiste en entrenar diferentes árboles de decisión con diferentes particiones, tanto en cantidad de sujetos como en cantidad de covariables, de la base original, proceso conocido como "Bootstrap", sin realizar ningún ejercicio de poda en las ramas; la predicción final se obtiene al promediar todas las predicciones de cada árbol individual. Para el caso del análisis de supervivencia, [Hothorn and Zeileis, 2016] utiliza la misma metodología y además para cada nodo terminal de cada árbol construye una función de supervivencia en la covariable X a partir del estimador Kaplan-Meier; para lograr dicha función utiliza un set de datos recolectados de cada árbol y de cada partición de la muestra utilizada para construir el mismo y que caen en el mismo nodo terminal para la covariable X; Esta aproximación será utilizada en el paquete de R *partykit* [Hothorn and Zeileis, 2016] utilizado en el presente trabajo.

USO Y ANÁLISIS DE INFORMACIÓN

Descripción de los datos

Se cuenta con la información de los clientes de una compañía de telecomunicaciones especializada en prestación de servicios de televisión por suscripción. La base se compone de 520.415 clientes activos al 1 de mayo de 2015 (inicio de la ventana de observación), dichos clientes son clientes de modalidad suscripción contractual pospago, es decir, que se factura mensualmente al cliente los servicios de televisión. Se considera como fin de la ventana de observación el 30 de septiembre de 2017 (fin de la ventana de observación).

Se recopiló información asociada a dichos clientes durante la ventana de observación y se cuenta con un total de 44 variables que fueron entregadas por la compañía para realizar el análisis, de estas variables 3 son variables relacionadas con la identificación del cliente (Código interno del cliente, la fecha de activación y la fecha de Churn) y 2 son variables objetivo (Duración y Flag de Censura) y 39 son varibles que se utilizaran como predictoras.

Dentro de las variables predictoras se encuentran algunas tipificadas como 'variables transaccionales', es decir, son variables relacionadas con la experiencia del cliente y por ende cambian de acuerdo a la ventana de medición que se defina dado un criterio de negocio; en este estudio y acordado con la compañía que proporciono los datos se define que dicha ventana de medición de variables debe ser de 6 meses y se tomará como punto de referencia la fecha en la que se presentó el evento de interés y si no se presentó el evento de interés se toman los 6 últimos meses antes del fin de la ventana de observación.

En el anexo 1 del presente documento se relaciona una tabla detallada de las variables con las que se cuenta.

Variables de Modelo

Evento de Interés: Churn. Durante la ventana de observación se determinaron que clientes presentaron el evento de interés, que para este caso, es que el cliente Cancele voluntariamente (Churn Voluntario) o se desconecte por no pago y le sean recogidos los decodificadores (Churn Involuntario). Para el presente análisis se asume que el cliente solo presenta una vez el evento de interés y dicho evento no puede ser repetitivo. Si el cliente decide volver a la compañía ingresa como un cliente nuevo, por lo que para el presente análisis no se asume este tipo de comportamiento dentro de los factores que predicen la duración del cliente

Variable Objetivo 1: Duración. La variable objetivo para el presente análisis es la variable "Duración". Esta variable se calculó como la diferencia en meses entre la fecha de activación del suscriptor y la fecha de Churn (para el cliente que presento el evento) o la fecha de fin de la ventana de observación (30 de septiembre de 2017). En la Figura 8 se muestra el histograma de frecuencias para la variable Duración

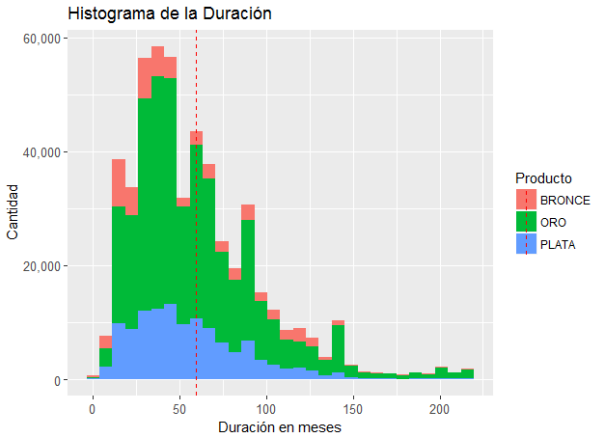

Figura 8. Histograma de la Variable Objetivo: Duración. Elaboración Propia.

La duración de los clientes tiene una media de 59,45 (línea punteada en la Figura 8), esto quiere decir que al final de la ventana de observación en promedio los clientes tienen una duración de 5 años aproximadamente. En la Figura 8 se distingue el producto que tiene el cliente (ORO, PLATA, BRONCE) y se puede ver que los clientes con el producto ORO tienen una duración mayor con respecto los demás, de modo contrario, los clientes PLATA tienen una duración notablemente menor. También se puede observar en el histograma como hay la presencia de una cola pesada en la distribución de la duración de los clientes sobre todo después del mes 150 y alcanzando una duración máxima de 217 meses.

Variable Objetivo 2: Censura. Esta variable determina si la duración del cliente está censurada o no. En este caso, la duración está censurada en el caso en el que el cliente terminó activo al final del periodo y no lo estará si el cliente pre-

sentó el evento de interés, es decir, hizo Churn. Para nuestro caso del total de clientes en la base proporcionada 58.82 % (306.122) se encuentran censuradas y el 41.18 % (214.293) presento el evento de interés.

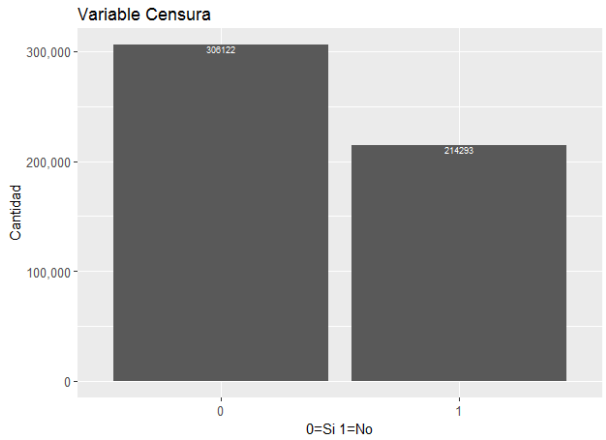

Figura 9. Participación de datos censurados en base de datos. Elaboración Propia.

Análisis de variables

De las 39 variables que entraron en los modelos para ser usadas como covariables se tienen 32 variables categóricas y 7 variables continuas. El primer paso para entender como se relacionan dichas variables con la variable objetivo 'Duración' (la cual es continua) se realizan gráficos BoxPlot para las variables categóricas con el fin de determinar si la distribución de la duración cambia con alguna de las categorías presentes en dichas variables; de igual forma se realizan histogramas para las variables continuas.

En la Figura 11 se hay una colección de gráficos BoxPlot para la mayoría de variables y hay algunos bordeados con una linea roja, esto indicando aquellas variables que en sus diferentes categorías muestran una distribución aparentemente diferente para la duración por lo que posiblemente sirvan como predictoras de la misma. Para empezar, se puede ver que en el BoxPlot No. 3 el canal de ventas aparentemente puede ser influyente para la duración, ya que los canales Directos, Dealers y Televentas tienen duraciones inferiores a canales como Alianzas y Otros. El BoxPlot No. 4 muestra que el tipo de suscripción también determina comportamientos diferentes en la duración, como se ve, los clientes 'Only TV' y 'Cross Selling' tienen duraciones superiores. La variable 'Convive ETB' representada en el BoxPlot No. 5 permite ver que la duración de un cliente cambia si tiene productos de dicho competidor. Variables descriptivas de la configuración del cliente como 'Producto' (BoxPlot No. 9), 'Tecnología' (BoxPlot No. 10) y 'Cantidad de decodificadores' (BoxPlot No. 11) tienen duraciones diferentes en cada categoría, por ejemplo los clientes con tecnología 'HD Only' y 'SD' tienen una duración inferior a los clientes con otras tecnologías y también los clientes con 3 o mas decodificadores tienen una duración promedio mayor que el resto de clientes; este comportamiento en estas variables tiene sentido desde un punto de vista de negocio ya que la configuración del producto es relevante para un cliente según investigaciones hechas por la compañía [2]. El análisis de los BoxPlots en la Figura 11 los clientes con productos como 'FOX+', ḤBO' y 'Adultos' (BoxPlot 14, 15 y 16 respectivamente), productos de categoría premium, tienen duraciones superiores a clientes que no tienen dichos productos.

Por último, algunas variables de experiencia como la del BoxPlot No. 20 'Presentó Intención', la del BoxPlot No. 22 'Presentó desconexiones' y la del BoxPlot No. 24 'Presentó Servicios Técnicos' muestran diferenciaciones claras con respecto a la duración del cliente, lo cual nuevamente hace sentido con la realidad de la compañía ya que dichas variables son generadoras del evento de interés (Churn). El histograma en el panel No. 17 representa el comportamiento de la variable 'Facturación Promedio' del cliente, si bien esta variable no es categórica puede influir en cuanto dura el cliente el valor de la factura que este paga ya que tiene que ver directamente con el comportamiento y la satisfacción del consumidor del producto.

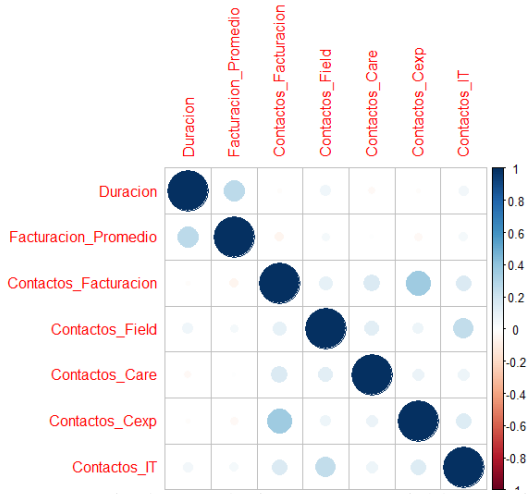

Figura 10. Matriz de correlaciones para variables continuas. Elaboración Propia.

Para las variables continuas se realizó una matriz de correlación para determinar cuales tienen correlación con la variable de duración. En la Figura 10 se puede observar como las variables de llamadas no tienen una relación fuerte con la variable objetivo pero aun así se decide incluirlas en los próximos modelos con el fin de determinar si efectivamente no tienen relación a la hora de determinar la duración de un cliente.

[2]Estas investigaciones no están documentadas públicamente pero se mencionan con autorización de los expertos de la compañía

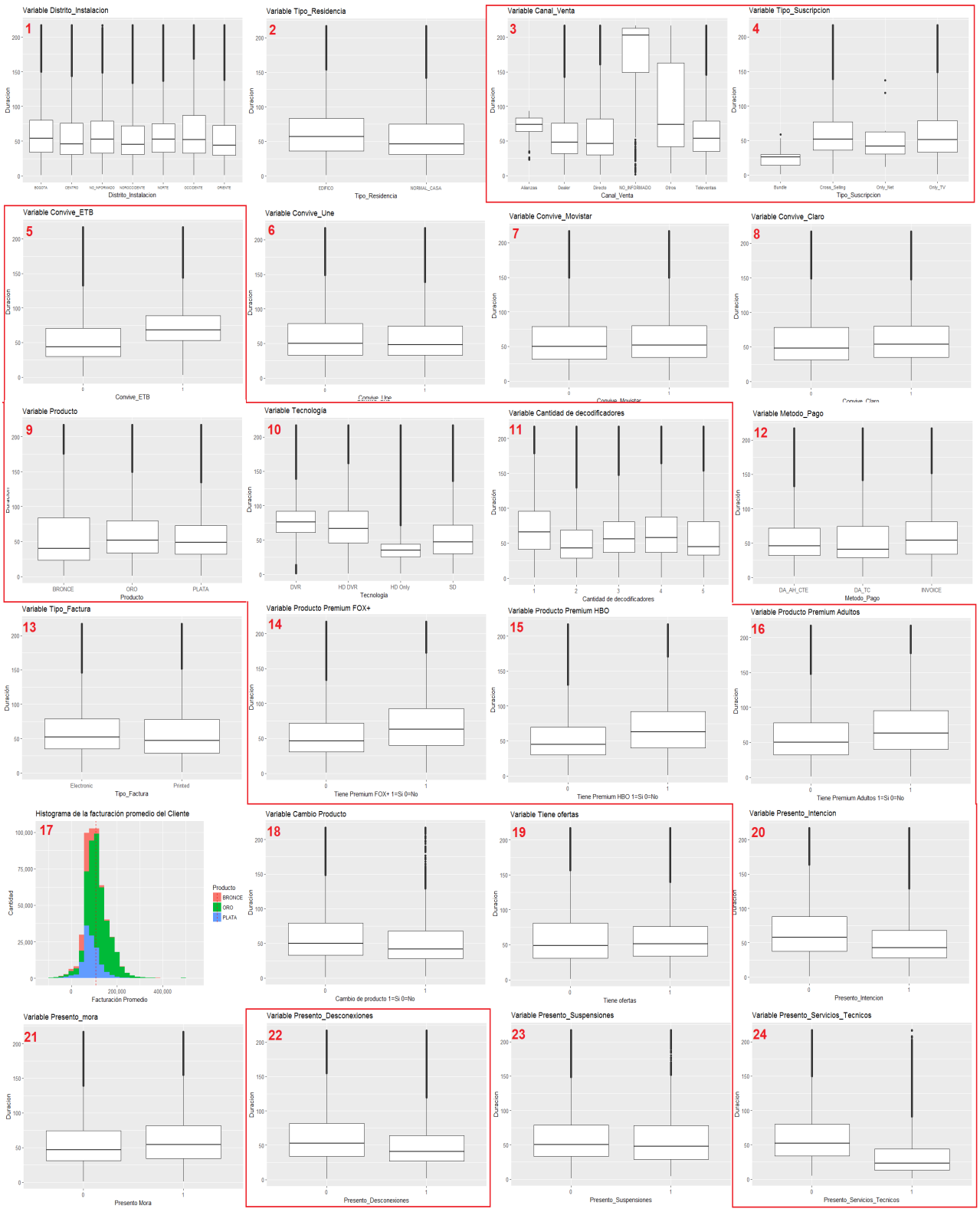

Figura 11. Boxplot de algunas de las variables proporcionadas por la compañía, resaltadas las variables que visualmente muestran una diferenciación

APLICACIÓN DE METODOLOGÍAS

Descripción de la Metodología

Como se mencionó en la introducción de este documento, el objetivo del presente trabajo es estimar la duración de un cliente de suscripción pospago en una compañía de telecomunicaciones especializada en servicios de televisión. Para ello se busca aplicar cuatro metodologías y compararlas entre ellas para así obtener la mejor estimación de la variable de interés. Todas estas metodologías serán implementadas en los datos mencionados anteriormente con la ayuda del paquete estadístico R.

Para los datos proporcionados se aplicaron metodologías estadísticas tradicionales: Kaplan-Meier y la Regresión de Cox y las metodologías de Machine learning: Survival Trees y Random Survival Forest. A continuación se detallan los resultados obtenidos para cada uno de las metodologías y los principales aspectos relacionados con las mismas.

Estimador Kapplan-Meier

Para la aplicación de esta metodología se utiliza el paquete *Survival* [Therneau, 2017] del software R con los datos mencionados anteriormente. Lo primero que se procede a realizar es la estimación de la curva de supervivencia para toda la base sin ningún tipo de estratificación. En la Figura 12 se muestra la curva estimada para la base de clientes de la compañía de telecomunicaciones en estudio. De esta curva se puede concluir que en los primeros 12 meses de duración hay una probabilidad de fuga de clientes no tan acelerada, pero una vez supera este punto dicha curva se acelera y de ahí en adelante la tasa de probabilidad de supervivencia se comporta relativamente constante pero debajo de la diagonal lo que se interpreta como una tasa de Churn acelerada. El detalle del estimador se encuentra en el apendice B de este documento.

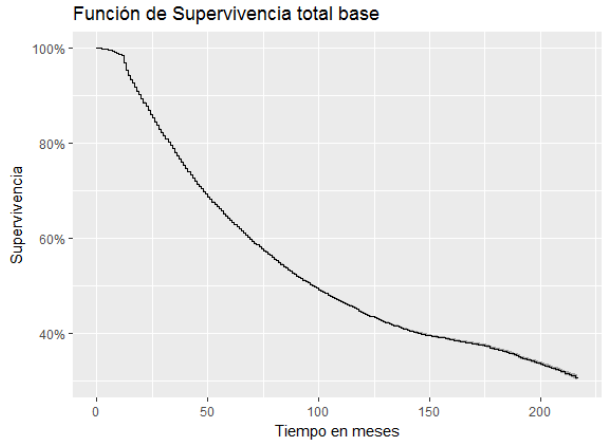

Figura 12. Curva de supervivencia para la base de clientes de una compañía de telecomunicaciones calculada con el estimador Kaplan-Meier. Elaboración Propia.

Para corroborar lo anterior y tener una mejor observación de los momentos de riesgo de la base de clientes, se procede a calcular y graficar la curva Hazard acumulada; esta nos permite observar en que momentos se presentan cambios en los riesgos de los clientes. En la Figura 13 se puede observar la curva Hazard acumulada para la base de clientes proporcionada, nuevamente se observa un cambio drástico en aproximadamente el mes 12, lo que confirma la teoría de que en dicho punto hay un cambio significativo en las probabilidades de supervivencia del cliente. Esta tendencia se mantiene un poco acelerada hasta el mes 150 donde dicha curva cambia mostrando que después de ese momento la probabilidad de supervivencia aumenta y la tasa de riesgo Hazard disminuye.

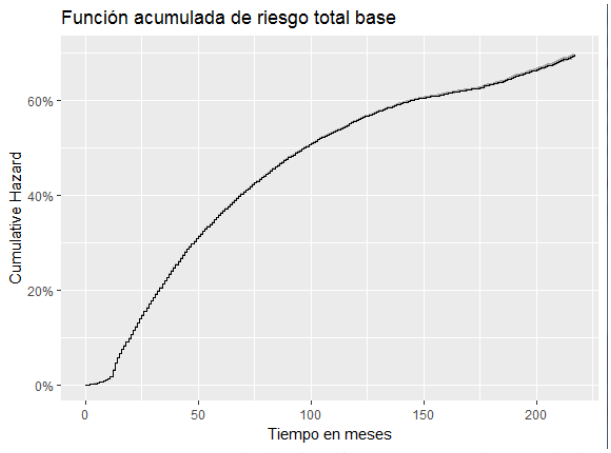

Figura 13. Curva acumulada de Riesgo (Hazard) para la base de clientes de una compañía de telecomunicaciones calculada con el estimador Kaplan-Meier. Elaboración Propia.

Una forma de inferir sobre que variables pueden afectar la supervivencia de los clientes se es calculando el estimador Kaplan-Meier y la curva de supervivencia pero realizando un split para cada categoría de determinada variable. Con la misma herramienta de R se procede a realizar el Split de la curva de supervivencia general para varias covariables proporcionadas por la entidad, estas curvas se encuentran en la Figura 14.

En la Figura 14A se realiza el Split con la variable Producto, como se puede observar, para el producto ORO se ve una diferenciación con respecto a los productos PLATA y BRONCE ya que tiene una curva de supervivencia superior lo cual parece indicar que dicha variable si afecta la duración de los clientes. En la Figura 14B se realizó el Split para la tecnología del decodificador asociada a dichos clientes, se pueden observar 4 curvas de supervivencia, pero aparentemente hay dos grupos, los clientes que tienen la tecnología DVR y HDDVR (aquellas que pueden grabar) vs los clientes que tienen las tecnologías SD y HDOnly (aquellas que no pueden grabar), otro indicio de que esta variable puede ser significativa a la hora de determinar impactos en la duración del cliente.

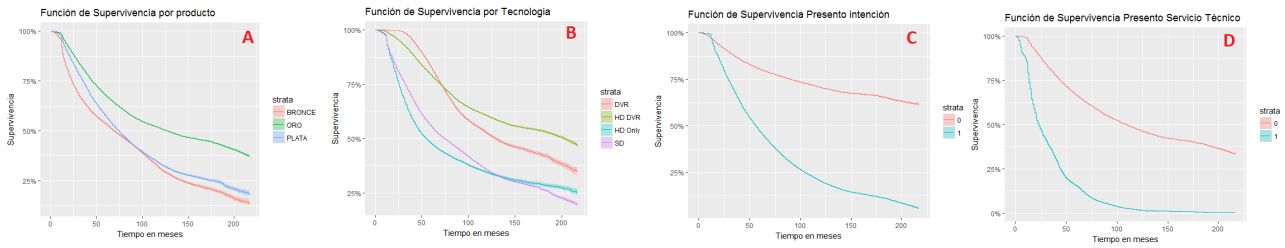

Figura 14. Curvas de supervivencia para la base de clientes de una compañía de telecomunicaciones calculada con el estimador Kaplan-Meier con diferentes splits con variables categóricas; A-Split con Producto, B-Split con Tecnología, C-Split con Presento Intención de cancelación, D-Split con Presento Servicio Técnico. Elaboración Propia.

También se decide realizar Splits con variables, que según el análisis de variables, pueden influir sobre la duración de un cliente, estas variables son si el cliente presento intención de cancelación en los últimos 6 meses o si el cliente solicito un servicio técnico en los últimos 6 meses. Para la primera en la Figura 14C es clara una diferenciación entre las dos curvas, se puede concluir que esta es una variable clave a la hora de determinar la duración de un cliente. En la Figura 14D se realiza el Split para servicios técnicos, también se observa una clara diferenciación en la curva de supervivencia, esta es otra variable que es determinante a la hora de calcular la duración de un cliente.

Regresión de COX

La regresión de COX o Modelo de riesgos proporcionales de COX es ejecutada con el paquete *Survival* [Therneau, 2017]. Para la correcta selección de variables se utiliza el método de penalización 'lasso' [3] el cual permite encontrar un penalizador λ para cada coeficiente β de la variable p obtenido por la regresión COX y así encontrar un modelo que optimice el uso de la información de todas las variables sin generar overfitting, el penalizador lasso esta definido por:

$$\lambda \sum_{P}^{p=1} |\beta_p| \tag{17}$$

Con la ayuda del paquete *glmnet* en R se puede obtener el parámetro λ óptimo para ser aplicado a los coeficientes. La Figura 16 muestra que luego de una aplicar 'cross validation' el parámetro ideal (el que minimiza el error) es $\lambda = 0{,}00104$ con este se prodece a estimar los coeficientes de la regresión de COX para obtener la curva de supervivencia estimada con esta metodología, en la Tabla 3 se pueden ver los valores de los coeficientes mas representativos en el conjunto de coeficientes de la regresión COX y en el apéndice 3 se puede encontrar la información completa para todas las variables.

De acuerdo a los resultados obtenidos se puede observar que las variables mas relevantes son la de Facturación Promedio con signo negativo, esto quiere decir, que el aumentar la facturación promedio una unidad reducirá el riesgo de

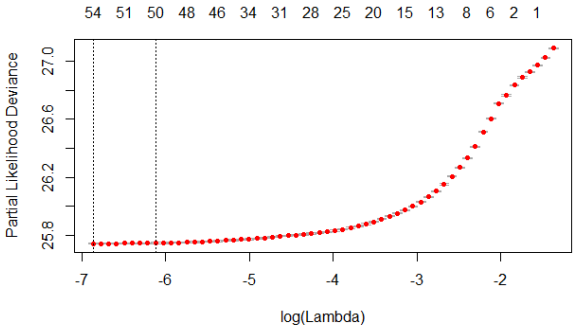

Figura 15. Resultados de la estimación de λ con cross validation. El lambda óptimo muestra 54 variables. Elaboración Propia.

Variable	Categ.	Cat. Ref.	Coef
Fact_Promedio			(2.2215)
Tipo_Suscripcion	Cross_Sell		(1.5065)
Tipo_Suscripcion	Only_Net	Bundle	(0.3783)
Tipo_Suscripcion	Only_TV		(1.1564)
Tecnologia	HD DVR		0.2066
Tecnologia	HD Only	DVR	1.0277
Tecnologia	SD		0.4554
Compro_Premium	Si	No	0.9403
Presento_Intencion	Si	No	1.5957
Presento_Retencion	Si	No	(1.1242)
Desconexiones	Si	No	0.6872
Servicios_Tecnicos	Si	No	1.0486

Tabla 3
Coeficientes mas relevantes generados en la regresión de COX con penalizador Lasso. Elaboración Propia.

falla del sujeto 2.22 veces si el resto de variables se mantiene constante; similarmente, la variable Tipo de suscripción muestra que el riesgo de falla del sujeto se reducirá si tiene un tipo de suscripción diferente a 'Bundle' manteniendo el resto de variables sin variaciones, esto porque todos los co-

[3] Least Absolut Shrinkage and Selection Operator

eficientes son negativos, ser 'Cross Sell' disminuye el riesgo en 1.50 veces, ser 'Only TV' lo reduce en 1.15 veces y 'Only Net' en 0.37 veces. La tecnología es otra de las variables relevantes, para este caso, si el cliente tiene una tecnología diferente a 'DVR' aumenta el riesgo del cliente pero quien mas riesgo tiene son los clientes 'HD Only' ya que su riesgo aumenta en 0.45 veces. Es importante ver como el hecho de que el cliente haya presentado una intención de cancelación en los ultimos 6 meses aumenta el riesgo 1.59 veces con respecto a los que no y si tuvieron un servicio técnico también aumenta el riesgo en 1.04 veces lo que indica que muestra que estas son variables muy relevantes a la hora de estimar la duración de un cliente.

Por ultimo se procede a estimar la curva de supervivencia con los coeficientes obtenidos para la base de clientes proporcionada por la compañía de telecomunicaciones, en la Figura 16 se puede observar dicha curva.

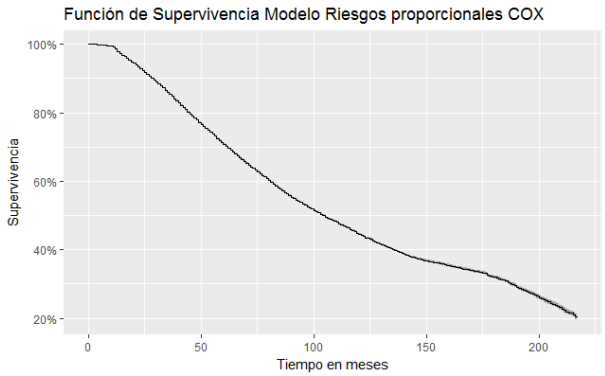

Figura 16. Curva de supervivencia para para la base de clientes de una compañía de telecomunicaciones calculada con regresión de COX. Elaboración Propia.

Survival Tree

Esta metodología es aplicada mediante el paquete *rpart* [Therneau et al., 2017]. Debido a que los árboles de decisión son flexibles en cuanto a la cantidad de variables que son incluidas para su estimación se decide incluir todas las variables predictoras en un primer paso para luego obtener un ranking de importancia de variables y calibrar el modelo.

La calibración se hace con un 'Parámetro de complejidad' (cp) implementado en R a través del paquete estadístico *rpart* [Therneau and Atkinson, 2017] y el cual funciona como parámetro de poda para el árbol, este parámetro esta definido como factor de penalización de la cantidad de nodos terminales $|T|$ el cual mide el 'costo' (Suma de error residual) $R(T)$ de añadir otra variable al modelo:

$$R_{cp}(T) = R(T) + cp\,|T| \tag{18}$$

Por lo anterior se ejecuta el árbol de decisión estableciendo $cp = 0$ y así lograr el máximo nivel de profundidad del árbol, es decir, un árbol saturado. En la Figura 17 se muestra el error relativo logrado con cada nodo nuevo.

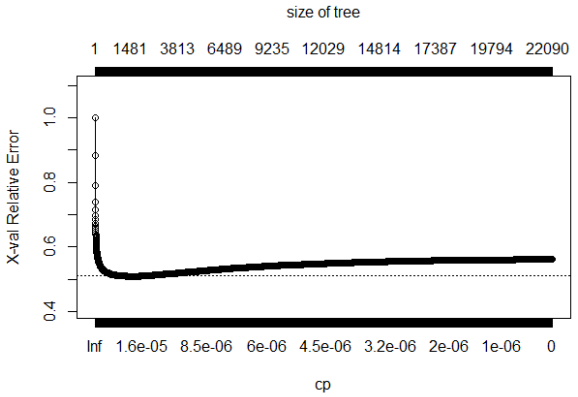

Figura 17. Curva de Errores relativos vs. el Parámetro de Complejidad (cp) para el Árbol de decisión saturado. La linea punteada corresponde al valor mínimo de error relativo logrado con el árbol . Elaboración Propia.

Con este proceso se logra determinar que el valor del parámetro de complejidad óptimo para reducir el error relativo es $cp = 0{,}003871$. Una vez determinado el cp óptimo para el calculo del árbol se procede a ejecutar nuevamente el árbol con dicho parámetro y así lograr el mejor árbol supervivencia calibrado para la base de una empresa de telecomunicaciones. En la Figura 18 se muestra como con el valor de cp establecido se logra alcanzar el error mínimo

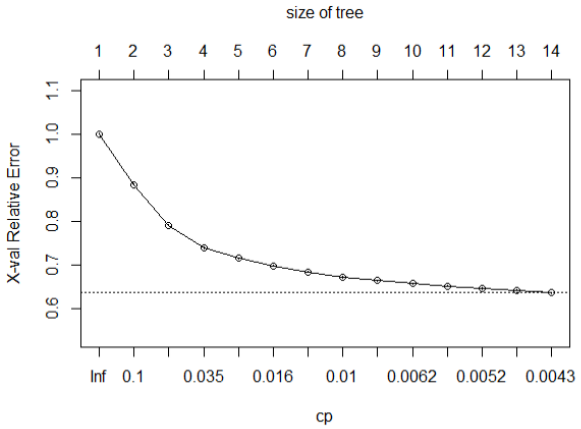

Figura 18. Curva de Errores relativos Vs. el Parámetro de Complejidad (cp) para el Árbol de decisión calibrado. La linea punteada corresponde al valor mínimo de error relativo logrado con el árbol . Elaboración Propia.

Con este nuevo árbol, la importancia de variables se recalcula, en la Figura 21 se muestran las variables mas importantes para el árbol calibrado, estas son si el cliente fue retenido, si presento una intención de cancelación, si tiene una oferta, la facturación promedio, estas variables tienen cierta concor-

dancia si se tienen en cuenta los resultados de los modelos tradicionales.

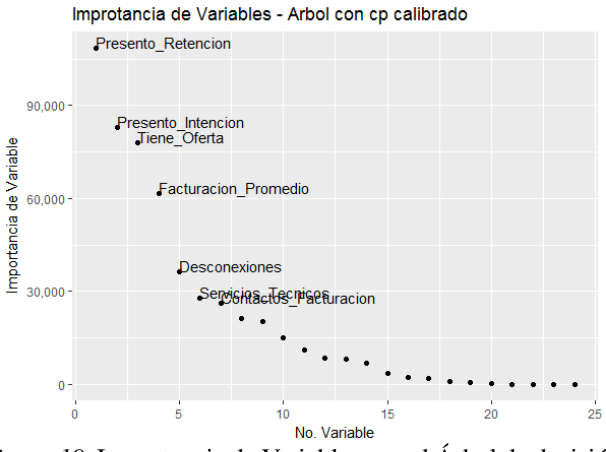

Figura 19. Importancia de Variables con el Árbol de decisión saturado. Elaboración Propia.

Por ultimo con la ayuda del paquete *partykit* [Hothorn and Zeileis, 2016] se procede a graficar el árbol de supervivencia calibrado. Ver la Figura 21; adicionalmente en la Figura 20 las curvas de supervivencia para cada uno de los nodos hoja logrados con el árbol calibrado

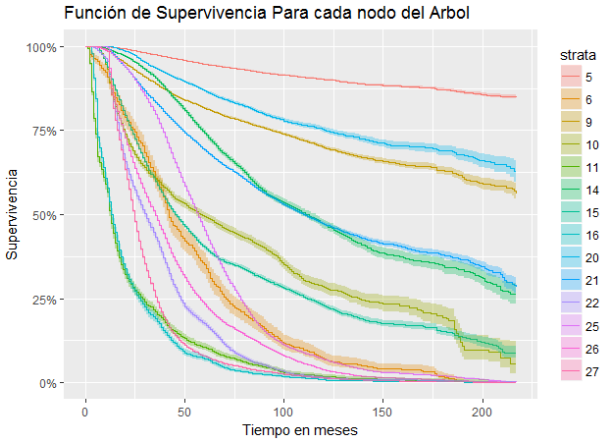

Figura 20. Curvas de supervivencia para la base de clientes de una compañía de telecomunicaciones obtenidas con el árbol de supervivencia. Elaboración Propia.

En la Tabla 4 se realiza una revisión de la duración promedio por cada uno de los nodos. Encontramos que aquellos con duración promedio mas alta son los nodos número 5, 14, 20, 9 y 25 en ese orden; por otro lado, los nodos con peores duraciones son el nodo 16, 11, 27, 22 y 26 y se puede ver como la media de la duración no necesariamente es un estimador correcto de la supervivencia de los clientes sobretodo porque hay nodos que son riesgosos según su curva de supervivencia pero su duración es alta, por ejemplo el nodo 25.

Nodo	Cantidad	Duración Promedio
5	135,921	77.3
6	824	53.9
9	73,131	62.1
10	6,664	48.7
11	3,234	22.6
14	13,787	70.9
15	30,330	43.2
16	3,410	21.6
20	33,015	65.2
21	90,382	58.3
22	7,393	36.8
25	38,622	59.9
26	46,878	43.0
27	36,824	29.2
Overall	520,415	59.5

Tabla 4
Duración promedio por nodo terminal en el árbol de supervivencia calibrado. Elaboración Propia.

Random Survival Forest

Esta metodología es aplicada con la ayuda del paquete *randomForestSRC* el cual nos permite calcular una cantidad determinada de árboles para estimar la curva de supervivencia de cada uno de los clientes con una mejor predicción ya que se promedian los resultados de varios árboles de supervivencia. Para nuestro caso se realiza la predicción de 100 árboles ya que con esta cantidad de árboles se llega al nivel de error mínimo, es decir, mas árboles no reducirían dicho nivel. En la Figura 22 se muestra el nivel de error frente a la cantidad de árboles.

Luego de ejecutado el proceso se obtiene la importancia de las variables la cual también es entregada por los árboles de supervivencia construidos en el Random Survival Forest, esta es una de las únicas salidas que ofrece el proceso pero que es clara para concluir sobre las variables que predicen la duración de un cliente.

Como se puede observar en la Figura 23 la variable mas importante en el Random Survival Forest es 'Presentó Intención' igual que en el árbol de Supervivencia construido anteriormente, es importante ver como los dos modelos tienen relativamente las mismas variables en orden de importancia. También es interesante ver como variables de competencia como 'Convive con Movistar' prácticamente no tiene nada que ver con la duración del cliente.

Comparación de modelos

Para comparar las metodologías se utiliza el paquete de R *PEC* el cual permite evaluar las estimaciones de los modelos de analisis de superviviencia implementados utilizando la medida conocida como Integrated Brier Score (IBR) la cual es un promedio ponderado de las distancias al cuadrado entre la función de supervivencia observada y la función de

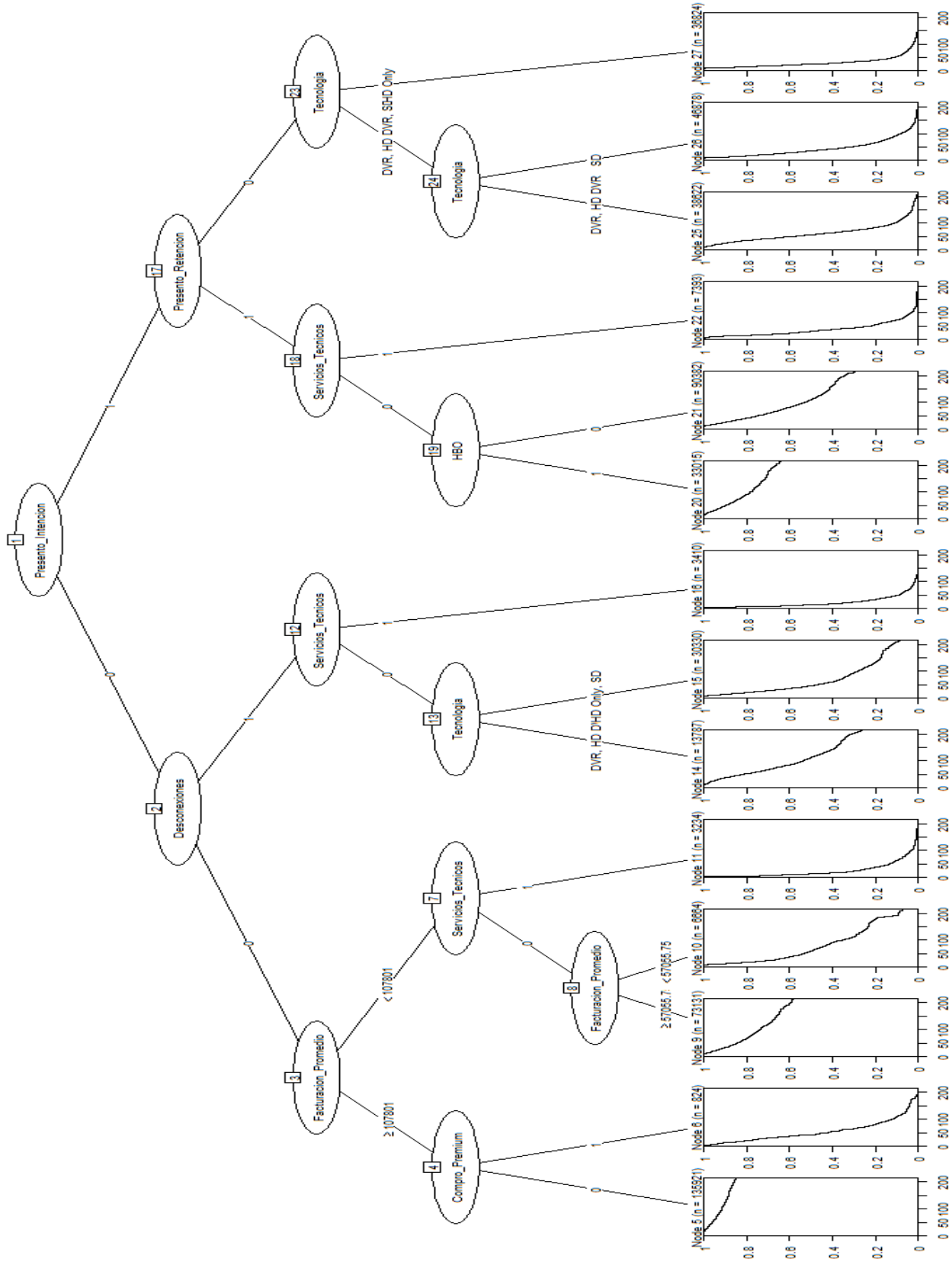

Figura 21. Árbol de decisión para la base de clientes de una compañía de telecomunicaciones. Elaboración Propia.

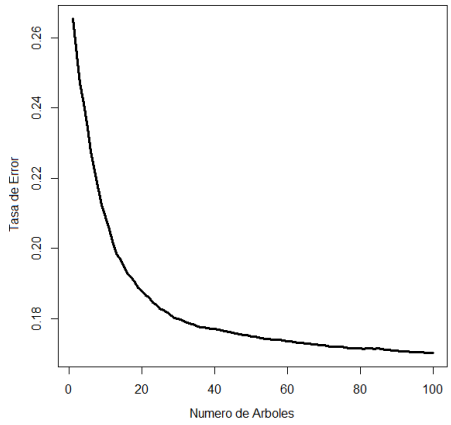

Figura 22. Tasa de error de Random Survival Forest frente a la cantidad de árboles. Elaboración Propia.

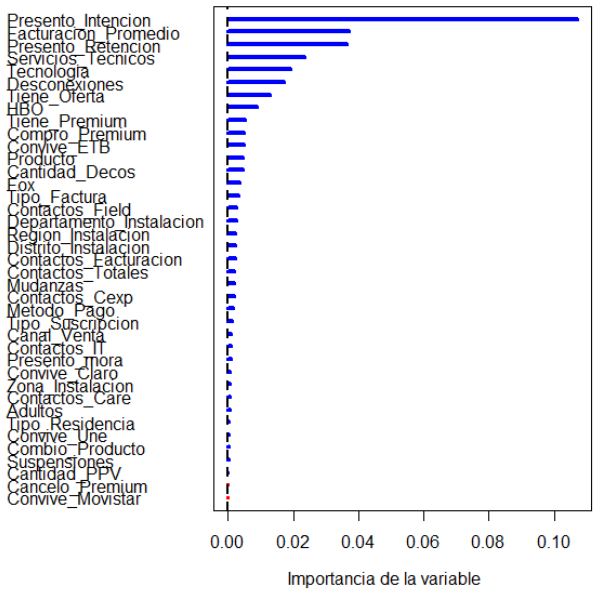

Figura 23. Importancia de variables en Random Survival Forest para una compañía de telecomunicaciones. Elaboración Propia.

supervivencia estimada. Este paquete utiliza metodologías de medición como Cross Validation o Bootstrap.

En nuestro caso se utilizará Bootstrap de 20 muestras seleccionadas sobre la muestra original y sobre ellas evaluar el rendimiento de la predicción de los modelos de Kaplan-Meier, Regresión de COX y Random Survival Forest.

En la figura 24 se puede observar la medición de las metodologías en cada uno de los momentos de la curva de supervivencia. Como se puede observar, Random Survival Forest presento un comportamiento mejor versus los modelos de Kaplan-Meier y Regresión de COX; en esta medición, el error de predicción va aumentando a medida que aumenta el tiempo, esto sucede debido a que la cantidad de clientes con censura aumentan a lo largo del estudio haciendo que la metodología tenga un error de predicción mayor, luego del mes 155 (aproximadamente) el error de predicción disminuye, esto se explica ya que la cantidad de clientes con duración superior a 155 meses es pequeña de acuerdo a la Figura 8.

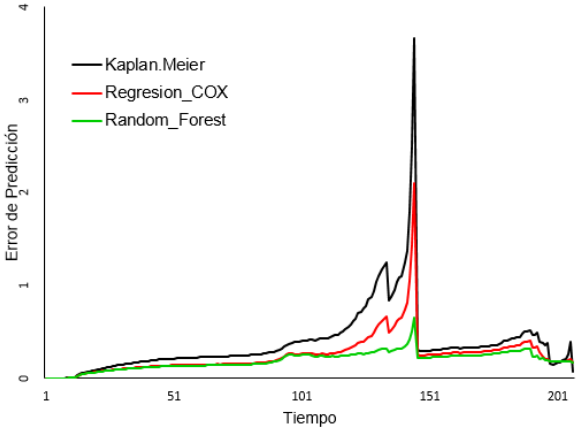

Figura 24. Rendimiento de la predicción para las metodologías aplicadas. Elaboración Propia.

CONCLUSIONES Y RECOMENDACIONES

Discusión de los Resultados

Dentro de los resultados obtenidos es importante ver como desde el análisis de variables hasta el desarrollo del modelo con mejores estimaciones (Random Survival Forest) las variables que son relevantes a la hora de estimar la duración tienden a ser las mismas, en la tabla 5 se resumen las variables mas importantes para cada una de las metodologías y puede ver que para este caso y de acuerdo con los datos proporcionados por la compañía de telecomunicaciones se obtuvo que las variables mas relevantes son: Si el cliente presentó intención de cancelación en los últimos 6 meses, si presentó un servicio técnico en los últimos 6 meses, su facturación promedio, la tecnología de su producto, si fue retenido en últimos 6 meses, si se desconecto por mora en los últimos 6 meses y otras variables adicionales.

Con las cuatro metodologías se obtuvo como resultado una curva de supervivencia $S(t)$ para cada individuo de tal forma que es posible establecer cual es la probabilidad de que dicho individuo sobreviva al tiempo t, esto permite realizar varias revisiones sobre dicha estimación, la primera es calcular la duración estimada del sujeto basándose en la siguiente ecuación:

$$\text{Duración estimada promedio} = \frac{\sum_{n}^{1} \sum_{t=1}^{T} s(t)}{n} \tag{19}$$

En donde $S(t)$ es la probabilidad de sobrevivir al tiempo t y n es la cantidad de clientes en la base. Básicamente lo que

Variable	KM	COX	ST	RF
Presento_Intencion	*	*	*	*
Servicios_Tecnicos	*	*	*	*
Facturacion_Promedio		*	*	*
Tecnologia	*	*		*
Presento_Retencion		*	*	*
Desconexiones		*	*	*
HBO			*	*
Tiene_Oferta			*	
Contactos_Facturacion			*	
Producto	*			
Tipo_Suscripcion		*		
Compro_Premium		*		

Tabla 5
Lista de Variables mas importantes para las metodologías aplicadas. Elaboración Propia.

Modelo	Dur.Prom	Nodo	Dur.Prom	% Cltes
KM	117			
COX	118			
RF	120			
		5	199	26 %
		6	51	0 %
		9	157	14 %
		10	73	1 %
		11	22	1 %
		14	119	3 %
ST	125	15	71	6 %
		16	19	1 %
		20	165	6 %
		22	119	17 %
		25	36	1 %
		26	61	7 %
		27	44	9 %

Tabla 6
Duraciones estimadas Promedio en la base de clientes. Elaboración Propia.

se realiza es sumar todas las probabilidades en la curva de supervivencia de cada cliente y luego esta duración estimada se promedia para estimar la duración de los clientes de la compañía.

En la tabla 6 se puede observar la duración estimada promedio para los 4 modelos, como se puede ver, con la metodología Random Survival Forest se obtiene una duración estimada promedio superior de 120 meses y está se puede interpretar como la mejor estimación de la duración total de los clientes. Como ya se concluyo en la sección de comparación de metodologías, la mejor duración estimada es la de Random Survival Forest.

Al realizar una comparación de la duración estimada promedio con la metodología Random Survival Forest con la duración promedio actual de la base de clientes encontramos que hay una desviación con respecto al promedio de duraciones (59 Meses vs 120 Meses) esto debido a que la metodología de análisis de supervivencia utilizada tiene en cuenta las posibles duraciones futuras que va a tener el suscriptor basándose en los datos de otros individuos, aun así al realizar un análisis con los expertos de negocio se considera que la duración promedio estimada puede ser muy alta por lo que se sugiere realizar algún tipo de ajuste a dicha estimación, lo que se dejara para futuros estudios.

Conclusiones

En el desarrollo del presente trabajo se aplicaron cuatro metodologías para estimar la duración promedio del cliente de una empresa de telecomunicaciones. Una vez aplicados se encontró que en cuanto a reducción de error de predicción la metodología Random Survival Forest es mas robusta a la hora de estimar la duración comparada con las otras metodologías.

Las metodologías tradicionales como Kaplan-Meier y Regresión de Cox son facilmente aplicables pero pueden presentar altos niveles de error a la hora de predecir por lo que se concluye que es mejor utilizar metodologías de Machine Learning como Survival Trees o Random Survival Forest; aunque esta última si bien es cierto es mejor predictora, tiene unas desventajas como altos tiempos de procesamiento y además no es fácilmente entendible como si lo es un Survival Tree.

Por otro lado se encuentra que las variables relacionadas con la experiencia del cliente como Intenciones de cancelación, servicios técnicos y desconexiones así como variables de configuración como Facturación Promedio, Tecnología son variables que describen la duración del mismo. Estas variables describen fuertemente el tiempo en que el cliente va a presentar el evento de interés (Churn).

Recomendaciones y Próximos desarrollos

Con los resultados del presente análisis se recomienda establecer medidas de reducción de riesgo para los clientes en los momentos en los que el riesgo de presentar el evento de Churn sea mayor (Ver curvas Hazard Acumuladas) por medio de metodologías de Lealtad, esto permitirá la reducción de fuga de clientes en el mediano plazo. Por otro lado se recomienda hacer una actualización de la información por lo menos cada 6 meses con el fin de contemplar comportamientos de mercado no incluidos en el presente analisis.

Dentro de los pasos siguientes para la compañía de telecomunicaciones es implementar las curvas de supervivencia obtenidas para cada cliente dentro del calculo del Life time value, esto con el fin de contar con el valor del cliente mas acertado posible. Por otro en el marco de la investigación asociada al presente documento se propone realizar una validación vía simulaciones discretas sobre la duración estimada aquí, esto con el fin de cerrar la brecha entre la duración estimada y la real con el fin de afinar el resultado del estudio.

Referencias

[Blattberg et al., 2008a] Blattberg, R., Kim, B.-D., and Neslin, S. (2008a). *Churn Management*, pages 607–633. Springer New York, New York, NY.

[Blattberg et al., 2008b] Blattberg, R., Kim, B.-D., and Neslin, S. (2008b). *Customer Lifetime Value: Fundamentals*, pages 105–131. Springer New York, New York, NY.

[Bou-Hamad et al., 2011] Bou-Hamad, I., Larocque, D., and Ben-Ameur, H. (2011). A review of survival trees. *Statist. Surv.*, 5:44–71.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

[Breiman et al., 1984] Breiman, L., Friedman, J., Stone, C., and Olshen, R. (1984). *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis.

[Ciampi et al., 1986] Ciampi, A., Thiffault, J., Nakache, J.-P., and Asselain, B. (1986). Stratification by stepwise regression, correspondence analysis and recursive partition: a comparison of three methods of analysis for survival data with covariates. *Computational Statistics & Data Analysis*, 4(3):185 – 204.

[Godoy, 2009] Godoy, A. (2009). Introducción al análisis de supervivencia con r. diploma thesis, Universidad Nacional Autónoma de México.

[Gordon and Olshen, 1985] Gordon, L. and Olshen, R. (1985). Tree-structured survival analysis. *Cancer treatment reports*, 69(10):1065—1069.

[Hothorn and Zeileis, 2016] Hothorn, T. and Zeileis, A. (2016). *Package 'partykit'*. R-Project.

[Kaplan and Meier, 1958] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481.

[Khan, 2012] Khan, I. (2012). Impact of customers satisfaction and customers retention on customer loyalty. *International Journal of Scientific and Technology Research*, 1.

[Kleinbaum and Klein, 2012] Kleinbaum, D. and Klein, M. (2012). *Introduction to Survival Analysis*, pages 1–54. Springer New York, New York, NY.

[LeBlanc and Crowley, 1992] LeBlanc, M. and Crowley, J. (1992). Relative risk trees for censored survival data. *Biometrics*, 48(2):411–425.

[Lee and Wang, 2013] Lee, E. and Wang, J. (2013). *Functions on survival time*, pages 8–16. John Wiley and Sons, Inc, Hoboken, New Jersey.

[MarketLine, 2017] MarketLine (2017). Telecomunication services in colombia. Technical report, MarketLine, London, United Kingdom,.

[Morgan and Sonquist, 1963] Morgan, J. N. and Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, 58(302):415–434.

[Rosset et al., 2003] Rosset, S., Neumann, E., Eick, U., and Vatnik, N. (2003). Customer lifetime value models for decision support. *Data Mining and Knowledge Discovery*, 7(3):321–339.

[Therneau, 2017] Therneau, T. M. (2017). *Package 'survival'*. R-Project.

[Therneau et al., 2017] Therneau, T. M., Atkinson, B., and Ripley, B. (2017). *Package 'rpart'*. R-Project.

[Therneau and Atkinson, 2017] Therneau, T. M. and Atkinson, E. (2017). An introduction to recursive partitioning using the rpart routines. *R-Project*.

[Wang et al., 2017] Wang, P., Li, Y., and Reddy, C. (2017). Machine learning for survival analysis: A survey. *ACM Computing Surveys*, 1.

[Zhou and McArdle, 2015] Zhou, Y. and McArdle, J. J. (2015). Rationale and applications of survival tree and survival ensemble methods. *Psychometrika*, 9:811–833.

Apéndice A

TABLA DE VARIABLES DISPONIBLES

Variable	Descripción	Tipo	Grupo
Código Cliente	Identificador único para cada cliente	Numérica	Base
Fecha Activación	Fecha de activación del servicio de televisión	Fecha	Base
Código Estado Producto	Código de estado del Cliente (4 = *Activo*, 6 = *Cancelado*, 7 = *Desconectado*).	Categórica	Base
Fecha Churn	Fecha en que el cliente hizo Churn (Paso de estado 4 a 6 o 7)	Fecha	Base
Duración	Duración del cliente, diferencia entre la fecha de activación y la Fecha de Churn. Expresada en Meses	Numérica	Objetivo
Censura	Indicador de si el cliente experimentó el evento de interés (Churn) o si es censurado (Quedo activo al final del experimento) (1 = *Censurado*, 0 = *NoCensurado*)	Binaria	Objetivo
Zona Instalación	Zona en donde se instalo el producto, Se refiere al perímetro. (*Urbano*, *Rural*, *Extrarural*)	Categórica	Configuración
Depto Instalación	Departamento del país en donde se instalo el producto de televisión del cliente	Categórica	Configuración
Región Instalación	Región del país en donde se instaló el producto, esta región la definió la compañía	Categórica	Configuración
Distrito Instalación	Distrito del país en donde se instaló el producto, estos los definió la compañía	Categórica	Configuración
Ciudad Instalación	Ciudad del país en donde se instalo el producto de televisión del cliente	Categórica	Configuración
Tipo de Residencia	Tipología de la edificación en la que vive el cliente	Categórica	Configuración
Canal de Venta	Canal por medio del cual se le vendió el producto al cliente	Categórica	Configuración
Tipo de Suscripción	Tipo de suscripción del cliente (*Bundle*, *CrossSelling*, *OnlyNet*, *OnlyTV*)	Categórica	Configuración
Convive con ETB	Convive con el servicio de ETB al final de la ventana de observación (1 = *Si*, 2 = *No*)	Binaria	Competencia
Convive con Une	Convive con el servicio de Une al final de la ventana de observación (1 = *Si*, 2 = *No*)	Binaria	Competencia
Convive con Movistar	Convive con el servicio de Movistar al final de la ventana de observación (1 = *Si*, 2 = *No*)	Binaria	Competencia
Convive con Claro	Convive con el servicio de Claro al final de la ventana de observación (1 = *Si*, 2 = *No*)	Binaria	Competencia
Producto	Descripción del producto de televisión del cliente al inicio de la ventana de observación (*Oro*, *Plata*, *Bronce*, *etc*.)	Categórica	Producto
Fox	Si tiene el canal FOX+ al final de la ventana de observación (1 = *Si*, 2 = *No*)	Binaria	Producto
HBO	Si tiene el canal HBO al final de la ventana de observación (1 = *Si*, 2 = *No*)	Binaria	Producto
Adultos	Si tiene el canal Adultos al final de la ventana de observación (1 = *Si*, 2 = *No*)	Binaria	Producto
Tiene Premium	Si tiene cualquier canal premium al final de la ventana de observación (1 = *Si*, 2 = *No*)	Binaria	Producto
Tecnología	Descripción de la tecnología del cliente al inicio de la ventana de observación (*Regular*, *HDcongrabacin*, *HD*, *Grabacin*)	Categórica	Producto
Cantidad Decos	Cantidad de decodificadores que el cliente tiene al inicio de la venta de observación	Categórica	Producto
Método de Pago	Método de pago del cliente (Invoice, Débito Tarjeta Crédito, Débito Cuenta ahorro)	Categórica	Facturación
Tipo de Factura	Tipo de factura del cliente (Impresa, Email)	Categórica	Facturación
Facturación Promedio	Facturación promedio en los 6 últimos meses antes de que se terminara el experimento	Numérica	Facturación
Cambio de Producto	Indicador de si el cliente cambio el producto en los 6 últimos meses (1 = *Si*, 2 = *No*)	Binaria	Experiencia
Compró Premium	Indicador de si el cliente compró algún canal premium en los 6 últimos meses (1 = *Si*, 2 = *No*)	Binaria	Experiencia
Canceló Premium	Indicador de si el cliente canceló algún canal premium en los 6 últimos meses (1 = *Si*, 2 = *No*)	Binaria	Experiencia
Cantidad PPV	Indicador de si el cliente compro una PPV en los últimos 6 meses (1 = *Si*, 2 = *No*)	Binaria	Experiencia
Presento Mora	Indicador de si el cliente entro en mora en los últimos 6 meses (1 = *Si*, 2 = *No*)	Binaria	Experiencia
Contactos de Factura	Número de contactos por facturación en los últimos 6 meses	Numérica	Experiencia
Contactos de Field	Número de contactos Field Services en los últimos 6 meses	Numérica	Experiencia
Contactos de Care	Número de contactos Customer care en los últimos 6 meses	Numérica	Experiencia
Contactos de Exp	Número de contactos Customer experience en los últimos 6 meses	Numérica	Experiencia
Contactos IT	Número de contactos IT en los últimos 6 meses	Numérica	Experiencia
Total de Contactos	Número de contactos totales en los últimos 6 meses	Numérica	Experiencia
Tiene Oferta	Indicador de si el cliente tiene alguna oferta al final de la ventana de observación (1 = *Si*, 2 = *No*)	Binaria	Churn
Presento Intención	Indicador de si el cliente llamo a cancelar en los últimos 6 meses (1 = *Si*, 2 = *No*)	Binaria	Churn
Presento Retenciones	Indicador de si el cliente fue retenido en los últimos 6 meses (1 = *Si*, 2 = *No*)	Binaria	Churn
Desconexiones	Indicador de si el cliente se desconecto en los últimos 6 meses (1 = *Si*, 2 = *No*)	Binaria	Churn
Suspensiones	Indicador de si el cliente se suspendió en los últimos 6 meses (1 = *Si*, 2 = *No*)	Binaria	Churn
Servicios técnicos	Indicador de si el cliente tuvo ordenes de servicio generadas en los últimos 6 meses (1 = *Si*, 2 = *No*)	Binaria	Experiencia
Mudanzas	Indicador de si el cliente tuvo ordenes de MUDANZA generadas en los últimos seis meses (1 = *Si*, 2 = *No*)	Binaria	Experiencia

Fin de la tabla

Apéndice B

TABLA DE COEFICIENTES DE LA REGRESIÓN DE COX CON EL PENALIZADOR LASSO

Variable	Categoría	Categ. Referencia	Coeficiente
Zona_Instalacion	NO_INFORMADO	EXTRARURAL	0.0036
Zona_Instalacion	RURAL		0.0104
Zona_Instalacion	URBANO		0.3958
Distrito_Instalacion	CENTRO	BOGOTA	(0.1394)
Distrito_Instalacion	NO_INFORMADO		0.0467
Distrito_Instalacion	NOROCCIDENTE		(0.2705)
Distrito_Instalacion	NORTE		(0.2457)
Distrito_Instalacion	OCCIDENTE		(0.3890)
Distrito_Instalacion	ORIENTE		(0.1666)
Tipo_Residencia	NORMAL_CASA	EDIFICIO	0.0607

Continua...

...Viene

Variable	Categoría	Categ. Referencia	Coeficiente
Canal_Venta	Directo	Dealers	0.0456
Canal_Venta	NO_INFORMADO		(1.1691)
Canal_Venta	Otros		(0.3890)
Canal_Venta	Televentas		0.0320
Tipo_Suscripcion	Cross_Selling	Bundle	(1.5065)
Tipo_Suscripcion	Only_Net		(0.3783)
Tipo_Suscripcion	Only_TV		(1.1564)
Convive_ETB	Si	No	(0.5799)
Convive_Une	Si	No	0.0472
Convive_Movistar	Si	No	(0.0106)
Convive_Claro	Si	No	(0.0047)
Producto	ORO	BRONCE	(0.0717)
Producto	PLATA		0.0426
Fox	Si	No	(0.1927)
HBO	Si	No	(0.3788)
Adultos	Si	No	(0.2735)
Tiene_Premium	Si	No	(0.0351)
Tecnologia	HD DVR	DVR	0.2066
Tecnologia	HD Only		1.0277
Tecnologia	SD		0.4554
Cantidad_Decos	2	1	0.4241
Cantidad_Decos	3		0.2275
Cantidad_Decos	4		0.1937
Cantidad_Decos	5		0.2479
Metodo_Pago	INVOICE	TC	(0.2204)
Tipo_Factura	Printed	Email	0.2181
Facturacion_Promedio			(2.2215)
Combio_Producto	Si	No	0.0699
Compro_Premium	Si	No	0.9403
Cantidad_PPV	Si	No	(0.0001)
Presento_mora	Si	No	(0.0825)
Contactos_Facturacion			(0.0369)
Contactos_Field			(0.1587)
Contactos_Care			0.0065
Contactos_Cexp			(0.0219)
Contactos_IT			(0.0796)
Contactos_Totales			0.0300
Tiene_Oferta	Si	No	(0.2749)
Presento_Intencion	Si	No	1.5957
Presento_Retencion	Si	No	(1.1242)
Desconexiones	Si	No	0.6872
Suspensiones	Si	No	(0.0682)
Servicios_Tecnicos	Si	No	1.0486
Mudanzas	Si	No	0.1740

Fin de la tabla

Apéndice C

TABLA SUPERVIVENCIA CONSTRUIDA CON LOS MODELOS KAPLAN-MEIER, REGRESIÓN DE COX Y ÁRBOL DE SUPERVIVENCIA

			Árbol de Supervivencia													
Time	KM	COX	5	6	9	10	11	14	15	16	20	21	22	25	26	27
1	100 %	100 %	100 %	100 %			99 %								100 %	
2	100 %	100 %	100 %	98 %			95 %			100 %			100 %	100 %	100 %	100 %
3	100 %	100 %	100 %	97 %			87 %			98 %			100 %	100 %	100 %	100 %
4	100 %	100 %	100 %	96 %			79 %			91 %			100 %	100 %	100 %	100 %
5	100 %	100 %	100 %	96 %	100 %	100 %	74 %	100 %	100 %	85 %		100 %	99 %	100 %	100 %	100 %
6	100 %	100 %	100 %	96 %			67 %		100 %	73 %			99 %	100 %	100 %	100 %
7	99 %	100 %	100 %	94 %	100 %	99 %	66 %	100 %	99 %	70 %		100 %	98 %	100 %	100 %	100 %
8	99 %	100 %	100 %	93 %	100 %	98 %	63 %	100 %	98 %	67 %	100 %	100 %	98 %	100 %	99 %	99 %
9	99 %	100 %	100 %	93 %	100 %	97 %	61 %	99 %	97 %	64 %	100 %	100 %	97 %	100 %	99 %	99 %
10	99 %	99 %	100 %	92 %	99 %	95 %	59 %	99 %	95 %	61 %	100 %	100 %	96 %	100 %	99 %	99 %
11	98 %	99 %	100 %	91 %	99 %	93 %	57 %	99 %	94 %	58 %	100 %	100 %	96 %	100 %	98 %	98 %
12	97 %	99 %	100 %	89 %	98 %	89 %	51 %	99 %	92 %	54 %	100 %	99 %	91 %	99 %	93 %	92 %
13	95 %	98 %	100 %	87 %	98 %	86 %	47 %	99 %	90 %	50 %	100 %	98 %	87 %	98 %	88 %	86 %
14	94 %	97 %	100 %	86 %	97 %	84 %	44 %	98 %	88 %	47 %	100 %	98 %	84 %	97 %	85 %	82 %
15	93 %	97 %	100 %	85 %	97 %	82 %	41 %	98 %	87 %	45 %	99 %	97 %	81 %	97 %	83 %	78 %
16	93 %	96 %	100 %	84 %	97 %	80 %	39 %	98 %	85 %	42 %	99 %	96 %	77 %	96 %	81 %	75 %

Continua...

...Viene

			Árbol de Supervivencia													
Time	KM	COX	5	6	9	10	11	14	15	16	20	21	22	25	26	27
17	92 %	96 %	99 %	83 %	96 %	79 %	37 %	98 %	84 %	39 %	99 %	96 %	74 %	96 %	79 %	72 %
18	91 %	95 %	99 %	81 %	96 %	77 %	35 %	98 %	82 %	37 %	99 %	95 %	71 %	95 %	77 %	69 %
19	90 %	95 %	99 %	80 %	95 %	75 %	34 %	97 %	81 %	35 %	99 %	94 %	68 %	94 %	75 %	66 %
20	89 %	94 %	99 %	79 %	95 %	74 %	32 %	97 %	80 %	34 %	99 %	94 %	66 %	94 %	73 %	63 %
21	89 %	94 %	99 %	78 %	95 %	72 %	31 %	97 %	78 %	32 %	98 %	93 %	64 %	93 %	71 %	60 %
22	88 %	93 %	99 %	77 %	94 %	71 %	30 %	96 %	77 %	31 %	98 %	92 %	62 %	92 %	69 %	57 %
23	87 %	93 %	99 %	76 %	94 %	70 %	29 %	96 %	75 %	29 %	98 %	92 %	60 %	92 %	67 %	54 %
24	86 %	92 %	99 %	75 %	93 %	69 %	28 %	96 %	74 %	28 %	97 %	91 %	59 %	91 %	65 %	51 %
25	85 %	92 %	99 %	74 %	93 %	68 %	28 %	95 %	72 %	27 %	97 %	90 %	57 %	90 %	64 %	48 %
26	85 %	91 %	98 %	73 %	92 %	67 %	27 %	95 %	71 %	26 %	97 %	90 %	56 %	89 %	62 %	45 %
27	84 %	91 %	98 %	72 %	92 %	66 %	26 %	94 %	70 %	25 %	96 %	89 %	54 %	88 %	61 %	42 %
28	83 %	90 %	98 %	71 %	92 %	65 %	25 %	94 %	68 %	24 %	96 %	88 %	53 %	87 %	59 %	40 %
29	82 %	89 %	98 %	70 %	91 %	65 %	25 %	94 %	67 %	23 %	96 %	87 %	51 %	86 %	58 %	37 %
30	82 %	89 %	98 %	68 %	91 %	64 %	24 %	93 %	66 %	22 %	95 %	87 %	50 %	85 %	57 %	35 %
31	81 %	88 %	98 %	67 %	90 %	63 %	23 %	93 %	65 %	22 %	95 %	86 %	49 %	84 %	55 %	33 %
32	80 %	88 %	98 %	66 %	90 %	62 %	23 %	92 %	63 %	21 %	95 %	85 %	48 %	83 %	54 %	31 %
33	80 %	87 %	98 %	64 %	90 %	62 %	22 %	92 %	62 %	20 %	94 %	85 %	46 %	82 %	52 %	29 %
34	79 %	87 %	98 %	63 %	89 %	61 %	22 %	91 %	61 %	20 %	94 %	84 %	45 %	81 %	51 %	27 %
35	78 %	86 %	97 %	61 %	89 %	61 %	21 %	91 %	60 %	19 %	94 %	83 %	43 %	80 %	50 %	26 %
36	77 %	85 %	97 %	60 %	88 %	60 %	20 %	90 %	59 %	18 %	93 %	83 %	42 %	79 %	48 %	24 %
37	77 %	85 %	97 %	58 %	88 %	60 %	20 %	89 %	58 %	17 %	93 %	82 %	40 %	77 %	47 %	22 %
38	76 %	84 %	97 %	57 %	88 %	59 %	19 %	89 %	57 %	17 %	93 %	81 %	39 %	76 %	46 %	21 %
39	75 %	84 %	97 %	55 %	87 %	58 %	19 %	88 %	56 %	16 %	92 %	81 %	37 %	74 %	44 %	20 %
40	75 %	83 %	97 %	53 %	87 %	58 %	18 %	87 %	55 %	15 %	92 %	80 %	36 %	73 %	43 %	19 %
41	74 %	82 %	97 %	51 %	87 %	57 %	17 %	87 %	54 %	14 %	92 %	79 %	34 %	71 %	42 %	18 %
42	73 %	82 %	97 %	49 %	86 %	57 %	17 %	86 %	53 %	13 %	91 %	79 %	32 %	70 %	40 %	16 %
43	73 %	81 %	97 %	48 %	86 %	56 %	16 %	85 %	52 %	13 %	91 %	78 %	31 %	68 %	39 %	15 %
44	72 %	80 %	96 %	47 %	86 %	55 %	16 %	85 %	51 %	12 %	91 %	78 %	29 %	67 %	37 %	15 %
45	71 %	80 %	96 %	46 %	85 %	55 %	15 %	84 %	50 %	12 %	91 %	77 %	28 %	65 %	36 %	14 %
46	71 %	79 %	96 %	46 %	85 %	54 %	15 %	84 %	49 %	11 %	90 %	77 %	27 %	64 %	35 %	13 %
47	70 %	78 %	96 %	45 %	85 %	54 %	14 %	83 %	49 %	10 %	90 %	76 %	26 %	63 %	34 %	13 %
48	70 %	78 %	96 %	43 %	85 %	54 %	14 %	82 %	48 %	10 %	90 %	75 %	25 %	61 %	33 %	12 %
49	69 %	77 %	96 %	43 %	84 %	53 %	13 %	82 %	47 %	9 %	90 %	75 %	24 %	60 %	32 %	12 %
50	69 %	77 %	96 %	42 %	84 %	53 %	13 %	81 %	46 %	9 %	89 %	74 %	23 %	58 %	31 %	11 %
51	68 %	76 %	96 %	42 %	84 %	53 %	13 %	80 %	46 %	9 %	89 %	74 %	22 %	57 %	30 %	11 %
52	68 %	75 %	96 %	41 %	84 %	52 %	12 %	80 %	45 %	8 %	89 %	73 %	21 %	56 %	29 %	10 %
53	67 %	75 %	95 %	41 %	83 %	52 %	12 %	79 %	44 %	8 %	89 %	73 %	21 %	55 %	28 %	10 %
54	67 %	74 %	95 %	40 %	83 %	51 %	12 %	78 %	44 %	8 %	88 %	72 %	20 %	53 %	27 %	10 %
55	66 %	74 %	95 %	39 %	83 %	51 %	11 %	77 %	43 %		88 %	72 %	20 %	52 %	27 %	9 %
56	66 %	73 %	95 %	38 %	83 %	51 %	11 %	77 %	43 %	8 %	88 %	71 %	19 %	51 %	26 %	9 %
57	65 %	73 %	95 %	37 %	82 %	50 %	11 %	76 %	42 %	8 %	88 %	71 %	19 %	49 %	25 %	9 %
58	65 %	72 %	95 %	36 %	82 %	50 %	11 %	75 %	41 %	8 %	87 %	70 %	18 %	48 %	25 %	8 %
59	64 %	71 %	95 %	36 %	82 %	50 %	11 %	75 %	41 %	7 %	87 %	70 %	18 %	47 %	24 %	8 %
60	64 %	71 %	95 %	35 %	82 %	49 %	11 %	74 %	40 %	7 %	87 %	69 %	17 %	45 %	23 %	8 %
61	63 %	70 %	95 %	34 %	81 %	49 %	10 %	74 %	40 %	7 %	86 %	69 %	17 %	44 %	23 %	8 %
62	63 %	70 %	94 %	33 %	81 %	49 %	10 %	73 %	39 %	7 %	86 %	68 %	16 %	42 %	22 %	7 %
63	62 %	69 %	94 %	31 %	81 %	48 %	10 %	72 %	39 %	6 %	86 %	68 %	16 %	41 %	21 %	7 %
64	62 %	69 %	94 %	30 %	81 %	48 %	9 %	72 %	38 %	6 %	86 %	67 %	15 %	40 %	21 %	7 %
65	62 %	68 %	94 %	29 %	80 %	48 %	9 %	71 %	38 %	6 %	85 %	67 %	15 %	39 %	20 %	7 %
66	61 %	67 %	94 %	29 %	80 %	47 %	9 %	70 %	37 %	6 %	85 %	66 %	14 %	38 %	20 %	6 %
67	61 %	67 %	94 %	28 %	80 %	47 %	9 %	70 %	37 %	5 %	85 %	66 %	13 %	36 %	19 %	6 %
68	60 %	66 %	94 %	27 %	80 %	47 %	8 %	69 %	37 %	5 %	85 %	65 %	13 %	35 %	19 %	6 %
69	60 %	66 %	94 %	26 %	80 %	46 %	8 %	68 %	36 %	5 %	84 %	65 %	12 %	34 %	18 %	6 %
70	59 %	65 %	94 %	25 %	79 %	46 %	8 %	68 %	36 %	5 %	84 %	65 %	12 %	33 %	18 %	6 %
71	59 %	65 %	94 %	25 %	79 %	46 %	8 %	67 %	36 %	4 %	84 %	64 %	11 %	32 %	17 %	6 %
72	59 %	64 %	94 %	24 %	79 %	45 %	8 %	67 %	36 %	4 %	84 %	64 %	10 %	31 %	17 %	5 %
73	58 %	64 %	94 %	23 %	79 %	45 %	7 %	66 %	35 %	4 %	84 %	63 %	10 %	30 %	17 %	5 %
74	58 %	63 %	93 %	23 %	79 %	45 %	7 %	66 %	35 %	4 %	83 %	63 %	9 %	29 %	16 %	5 %
75	57 %	63 %	93 %	22 %	78 %	44 %	7 %	65 %	35 %	4 %	83 %	63 %	9 %	28 %	16 %	5 %
76	57 %	62 %	93 %	22 %	78 %	44 %	7 %	65 %	35 %	3 %	83 %	62 %	9 %	27 %	16 %	5 %
77	57 %	62 %	93 %	22 %	78 %	44 %	7 %	64 %	35 %	3 %	83 %	62 %	8 %	26 %	15 %	5 %
78	56 %	61 %	93 %	21 %	78 %	44 %	6 %	64 %	34 %	3 %	83 %	61 %	8 %	25 %	15 %	5 %
79	56 %	61 %	93 %	21 %	78 %	43 %	6 %	63 %	34 %	3 %	82 %	61 %	8 %	24 %	15 %	5 %
80	56 %	60 %	93 %	20 %	77 %	43 %	6 %	62 %	34 %		82 %	60 %	7 %	23 %	14 %	4 %
81	55 %	60 %	93 %	20 %	77 %	43 %	6 %	62 %	33 %	3 %	82 %	60 %	7 %	22 %	14 %	4 %
82	55 %	59 %	93 %	19 %	77 %	43 %	6 %	61 %	33 %	3 %	82 %	60 %	7 %	21 %	13 %	4 %
83	54 %	59 %	93 %	19 %	77 %	42 %	6 %	61 %	33 %	3 %	81 %	59 %	7 %	20 %	13 %	4 %
84	54 %	58 %	93 %	19 %	77 %	42 %	5 %	60 %	32 %	3 %	81 %	59 %	6 %	19 %	13 %	4 %

Continua...

...Viene

			Árbol de Supervivencia													
Time	KM	COX	5	6	9	10	11	14	15	16	20	21	22	25	26	27
85	54 %	58 %	93 %	18 %	76 %	42 %	5 %	59 %	32 %	3 %	81 %	58 %	6 %	19 %	13 %	4 %
86	53 %	57 %	92 %	18 %	76 %	41 %	5 %	59 %	32 %	3 %	81 %	58 %	6 %	18 %	12 %	4 %
87	53 %	57 %	92 %	17 %	76 %	41 %	5 %	58 %	32 %	3 %	80 %	57 %	6 %	17 %	12 %	4 %
88	53 %	56 %	92 %	17 %	76 %	41 %	5 %	58 %	31 %	3 %	80 %	57 %	5 %	16 %	12 %	4 %
89	52 %	56 %	92 %	17 %	76 %	40 %	5 %	57 %	31 %	2 %	80 %	57 %	5 %	16 %	11 %	3 %
90	52 %	55 %	92 %	16 %	76 %	40 %	5 %	57 %	31 %	2 %	80 %	56 %	5 %	15 %	11 %	3 %
91	52 %	55 %	92 %	16 %	75 %	40 %	5 %	56 %	30 %	2 %	80 %	56 %	5 %	15 %	11 %	3 %
92	51 %	55 %	92 %	15 %	75 %	39 %	4 %	56 %	30 %	2 %	79 %	56 %	5 %	14 %	10 %	3 %
93	51 %	54 %	92 %		75 %	39 %	4 %	56 %	30 %	2 %	79 %	55 %	5 %	14 %	10 %	3 %
94	51 %	54 %	92 %	14 %	75 %	38 %	4 %	55 %	30 %	2 %	79 %	55 %	4 %	13 %	10 %	3 %
95	51 %	53 %	92 %	14 %	75 %	38 %	4 %	55 %	29 %		79 %	55 %	4 %	13 %	9 %	3 %
96	50 %	53 %	92 %	14 %	75 %	38 %	4 %	55 %	29 %	2 %	79 %	54 %	4 %	13 %	9 %	3 %
97	50 %	53 %	92 %	13 %	74 %	37 %	4 %	54 %	29 %		79 %	54 %	4 %	12 %	9 %	3 %
98	50 %	52 %	92 %	13 %	74 %	36 %	3 %	54 %	29 %		78 %	54 %	4 %	12 %	9 %	3 %
99	49 %	52 %	91 %	12 %	74 %	36 %	3 %	54 %	28 %	2 %	78 %	53 %	3 %	12 %	8 %	3 %
100	49 %	51 %	91 %	12 %	74 %	35 %	3 %	53 %	28 %	2 %	78 %	53 %	3 %	11 %	8 %	2 %
101	49 %	51 %	91 %	12 %	74 %	35 %	3 %	53 %	28 %	2 %	78 %	53 %	3 %	11 %	8 %	2 %
102	49 %	51 %	91 %	11 %	74 %	34 %	3 %	53 %	28 %	2 %	77 %	52 %	3 %	11 %	8 %	2 %
103	48 %	50 %	91 %	11 %	73 %	34 %	3 %	53 %	27 %	2 %	77 %	52 %	3 %	10 %	7 %	2 %
104	48 %	50 %	91 %	11 %	73 %	33 %	3 %	52 %	27 %	2 %	77 %	52 %	3 %	10 %	7 %	2 %
105	48 %	50 %	91 %	11 %	73 %	33 %	3 %	52 %	27 %	2 %	77 %	51 %	3 %	10 %	7 %	2 %
106	48 %	49 %	91 %		73 %	32 %	3 %	52 %	27 %	1 %	77 %	51 %	2 %	10 %	6 %	2 %
107	47 %	49 %	91 %		73 %	32 %		51 %	26 %	1 %	77 %	51 %	2 %	9 %	6 %	2 %
108	47 %	49 %	91 %	10 %	72 %	32 %	2 %	51 %	26 %	1 %	77 %	51 %	2 %	9 %	6 %	2 %
109	47 %	48 %	91 %		72 %	31 %	2 %	51 %	26 %	1 %	77 %	50 %	2 %	9 %	6 %	2 %
110	47 %	48 %	91 %	10 %	72 %	30 %	2 %	50 %	26 %		77 %	50 %	2 %	9 %	6 %	2 %
111	46 %	48 %	91 %	10 %	72 %	30 %	2 %	50 %	25 %	1 %	76 %	50 %	2 %	9 %	6 %	2 %
112	46 %	47 %	91 %	9 %	72 %	30 %	2 %	50 %	25 %	1 %	76 %	49 %	2 %	8 %	5 %	2 %
113	46 %	47 %	91 %	9 %	71 %	30 %	2 %	50 %	25 %		76 %	49 %	2 %	8 %	5 %	2 %
114	46 %	47 %	91 %	9 %	71 %	30 %	2 %	49 %	24 %	1 %	76 %	49 %	2 %	8 %	5 %	2 %
115	45 %	46 %	90 %	8 %	71 %	29 %	2 %	49 %	24 %		76 %	49 %		8 %	5 %	2 %
116	45 %	46 %	90 %	8 %	71 %	29 %	2 %	48 %	24 %	1 %	76 %	48 %	2 %	8 %	5 %	2 %
117	45 %	46 %	90 %	8 %	71 %	29 %		48 %	24 %	1 %	76 %	48 %	2 %	7 %	4 %	1 %
118	45 %	45 %	90 %	8 %	70 %	29 %	2 %	48 %	23 %	1 %	76 %	48 %	2 %	7 %	4 %	1 %
119	45 %	45 %	90 %	7 %	70 %	29 %	2 %	47 %	23 %	1 %	75 %	47 %	2 %	7 %	4 %	1 %
120	44 %	45 %	90 %	7 %	70 %	28 %	2 %	47 %	23 %	1 %	75 %	47 %	2 %	7 %	4 %	1 %
121	44 %	44 %	90 %	7 %	70 %	28 %	2 %	47 %	22 %	1 %	75 %	47 %	1 %	7 %	4 %	1 %
122	44 %	44 %	90 %	7 %	69 %	28 %		46 %	22 %	1 %	75 %	47 %	1 %	7 %	4 %	1 %
123	44 %	43 %	90 %	6 %	69 %	28 %		46 %	22 %		75 %	47 %	1 %	6 %	3 %	1 %
124	43 %	43 %	90 %	6 %	69 %	28 %		46 %	22 %		74 %	46 %	1 %	6 %	3 %	1 %
125	43 %	43 %	90 %		69 %	28 %	2 %	45 %	22 %		74 %	46 %	1 %	6 %	3 %	1 %
126	43 %	43 %	90 %		69 %	27 %	2 %	45 %	21 %	1 %	74 %	46 %	1 %	6 %	3 %	1 %
127	43 %	42 %	90 %		69 %	27 %	2 %	45 %	21 %	1 %	74 %	46 %	1 %	6 %	3 %	1 %
128	43 %	42 %	90 %	6 %	69 %	27 %	1 %	44 %	21 %	1 %	74 %	45 %	1 %	6 %	3 %	1 %
129	42 %	42 %	89 %	6 %	68 %	27 %		44 %	21 %		74 %	45 %	1 %	5 %	3 %	1 %
130	42 %	41 %	89 %		68 %	27 %	1 %	44 %	20 %		74 %	45 %	1 %	5 %	3 %	1 %
131	42 %	41 %	89 %		68 %	26 %	1 %	44 %	20 %		73 %	45 %		5 %	3 %	1 %
132	42 %	41 %	89 %		68 %	26 %		43 %	20 %		73 %	44 %	1 %	5 %	3 %	1 %
133	42 %	41 %	89 %	6 %	68 %	26 %	1 %	43 %	20 %		73 %	44 %	1 %	5 %	2 %	1 %
134	42 %	40 %	89 %		68 %	25 %		42 %	20 %	1 %	73 %	44 %		5 %	2 %	1 %
135	41 %	40 %	89 %		68 %	25 %		42 %	20 %		73 %	44 %	1 %	5 %	2 %	1 %
136	41 %	40 %	89 %	5 %	67 %	25 %		42 %	20 %		73 %	44 %	1 %	4 %	2 %	1 %
137	41 %	39 %	89 %	5 %	67 %	25 %	1 %	42 %	19 %	1 %	73 %	44 %		4 %	2 %	1 %
138	41 %	39 %	89 %	5 %	67 %	25 %	1 %	41 %	19 %		73 %	43 %	1 %	4 %	2 %	1 %
139	41 %	39 %	89 %	5 %	67 %	25 %		41 %	19 %	1 %	72 %	43 %	1 %	4 %	2 %	1 %
140	41 %	39 %	89 %	5 %	67 %	24 %		41 %	19 %		72 %	43 %	1 %	4 %	2 %	1 %
141	40 %	38 %	89 %		67 %	24 %	1 %	41 %	19 %		72 %	43 %	1 %	4 %	2 %	1 %
142	40 %	38 %	89 %	5 %	67 %	24 %		40 %	19 %	1 %	72 %	42 %	1 %	4 %	2 %	1 %
143	40 %	38 %	89 %	5 %	67 %	24 %		40 %	18 %		72 %	42 %	1 %	3 %	2 %	1 %
144	40 %	38 %	89 %		67 %	24 %	1 %	39 %	18 %		72 %	42 %	1 %	3 %	2 %	1 %
145	40 %	38 %	88 %	4 %	66 %	24 %		39 %	18 %	0 %	72 %	42 %		3 %	2 %	1 %
146	40 %	37 %	88 %	4 %	66 %	24 %		39 %	18 %		72 %	42 %		3 %	2 %	1 %
147	40 %	37 %	88 %	4 %	66 %		1 %	39 %	18 %		71 %	42 %		3 %	2 %	1 %
148	40 %	37 %	88 %	4 %	66 %	24 %	1 %	39 %	18 %		71 %	42 %		3 %	2 %	1 %
149	40 %	37 %	88 %		66 %	24 %		39 %	18 %	0 %	71 %	41 %		3 %	2 %	
150	39 %	37 %	88 %		66 %	23 %	1 %	38 %	18 %		71 %	41 %		3 %	2 %	1 %
151	39 %	37 %	88 %		66 %	23 %		38 %	18 %		71 %	41 %		3 %	1 %	1 %
152	39 %	36 %	88 %		66 %	23 %		38 %	18 %		71 %	41 %		3 %	1 %	

Continua...

...Viene

			Árbol de Supervivencia													
Time	KM	COX	5	6	9	10	11	14	15	16	20	21	22	25	26	27
153	39 %	36 %	88 %		66 %	23 %		38 %	17 %		71 %	41 %		3 %	1 %	1 %
154	39 %	36 %	88 %		66 %		1 %	38 %	17 %		71 %	41 %		3 %	1 %	1 %
155	39 %	36 %	88 %		65 %			38 %	17 %	0 %	70 %	41 %		3 %	1 %	1 %
156	39 %	36 %	88 %		65 %			38 %	17 %		70 %	41 %		3 %	1 %	1 %
157	39 %	36 %	88 %		65 %			38 %	17 %		70 %	41 %		3 %	1 %	1 %
158	39 %	36 %	88 %		65 %	23 %	1 %	38 %	17 %		70 %	41 %		3 %	1 %	0 %
159	39 %	35 %	88 %	4 %	65 %			38 %	17 %		70 %	40 %		3 %	1 %	0 %
160	39 %	35 %	88 %		65 %	23 %		37 %	17 %		70 %	40 %		3 %	1 %	0 %
161	38 %	35 %	88 %		65 %		1 %	37 %	17 %		70 %	40 %	1 %	3 %	1 %	
162	38 %	35 %	88 %		65 %	23 %		37 %	17 %		70 %	40 %	1 %	3 %	1 %	0 %
163	38 %	35 %	88 %	4 %	65 %	22 %		37 %	17 %		70 %	40 %	1 %	3 %	1 %	0 %
164	38 %	35 %	88 %		65 %	22 %		37 %	17 %		70 %	39 %		3 %	1 %	0 %
165	38 %	35 %	88 %		64 %	22 %		36 %	17 %		70 %	39 %	1 %	3 %	1 %	0 %
166	38 %	34 %	88 %		64 %			36 %	17 %		70 %	39 %		3 %	1 %	0 %
167	38 %	34 %	88 %		64 %	22 %		36 %	17 %		70 %	39 %	1 %	2 %	1 %	0 %
168	38 %	34 %	88 %		64 %	22 %		36 %	17 %		70 %	39 %		2 %	1 %	0 %
169	38 %	34 %	88 %	4 %	64 %	21 %		36 %			70 %	39 %	1 %	2 %	1 %	0 %
170	38 %	34 %	88 %	3 %	64 %	21 %		36 %	17 %		70 %	39 %		2 %	1 %	0 %
171	38 %	34 %	88 %		64 %			36 %	16 %		70 %	39 %		2 %	1 %	0 %
172	38 %	34 %	88 %		64 %			36 %	16 %		70 %	39 %		2 %	1 %	
173	38 %	33 %	88 %	3 %	64 %			36 %	16 %		70 %	39 %		2 %	1 %	
174	37 %	33 %	88 %	3 %	64 %	20 %	1 %	36 %	16 %		70 %	39 %		2 %	1 %	0 %
175	37 %	33 %	88 %		64 %	20 %		36 %	16 %		70 %	39 %		2 %	1 %	0 %
176	37 %	33 %	87 %	3 %	64 %			35 %	16 %		70 %	39 %		2 %	1 %	
177	37 %	32 %	87 %	2 %	63 %	20 %		35 %	16 %	0 %	69 %	38 %	1 %	2 %	1 %	0 %
178	37 %	32 %	87 %	2 %	63 %			35 %	16 %	0 %	69 %	38 %	1 %	2 %	1 %	
179	37 %	32 %	87 %		63 %			35 %	15 %		69 %	38 %	1 %	2 %	1 %	0 %
180	37 %	32 %	87 %	2 %	63 %	19 %		35 %	15 %		69 %	38 %	0 %	2 %	1 %	0 %
181	37 %	32 %	87 %	2 %	63 %	19 %		34 %	15 %		69 %	38 %	0 %	2 %	1 %	0 %
182	36 %	32 %	87 %	1 %	63 %	19 %	1 %	34 %	15 %	0 %	69 %	38 %		2 %	1 %	
183	36 %	31 %	87 %	1 %	63 %	18 %		34 %	15 %		69 %	38 %		2 %	1 %	0 %
184	36 %	31 %	87 %	1 %	63 %	18 %		34 %	15 %		69 %	37 %		2 %	1 %	0 %
185	36 %	31 %	87 %		62 %		0 %	34 %	15 %	0 %	69 %	37 %		2 %	1 %	
186	36 %	30 %	87 %	1 %	62 %	16 %	0 %	34 %	14 %		68 %	37 %	0 %	2 %	1 %	
187	36 %	30 %	87 %	1 %	62 %	15 %		34 %	14 %		68 %	37 %		2 %	1 %	0 %
188	35 %	30 %	87 %		62 %	14 %		34 %	14 %		68 %	37 %		2 %	1 %	0 %
189	35 %	30 %	86 %		62 %	12 %		34 %	14 %		68 %	37 %	0 %	2 %	1 %	0 %
190	35 %	29 %	86 %	0 %	61 %	11 %		34 %	14 %		68 %	37 %		1 %	1 %	
191	35 %	29 %	86 %	0 %	61 %	10 %		33 %	13 %	0 %	68 %	36 %		1 %	1 %	
192	35 %	28 %	86 %		60 %		0 %	33 %	13 %		68 %	36 %		1 %	1 %	
193	35 %	28 %	86 %		60 %			33 %	13 %		67 %	36 %	0 %	1 %	0 %	
194	34 %	28 %	86 %		60 %			33 %	13 %	0 %	67 %	36 %		1 %	0 %	
195	34 %	28 %	86 %		60 %	10 %		33 %	13 %		67 %	35 %		1 %	0 %	0 %
196	34 %	27 %	86 %		60 %			32 %	13 %		67 %	35 %	0 %	1 %	0 %	0 %
197	34 %	27 %	86 %		60 %			32 %	12 %		66 %	35 %		1 %	0 %	
198	34 %	27 %	86 %		59 %	10 %		32 %	12 %		66 %	35 %		1 %	0 %	0 %
199	34 %	26 %	86 %		59 %			31 %	12 %	0 %	66 %	34 %		1 %	0 %	0 %
200	34 %	26 %	86 %		59 %			31 %	12 %		66 %	34 %		1 %	0 %	
201	33 %	26 %	86 %		59 %		0 %	30 %	12 %		66 %	34 %		1 %	0 %	0 %
202	33 %	25 %	85 %		59 %	10 %		30 %	12 %		66 %	33 %		1 %	0 %	0 %
203	33 %	25 %	85 %		59 %	9 %	0 %	30 %	12 %		66 %	33 %	0 %	1 %	0 %	0 %
204	33 %	25 %	85 %		59 %			29 %	11 %		66 %	33 %		1 %	0 %	
205	33 %	24 %	85 %		59 %	9 %		29 %	11 %		66 %	33 %		1 %	0 %	
206	33 %	24 %	85 %		59 %			29 %	11 %		65 %	33 %		1 %	0 %	
207	32 %	24 %	85 %		59 %	9 %		29 %	11 %		65 %	33 %	0 %	1 %	0 %	
208	32 %	24 %	85 %		59 %			29 %	10 %		65 %	33 %	0 %	1 %	0 %	
209	32 %	23 %	85 %		59 %	7 %		28 %	10 %		65 %	32 %		0 %	0 %	0 %
210	32 %	23 %	85 %		59 %			27 %	9 %		65 %	31 %		0 %	0 %	0 %
211	32 %	22 %	85 %		58 %			27 %	9 %		65 %	30 %		0 %	0 %	0 %
212	31 %	22 %	85 %		58 %	7 %		27 %	9 %		64 %	30 %		0 %	0 %	0 %
213	31 %	22 %	85 %		58 %			27 %	9 %		64 %	30 %		0 %	0 %	0 %
214	31 %	21 %	85 %		58 %	6 %		26 %	9 %		64 %	30 %		0 %	0 %	
215	31 %	21 %	85 %		58 %	6 %		26 %	9 %		64 %	29 %		0 %	0 %	0 %
216	31 %	21 %	85 %		57 %			26 %	9 %		63 %	29 %	0 %	0 %	0 %	
217	30 %	20 %	85 %		56 %	6 %		26 %	9 %	0 %	63 %	28 %	0 %	0 %	0 %	0 %

Fin de la tabla