



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

**MODELAMIENTO DE PREDICTOR DE
FUGA DE CLIENTES, CON LA
UTILIZACIÓN DEL VALOR DE VIDA DEL
CLIENTE Y REDES NEURONALES
RECURRENTES.**

PEDRO AGUSTÍN HOJAS LORET

Tesis para optar al grado de
Magister en Ciencias de la Ingeniería

Profesor Supervisor:
CONSTANZA MIRANDA MENDOZA

Santiago de Chile, Diciembre 2018
2018, Pedro Hojas



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

**MODELAMIENTO DE PREDICTOR DE
FUGA DE CLIENTES, CON LA
UTILIZACIÓN DEL VALOR DE VIDA DEL
CLIENTE Y REDES NEURONALES
RECURRENTES.**

PEDRO AGUSTÍN HOJAS LORET

Tesis presentada a la Comisión integrada por los profesores:

CONSTANZA MIRANDA MENDOZA

DENIS PARRA SANTANDER

FELIPE CAMPOSANO LORENZINI

CRISTIÁN ESCAURIAZA MESA

A Agustín, por todo.

AGRADECIMIENTOS

Quiero agradecer a Empresas Dimerc por involucrarse en este proyecto: particularmente a Felipe Camposano quien apoyó el desarrollo de esta Tesis desde su comienzo. Desde una activa participación en la definición del problema y facilitar la información para el análisis de las transacciones históricas.

INDICE GENERAL

	Pág.
DEDICATORIA.....	ii
AGRADECIMIENTOS	iv
INDICE GENERAL.....	v
INDICE DE TABLAS	vii
INDICE DE FIGURAS.....	viii
RESUMEN.....	x
ABSTRACT	xi
1 Introducción.....	1
1.1 La Importancia del Cliente.....	1
1.2 En busca del futuro.....	4
1.3 Objetivos y preguntas de investigación.....	5
1.4 Caso de estudio.....	6
2 Marco Teórico.....	7
2.1 Valorización de los clientes.....	7
2.1.1 Valor de vida.....	7
2.1.2 Segmentación según Recencia, Frecuencia y Monetización	8
2.2 ¿Qué es la fuga?	10
2.2.1 Tipo de relación comercial.....	10
2.2.2 Distintos enfoques de valorización de la fuga no contractual	11
2.3 Predicción en series de tiempo	13
2.3.1 Tiempo para el siguiente evento	14
2.3.2 Análisis con datos censurados	15
2.3.3 La distribución de Weibull	16
2.4 Redes Neuronales Recurrentes.....	19

3	Metodología.....	p22
3.1	Descripción de los datos.....	22
3.2	Modelo propuesto.....	22
3.3	Etapas	23
3.3.1	Preparación de Datos	23
3.3.2	WTTE-RNN.....	26
3.3.3	Predicción de Fuga.....	27
3.4	Evaluación del modelo	28
3.4.1	Información histórica	28
3.5	Diferentes formas de evaluar el modelo.....	29
4	Resultados.....	31
4.1	Valores de αt y βt en el modelo wtte-rnn	31
4.2	Validación del modelo	32
4.2.1	Como predictor de fuga según CLV	32
4.2.2	Como predictor de fuga según RFM.....	33
5	Discusión	35
5.1	El modelo como predictor de fuga	35
5.1.1	Limitaciones del modelo.....	35
5.1.2	Optimizaciones del modelo	36
5.1.3	Conclusiones.....	37
	Bibliografía.....	38
	Anexos	41
	Datos transaccionales	41
	Datos <i>CRM</i>	42

INDICE DE TABLAS

	Pág.
Tabla 1: Escalamiento de los atributos <i>RFM</i> como quintil del número de clientes.	9
Tabla 2: Esquema de datos de la información históricos de compra.	41
Tabla 3: Esquema de datos de la información contextual de los clientes.	42

INDICE DE FIGURAS

	Pág.
Figura 1: Se ilustra el mapa contextual de los conceptos que se enmarca la investigación..	2
Figura 2: : Ilustración representativa de una predicción de compra.	4
Figura 3: Ilustración que explica el concepto de TTE sin censura y_t , y la predicción y_t con censura.....	16
Figura 4: Gráfica de la Función de Distribución de Probabilidad (<i>pdf</i>) y la Tasa de Falla (<i>hazard</i>) para distintos valores de α y β . Fuente: de Martisson 2016.....	18
Figura 5: Diagrama de una Red Neuronal Recurrente genérica, donde se aprecia que los parámetros de cada estado (x_t) y la información del estado anterior (h_t) son los los parámetros de entrada de cada paso. Fuente: Martinsson 2016.	19
Figura 6: Estructura de una WTTE-RNN. Se muestra como parámetros de salida para cada instante t , los valores α_t y β_t , que determinan la distribución Weibull. Fuente . Fuente: Martinsson, 2016.....	20
Figura 7: Gráfico de flujo del Modelo predictor	23
Figura 8: Gráfico de barras que muestra la cantidad de transacciones que tienen la misma cantidad de días desde la última compra. Fuente: Elaboración Propia.	25
Figura 9: Matriz de one-hot-encoded de las compras realizadas por 80 clientes elegidos al azar. En el eje x se ven los 220 periodos que se utilizaron para entrenar el modelo; en el Y son 80 clientes elegidos aleatoriamente. Los cuadros grices representan 1's en la matriz <i>one-hot-encoded</i>	25
Figura 10: Mapa de calor del Tiempo al siguiente evento de los mismos clientes de la figura anterior. En esta visualización, mientras más rojo es el pixel, mayor es el <i>TTE</i> para ése cliente.	26
Figura 11: Línea de tiempo que ilustra la cantidad de tiempo sobre los cuales se recolectó información. Fuente: Elaboración Propia.....	29
Figura 12: Ilustración que muestra cómo varían los valores de t y t_a lo largo del periodo de entrenamiento para un cliente específico. Fuente: Elaboración Propia.	31
Figura 13: Curva de ROC para cada una de las semanas del períodos de predicción.	33

Figura 14: Curva de ROC para cada una de las semanas del períodos de predicción. 34

RESUMEN

Para todas las industrias, el manejo estratégico de los clientes es vital para conservar la competitividad y la supervivencia empresarial en el tiempo. En este contexto existen dos enfoques estratégicos principales: adquisición y retención de clientes. Debido a que las segundas son las más rentables para empresas consolidadas, este estudio se centrará en aquellas, particularmente aquellas que buscan la *prevención de fuga de clientes*. En particular, esta tesis se centrará en el contexto de la venta orientada a empresas –o *B2B*– tiene un mercado objetivo se concentra en menos actores, con valores de compra más elevados: por lo que este ámbito de estudio es particularmente importante.

De esta manera, se implementó un modelo predictor de fuga de clientes en base a información histórica de transacciones, en un contexto de *B2B*. Esto se hizo en dos partes: primero fue implementado un modelo de predicción secuencial de *tiempo a la siguiente compra*: denominado *Weibull Time To Event Recurrent Neural Network (wtte-rnn)*. Este, tiene tanto las bondades de las Redes Neuronales Recurrente –procesar patrones temporales complejos– como la versatilidad de la distribución de *Weibull* para la predicción de eventos. En una segunda instancia, se utiliza el cálculo del Valor del Ciclo de Vida del Cliente, donde se puede valorar la información histórica con la predicción anterior.

Para validar la implementación del modelo se utilizará la información transaccional de una empresa de insumos no estratégicos orientado a la venta de empresas (*B2B*) con información de más de 25.000 clientes anuales. En particular se analizarán más de 3 millones de transacciones ocurridas entre los años 2014 y 2018.

Con esta información y dependiendo del escenario se logra tener un predictor de fuga cuya ganancia sobre un predictor aleatorio es considerable, con un *AUC* entre 0,61 a 0,69.

Palabras Claves: Redes Neuronales Recurrentes, Aprendizaje de Máquinas, Predicción de Fuga, B2B, Ciclo de Vida del Cliente, Tiempo para Evento, Análisis de Supervivencia

ABSTRACT

For every industry, the management of customers is one of the main tasks to keep the company alive over time. In this context, there are two main strategic approaches: customer **acquisition** and **retention**. Because the latter are the most profitable for consolidated companies, this study will focus on those: more specifically, at the ones that seek prevention of **customer churn**. In particular, this thesis will focus on the context of business among companies -or B2B-, which's target market is concentrated in fewer actors, with higher purchase values.

The implementation of a customer churn prediction model was implemented based on historical transaction information of each specific client. This was done in two parts: in first place, a sequential time prediction model was implemented for the next purchase: called Weibull Time To Event Recurrent Neural Network (wtte-rnn). This has both the benefits of Recurrent Neural Networks -process complex temporal patterns- and the versatility of the *Weibull* as distribution for time to event prediction.

In a second instance, the calculation of the Customer Lifetime Value -CLV- based on the model output was used, where changes on historical patterns are used to predict future churning behavior.

To validate the implementation of the model, transactional information of an office supply store oriented in the business-to-business market, with more than 25,000 customers per year. In particular, more than 3 million transactions will be analyzed between 2014 and 2018.

This results show that the model performs better than random, with an AUC near to 0.6 or almost 0.7 depending on which churn definition was used.

Keywords: recurrent neural networks, machine learning, churn prediction, B2B, customer lifetime value, time to event, weibull, survival analysis

1 INTRODUCCIÓN

1.1 La Importancia del Cliente

Para las empresas de *retail*, los clientes son su activo más valioso: una empresa sin clientes, no tiene ingresos y por lo tanto, carece de valor. En este contexto, la gestión estratégica de la relación que mantienen las empresas con sus clientes, se conoce como *Customer Relationship Management* o *CRM*. Si bien el concepto fue acuñado a fines de los 90s bajo el alero de las TI's (Wübben, 2008), hoy en día se entiende como estrategia integral a la empresa, para construir, administrar y fortalecer relaciones basadas en la lealtad y largoplacistas con los clientes (Vafeiadis et al, 2015).

Es importante destacar que en la gestión estratégica de los clientes, los conceptos de **adquisición** y **retención** aparecen arraigados de manera transversal como los principales objetivos del *CRM* (D'Haen & Van den Poel, 2013). Si bien ambos conceptos son relevantes para los negocios, las firmas ya consolidadas debieran enfocarse en el segundo: ya que la retención conlleva una ganancia de última línea: según el estudio transversal realizado Gupta et. al (2004), se demuestra que en diferentes empresas, la elasticidad promedio de retención es 4.9. Lo que significa que un 1% de aumento en la tasa de retención puede significar cerca de un 5% de aumento en el valor de la cartera de clientes. Por otro lado, en algunos casos el valor de capturar a un nuevo cliente puede ser hasta 20 veces mayor que el costo de retenerlo (Vafeiadis et al, 2015).

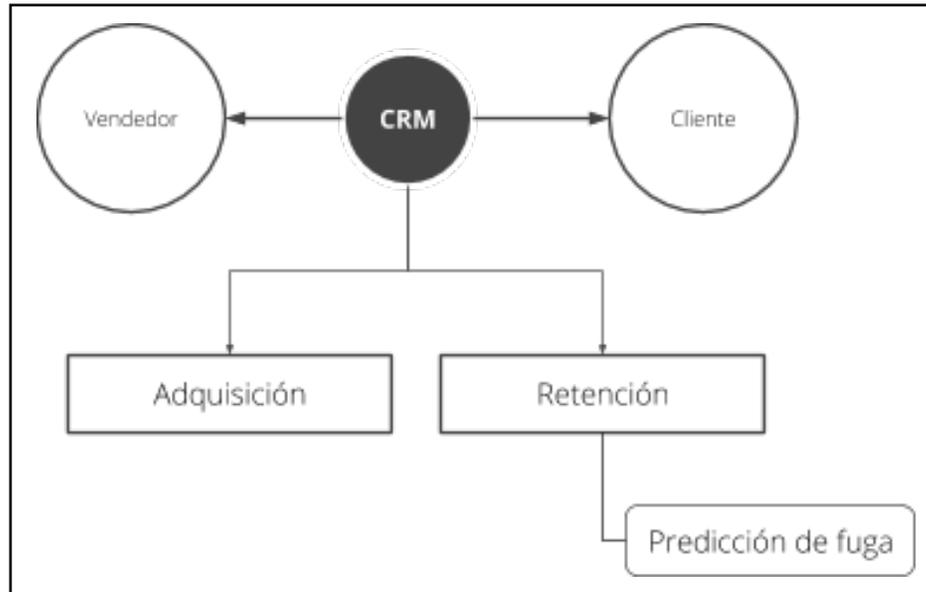


Figura 1: Se ilustra el mapa contextual de los conceptos que se enmarca la investigación.

Fuente: Elaboración Propia.

Este hecho es aún más claro en el contexto de ventas orientadas a empresas –o *B2B*– donde el mercado objetivo tiende a estar más saturado: con menos clientes que, a su vez, tienden a hacer compras más grandes y frecuentes (Rauyrue y Miller, 2007). Esto hace que cada uno de los clientes en la cartera, de manera individual, tengan un valor transaccional mayor al de los negocios de venta orientados a clientes o consumidores finales – o *B2C* (Rauyrue y Miller, 2007). Como consecuencia directa, en el contexto *B2B* el costo asociado para la empresa de la fuga de un cliente, es sustantivamente mayor que en el caso del *B2C*.

Dentro de las estrategias de retención existen diferentes enfoques como la satisfacción y experiencia del cliente, el desarrollo de nuevos productos o estrategias competitivas de precio (Vafeiadis et al, 2015). Todas estas pueden ser potenciadas con la utilización de herramientas de *Inteligencia de Negocios*. En particular, en esta tesis se explorará el avance en predictores de fuga con el fin de hacer más rentables las estrategias de retención a través de la focalización de recursos en clientes que se presenten como más propensos a la fuga.

Si bien en el lenguaje cotidiano existe una concepción de lo que es la **fuga de un cliente**, la sistematización de este concepto es difícil ya que involucra entender profundamente el

modelo de negocios de cada empresa. Para efectos de esta tesis, se tomará definición inicial de Kamakura et al. (2005) para fijar un punto de inicio; “*Se denomina fuga a la tendencia de un cliente para desertar o dejar de hacer negocios con una empresa*”.

Dentro de esta definición, existen dos elementos que debemos ser capaces de medir. Por un lado, valorar lo que significa “*hacer negocios*”: ya que cada empresa tiene un modelo de generación de valor específico. Y por otro, encontrar una métrica que sea capaz de entender lo que significa “*tendencia de hacer negocios*”, ya que existen clientes con patrones de compra disímiles. Así:

- Para lo primero, es necesario contar una métrica capaz de medir los negocios que una empresa en un horizonte temporal específico: En este estudio se utilizarán dos métricas de valorización, la primera, **Valor de Vida del Cliente** –*CLV*, por sus siglas en inglés– gracias a su flexibilidad y dinamismo (Gupta et al, 2006). Y la segunda, **Recencia, Frecuencia y Monetización** –o *RFM*–, debido a ser la más utilizada por las empresas (Wei et al, 2010).
- Lo segundo, se construirá sobre el trabajo de Glady et al. (2009) y Ljunghed (2016) en donde se trabaja sobre los delta del *CLV* para determinar si un cliente es candidato a fuga o no.

1.2 En busca del futuro

Anteponerse a la fuga de un cliente es un desafío transversal para todas industrias, ya que – como fue expuesto en la introducción– su beneficio es enorme. La literatura en gestión empresarial determina que la retención es una de las actividades con mejor retorno para las empresas. Según Chen et al. (2013) para las industrias tradicionales, un 5% en la disminución de fuga puede aumentar un 20% la rentabilidad de la empresa. Esto ha hecho que una parte importante del avance logrado por las tecnologías de aprendizaje de máquina, haya sido en la predicción de fuga: donde estos métodos han logrado superar a aquellos que se basan en la estadística tradicional (Lemmens y Croux, 2006).

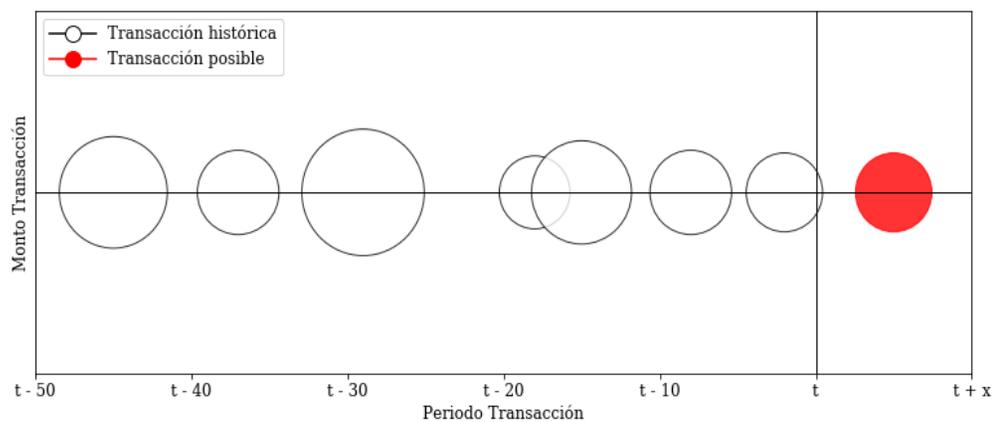


Figura 2: : Ilustración representativa de una predicción de compra.

Fuente: elaboración propia

Si bien la literatura está avanzada en este concepto, actualmente existe una **brecha entre los modelos de negocio B2B y B2C**: El origen de ésta –según Wiersema (2013)– se basa en que:

“(…) los modelos de Machine Learning, en contextos de venta orientados a empresa, no se han empujado con el mismo entusiasmo: los motivos son variados pero principalmente por la dificultad de encontrar volúmenes de datos que importantes para alimentar modelos de Big Data.”

Un enfoque particularmente poderoso para la predicción de fuga dentro del aprendizaje de máquina son las Redes Neuronales Artificiales. Un enfoque que está tomando mayor importancia dentro de estas, son las **Redes Neuronales Recurrentes**: Su gran virtud radica en que son capaces de almacenar información del contexto temporal dentro del análisis de series de tiempo. Si bien los beneficios han sido evidenciados hace más de 30 años, es sólo que hasta hace poco, que se han convertido en el estado del arte de los modelos predictivos temporales. Esto se debe al avance en las arquitecturas, los modelos de entrenamiento, y la capacidad de paralelizar los cómputos. (Lipton et al, 2015)

1.3 Objetivos y preguntas de investigación

Esta tesis busca contribuir a la literatura sobre la predicción de fuga de clientes, a través del uso de modelos de aprendizaje de máquina aplicado a la industria *B2B* –o venta entre empresas. Debido a la configuración del problema, se investiga sobre la aplicación del concepto de Valor de Vida del Cliente en series de tiempo con el fin de extender esta métrica base, sobre la cual generar un modelo de Redes Neuronales Recurrentes para su predicción. Así el **objetivo** principal de esta investigación es:

- Idear un modelo capaz de predecir oportunamente qué cliente es un candidato a fuga en el contexto de venta *B2B*.

Con los objetivos anteriores expuestos, las **preguntas de investigación** asociadas son

1. ¿Es posible modelar la propensión a la fuga de clientes –disímiles entre si– a través del uso de *RNN*?
2. ¿Ambas definiciones de valorización de cliente –*CLV* y *RFM*– se comportan igual ante el predictor?

Finalmente, la **hipótesis** del trabajo se constituye de la siguiente manera:

1. Ante la primera pregunta de investigación, se propone la hipótesis: Se puede elaborar un modelo, que mediante la utilización de Redes Neuronales Recurrentes, pueda tener una predicción de fuga que tenga un rendimiento de AUC mejor que aleatorio.
2. Ante la segunda, se propone la siguiente hipótesis: Ambas definiciones deberían tener un rendimiento relativamente similar.

1.4 Caso de estudio

Para realizar el modelo se utilizará la información transaccional de *Dimerc Office*, una empresa de venta y distribución de insumos de oficina, situada en Chile y Perú con una cartera de clientes anual que supera los 25.000 y 10.000 respectivamente. Esta empresa tiene solamente clientes-empresa, por lo cual es B2B y su monto de ventas es cercano a los 220 MUSD anuales. Para el estudio se tomaron cerca de 10.000 *clientes-empresa* chilenos, que pertenecen al segmento de Pequeñas y Medianas según la información registrada en el Servicio de Impuestos Internos. Para el análisis de compras se utilizó la información de facturación total entre los años 2014 y 2017, que es una fuente de datos fidedigna y confiable, ya que es la base para la declaración tributaria de cada cliente.

Todos los productos que se transan en este estudio corresponden a aquellos que generalmente tienen poco valor y costo para el cliente lo que refleja una relación transaccional no contractual con dichos clientes.

2 MARCO TEÓRICO

En este capítulo se presenta la revisión y comparación de la literatura y teoría que existe en la predicción de fuga. Particularmente se exploran conceptos de *marketing* estratégico e ingeniería industrial.

2.1 Valorización de los clientes

Para la gestión estratégica de clientes, la valorización de estos es un elemento preponderante, y cuantificar su valor se ha abordado múltiples ángulos. En esta sección se expondrá el concepto de Valor de Vida del Cliente (*CLV*) y la segmentación de clientes en función de su Recencia, Frecuencia y Monetización (*RFM*).

2.1.1 Valor de vida

Para solucionar el problema anteriormente enunciado, uno de los aspectos esta centrales de esta tesis, es trabajar con el concepto de *Customer Lifetime Value* o *CLV* para poder tener una métrica de valorización de la relación de negocios de un cliente. Según Blattberg et al. (2008) se puede entender como:

El valor presente de todos los ingresos relacionados a un cliente específico, una vez que éste ha sido adquirido, después de descontar todos los gastos de venta, marketing, producción y servicio dentro del ciclo de vida del cliente.

Si bien las interpretaciones de cada uno de estos valores, son propias de cada negocio –con variaciones entre industrias y empresas– es una métrica ampliamente utilizada, entendida y escalable. Para materializar esta definición utilizaremos la ecuación descrita en (Gupta et al, 2006):

$$CLV = \sum_{t=1}^H \frac{(p_t - c_t) * r_t}{(1+i)^t} - AC \quad (1)$$

Donde:

- p_t Beneficio neto del cliente en el periodo t
- c_t Costo de servir al cliente en el periodo t
- t_t Probabilidad de que el cliente haga una compra en en t
- i Tasa de descuento del capital
- AC Costo de adquisición del cliente
- H Horizonte de tiempo

Como el objeto del estudio se basa en empresas consolidadas, con clientes que ya se encuentran activos, el valor de adquisición es igual a cero, pues este ya se utilizó. Además, se denominará CF_t – debido a las siglas en inglés de *Cash Flow*– a la resta $(p_t - c_t)$, que es el ingreso de dinero que genera la transacción para en el instante de tiempo t .

$$CLV = \sum_{t=1}^H \frac{r_t}{(1+i)^t} * CF_t \quad (2)$$

2.1.1.1 Tasa de descuento

Para efectos de este estudio, el valor de la tasa de descuento i será tomada de la tasa de descuento utilizada en Chile para las empresas de similares a la del caso de estudio. El cálculo preciso de esta, no está dentro de los alcances de esta tesis, ya que al ser escalas de tiempo pequeñas, variaciones porcentuales no inciden mayormente dentro de la efectividad del modelo predictor.

2.1.2 Segmentación según Recencia, Frecuencia y Monetización

Otra forma de medir el valor de un cliente, es a través de la segmentación de Recencia, Frecuencia y Monetización. De acuerdo a algunos autores, esta aproximación es la más utilizada por las empresas, de manera transversal en las distintas industrias (Wei et al, 2010). Según las definiciones del mismo autor, se entiende cada uno de los parámetros base para la segmentación como:

- **Recencia:** Corresponde a la cantidad de periodos que han pasado desde la última compra. En otras palabras, es el intervalo de tiempo que ha pasado desde la última compra (sean días o meses).
- **Frecuencia:** Es la cantidad de compras que se han realizado en una ventana de tiempo específica.
- **Monetización:** Responde a la cantidad media de dinero gastado por el cliente en la misma ventana de tiempo.

Si bien existen numerosas formas de interpretar estos datos, en función del tipo de empresa o modelo de negocios que se quiera estudiar, en esta investigación será utilizado el modelo de Cheng & Chen (2009):

La metodología se basa en escalar las variables de recencia, frecuencia y monetización en parámetros discretos que toman valores del 1 al 5, en función del valor del quintil al que representan. De esta manera, para los valores de frecuencia y monetización, quedan cinco conjuntos con igual número de clientes. Donde el valor es igual a 5 lo tienen aquellos compradores que se encuentran en el 20% superior, según su contribución o número de compras realizadas. Por el contrario, aquellos clientes que tienen un tiempo desde la última compra menor, se encuentran en el segmento de recencia más alto, que es igual a 5 y aquellos que tienen un período desde la última compra mayor, se encuentran en el segmento valorado como 1.

De esta manera se obtienen tres valores para cada cliente, en cada periodo o horizonte de estudio.

Tabla 1: Escalamiento de los atributos *RFM* como quintil del número de clientes.

Valor	R - Recencia (%)	F - Frecuencia (%)	M- Monetización (%)
5	80-100	0-20	0-20
4	60-80	20-40	20-40
3	40-60	40-60	40-60
2	20-40	60-80	60-80

Finalmente se yuxtaponen los valores correspondientes a cada uno de los valores que se obtienen de *Recencia*, *Frecuencia* y *Monetización*, con lo que se obtiene un segmento específico. De modo ilustrativo, ténganse en cuenta el siguiente ejemplo: En una determinada empresa existe un cliente que compró hace poco tiempo (por lo que pertenece al quinto quintil de *Recencia*, que representa a los clientes más recientes). Además, este cliente suele comprar muy seguido (de esta manera, pertenece al quinto quintil de *Frecuencia*, donde están aquellos clientes que compran con mayor recurrencia), pero generalmente compra montos muy bajos (lo que hace que su quintil de *Monetización* sea el primero). De esta manera, este es un la clasificación $RFM = (5,5,1)$.

2.2 ¿Qué es la fuga?

Como se mencionó en la introducción, en el desarrollo de esta tesis se comprenderá la fuga como “la tendencia de un cliente para desertar o dejar de hacer negocios con una empresa”. (Kamakura et al, 2005) Pero esta definición deja algunos espacios a interpretación del lector, ya que, en última instancia ¿Qué significa dejar de hacer un negocios? ¿Son todos los negocios iguales?

En los párrafos que suceden se explicarán detalladamente algunos elementos importantes para el desarrollo del concepto.

2.2.1 Tipo de relación comercial

El primer punto para construir una definición de fuga, es determinar existe un hito de fuga: Por ejemplo, darse de baja de una suscripción o cerrar una cuenta de banco. Estos casos se dan cuando la relación comercial está en un contexto **contractual** y su cancelación marca el momento de fuga.

En otros –como el *retail*– dicho hito no existe, y su definición se vuelve más difusa. A modo ilustrativo: ¿Después de cuántos días que un cliente no compra en un supermercado se considera fugado? ¿Cuántos productos *menos* tiene que comprar para que se considere fugado?. Cualquiera sea la respuesta para estos escenarios, comienzan a presentar problemas cuando se enfrenta a clientes disímiles en su comportamiento. Para profundizar aún más lo desafiante de la idea, se presentan los siguientes ejemplos:

Un cliente de una empresa de *retail*, ha comprado durante varios meses un *mix* de 4 productos para el hogar. Durante sus dos últimas compras ha dejado de comprar dos de los productos de dicho *mix*.

- ¿Es este un cliente fugado?
- ¿Es un cliente *más fugado* que uno que siempre ha comprado los dos productos que adquirió en sus últimas compras?
- Un cliente que compra cada quince días durante un varios años, pero que no ha comprado hace seis meses: ¿Es, en comparación, un cliente *más fugado* que uno que compra solamente en Navidad y llevan también meses sin comprar?

Escenarios como los anteriores buscan ilustrar que la definición de fuga no necesariamente es única –incluso para un mismo modelo de negocios– sino que depende, además, del **comportamiento de cada cliente**. Ejemplos como los anteriores se dan en contextos **no-contractuales**. En otras palabras, donde no existe un evento que defina la fuga, sino que son determinaciones que se tienen que tomar desde un punto de vista de la gestión estratégica, tomando en cuenta el contexto donde se desarrolla el negocio.

De esta manera, al existir un contexto **no-contractual** se tiene la necesidad de definir reglas de negocio medibles que determinan cuándo un cliente será considerado fugado. Esta es la base sobre la cual construir modelos de predicción de fuga y sobre estos, estrategias de retención más efectivas.

2.2.2 Distintos enfoques de valorización de la fuga no contractual

Esta tesis se enmarca dentro aquellas relaciones comerciales que no están mediadas por un contrato, particularmente aquellas que son recurrentes y suceden reiterativamente entre un

cliente y un proveedor, por este motivo el estudio cobra alta importancia en la venta orientada a empresas o B2B. De esta manera, se entiende que **la predicción de fuga** es el proceso de cálculo de la probabilidad de que un cliente en la base de datos de una empresa, tenga un futuro comportamiento o conducta que refleje la fuga, a través de un modelo predictivo: que utiliza información del pasado o comportamiento histórico (Coussement & De Bock, 2013). Como consecuencia, la fuga se entiende como un concepto dinámico y específico para cada cliente: donde cada instante de tiempo t que transcurra, aporta información importante para del comportamiento de un cliente. Con esto se logra tener más información valiosa para el manejo de las estrategias CRM de la empresa; pudiendo construir estrategias de retención oportunas y más efectivas.

2.2.2.1 La fuga según el Valor de Vida del Cliente

Anteriormente, se evidenciaron las virtudes del CLV para medir la valorización de un cliente. En esta tesis, se calculará esta métrica se periódicamente: en cada instante de tiempo t y para cada cliente con el fin de medir la evolución de cada uno de estos durante un periodo de estudio. Con la definición de Kamakura et al (2005) anteriormente descrita y construyendo sobre el trabajo de Ljunghed, (2017), se puede entender que al existir una evolución negativa, de su valor estamos en un cliente fugado, tal como se muestra en la Ecuación 3.

$$CLV_t = \alpha_t * CLV_{t-n} \quad (3)$$

De esta manera, dependiendo del valor que tome la variable t , se puede determinar si el cliente es candidato a fuga en el periodo t .

$$fuga_t = \begin{cases} 1, & \alpha_t < 1 \\ 0, & \alpha_t \geq 1 \end{cases} \quad (4)$$

2.2.2.2 Según la segmentación R-F-M

La lógica para definir la fuga de un cliente según los parámetros de Recencia, Frecuencia y Monetización, es análoga a la que se usó anteriormente. Primero, para cada instante de tiempo t , y para cada cliente, se hace un cálculo de sus parámetros de RFM sobre un horizonte de tiempo. Con esto, se obtiene una valoración según estos parámetros.

Segundo, se tiene que evaluar si entre t y $t+1$ hubo un cambio en el valor de RFM . Para lograr esto, hay que resolver una valorización para cada segmento. En otras palabras ¿Los clientes $RFM_t = (1,5,5)$, $RFM_t = (5,1,5)$, y $RFM_t = (5,5,1)$, son igualmente valiosos para la empresa?

Con el fin de resolver esto, en el trabajo de Cheng y Chen (2009) se expone que desde los años 90, la discusión sobre el peso de cada variable en la segmentación final ha sido una discusión importante en la literatura. Sin embargo, concluye que el peso de las tres variables debiese ser igual.

De esta manera –por ejemplo– se puede definir como cliente fugado a un cliente que en el periodo t se segmenta como $RFM_t = (5,4,5)$ y en el periodo $t+1$ cae en el segmento de $RFM_{t+1} = (5,4,4)$. Ya que este decayó su monetización, y no así, si este mismo cliente hubiese sido segmentado como $RFM_{t+1} = (5,4,4)$ ya que si bien su monetización bajó, su frecuencia se vió incrementada, por lo que este no es considerado una fuga, sino un cambio de comportamiento.

$$fuga_t = \begin{cases} 1, & RFM_t < RFM_{t-n} \\ 0, & RFM_t \geq RFM_{t-n} \end{cases} \quad (5)$$

2.3 Predicción en series de tiempo

Poder predecir eventos es uno de los desafíos centrales de la estadística tradicional (Lee et al, 2010) El interés sobre este tema abarca una miríada de aplicaciones: como la predicción del tiempo entre terremotos, recesiones económicas, el tiempo de vida que le queda a un paciente que recibe un determinado tratamiento o el tiempo antes de que una máquina falle (Martinsson, 2016). Para el caso de estudio de esta investigación, los eventos de interés son compras.

Como se busca predecir la si existirá una fuga de un cliente en un período específico –un mes, semana, día o segundo determinado– sólo se cuenta con la información histórica disponible en ese momento: las cuales pueden ocurrir múltiples veces para cada cliente. Lo particularmente importante es el tiempo que ha pasado desde el último evento hasta el momento que se quiere hacer una predicción, ya que la información es incompleta: No se sabe cuánto tiempo más pasará para la siguiente compra, ni si quiera si es que la habrá. Esto se se conoce como información **truncada** o **censurada** (Klein y Moeschberger, 2003). Particularmente, para el caso descrito se entiende como **información censurada por la derecha** ya que no sabemos qué pasará a la derecha –que significa *después*, según eje de tiempo X– del instante de corte.

2.3.1 Tiempo para el siguiente evento

Para el caso de este estudio, parte importante del problema se basa en modelar el tiempo que pasará para la siguiente compra, si es que esta existe. Para la literatura específica en *predicción de fuga* existen distintas formas de aproximarse al problema: Transversalmente, este se reduce a poder predecir –si existe– una siguiente instancia de compra, teniendo solamente la información histórica que se tiene en un instante de tiempo. Esta información la podemos dividir en dos tipos:

2.3.1.1 Información contextual

Información de las características del cliente, así como: tamaño, ubicación geográfica, nombre entre otros. Esta información es relativamente estática, ya que sólo cambia en horizontes de tiempo relativamente largos. Para efecto de este estudio serán considerados como información estática.

2.3.1.2 Información transaccional

La *información transaccional* es aquella que se puede obtener al momento de ejecutar una interacción entre las partes involucradas. En el caso de las empresas de *retail*, esta incluye información sobre el momento en que se realiza la compra, el canal, el tipo de productos

comprado, así como también datos comerciales; como el precio y el costo de venta. Además, se puede recabar información más allá de la venta misma: por ejemplo, si la interacción es a través de un comercio electrónico, este canal puede agregar información de comportamiento, como la interacción con el sitio web: La cantidad de veces que se visitó un producto, si agregó al carro o la interacción con una pieza de *marketing*.

Independiente de la forma, este tipo de información es dinámica y va cambiando con cada interacción que tiene el cliente con la empresa. Cabe destacar que esta información, como tal, no es valiosa de por sí, ya que debe ser procesada e interpretada.

2.3.2 Análisis con datos censurados

El desafío principal que presenta la **información censurada por la derecha** es que desde el último evento que se tiene información hasta el momento en que se busca hacer la predicción del siguiente, sólo se tiene información sobre la ausencia de eventos. Esto requiere el uso de técnicas específicas para lidiar con este tipo de problemas

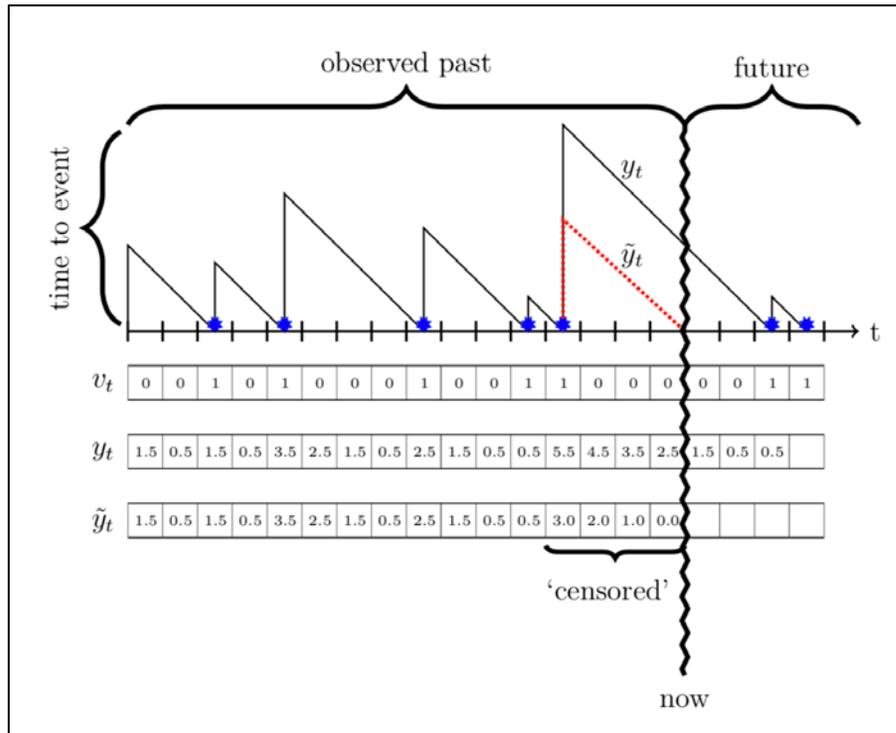


Figura 3: Ilustración que explica el concepto de TTE sin censura y_t , y la predicción \tilde{y}_t con censura
Fuente: Martisson, 2016.

Para el caso de la fuga de clientes, el evento de interés son las compras recurrentes realizadas por un cliente. Se tiene información histórica de cómo cada cliente se ha comportado anteriormente y se busca ser capaces de predecir –de manera oportuna– cuándo un cliente tendrá un TTE más alto que el anterior – o incluso infinito– lo que se puede interpretar como un candidato a fuga.

2.3.3 La distribución de Weibull

El objetivo que se busca con este modelo es predecir el *tiempo al siguiente evento* para un evento de interés particular que ocurrirá – o no– en un futuro. En particular, el evento es la posible compra que efectuará un cliente. Este modelo fue introducido el año 2016 por Martinsson. Y para entenderlo lo haremos en dos partes:

2.3.3.1 Distribución de *Weibull* como función de supervivencia.

La distribución de *Weibull* ha sido ampliamente usada en el análisis de supervivencia debido a su flexibilidad, particularmente para su uso como función que modela tasa de falla. En particular, se asume que el tiempo al siguiente evento distribuye:

$$\text{TTE} \sim \text{Weibull}(\alpha, \beta) \quad (6)$$

Algunas de las parametrizaciones más relevantes para la función de *Weibull* se muestran a continuación:

$$F(x) = 1 - \exp\left[-\left(\frac{t}{\alpha}\right)^\beta\right] \quad (7)$$

$$f(x) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^\beta\right] \quad (8)$$

$$\Lambda(x) = \left(\frac{t}{\alpha}\right)^\beta \quad (9)$$

$$\lambda(x) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1} \quad (10)$$

Donde:

- Ecuación 7: corresponde a la Función de Densidad de Probabilidad –o *PDF*– por sus siglas en inglés: *Probability Density Function*.
- Ecuación 8: hace referencia a la Función de Densidad Acumulada –o *CDF*– por sus siglas en inglés: *Cumulative Distribution Function*.
- Ecuación 9: se refiere a la Tasa de Falla Acumulada –o *CHF*– por sus siglas en inglés: *Cumulative Hazard Function*.
- Ecuación 10: sistematiza la Tasa de Falla –o *HF*– por sus siglas en inglés: *Hazard Function*.

Mientras que los parámetros de cada una de las funciones anteriormente enunciadas :

- α Se denomina parámetro de escala, en el contexto de esta tesis, determina **cuándo** ocurrirá un evento.

- β Se denomina parámetro de forma o pendiente de Weibull. Para efecto de la predicción de *TTE*, está relacionado al **qué tan seguro está el modelo de la predicción.**

Para efectos de esta tesis, las dos parametrizaciones más importantes son la función de distribución de probabilidad (*PDF*) y la tasa de falla (*Hazard Function*). Con este fin se anexa la Figura 4, donde se puede entender por qué la distribución Weibull es tan importante en el uso de funciones de supervivencia. Como se puede observar, para la tasa de falla, tenemos que la función se comporta de distintas maneras: si $\beta > 1$ la distribución queda descrita con una función monótona creciente; si es $\beta < 1$, esta es monótona decreciente y si $\beta = 1$ la función es constante.

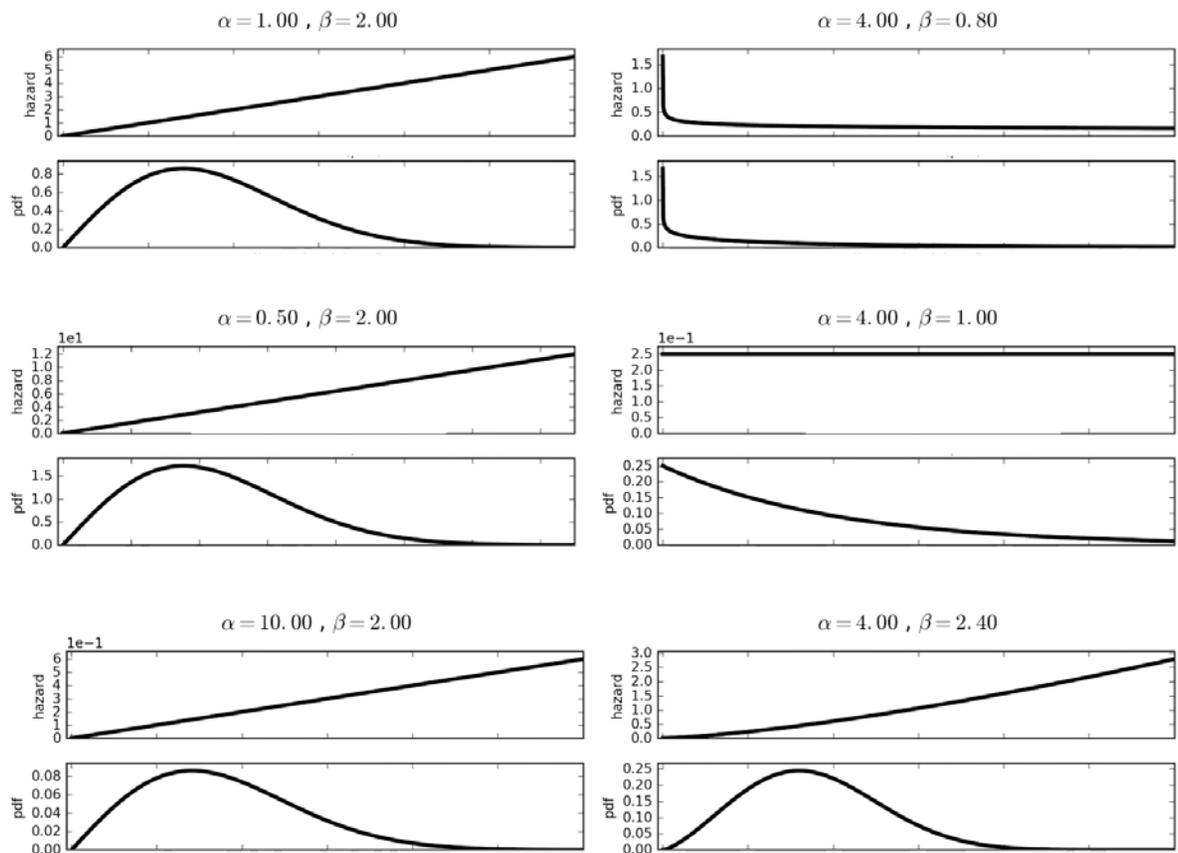


Figura 4: Gráfica de la Función de Distribución de Probabilidad (*pdf*) y la Tasa de Falla (*hazard*) para distintos valores de α y β . Fuente: de Martisson 2016

Gracias a esta flexibilidad, la función de *Weibull* ha sido ampliamente utilizada para la elaboración de modelos de supervivencia.

2.4 Redes Neuronales Recurrentes

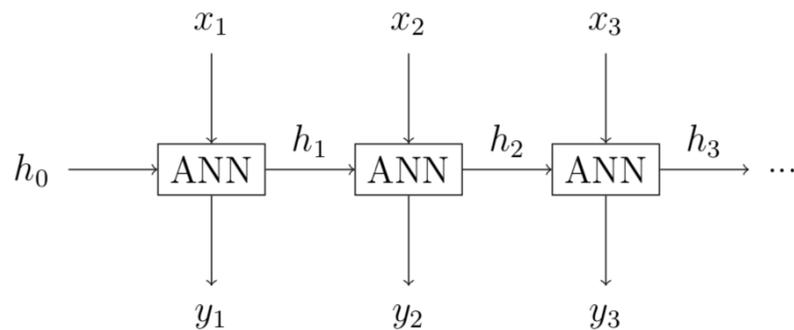


Figura 5: Diagrama de una Red Neuronal Recurrente genérica, donde se aprecia que los parámetros de cada estado (x_t) y la información del estado anterior (h_t) son los los parámetros de entrada de cada paso. Fuente: Martinsson 2016.

Las **Redes Neuronales Recurrentes** (*RNN*) son una aplicación de una Red Neuronal Artificial (*ANN*) cuya virtud específica es que son capaces de almacenar un estado interno temporal (Kasiran et al, 2018). Por esta razón son utilizadas en interpretación de elementos secuenciales o contextuales, entre los cuales destacan el modelamiento de problemas de series de tiempo y interpretación del lenguaje natural.

Una aplicación de estas, desarrollada por Martinsson (2016) es utilizarlas para poder predecir los parámetros α_t y β_t de una distribución *Weibull* que determina el tiempo que transcurrirá hasta un siguiente evento, que para los alcances de este trabajo, es una compra. Así, se logra tener para cada instante específico, una función que modela el tiempo que pasará hasta la siguiente compra, denominado *TTE*, para efectos de este trabajo.

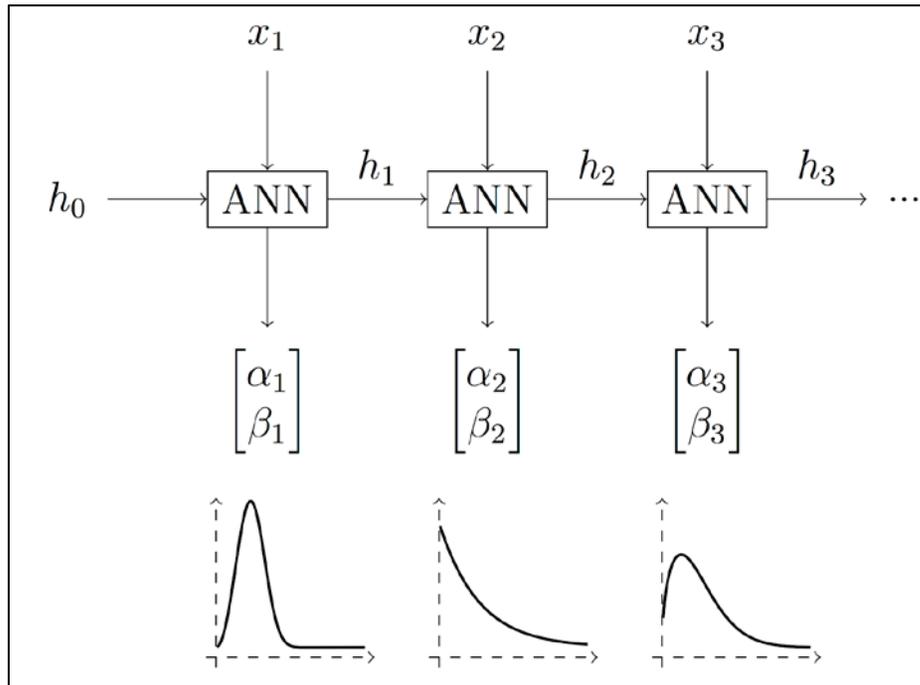


Figura 6: Estructura de una WTTE-RNN. Se muestra como parámetros de salida para cada instante t , los valores α_t y β_t , que determinan la distribución Weibull. Fuente . Fuente: Martinsson, 2016

De esta manera, para cada instante de tiempo, se tiene una distribución de probabilidad de *cuándo* y *qué tan seguro* está el modelo del TTE para la siguiente compra de cada cliente. Además, este modelo tiene las siguientes características útiles para la predicción de fuga:

- Puede manejar tiempo discreto o continuo: Para el caso de esta investigación se utilizará solamente el tiempo discreto, donde el paso será semanal.
- Entrenarse con información censurada: El modelo expuesto por (Martinsson, 2016) tiene la virtud de ser capaz de utilizar la información censurada dentro de su entrenamiento. Esto le agrega robustez al modelo.
- Capaz de reconocer parámetros temporales con un horizonte de largo plazo: Es capaz de reconocer tendencias de compra de cada cliente en específico al momento de hacer la predicción

El modelo tiene las siguientes suposiciones:

- Asume que el tiempo entre eventos distribuye *Weibull*

- Asume que censura no es informativa.

Ambas suposiciones tienen sentido en el contexto en el que se describe el problema, ya que por un lado, la distribución de *Weibull*, por su flexibilidad en el análisis de supervivencia es aplicable al modelamiento del tiempo a la siguiente compra, y por otro, la información de censura no presenta una información en sí misma.

3 METODOLOGÍA

En esta sección se describe el modelo que fue elegido para su implementación. Con el fin de lograr un predictor de fuga de clientes en el contexto de una empresa de *retail* orientado a empresas – o *B2B*–. También se describen los datos que fueron utilizados.

3.1 Descripción de los datos

Los datos utilizados para el desarrollo de esta tesis devienen del ejercicio comercial de la empresa *Dimerc Office Chile*. Cuyo fin es ser analizados para implementar un sistema que sea capaz de predecir oportunamente la fuga de sus clientes-empresa. La información sobre la cual se construyó este modelo, se comprende dos fuentes de datos:

i- Información transaccional

Para el desarrollo del modelo se utilizará un *set* de datos es generados por las ventas realizadas durante un periodo de 4 años completos: comprendidos de enero del 2014 y diciembre de 2017. Además, para la validación de este modelo, se utilizará un set de datos que incluye los primeros 6 meses de venta de la empresa durante el 2018, lo que comprende desde el 1 de enero al último día de junio.

ii- Información contextual

Los datos de caracterización de la empresa son obtenidos del *CRM* que maneja la empresa. Estos contemplan números de empleados, facturación histórica, clasificación según el Servicio de Impuestos Internos, entre otras.

Esta información comprende un set de datos de más de 25 millones de líneas de venta, cerca de 3 millones de transacciones asociadas a cerca de cuarenta mil clientes-empresa.

3.2 Modelo propuesto

El modelo que se implementa en esta investigación se basa en generar un predictor de fuga que se basa en el cálculo de valor de vida de un cliente o *CLV* de manera periódica. Este

valor se calcula a través del modelo de wtte-rnn de (Martinsson, 2016): donde una Red Neuronal Recurrente que es capaz de reconocer los parámetros de comportamiento únicos de cada cliente y es capaz de predecir los valores α_t y β_t una distribución de probabilidad *Weibull* que determina el tiempo que pasará hasta una siguiente compra. Con esta distribución, el modelo es capaz de obtener la probabilidad de realización de una compra en un horizonte de tiempo específico. Finalmente, con dicho valor se puede determinar cuándo un cliente es candidato a fuga, dependiendo de la tendencia en su proyección de *CLV*.

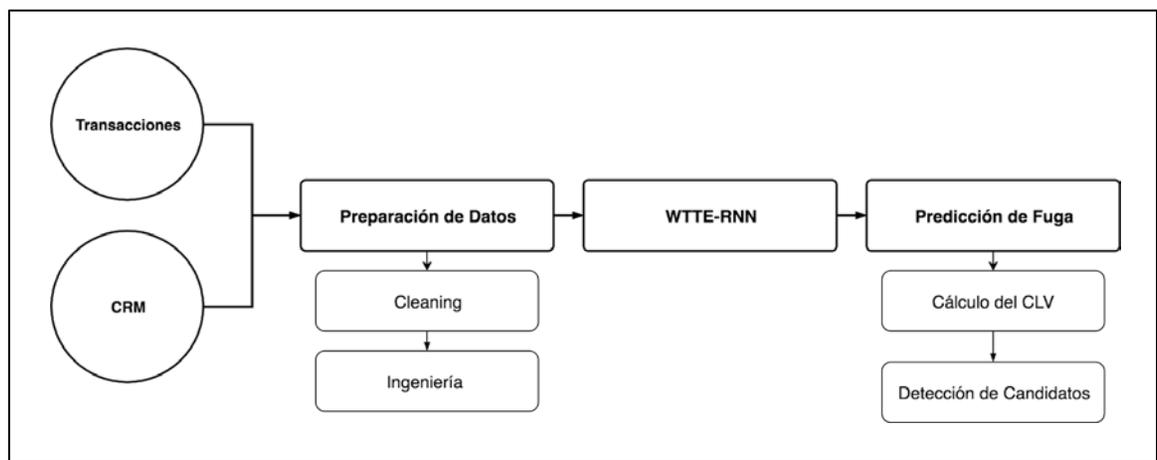


Figura 7: Gráfico de flujo del Modelo predictor

3.3 Etapas

3.3.1 Preparación de Datos

3.3.1.1 Limpieza de datos

El objetivo de esta tesis es obtener un predictor de fuga confiable y dinámico en el contexto en que se inscribe. De esta manera, se siguieron los siguientes criterios para su modelamiento y medición solamente:

- Se incorporan al estudio sólo aquellos clientes pertenecientes al segmento de pequeñas y medianas empresas. Esta razón es porque aquellas que son de mayor tamaño, tienden a tener un modelo de negocios que sigue la lógica contractual, lo que está fuera del alcance de este trabajo.
- Se incorporan en el modelamiento, aquellos clientes que son considerados **compradores recurrentes**. De esta manera, sólo fueron incluidos aquellos que presentaron al menos 5 compras dentro del periodo de estudio.
- Se consideran en este estudio aquellos clientes que presentaron al menos una compra dentro de los últimos tres meses del periodo de estudio, con el fin de que aquellos que ya se presentaban como fugados, no afecten las métricas de evaluación.

Con los criterios antes descritos se logra tener una base de datos de cerca de diez mil clientes para ser analizados.

3.3.1.2 Ingeniería de características.

Para poder implementar este modelo, se deben preparar las fuentes de datos de manera que sean utilizables por los modelos seleccionados. Esto se hizo a través de dos transformaciones importantes.

3.3.1.2.1 Agregación de datos

Para que el modelo de *wtte-rnn* descrito por Martinsson (2016) funcione correctamente, debe ser alimentado con series de temporales discretas. De esta manera se realizó una **agregación semanal de datos** de compras. El motivo de esta selección radica en que la semana es el comportamiento más recurrente de los periodos de re-compra, como se puede ver en el siguiente gráfico.

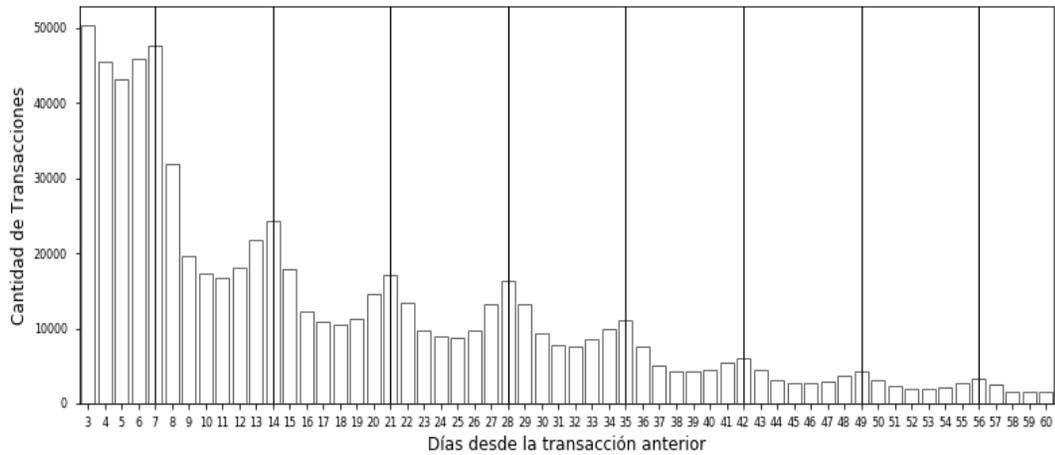


Figura 8: Gráfico de barras que muestra la cantidad de transacciones que tienen la misma cantidad de días desde la última compra. Fuente: Elaboración Propia.

3.3.1.2.2 One-hot-encode

Una vez definido el intervalo de tiempo que se elige para el análisis, se necesita tener la fecha de compra con la codificación *one-hot* para el período específico: De esta manera se obtiene una matriz cuyas columnas representan instantes discretos y sus valores son 1 o 0, según si hubo o no una compra en ese periodo. Particularmente, en el estudio se agregaron los datos de manera semanal.

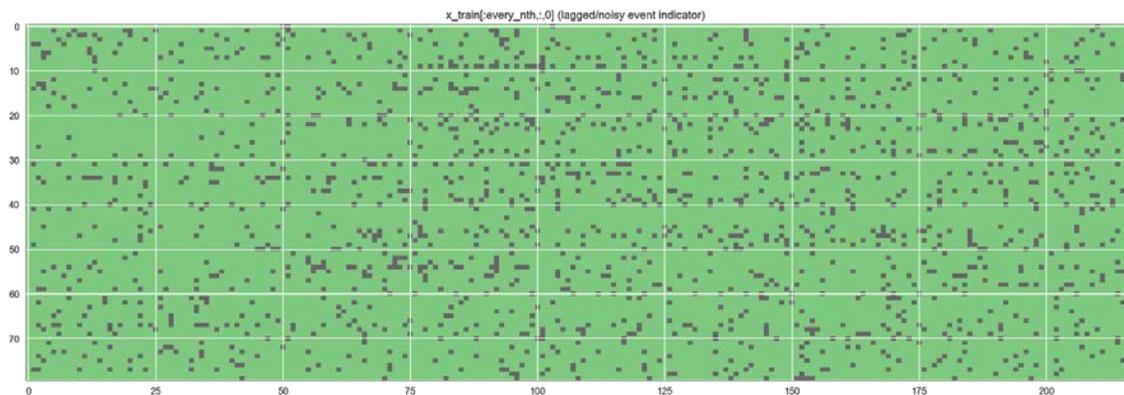


Figura 9: Matriz de one-hot-encoded de las compras realizadas por 80 clientes elegidos al azar. En el eje x se ven los 220 periodos que se utilizaron para entrenar el modelo; en el Y son 80 clientes elegidos aleatoriamente. Los cuadros grises representan 1's en la matriz *one-hot-encoded*.

3.3.1.2.3 Valor objetivo

Para el entrenamiento del modelo, se necesita tener el tiempo al siguiente evento de cada uno de los clientes. Así, se obtiene una matriz, para cada instante t , para cada cliente que contiene el tiempo que falta para el siguiente evento –de compra, para los alcances de la investigación–. Esta es posible visualizarla de como muestra la Figura 10, donde se tiene un mapa de calor para los periodos que faltan para la siguiente compra.

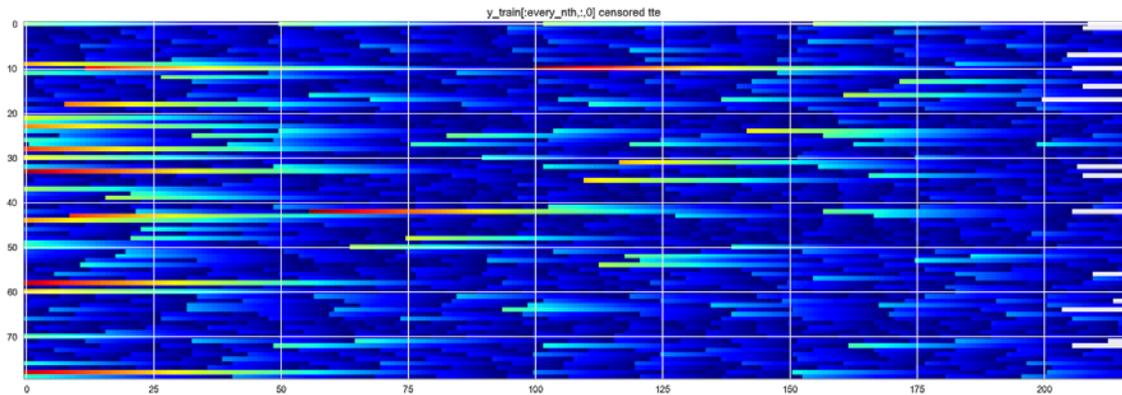


Figura 10: Mapa de calor del Tiempo al siguiente evento de los mismos clientes de la figura anterior. En esta visualización, mientras más rojo es el pixel, mayor es el TTE para ése cliente.

3.3.2 WTTE-RNN

El modelo de *wtte-rnn* descrito por Martinsson (2016) se basa en la utilización de Redes Neuronales Recurrentes para la predicción del tiempo al siguiente evento – o TTE , por sus siglas en inglés–. Esto se hace con una Red Neuronal Recurrente, cuyo *output* es un vector que contiene los parámetros α_t y β_t de una función de distribución de *Weibull*. Esta describe el comportamiento del tiempo que pasará hasta una ocurrencia del siguiente evento, en este caso: una compra. Con esta información se construye una distribución de probabilidad – ampliamente utilizada en el análisis de supervivencia– con lo que es posible obtener el valor esperado del CLV_t en cada instante de tiempo.

De esta manera, es posible calcular la probabilidad de compra en un horizonte de tiempo para cada cliente, mediante la función de probabilidad acumulada de la distribución *Weibull* que el modelo entregó para ese cliente en específico, para el instante preciso.

$$\Pr(Y_t \leq \tau) = \theta_t \quad (7)$$

Esta probabilidad de que exista una compra en el horizonte indicado, para cada periodo de tiempo, permite que sea posible aplicar la función de Ciclo de Vida del Cliente en cada instante: abriendo las puertas a analizar cambios de este valor en cada periodo.

3.3.3 Predicción de Fuga

3.3.3.1 Cálculo del *CLV*

Una vez que se tiene la probabilidad de compra en un horizonte establecido para cada cliente, es posible reemplazar la probabilidad de que haya un evento en el horizonte de tiempo establecido (θ_t), como probabilidad de compra (r_i), en el cálculo del CLV_t descrito en la **Ecuación 2**. Esto se hace asumiendo dos cosas:

- El valor de los flujos será constantes, iguales al último monto de compra, desde el periodo t+1 en adelante.
- El valor del será constante desde el periodo t+1 en adelante.

Para los otros valores se utilizaron:

- Un horizonte de tiempo = 14 semanas.
- Por los motivos que se mencionaron en el Marco Teórico la tasa de descuento que se utiliza para el cálculo será del 10%.

3.3.3.2 Detección de candidatos a fuga

Una vez que se calcula el CLV_t para cada cliente, en cada instante de tiempo t, el modelo es capaz de evaluar cambios en la pendiente de dicho valor donde un valor mayor a uno de α_t implica que es un candidato a fuga, en el horizonte de tiempo H. Esta forma de evaluar si un cliente es candidato a fuga o no, deviene del trabajo de Gladys et al. (2009) y . Para sacar la propensión a fuga, se mide durante cuántos periodos se presenta un t negativo, con lo que se pueden tener distintos umbrales, discutidos en los resultados.

3.4 Evaluación del modelo

3.4.1 Información histórica

El modelo fue entrenado con la información transaccional de clientes pertenecientes a un segmento específico explicado en la preparación de datos, quienes mostraron compras entre los meses de Enero del 2014 y Diciembre del 2017. Para la evaluación de las predicciones del modelo, se utilizaron transacciones ocurridas entre los meses de enero y junio del 2018, ya que esta nunca fue utilizada para el entrenamiento del mismo. Dentro de este periodo de evaluación, a su vez, se separa en dos horizontes temporales: el primero, el periodo de testeo—comprendido entre enero y marzo, corresponde a aquellos periodos sobre los cuales se hizo una predicción de fuga; el segundo —comprendido entre abril y junio— corresponde a los periodos que validan si es que la fuga efectivamente existió dentro del horizonte temporal.

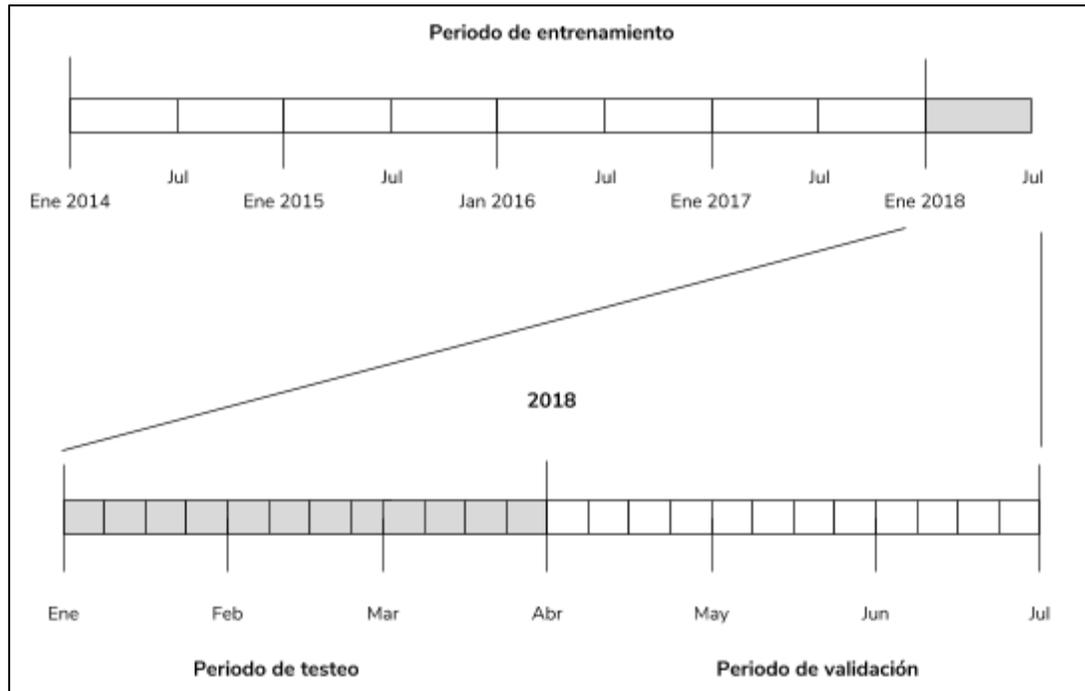


Figura 11: Línea de tiempo que ilustra la cantidad de tiempo sobre los cuales se recolectó información.
Fuente: Elaboración Propia.

3.5 Diferentes formas de evaluar el modelo

Como la definición misma de la fuga representa un desafío, se compara el modelo frente a dos definiciones diferentes de fuga anteriormente descritas. De esta manera, para el instante t se calculó el valor de CLV_t y se calculó el t de requerido para satisfacer la Ecuación 3. Con esto, se predice la propensión a fuga de cada cliente. Esta se compara con dos formas de calcular si un cliente efectivamente presentó una fuga o no, según los dos modelos anteriormente descritos.

3.5.1.1 Evaluación según CLV

En una primera instancia, se compara el valor de cambio en el CLV , α_t entregado por el modelo y se estima la propensión a fuga de un cliente en el futuro. Esto se hizo a través de

la comparación del valor de cambio obtenido en t con los valores α_{t+i} donde i es el pertenece a todos los valores del horizonte H . En caso de existir un valor negativo en el cambio de los valores del parámetro de cambio del CLV , se tiene una fuga positiva.

3.5.1.2 Evaluación según R-F-M

Para evaluar la efectividad del modelo, se utilizó el valor de segmentación de RFM para cada instante de tiempo, para cada cliente. Así, similar a la evaluación según CLV , en caso de existir una disminución en la valorización de RFM dentro del horizonte de tiempo H a evaluar por parte del cliente, este se consideró como fugado en ése periodo.

4 RESULTADOS

En esta sección se presentan los resultados obtenidos del entrenamiento del modelo y su utilización como predictor de fuga.

4.1 Valores de α_t y β_t en el modelo wtte-rnn

El modelo se basa en la utilización de una distribución de *Weibull* como predictor del tiempo hasta el siguiente evento, La Red Neuronal Recurrente, entrega dos parámetros $-\alpha_t$ y β_t para cada instante de tiempo, para cada cliente. Con estos valores, el modelo es capaz de crear una función $Weibull(\alpha_t, \beta_t)$ que describe el comportamiento del *tiempo al siguiente evento*. Como se puede ver en **Figura 12**, estas variables van cambiando su valor en cada instante de tiempo, por lo que la función de probabilidad que describen se actualiza periódicamente, con toda la información que el modelo puede adquirir.

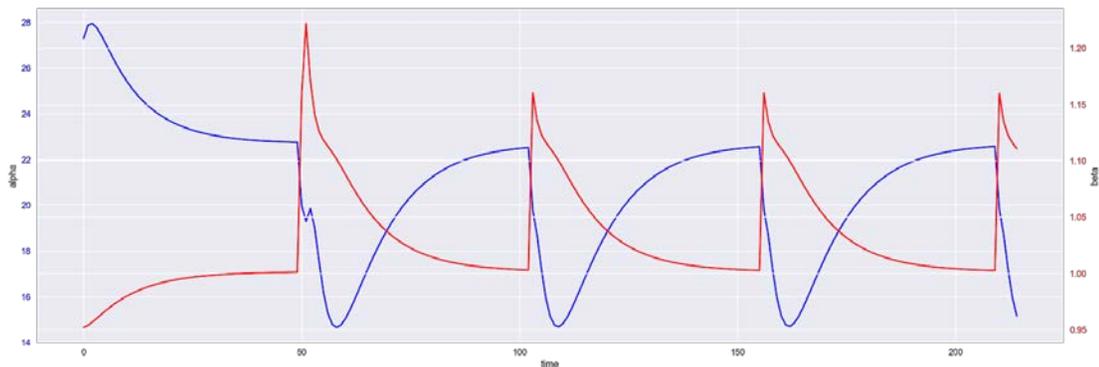


Figura 12: Ilustración que muestra cómo varían los valores de α_t y β_t a lo largo del periodo de entrenamiento para un cliente específico. Fuente: Elaboración Propia.

Para los alcances de esta tesis, se utilizó la Función de Distribución Acumulada con el fin de obtener “el valor de que exista un evento en el horizonte H ”, ya que con esta información, es posible utilizar la definición del *CLV* descrita en la **Ecuación 2**.

4.2 Validación del modelo

Una vez que se el modelo entrega los valores de α_t y β_t para la distribución, se puede calcular la *probabilidad de una compra dado el intervalo H*. Como fue explicado en la sección de información histórica, el modelo fue alimentado con datos transaccionales generados entre los años 2014 al 2017. Dicha información fue utilizada para el entrenamiento del modelo, mientras que la información del primer semestre del 2018 fue utilizada para testear el modelo. Este testeo sigue dos metodologías para validar la efectividad del uso de *wtte-rnn* en combinación con el cálculo del Valor de Vida del Cliente.

4.2.1 Como predictor de fuga según CLV

Según fue explicado en la metodología de evaluación, se contrastó la propensión a fuga de un cliente en comparación con los valores del *CLV* obtenidos durante el horizonte, así se comparó el t de la Ecuación 3 que se obtuvo a través de la predicción del Valor de Vida del Cliente, con lo que se obtuvo el siguiente gráfico de Curva Característica Operativa del Receptor o *ROC*, por sus siglas en inglés. Donde el área de *AUC* de la curva del predictor fue, en promedio de las 14 semanas a evaluar, de 0,61.

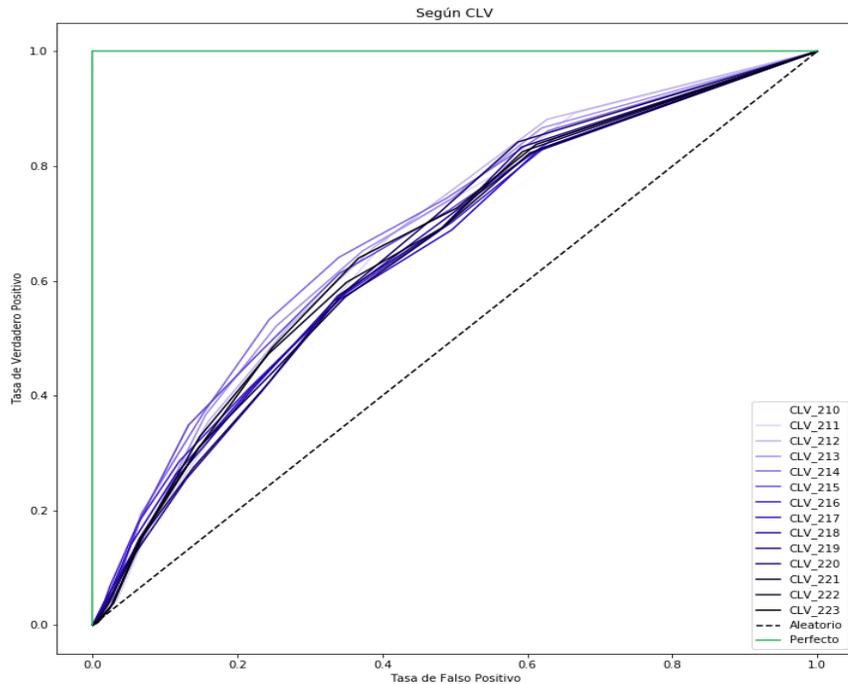


Figura 13: Curva de ROC para cada una de las semanas del períodos de predicción.

4.2.2 Como predictor de fuga según RFM

Símil al caso anterior, se determinó que un cliente presentó una fuga si hubo una disminución en su segmentación de *RFM*. En este caso, el predictor funcionó bastante mejor, logrando un *AUC* de 0.69 en promedio de los 14 intervalos de tiempo que fue utilizado.

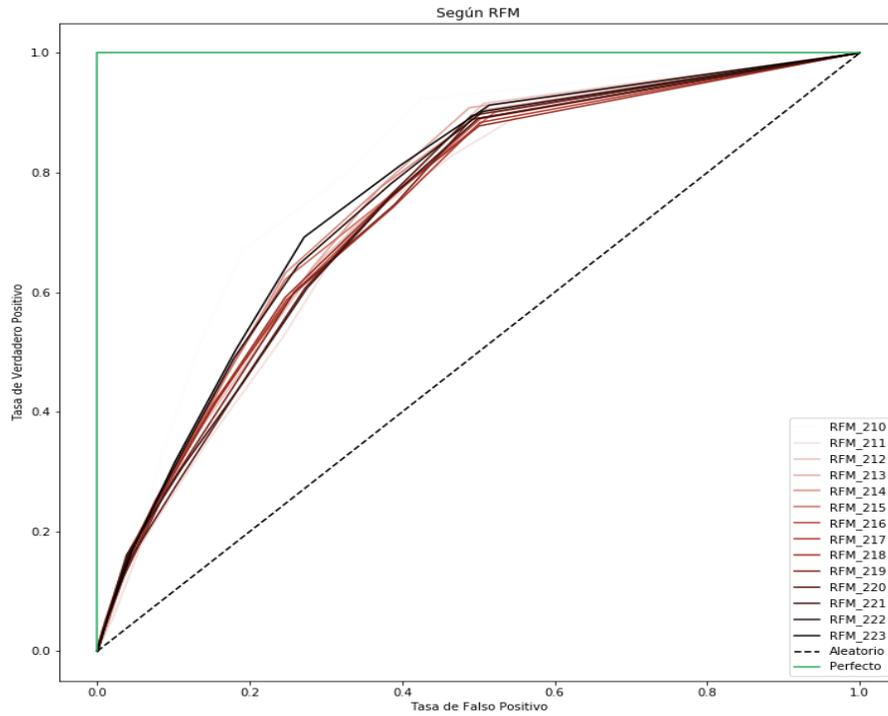


Figura 14: Curva de ROC para cada una de las semanas del períodos de predicción.

5 DISCUSIÓN

Esta sección busca darle al lector una perspectiva del significado de este trabajo, cómo contribuye a la literatura existente y cómo puede complementarse en un futuro.

5.1 El modelo como predictor de fuga

El trabajo realizado en esta tesis es un aporte a la literatura existente en la utilización de Redes Neuronales Recurrentes –particularmente del modelo *wtte-rnn* introducido por Martisson el 2016– para la predicción de fuga de Clientes. Más aún, busca profundizar en el desarrollo de modelos de Aprendizaje de Máquina en el negocio de venta entre empresas –o *B2B*– ya que como fue expuesto en la Introducción, existe una gran brecha en la literatura de Gestión de Clientes en relación al *retail* orientado a consumidores finales –o *B2C*.

Si bien el modelo presenta una mejora modesta frente a una predicción aleatoria en el caso de ser evaluado con la definición de fuga de *CLV*, la mejora es considerable si se utiliza la definición de fuga de *RFM*.

Otro aspecto relevante es que el modelo puede ser mejorado y complementado de múltiples maneras discutidas en la siguiente sección. Además de esto, tiene la virtud de ser escalable no sólo a distintas empresas del *B2B*, sino que puede ser aplicado en cualquier modelo que involucre una relación comercial *no contractual* a largo plazo con compras periódicas. Ejemplos de esto pueden ser modelos de negocio de compra recurrente y no contractual, algunos de los cuales están tomando gran importancia en Chile donde actores descriptores como *Cornershop* han tenido un posicionamiento importante.

5.1.1 Limitaciones del modelo.

5.1.1.1 Información requerida

Una de las falencias más importantes del modelo radica en que requiere mucha información para poder hacer predicciones sobre la fuga. Esto hace que sea un predictor de fuga para

clientes que ya se encuentran en la compañía. Este elemento debe ser considerado al momento de implementarse, ya que no es útil para una empresa que esté en fase de expansión, o que lleve poco tiempo en el mercado.

5.1.2 Optimizaciones del modelo

5.1.2.1 Segmentación previa

Una forma de mejorar el rendimiento del modelo, es aplicar una clusterización previa en función del comportamiento del cliente. Así, se puede obtener una mejor información respecto al patrón de compra al que pertenece un tipo específico de clientes. La mayor contribución que tendría esto, es que para compañías con clientes muy disímiles, se puede ajustar cada segmentación de manera independiente, sin correr el riesgo de sobreajuste el modelo para un resultado único sobre todos los datos.

5.1.2.2 Complementar con información más allá de la compra

Si bien el modelo no se presenta como un predictor excelente, este puede ser optimizado de múltiples maneras. Una de ellas es utilizar información del cliente más allá de la compra. Como el contexto de esta Tesis es en el rubro del *retail* orientado a empresas, la *RNN* puede ser alimentada con múltiples fuentes de información. Algunos ejemplos de información fuera del *CRM* son:

- Para empresas que transen en bolsa, se puede nutrir la *RNN* con información del valor de la acción de la empresa.
- Para empresas gubernamentales, se puede obtener información en portales de transparencia de gastos públicos. Como por ejemplo en el caso de Chile existe el sitio <https://www.mercadopublico.cl> donde puede descargarse todo el historial de gastos de cada institución pública.

Otros ejemplos de información contenida dentro del *CRM* de la compañía pueden ser, por ejemplo:

- Información sobre la interacción del cliente con los distintos canales: como: veces que el cliente inicia sesión en el sitio web, o llamadas telefónicas.

- Información sobre la cantidad de transacciones del mismo rubro o segmento.

5.1.3 Conclusiones

Como último punto, serán retomadas las **hipótesis** que inicialmente fueron presentadas en esta Tesis.

- **Hipótesis 1:** Se puede elaborar un modelo, que mediante la utilización de Redes Neuronales Recurrentes, pueda tener una predicción de fuga que tenga un rendimiento de AUC mejor que aleatorio.
 - Efectivamente el modelo presentó una mejora considerable frente a un predictor aleatorio. Logrando un AUC de 0,61 y 0,69 dependiendo de la métrica de valorización. Como fue presentado en la Discusión, este resultado es modesto en comparación con el potencial que puede tener el modelo al incorporar más variables que puedan contribuir a una predicción más sólida.
- **Hipótesis 2:** Ambas definiciones deberían tener un rendimiento relativamente similar.
 - Contrario a lo que se propuso en la hipótesis, ambas definiciones de fuga se comportaron distinto. El motivo de esto, como se pudo ver en la discusión, es que la predicción y la valorización según CLV tiene un problema, ya que el periodo de testeó es acotado, truncando la valorización efectiva de los flujos reales del CLV, mientras el RFM si es capaz de capturar más información en la misma ventana de tiempo.

BIBLIOGRAFÍA

- Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., ... Schrift, R. (2018). In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions. *Customer Needs and Solutions*, 5(1–2), 65–81. <https://doi.org/10.1007/s40547-017-0080-0>
- Blattberg, R. C., Kim, P., & Neslin, S. A. (2008). *Database marketing: analyzing and managing customers*. New York: Springer.
- Chen, J., Wang, C., & Peng, X. (2013). E-Loyalty Simulation Based on Hidden Markov Model. In Z. Zhong (Ed.), *Proceedings of the International Conference on Information Engineering and Applications (IEA) 2012* (Vol. 216, pp. 317–323). London: Springer London. https://doi.org/10.1007/978-1-4471-4856-2_39
- Cheng, C.-H., & Chen, Y.-S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, 36(3), 4176–4184. <https://doi.org/10.1016/j.eswa.2008.04.003>
- Coussement, K., & De Bock, K. W. (2013). Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *Journal of Business Research*, 66(9), 1629–1636.
- D’Haen, J., & Van den Poel, D. (2013). Model-supported business-to-business prospect prediction based on an iterative customer acquisition framework. *Industrial Marketing Management*, 42(4), 544–551. <https://doi.org/10.1016/j.indmarman.2013.03.006>
- Gladly, N., Baesens, B., & Croux, C. (2009). Modeling churn using customer lifetime value. *European Journal of Operational Research*, 197(1), 402–411. <https://doi.org/10.1016/j.ejor.2008.06.027>

- Gupta, S., & Zeithaml, V. (2006). Customer Metrics and Their Impact on Financial Performance. *Marketing Science*, 25(6), 718–739. <https://doi.org/10.1287/mksc.1060.0221>
- Gupta, S., & Lehmann, D. (2004). Valuing Customers. *Journal of Marketing Research*, 41(1), 7–18
- Kamakura, W., Mela, C. F., Ansari, A., Bodapati, A., Fader, P., Iyengar, R., ... Wilcox, R. (2005). Choice Models and Customer Relationship Management, 13.
- Kasiran, Z., Ibrahim, Z., & Ribuan, M. S. M. (2018). Mobile Phone Customers Churn Prediction using Elman And Jordan Recurrent Neural Network.
- Klein, J. P., & Moeschberger, M. L. (2003). Survival analysis: techniques for censored and truncated data (2nd ed). New York: Springer.
- Lee, M.-L. T., Whitmore, G. A., & Rosner, B. A. (2010). Threshold regression for survival data with time-varying covariates. *Statistics in Medicine*, 29(7–8), 896–905. <https://doi.org/10.1002/sim.3808>
- Lemmens, A., & Croux, C. (2006). Bagging and Boosting Classification Trees to Predict Churn. *Journal of Marketing Research*, 43(2), 276–286. <https://doi.org/10.1509/jmkr.43.2.276>
- Lemmens, A., & Gupta, S. (2017). Managing Churn to Maximize Profits. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2964906>
- Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. *ArXiv:1506.00019 [Cs]*. Retrieved from <http://arxiv.org/abs/1506.00019>
- Ljunghed, J. (2017). Predicting Customer Churn Using Recurrent Neural Networks, 53.
- Martinsson, E. (2016). A model for sequential prediction of time-to-event in the case of discrete or continuous censored data, recurrent events or time-varying covariates, 103.

- Rauyruen, P., & Miller, K. E. (2007). Relationship quality as a predictor of B2B customer loyalty. *Journal of Business Research*, 60(1), 21–31. <https://doi.org/10.1016/j.jbusres.2005.11.006>
- Tamaddoni Jahromi, A., Stakhovych, S., & Ewing, M. (2014). Managing B2B customer churn, retention and profitability. *Industrial Marketing Management*, 43(7), 1258–1268. <https://doi.org/10.1016/j.indmarman.2014.06.016>
- Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55, 1–9. <https://doi.org/10.1016/j.simpat.2015.03.003>
- Wei, J.-T., Lin, S.-Y., & Wu, H.-H. (2010). A review of the application of RFM model. *Afr. J. Bus. Manage.*, 8.
- Wiersema, F. (2013). The B2B Agenda: The current state of B2B marketing and a look ahead. *Industrial Marketing Management*, 42(4), 470–488. <https://doi.org/10.1016/j.indmarman.2013.02.015>
- Wübben, M. (2008). *Analytical CRM: developing and maintaining profitable customer relationships in non-contractual settings* (1. ed). Wiesbaden: Gabler.

ANEXOS

Datos transaccionales

Los datos transaccionales de la empresa corresponden a la información transaccional de cada una de las compras de los clientes efectuadas entre los meses de enero a diciembre de los años 2014 a 2017. Esta información tiene la siguiente estructura:

Tabla 2: Esquema de datos de la información históricos de compra.

Variable	Descripción	Tipo
order_id	Identificador único de la factura	integer
product_id	Identificador de producto de cada línea de la factura	string
client_id	RUT del cliente sin dígito verificador	int
date	Fecha de la factura	datetime
price	Valor neto de la línea de facturación	integer
cost	Costo comercial de la línea de facturación	integer
product_qty	Cantidad de ítems en la línea de facturación	integer
asile	Familia a la que pertenece el producto.	category
cot_id	Si la compra fue sujeta a contrato (cotización)	integer

Datos CRM

Los datos provenientes del CRM son aquellos que guarda la empresa sobre los clientes. Estos se resumen en la siguiente tabla

Tabla 3: Esquema de datos de la información contextual de los clientes.

Variable	Descripción	Tipo
client_id	RUT del cliente sin dígito verificador	int
name	Razón social de la empresa	string
size	Tamaño de la empresa	integer
branches	Cantidad de sedes del comprador	integer
business	Giro comercial	category
sales_rep	Ejecutivo comercial que lo atiende	category