



**UNIVERSIDAD DE CHILE**

**FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS**

**DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

**PREDICCIÓN DE FUGA DE CLIENTES MEDIANTE EL  
REDISEÑO DEL PROCESO DE ATENCIÓN EN SEGUROS  
MASIVOS EN UNA COMPAÑÍA ASEGURADORA**

PROYECTO DE GRADO PARA OPTAR AL GRADO DE MAGÍSTER EN  
INGENIERÍA DE NEGOCIOS CON TECNOLOGÍAS DE INFORMACIÓN

**SERGIO PABLO LORCA FIGUEROA**

PROFESOR GUÍA:  
PATRICIO WOLFF ROJAS

MIEMBROS DE LA COMISION:  
ENRIQUE BRAVO CASTRO  
CRISTIÁN JULIO AMDAN

SANTIAGO DE CHILE  
2021

## RESUMEN EJECUTIVO

El trabajo se desarrolló en una compañía líder del mercado de seguros en Chile que ofrece soluciones de protección, ahorro y previsión a través de sus principales líneas de negocio: seguros de vida, seguros colectivos, rentas vitalicias, mutuaría y seguros masivos. Sobre esta última línea, la generación de nuevos productos y las actuales acciones de retención parecen no ser suficientes para evitar la constante fuga de clientes, siendo difícil mantener un grado de satisfacción y permanencia adecuado de los mismos. Una tasa cercana a un 55% de abandono de clientes al término del primer año, implica pérdidas significativas para la compañía y altos costos de mantención, llegando incluso al término de relaciones comerciales con sus principales socios estratégicos.

Con estos antecedentes, se hace cada vez más necesario estudiar el comportamiento de fuga de los clientes dada la alta competitividad del negocio y las nuevas regulaciones del mercado, obligando a actuar velozmente para lograr atraer a nuevos asegurados y sobre todo mantener a los actuales. El objetivo principal de este proyecto de tesis busca disminuir la tasa de fuga de clientes de seguros masivos (*attrition rate*) de pólizas *bancassurance*, por medio del rediseño al actual proceso de Análisis y Gestión de Clientes, aplicando la metodología de Ingeniería de Negocios. Este proceso es fundamental en la interacción entre la aseguradora y sus clientes, generando el espacio necesario para poder cautivar con una mejor experiencia o servicio y retenerlos en su deseo de abandono. El levantamiento de procesos indica la inexistencia de un área de retención para la línea, la que se plantea implementar mediante la incorporación de nuevos procesos y actividades junto a una propuesta tecnológica que dé el soporte y resultados necesarios para lograrlo.

Este trabajo comprende el desarrollo y evaluación de tres modelos de predicción de fuga altamente utilizados en la industria (y en literatura) para estas situaciones: regresión logística, árboles de decisión y *random forest*. Utilizando técnicas de minería de datos bajo la metodología del *framework* CRISP-DM, y una base aproximada de clientes sobre los 40.000 registros, se observan resultados de precisión muy favorables (88% en la predicción con el modelo de clasificación supervisado *random forest*). Demostrando que la fuga de clientes se produce fundamentalmente por factores como la edad o meses vigentes de los asegurados que, según configuraciones propias del producto, manifiestan una probabilidad mayor de abandonar la compañía.

Con estos resultados, es posible elaborar estrategias de marketing y/o actividades comerciales diferenciadas para diversos tipos de clientes a través del *cálculo de su valorización o "importancia"*, utilizando los ingresos esperados y los resultados de predicción, para así conseguir una retención efectiva antes de que se manifieste la alternativa de abandono.

A mi familia, mi señora y mi futuro hijo(a)...

# TABLA DE CONTENIDO

<b>CAPÍTULO 1: INTRODUCCIÓN Y CONTEXTO .....</b>	<b>1</b>
1.1 ANTECEDENTES DE LA INDUSTRIA .....	1
1.2 DESCRIPCIÓN GENERAL DE LA EMPRESA .....	2
1.3 PROBLEMA U OPORTUNIDAD IDENTIFICADA .....	3
1.4 OBJETIVOS Y RESULTADOS ESPERADOS DEL PROYECTO .....	5
1.4.1 <i>Objetivo General</i> .....	5
1.4.2 <i>Objetivos Específicos</i> .....	5
1.4.3 <i>Resultados Esperados</i> .....	5
1.5 ALCANCE.....	6
1.6 RIESGOS POTENCIALES .....	6
<b>CAPÍTULO 2: MARCO TEÓRICO .....</b>	<b>8</b>
2.1 METODOLOGÍA DE INGENIERÍA DE NEGOCIOS .....	8
2.2 METODOLOGÍA CRISP DM .....	8
2.3 <i>ATTRITION/CHURN RATE – MACHINE LEARNING</i> EN LA INDUSTRIA DE SEGUROS .....	9
2.4 MODELOS DE PREDICCIÓN DE DATOS.....	11
2.4.1 <i>Regresión Logística</i> .....	11
2.4.2 <i>Árbol de Decisión</i> .....	12
2.4.3 <i>Random Forest</i> .....	12
2.4.4 <i>Métodos de Ensamble para Modelos de Predicción</i> .....	12
2.4.5 <i>Métricas y Formas de Evaluación</i> .....	13
<b>CAPÍTULO 3: PLANTEAMIENTO ESTRATÉGICO Y MODELO DE NEGOCIOS</b>	<b>18</b>
3.1 POSICIONAMIENTO ESTRATÉGICO.....	18
3.2 BALANCED SCORECARD .....	19
3.3 MODELO DE NEGOCIOS.....	20
<b>CAPÍTULO 4: ANÁLISIS DE LA SITUACIÓN ACTUAL .....</b>	<b>22</b>
4.1 MODELAMIENTO DETALLADO DE PROCESOS.....	22
4.1.1 <i>Modelamiento IDEF0</i> .....	22
4.2 DIAGNÓSTICO DE LA SITUACIÓN ACTUAL.....	27
4.3 CUANTIFICACIÓN DEL PROBLEMA U OPORTUNIDAD .....	29
<b>CAPÍTULO 5: PROPUESTA DE DISEÑO DE PROCESOS.....</b>	<b>31</b>
5.1 DIRECCIONES DE CAMBIO.....	31
5.2 ARQUITECTURA DE PROCESOS TO BE.....	32
5.2.1 <i>Diseño en IDEF0 – Gestión de Modelos de Predicción de Fuga de Clientes</i> .....	33
5.2.2 <i>Diseño en BPMN</i> .....	35
<b>CAPÍTULO 6: LÓGICA DE NEGOCIO .....</b>	<b>37</b>
6.1 PREPARACIÓN Y CONSTRUCCIÓN DEL MODELO PREDICTIVO.....	37
6.1.1 <i>Selección de Datos (Set de Datos)</i> .....	37
6.1.2 <i>Transformaciones de Datos</i> .....	38
6.1.3 <i>Correlación de Datos</i> .....	38
6.1.4 <i>Exclusión de Datos</i> .....	39
6.1.5 <i>Set de Entrenamiento de los Datos</i> .....	39
6.1.6 <i>Parámetros Utilizados en la Ejecución de los Modelos de Predicción</i> .....	40
6.1.7 <i>Balanceo de Datos</i> .....	40
6.2 RESULTADOS DE MODELOS DE PREDICCIÓN .....	41

6.2.1	<i>Selección del Modelo de Predicción.....</i>	43
6.3	MEDIDAS A CONSIDERAR EN EL FUTURO .....	44
6.4	EL VALOR DE LOS CLIENTES.....	45
6.5	INFORMACIÓN POST-ATENCIÓN DE CLIENTES.....	47
<b>CAPÍTULO 7:</b>	<b>APOYO TECNOLÓGICO .....</b>	<b>49</b>
7.1	CASOS DE USO .....	50
7.2	DIAGRAMA DE SECUENCIA.....	50
7.3	DISEÑO DE PROTOTIPOS .....	51
7.3.1	<i>Módulo de Parámetros del Modelo Predictivo .....</i>	<i>51</i>
7.3.2	<i>Módulo de Ejecución y Resultados del modelo predictivo.....</i>	<i>52</i>
7.3.3	<i>Módulo de Valorización de Clientes.....</i>	<i>53</i>
7.3.4	<i>Módulo de Ingreso de Datos Post-Atención Clientes.....</i>	<i>54</i>
<b>CAPÍTULO 8:</b>	<b>EVALUACIÓN ECONÓMICA DEL PROYECTO .....</b>	<b>56</b>
8.1	SUPUESTOS UTILIZADOS .....	56
8.2	EVALUACIÓN ECONÓMICA (SITUACIÓN ACTUAL – MEJORADA – IMPLEMENTACIÓN DEL PROYECTO).....	58
8.3	FLUJO DE CAJA .....	59
8.4	ANÁLISIS DE SENSIBILIDAD.....	59
<b>CAPÍTULO 9:</b>	<b>CONCLUSIONES .....</b>	<b>61</b>
<b>CAPÍTULO 10:</b>	<b>BIBLIOGRAFÍA.....</b>	<b>64</b>
<b>CAPÍTULO 11:</b>	<b>ANEXOS.....</b>	<b>1</b>

## ÍNDICE DE TABLAS

Tabla 1: Utilidad y pérdida estimada por variaciones en la persistencia.....	4
Tabla 2: Matriz de clasificación de riesgos del proyecto .....	7
Tabla 3: Matriz de Confusión.....	14
Tabla 4: Cuantificación estimada del impacto de fuga clientes en pesos (\$).....	29
Tabla 6: Parámetros Utilizados en la ejecución de los Modelos Predictivos .....	40
Tabla 5: Cantidad de Clientes Vigentes/No Vigentes en la base .....	40
Tabla 7: Resultados y métricas de desempeño de Modelos de Predicción .....	41
Tabla 8: Matriz de Confusión - <i>Random Forest</i> .....	43
Tabla 8: 10 Predictores Preponderantes en el Modelo - <i>Random Forest</i> .....	44
Tabla 10: Factores de Siniestralidad Promedio para Cálculo de Valor del Cliente. ....	47
Tabla 11: Prima Promedio Mensual en pesos por Sponsor y Clúster Banca. ....	57
Tabla 12: Promedio Porcentual de Siniestralidad y Costos de Adm. y Ventas por Clúster Banca. .....	57
Tabla 13: Costo de mantención de Software más Costos operacionales Promedio en M\$.....	57
Tabla 14: Resumen resultados en M\$ de los escenarios económicos propuestos.....	59
Tabla 15: Evaluación Económica Implementación del Proyecto.....	59

## ÍNDICE DE ILUSTRACIONES

Ilustración 1: <i>MetLife</i> Chile, sus filiales y líneas de negocio. ....	2
Ilustración 2: Fases metodología CRISP – DM .....	9
Ilustración 3: Tipos de curvas ROC. ....	16
Ilustración 4: <i>Cross-Validation</i> representación gráfica. ....	16
Ilustración 5: Posicionamiento Estratégico <i>MetLife</i> según Modelo Delta.....	18
Ilustración 6: BSC objetivos estratégicos <i>MetLife</i> .....	19
Ilustración 7: Diagrama de Proceso BPMN de Macroprocesos (Nivel I) para <i>MetLife</i> Seguros de Vida.....	23
Ilustración 8: Macroprocesos Nivel II, Adm. Relación con el Cliente.....	25
Ilustración 9: Macroprocesos Nivel III: Marketing y Análisis del Mercado.....	26
Ilustración 10: Macroproceso Nivel IV, Análisis y Gestión del Comportamiento Clientes Seg. Masivos.....	27
Ilustración 11: Diagrama actual de subproceso para Gestión de Atención de Clientes de Seguros Masivos.....	28
Ilustración 12: Diagrama de Proceso BPMN para Gestión de Atención de Clientes en Seguros Masivos con Identificación de Procesos Clave Faltantes. ....	31
Ilustración 13: Nuevo Proceso de Gestión de Modelos de Predicción de Fuga de Clientes. ....	33
Ilustración 14: Actividades contenidas en el “Desarrollo de Modelos de Predicción” .....	34
Ilustración 15: Actividades contenidas en Calibración - Ejecución y Control de Modelos de Predicción de Fuga .....	34
Ilustración 16: Diagrama de Proceso BPMN para Gestión de Atención de Clientes en Seguros Masivos con nuevas actividades propuestas.....	35
Ilustración 17: Distribución de Datos por Cobertura y Número de Pólizas Vigentes/No Vigentes. ....	39
Ilustración 18: Resultados de balanceo de datos aplicando <i>oversampling</i> , en el set de entrenamiento. ....	41
Ilustración 18: Curva ROC de resultados a los modelos predictivos .....	42
Ilustración 20: Encuesta Post-Atención de Clientes.....	48
Ilustración 19: Esquema General de Solución.....	49
Ilustración 20: Diagrama de Casos de Uso.....	50
Ilustración 21: Diagrama de Secuencias.....	51
Ilustración 22: Prototipo Modulo de Parámetros de Modelo de Predicción .....	52
Ilustración 23: Prototipo Módulo Ejecución y Resultados Modelo Predicción .....	53
Ilustración 24: Prototipo Módulo Valorización de Clientes.....	54
Ilustración 25: Prototipo Modulo de Ingreso de Datos Encuesta Post-Atención de Clientes ...	55
Ilustración 26: Cálculo de la Tasa de Interés utilizando método CAPM. ....	58
Ilustración 27: Descripción de Escenarios para Análisis de Sensibilidad.....	60
Ilustración 28: Evaluación Económica ante análisis de Sensibilidad.....	60

## INDICE DE GRÁFICOS

Gráfico 1 : Prima directa de seguros de vida a Diciembre 2018. ....	2
Gráfico 2: Evolución Mensual de Pólizas Vigentes, Altas y Bajas para Mercado Masivo <i>MetLife</i> . .....	3
Gráfico 3: Evolución promedio mensual de persistencia de clientes para mercado masivo <i>MetLife</i> .....	4
Gráfico 4: Proyección Semestral de Pólizas por Clúster para Mercado Masivo en <i>MetLife</i> . ....	56



## GLOSARIO DE TÉRMINOS

---

**Bancassurance:** Es una relación entre un banco y una compañía de seguros que tiene como objetivo, ofrecer productos de seguros o beneficios de seguros a los clientes del banco.

**Bróker de Seguros:** Un *bróker* de seguros es una persona que actúa como intermediario de varias compañías aseguradoras, sin estar vinculado en exclusiva a ninguna de ellas, comercializando contratos de seguro a sus clientes.

**Devolución por Experiencia Favorable:** Ocurre cuando existe una buena siniestralidad de la cartera de clientes, que se traduce en retornos de comisiones al sponsor y mayor acceso a clientes de parte de la compañía de seguros.

**Camada:** Grupo de personas que participa en la venta de pólizas de seguros durante un periodo en particular, por ejemplo, mensual.

**Fuga:** Número de pólizas dadas de baja en un periodo (t) determinado.

**Persistencia (*attrition rate*):** Es la medida de cuánto tiempo una póliza o un bloque de pólizas se mantiene vigente a lo largo del tiempo. Mensualmente se calcula a través de la fórmula:

$$\text{Attrition Rate puntual}(t)\% = \frac{N^{\circ} \text{ de pólizas fugadas}(t)}{N^{\circ} \text{ de pólizas vigentes}(t)} * 100$$

Donde,

t= Periodo (mes) actual de cálculo.

**Póliza:** La póliza de seguro es un contrato entre un asegurado y una compañía de seguros, que establece los derechos y obligaciones de ambos, en relación con el seguro contratado.

**Prima Directa:** Corresponde a los ingresos por la venta de seguros efectuada por el total de las compañías, rebajadas las anulaciones. Debe considerarse la venta realizada directamente por los aseguradores y aquella efectuada a través de corredores de seguros.

**Seguros Masivos (o de Afinidad):** Línea de negocios que surge como resultado de la necesidad de desarrollar modelos de servicio y atención personalizada para clientes del *Sponsor* que tienen un grupo de afinidad.

**Sponsor:** Un sponsor es generalmente una entidad o individuo que financia algún tipo de actividad con el fin de obtener algún beneficio económico.

**Stock (pólizas vigentes):** Número de pólizas vigentes en un periodo (t) determinado.

**Siniestralidad:** Conjunto de los siniestros sufridos por una persona o cosa en una póliza o grupo de pólizas.

**UF:** Unidad de Fomento.

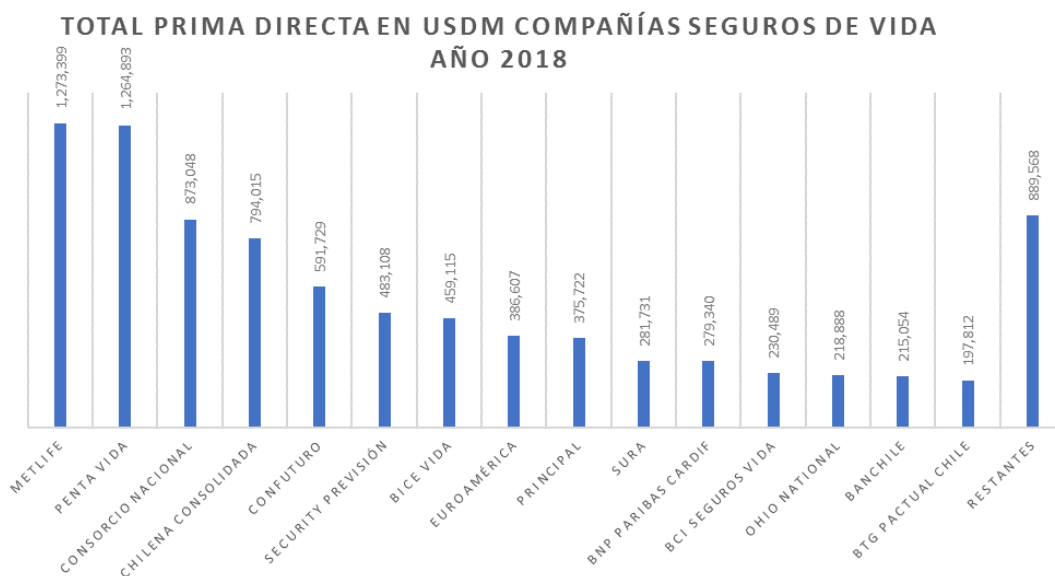
# CAPÍTULO 1: INTRODUCCIÓN Y CONTEXTO

*Metropolitan Life Insurance Company (MetLife)* nace en el año 1868 en la ciudad de Nueva York, transformándose en una de las primeras compañías aseguradoras del país americano. Hoy con más de 150 años cuenta con presencia global en 15 países, ofreciendo soluciones de protección, ahorro y previsión a más de 8 millones de clientes sólo en Chile, donde es la compañía líder del mercado asegurador (Comisión de Mercado Financiero CMF, 2020).

## 1.1 Antecedentes de la Industria

*MetLife* pertenece en Chile al sector financiero de Seguros y Previsión, donde se incluyen a todas las instituciones bancarias, aseguradoras y financieras, cuyo principal objetivo es captar fondos del público y colocarlos en forma de créditos o inversiones. El tamaño de este sector industrial es medido a través de los activos totales de las empresas participantes por medio de sus primas directas las cuales ascienden a más de USDM 8.814 al 31 de diciembre 2018 para seguros de vida y USDM 3.730 para seguros generales al mismo periodo. La industria aseguradora en Chile se compone de 37 compañías de seguros de vida que corresponden al 68.9% del mercado, y 32 compañías de seguros generales que corresponden al 31.1% del total de la industria. (AACH, 2018).

En el ranking de compañías de seguros de vida con mayor participación en prima directa se tiene que Penta Vida, Consorcio Nacional y Chilena Consolidada son los principales competidores de *MetLife*, con valores en de USDM 1.264, 873 y 794 respectivamente, versus USDM 1.273 para *MetLife* a diciembre del año 2018.

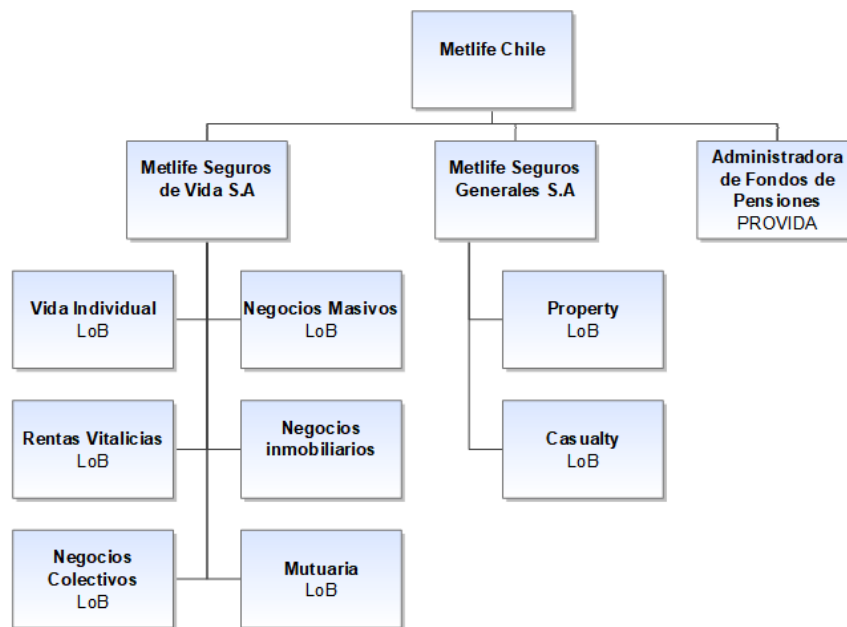


**Gráfico 1** : Prima directa de seguros de vida a Diciembre 2018.

Fuente: AACH, 2018

## 1.2 Descripción General de la Empresa

*MetLife* cuenta en Chile con tres principales filiales de negocios: Seguros de Vida, Seguros Generales y Administradora de Fondos de Pensiones. Este trabajo concentra sus esfuerzos sobre *MetLife* Seguros de Vida (en adelante SDV), donde seis líneas de negocio la componen dentro de sus operaciones: Vida Individual, Rentas Vitalicias & *Policy Loans*, Negocios Colectivos (*Employee Benefits & Worksite Marketing*), Negocios Masivos, Mutuaria y Negocios Inmobiliarios.



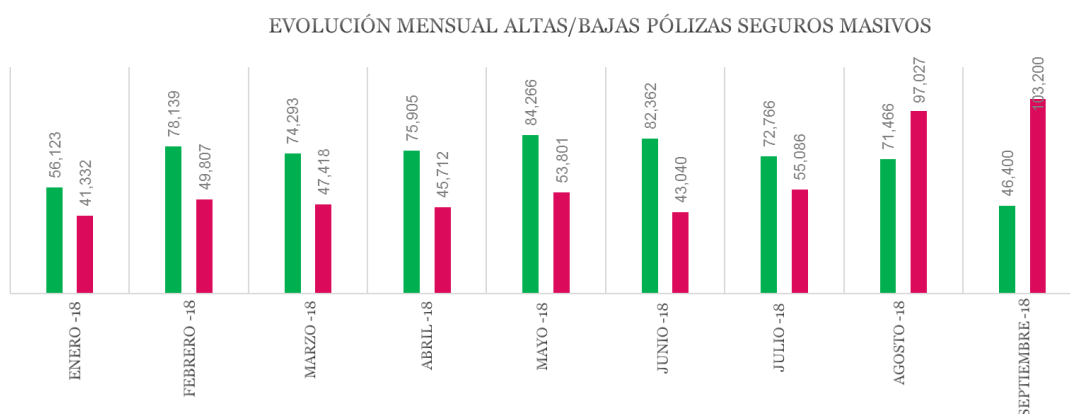
**Ilustración 1:** *MetLife* Chile, sus filiales y líneas de negocio.

Fuente: *MetLife*, Elaboración Propia.

*MetLife* SDV ofrece una variedad de productos pensados para personas, empresas y negocios, destacando: *seguros de vida*, orientados para personas naturales; *seguros colectivos*, para grupos numerosos de personas que pertenecen a empresas públicas o privadas; *seguros de salud*, enfocados a la atención médica u hospitalaria y soluciones de tipo *previsional*, donde la compañía tiene un 65% de participación de mercado en relación con la prima directa (Superintendencia de AFP's, 2019).

### 1.3 Problema u Oportunidad Identificada

La figura siguiente muestra la evolución mensual de pólizas para la línea de negocios *bancassurance* de seguros masivos, indicando el número de nuevas ventas (altas) y renunciaciones (bajas) en el periodo comprendido entre los meses de enero y septiembre 2018, denotando un marcado y progresivo aumento de clientes que abandonan sus productos con la compañía. Los motivos de abandono están caracterizados principalmente por renunciaciones voluntarias o no pago de primas, donde en la mayoría de los casos se desconocen los motivos concretos de fuga. Con ello se torna difícil elaborar medidas de mitigación/retención por parte de la compañía, comprendiendo que una gran cantidad de renunciaciones a lo largo del tiempo se traduce en una costosa fuga de clientes.



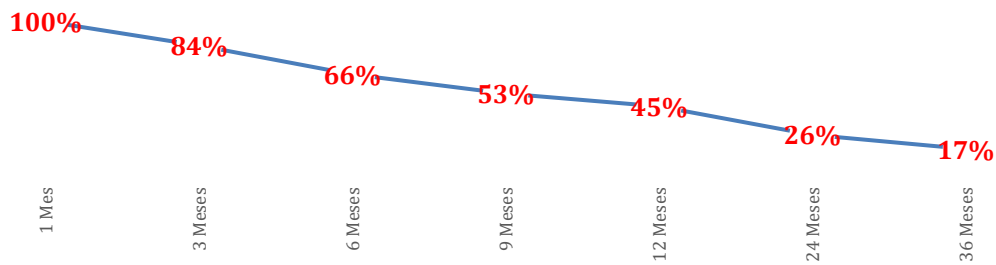
**Gráfico 2:** Evolución Mensual de Pólizas Vigentes, Altas y Bajas para Mercado Masivo *MetLife*.

Fuente: Elaboración Propia.

Esto último se sustenta a través de un análisis a la persistencia de clientes, la cual se define como la capacidad de mantener y retener a los consumidores en la compañía a lo largo del tiempo, transformándose en una medida fundamental en el cálculo de la rentabilidad del negocio. Si la persistencia se conserva dentro de márgenes estables, permite la sustentabilidad financiera del negocio ya que muestra la satisfacción de los clientes con la compañía, direcciona a los procesos y a la cultura interna de la misma.

Del porcentaje promedio de clientes de seguros masivos que permanece en el tiempo con sus productos, sólo un 84% de la cartera original se mantiene al final del tercer mes de vigencia del contrato y, a contar de los doce meses, este número se reduce a un 45%.

EVOLUCION PERSISTENCIA TOTAL PROMEDIO SEGUROS MASIVOS



**Gráfico 3:** Evolución promedio mensual de persistencia de clientes para mercado masivo *MetLife*.

Fuente: Elaboración propia

El impacto monetario de esta reducción sostenida en la persistencia puede calcularse de forma general, si se asume que el valor en promedio de una póliza de seguro masivo es de UF 0,235 (aprox. \$6.486) y existe un porcentaje de siniestralidad del 5%. Además, los gastos de administración del producto corresponden aproximadamente a un 15% de la prima. Con estos supuestos es posible calcular la utilidad esperada para un stock de pólizas determinado:

**Tabla 1:** Utilidad y pérdida estimada por variaciones en la persistencia.

Fuente: Elaboración Propia

	Utilidad después de Gastos y Siniestralidad						
	Primer Mes	3 Meses	6 Meses	9 Meses	12 Meses	24 Meses	36 Meses
Camada 1	596,19	471,94	306,60	229,01	191,00	121,77	76,76
Camada 2	616,00	480,59	328,30	260,23	219,17	136,96	86,35
Camada 3	532,25	413,32	259,00	202,37	168,11	108,26	64,19
Camada 4	556,18	442,68	287,21	233,21	198,94	115,33	81,58
Camada 5	463,47	379,61	258,58	203,53	164,57	96,56	66,99
Camada 6	556,29	471,17	325,08	231,67	183,51	111,62	76,19
Camada 7	735,80	610,19	426,30	276,36	247,73	149,00	103,15
Camada 8	853,41	680,93	477,82	283,19	248,75	134,09	93,35
Camada 9	770,04	639,17	458,95	366,31	308,46	171,85	121,87
Camada 10	639,28	521,53	392,63	316,16	271,46	170,77	127,54
Camada 11	838,32	681,69	500,71	401,17	338,98	201,01	149,94
Camada 12	825,65	654,85	475,65	373,49	321,13	177,66	135,91
<b>Total Utilidad (UF)</b>	<b>7.982,9</b>	<b>6.447,7</b>	<b>4.496,8</b>	<b>3.376,7</b>	<b>2.861,8</b>	<b>1.694,8</b>	<b>1.183,8</b>
<b>Variación Utilidad</b>	<b>100%</b>	<b>81%</b>	<b>56%</b>	<b>42%</b>	<b>36%</b>	<b>21%</b>	<b>15%</b>
<b>% Pérdida Estimada</b>	<b>0%</b>	<b>-19%</b>	<b>-44%</b>	<b>-58%</b>	<b>-64%</b>	<b>-79%</b>	<b>-85%</b>
<b>Pérdida Estimada (UF)</b>		<b>-1.535,2</b>	<b>-3.486,0</b>	<b>-4.606,2</b>	<b>-5.121,1</b>	<b>-6.288,0</b>	<b>-6.799,1</b>

Estos resultados determinan a priori que la persistencia tiene directa relación con la utilidad esperada, tomando el ejemplo anterior, al término del mes 12 se puede dejar de percibir más del 64% de rentabilidad del negocio (UF 5.121 ~ USD 203k aprox.) si el comportamiento de fuga se mantiene o aumenta a lo largo del tiempo.

El modelo de negocios de este tipo de seguros se sustenta a través de sus aliados comerciales (empresas y corredores), los que tienen la potestad de cancelar los acuerdos con *MetLife* si no se cumplen las métricas de venta, mantención y rentabilidad comprometidas de ambas partes.

Todo esto tiene un impacto directo con los objetivos estratégicos, en el que la fuga de clientes se convierte en una pérdida de la rentabilidad del negocio. Los clientes no están siendo el foco de prioridad declarado por la compañía y en el mediano plazo, los negocios acordados pueden no ser sustentables, pudiendo provocar la pérdida de relaciones comerciales con los *Sponsors*, incumpliendo el compromiso de los valores declarados por *MetLife* afectando su modelo de Negocios.

## **1.4 Objetivos y Resultados Esperados del Proyecto**

### **1.4.1 *Objetivo General***

Disminuir la tasa de abandono de clientes de seguros masivos por medio del rediseño del actual proceso de Atención de Clientes en *MetLife*.

### **1.4.2 *Objetivos Específicos***

- Rediseñar el actual proceso de atención de clientes para la incorporación de actividades que permitan la evaluación y retención de clientes.
- Diseñar un modelo de predicción de fuga para clientes de seguros masivos bancarios (*bancassurance*).
- Desarrollar una valoración del conocimiento de clientes, que será base para el logro de una buena fidelización (retención) tomando como parámetros los resultados del modelo de predicción de fuga y condiciones contractuales del negocio.
- Generar y proponer al menos 7 nuevas medidas para reducción de fuga de clientes.

### **1.4.3 *Resultados Esperados***

- Rediseño del *Proceso de Atención de Clientes* de seguros masivos, incorporando actividades claves de retención y mantención de modelo predictivo.
- Modelo de predicción de fuga de clientes *bancassurance*.
- Sistema de conocimiento de clientes que permitirá valorizar y determinar problemas en la gestión de la cartera.
- Generación de medidas de retención para aumento de persistencia de clientes en la línea de negocios.

## 1.5 Alcance

El enfoque de este trabajo aplica exclusivamente dentro de la línea de negocios de seguros masivos en *MetLife*, no considerando otras clases ni filiales de negocio ajenas a seguros de vida. Dentro de esta línea, el segmento elegido corresponde a seguros ligados a la banca o *bancassurance* y ubica el problema detectado en el proceso de “*Gestión de Atención de Clientes*” dentro de la dirección de operaciones de seguros masivos.

Los modelos que se desarrollan en este documento identificarán aquellos asegurados con mayor probabilidad de fuga y guiarán a la compañía a la priorización en la atención en base al valor de los clientes, bajo la adopción de un sistema de inteligencia para el conocimiento de los mismos. Este sistema estará basado en los resultados del modelo de predicción de fuga y condiciones propias de la línea de negocios.

Para el desarrollo de estos modelos se realiza una búsqueda y exploración de datos desde bases históricas de propiedad de la compañía, tomando más de 40.000 registros que contienen información relevante de clientes y sus pólizas de seguros. Los resultados serán de gran importancia en las acciones comerciales y gestiones de la cartera *bancassurance*, facultando la clasificación de aquellos clientes y productos con mayor probabilidad de fuga para la aplicación de estrategias de retención o, por otro lado, modificar las configuraciones de productos para hacerlos más atractivos y mantener la decisión de permanencia.

Sumado a lo anterior, el alcance del proyecto considera el diseño de políticas comerciales de retención no así su ejecución.

## 1.6 Riesgos Potenciales

El desarrollo de este trabajo conlleva diversos tipos de riesgos: técnicos, externos, de gestión y organizacional. Para la evaluación de éstos se utiliza la metodología basada en el análisis de falla y efecto *FMEA* (Mikulak R., 1996), que permite la identificación de fallas en el proceso de ejecución para cumplimiento de criterios, estimando el riesgo de causas específicas, evaluando y priorizando las acciones y planes de control que deben llevarse a cabo para disminuir los riesgos detectados.

Los distintos riesgos detectados para el cumplimiento de los objetivos son calificados utilizando el criterio del *número prioritario de riesgo* (*Risk Priority Number, RPN*), que se determina según severidad (estimación de la gravedad del efecto), ocurrencia (probabilidad de que una causa específica resulte en un modo de falla) y detección según muestra la tabla siguiente:

**Tabla 2:** Matriz de clasificación de riesgos del proyecto

Fuente: Elaboración propia

<b>Tipo de Riesgo</b>	<b>Impacto</b>	<b>Severidad</b>	<b>Ocurrencia</b>	<b>Detección</b>	<b>RPN</b>
Acceso a los datos	Impedimento	10	3	1	30
Calidad de los datos	Fracaso	10	8	1	80
Implementación	Falta de Interés	7	8	1	56
Recursos económicos para desarrollos en los sistemas (patrocinador)	Falta de Interés	7	6	2	84
Soporte y mantención de desarrollo tecnológico	Falta de Interés	7	6	5	210
Cambio de alcance	Impedimento	6	7	4	168
Resistencia al cambio	Fracaso	6	6	6	216

El cálculo del RPN se realiza multiplicando el valor asignado a cada factor, asignando de esta forma un puntaje de criticidad a cada uno de los riesgos listados. Los factores se definen numéricamente de 1 a 10 en base a la severidad del efecto (mientras más severo, mayor puntaje), la probabilidad ocurrencia (mayor probabilidad, mayor puntaje) y probabilidad de detección del efecto (mayor probabilidad, mayor puntaje). Los resultados obtenidos muestran que el mayor riesgo es correspondiente a la resistencia al cambio, que puede llevar al fracaso de la implementación de los nuevos procesos y cambios en los sistemas, pudiendo quedar inválida la solución que se propone. Luego, la falta de soporte y mantención de desarrollo tecnológico para llevarlo a cabo puede llevar a un retraso en la implementación del proyecto, teniendo presente que para manejar grandes volúmenes de datos es necesario el soporte de dichas herramientas para la evaluación y mantenimiento de los resultados.



## CAPÍTULO 2: MARCO TEÓRICO

---

### 2.1 Metodología de Ingeniería de Negocios

La ejecución de esta tesis comprende la utilización metodológica de la *Ingeniería de Negocios* (Barros, 2004), la cual comprende de las siguientes etapas secuenciales que describen en conjunto, el desarrollo de las actividades llevadas a cabo para su aplicación efectiva:

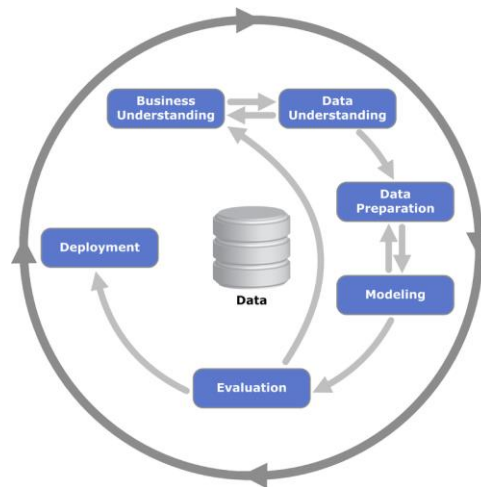
1. *Planteamiento Estratégico*: Identificación del planeamiento estratégico de la empresa, aplicando el modelo Delta (Hax, 2010).
2. *Definición del Modelo de Negocio*: Se definen los componentes necesarios del modelo de negocio según modelo CANVAS (Osterwalder, 2010). Con esto es posible identificar qué es lo que genera valor para los clientes.
3. *Diseño de Arquitectura de Procesos*: Se identifican macroprocesos relevantes y relaciones según los patrones establecidos dentro de la arquitectura de procesos. El método elegido para modelar estas relaciones, acciones y actividades es sobre IDEF 0.
4. *Diseño de Procesos*: Detalle y apertura de los macroprocesos identificados en la etapa anterior. Para lograrlo se utiliza la metodología IDEF 0 y notación BPMN (*Business Process Management Notation*).
5. *Diseño de la aplicación de apoyo*.
6. *Construcción e implementación*, utilizando la estructura de diseño definida anteriormente.
7. *Ejecución y control de los procesos diseñados*.

### 2.2 Metodología CRISP DM

Junto con el rediseño de procesos, es necesario la incorporación de un modelo predictivo de fuga de clientes empleando para ello, la metodología descrita en el *framework* CRISPDM “*The Cross-Industry Standard Process for Data Mining*” (Wirth, 2000) de minería de datos, la cual comprende de seis fases:

1. *Análisis del negocio*: Esta primera fase se centra en la comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial, para luego convertir este conocimiento en una definición del problema de minería de datos y un plan preliminar diseñado para alcanzar los objetivos.
2. *Análisis de los datos*: Esta fase comienza con una recopilación inicial de datos y continúa con actividades de “familiarización” de los mismos, a fin de identificar los problemas de calidad en los datos o detectar información interesante para formular hipótesis de información oculta.
3. *Preparación de los Datos*: Cubre todas las actividades de construcción del conjunto de datos finales a partir de los datos brutos iniciales.

4. *Modelamiento:* En esta fase se seleccionan y aplican diversas técnicas de modelización y se calibran parámetros a sus valores óptimos.
5. *Fase de Evaluación:* En esta etapa se evalúan en profundidad el o los modelos obtenidos y se revisan los pasos ejecutados para construirlo con el fin de asegurar el logro adecuado de los objetivos del negocio.
6. *Despliegue o Implementación:* La creación del modelo no suele ser el fin del proyecto en sí, incluso si su propósito es aumentar el conocimiento de los datos. Los conocimientos adquiridos deberán organizarse y presentarse de manera que el negocio/cliente pueda utilizarlos. (Chapman, 2008)



**Ilustración 2:** Fases metodología CRISP – DM

Fuente: Adaptado de Chapman et al., 2000

### 2.3 *Attrition/Churn Rate – Machine Learning en la Industria de Seguros*

Los clientes son el activo más importante de las organizaciones, por lo que estas deben centrar sus esfuerzos en mantener altos grados de satisfacción a través de productos y servicios atractivos en pos de su persistencia en el tiempo, traduciéndose en buenos índices de rentabilidad al largo plazo (Verbeke et al., 2011). Para las compañías de seguros esto debiera ser evidente dentro de su estrategia de negocios, entendiendo el alto costo de atraer nuevos clientes versus tratar de mantenerlos (Reinartz and Kumar, 2003). Con esto, se hace necesario realizar un análisis exhaustivo en el contexto de su gestión de la relación con los clientes (Staudt et al., 1998), yendo un paso más allá del análisis estadístico que las ayude a actuar rápida y anticipadamente frente a su intención de abandono. Por lo que se requiere contar con herramientas y métricas de gestión y satisfacción de clientes, que reflejen la manera en que se desarrollan las actividades de cara al cliente.

El *attrition rate* o *churn rate* es una unidad básica utilizada para describir la pérdida de clientes, y se define como la tasa bruta de abandono durante un periodo de tiempo determinado. Este indicador es muy utilizado/estudiado en diferentes áreas de negocios como: telecomunicaciones, *retail*, servicios financieros y comercio electrónico (Chen, Fan and Sun, 2012). Permite evaluar los esfuerzos de retención de clientes de un proveedor o industria de servicios, proporcionando una idea del crecimiento o disminución de la base de estos, así como la duración promedio de mantenimiento o satisfacción del servicio (Qian, Jiang & Tsui, 2006).

El abandono de los clientes con sus productos o servicios puede tener diversos orígenes, ya sea por voluntad propia debido a una mala gestión de venta p.ej., o por renunciaciones no voluntarias como problemas con el medio de pago o cuenta/tarjeta sin fondos. Este indicador tiene múltiples implicancias dentro de la compañía, ya que bajo él se configuran los acuerdos comerciales con grandes *sponsors* para asegurar el retorno de la rentabilidad mínima acordada entre ambos o también, como medida de gestión comercial de los agentes (*brokers*) de venta para el pago de comisiones, donde tiene un papel principal en la configuración de procesos y actividades internas que impactan tanto al cliente como a quienes están detrás de la creación del producto, comercialización y gestión de la venta.

Existen en la literatura numerosos estudios que demuestran el beneficio tangible dado en aquellas empresas con políticas de retención (Athanassopoulos, 2000), debido al aumento de los ingresos por el incremento/mantenimiento del número de pólizas vigentes. La utilización de *Machine Learning (ML)* como método de predicción de pérdidas en primas de seguros, muestran exitosas experiencias de su aplicación dentro de la industria aseguradora para la gestión de políticas centradas en la segmentación de clientes (López, 2005), solvencia financiera (Martínez, 2007) y predicción de fuga de clientes (Stucki, 2019).

La utilización de *ML* en grandes volúmenes de datos se ha incrementado enormemente en la actualidad gracias a las nuevas tecnologías de procesamiento, donde el problema de *clasificar* la información de que un cliente se fuge o no, puede tornarse interesante al poder utilizar diversa y gran cantidad de datos. Entre los métodos más usados en la clasificación supervisada para la predicción de fuga se encuentran (Sahar F. Sabbeh, 2018): regresión logística, árboles de decisión, *random forest*, *SGB (stochastic gradient boost)*, *SVM (support vector machines)*.

Una investigación realizada por Coussement & Van den Poel, (2008), refleja que los resultados obtenidos por el método *SVM* superan a las regresiones logísticas sólo si la selección de parámetros es adecuada, pero la utilización de los árboles de decisión siempre es más precisa. Otro estudio compara el desempeño en los resultados de clasificación entre modelos *SVM*, regresión logística y árboles de decisión, haciendo uso de potenciadores o *boosting*, encontrando que estos últimos tuvieron el mejor desempeño predictivo entre ellos (Tamaddoni, Stakhovych

y Ewing, 2016). Sin embargo, las diferencias entre los resultados de precisión de los métodos anteriores no fueron muy significativos.

Sumado a lo anterior, Günter et al. (2014) centra sus esfuerzos en la predicción de fuga de clientes en una compañía de seguros, donde sugiere que los modelos de regresión logística parecen ser los más populares por su simplicidad, desempeño e interpretabilidad. Esto permite agilizar y simplificar la toma de decisiones a la gestión futura del negocio, utilizando bases claras en los resultados que pueden interpretarse de cada uno de los modelos.

Las investigaciones anteriores avalan la decisión de utilizar y evaluar tres modelos de predicción de fuga: regresión logística, árboles de decisión y *random forest*.

## 2.4 Modelos de Predicción de Datos

Un modelo predictivo es un tipo de modelo estadístico que nos permite inferir la probabilidad de que ocurran determinados sucesos antes de su efectiva ocurrencia. Dentro de sus aplicaciones más utilizadas se encuentran aquellas que pueden predecir aquellos clientes que tienen mayor probabilidad de abandonar un producto/servicio o una compañía, cuantificar la satisfacción de clientes o identificar oportunidades comerciales que permitan tomar cierta ventaja competitiva frente al resto.

### 2.4.1 Regresión Logística

Los modelos de regresión logística permiten el análisis de resultados en términos explicativos y predictivos, siendo uno de los instrumentos más expresivos y versátiles que se disponen para el análisis de datos. Se utiliza para predecir la probabilidad de que una instancia corresponda a alguna clase particular. Para ello, utiliza la función *sigmoidea* para generar una probabilidad, donde esta última permite definir un umbral y convertirlo en un clasificador.

La probabilidad de que la instancia  $X_i$  sea de clase  $Y = 1$  se puede expresar con la forma funcional:

$$p(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Donde  $\beta_0, \beta_1 \dots \beta_n$  corresponden a los parámetros asociados para cada característica definida en  $x_1, x_2 \dots x_n$ . Los resultados óptimos para cada uno de los parámetros definidos se alcanzan a través de la función de estimación de máxima verosimilitud (MLE) para ajustar el modelo y estimar así de mejor forma sus parámetros.

## 2.4.2 Árbol de Decisión

Los árboles de decisión crean un modelo de clasificación basado en diagramas de flujo, clasificando casos en grupos o pronosticando valores de una variable dependiente (criterio) basada en valor de variables independientes (predictoras). A cada evento se le asignan probabilidades y a cada una de las ramas se le determina un resultado.

La mayoría de los algoritmos utilizados para construir un árbol son variaciones de uno genérico llamado “*Greedy algorithm*” que básicamente va desde la raíz hacia abajo (*Top-Down*) buscando de manera recursiva los atributos que generan el mejor árbol hasta encontrar el óptimo global con una estructura de árbol lo más simple posible. Existen diversos algoritmos aplicados a los árboles de decisión, donde se caracterizan por su capacidad de procesar un gran volumen de información de manera y manejo eficiente del ruido (error en los valores o en la clasificación de estos) que pudiese existir en los datos de entrenamiento (Dupouy C., 2014)

## 2.4.3 *Random Forest*

Los bosques aleatorios (*random forest*) son un esquema propuesto por *Breiman (2001)* para construir un conjunto de predicción con un conjunto de árboles de decisión que crecen en subespacios de datos seleccionados al azar (Biau, s. f.). Formalmente son un predictor de datos que consiste en una colección de árboles de regresión con base aleatoria  $\{r_n(x, \theta_m, D_n), m \geq 1\}$ , donde  $\theta_1, \theta_2, \dots, \theta_n$  son variables aleatorias independientes resultantes de una variable aleatoria  $\theta$ . Estos árboles aleatorios se combinan para formar la estimación de regresión agregada:

$$\bar{r}_n(X, D_n) = E_{\theta}[r_n(X, \theta, D_n)],$$

Donde  $E_{\theta}$  denota la expectativa con respecto al parámetro aleatorio condicionalmente en  $X$  y el conjunto de datos  $D_n$ .

## 2.4.4 Métodos de Ensamble para Modelos de Predicción

El modelo *Random Forest* pertenece a los denominados “métodos de ensamble”. Estos comprenden un conjunto de técnicas que combinan múltiples modelos predictivos, para lograr un equilibrio entre la distancia promedio de las predicciones de un modelo respecto a los valores reales, y la varianza. Los métodos más utilizados se denominan *bagging* y *boosting*:

- *Bagging*: En lugar de ajustar un único árbol, se ajustan muchos de ellos en paralelo formando una especie de “bosque”, en sentido figurado. En cada nueva predicción, todos los árboles que forman el bosque participan aportando su predicción. Como valor final, si se trata de variables continuas se toma la media de todas las predicciones o, por el contrario, si se trata de variables cualitativas se considera la clase más frecuente. Uno de los métodos de *bagging* más utilizados y conocidos es precisamente *Random Forest*.

- *Boosting*: Consiste en ajustar secuencialmente múltiples modelos sencillos, llamados *weak learners*, de forma que cada modelo aprende de los errores del anterior. Como valor final, de la misma forma que en *bagging*, se toma la media de todas las predicciones si se trata de variables continuas, o la clase más frecuente si se trata de variables cualitativas. Dos de los métodos de *boosting* más empleados son *AdaBoost*, *Gradient Boosting Trees*.

## 2.4.5 Métricas y Formas de Evaluación

Cada uno de estos modelos presenta resultados diversos, donde aquel con mejor desempeño en su ejecución, menor cantidad de errores en la clasificación y validaciones estadísticas se adapte a los datos utilizados para la clasificación y predicción, será el utilizado para la continuidad de las actividades de gestión y modelo de clientes propuesto.

Teniendo presente que el objetivo es poder predecir correctamente aquellos clientes que tienen mayor probabilidad de fugarse, se pueden presentar las siguientes situaciones posibles:

- El modelo predice correctamente la fuga de un cliente, pero en realidad no lo hace. A esto denominamos *Falso Positivo [FP] o false positive*.
- El modelo predice que no existe fuga de cliente, pero efectivamente lo hace. A esto denominamos *Falso Negativo [FN] o false negative*.

En el caso en que el modelo sea asertivo en sus resultados significa que:

- Habrá clientes que se contactarán porque el modelo ha predicho que se fugarían y en realidad si lo hacen (*Verdadero Positivo [TP] o true positive*).
- Habrá clientes que no se contactarán porque el modelo ha predicho que no se fugarían y en realidad no lo hacen (*Verdadero Negativo [TN] o true negative*).

### 2.4.5.1 Matriz de Confusión

Una matriz de confusión de un problema de *clase n* es una matriz  $n \times n$  donde las filas representan el ‘escenario real’ y las columnas las clases previstas por el modelo, siendo fuga/no fuga del cliente. Utilizando los conceptos descritos en la sección anterior, se tiene que la configuración de la matriz queda de la siguiente forma:

**Tabla 3: Matriz de Confusión**

Fuente: Elaboración Propia

		<b>Predicción del Modelo</b>	
		<i>No Fuga [0]</i>	<i>Fuga [1]</i>
<b>Realidad</b>	<i>No Fuga [0]</i>	<b>TN</b>	<b>FP</b>
	<i>Fuga [1]</i>	<b>FN</b>	<b>TP</b>

#### 2.4.5.2 Precisión (*Positive Predictive Value*)

La métrica de precisión permite medir la *calidad* del modelo en tareas de clasificación. Para calcular la precisión se utiliza la siguiente fórmula:

$$precision = \frac{TP}{TP + FP} = \frac{precisiones\ correctas}{numero\ total\ de\ predicciones\ positivas}$$

Esto indica que el modelo fallará un  $(1 - \text{precisión})\%$  de las veces cuando prediga si un cliente se fuga o no.

#### 2.4.5.3 Especificidad

De la misma manera que la precisión, la especificidad mide la proporción entre los valores negativos del modelo y que lo son realidad, y el total de valores negativos observados.

$$especificidad = \frac{TN}{TN + FN}$$

Esto indica que es la probabilidad de clasificar correctamente a un asegurado cuyo estado real sea clasificado como ‘no fuga’.

#### 2.4.5.4 Recall (Exhaustividad)

Esta métrica informa la cantidad que el modelo de clasificación es capaz de identificar, vale decir, indica el porcentaje de los clientes que se fuga (y no) y que el modelo es capaz de reconocer.

Para el cálculo del *recall* se tiene:

$$recall = \frac{TP}{TP + FN} = \frac{precisiones\ correctas}{numero\ total\ de\ predicciones}$$

#### 2.4.5.5 F-1 o *f-measure*

Este valor combina las medidas de precisión y *recall* en un solo valor, de manera de poder comparar el rendimiento combinado de ambas métricas entre varias soluciones.

El F-1 se calcula haciendo la media armónica entre la precisión y el *recall*:

$$F1 = 2 \times \frac{\textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}}$$

#### 2.4.5.6 Accuracy (Exactitud)

La exactitud mide el porcentaje de casos en que el modelo ha sido asertivo. Si bien se utiliza como parte de las métricas más utilizadas para la selección de modelos de clasificación, tiene la mala fama de que puede ser engañosa, haciendo creer que un modelo que se sabe es ‘malo’, parezca que no lo es. Para evitar estos casos las clases consideradas no deben estar desbalanceadas.

Esta métrica se calcula:

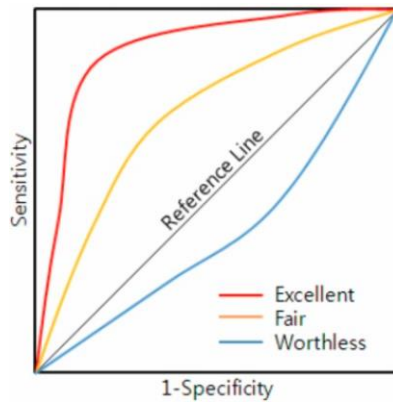
$$\textit{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

#### 2.4.5.7 Curva ROC (*Receiver Operating Characteristic*)

La curva ROC corresponde a una representación gráfica del rendimiento del modelo de clasificación a través de la distribución de las fracciones de verdaderos positivos [TP] y de falsos positivos [FP]. Esto es, una representación de la precisión frente a (1- especificidad), donde cada resultado de predicción representa un punto en el espacio de la curva ROC.

Esta curva permite comparar modelos a través del área bajo su curva, donde un modelo se considerará efectivo mientras recorra los bordes izquierdo y superior del gráfico, teniendo así una mayor cantidad de área.



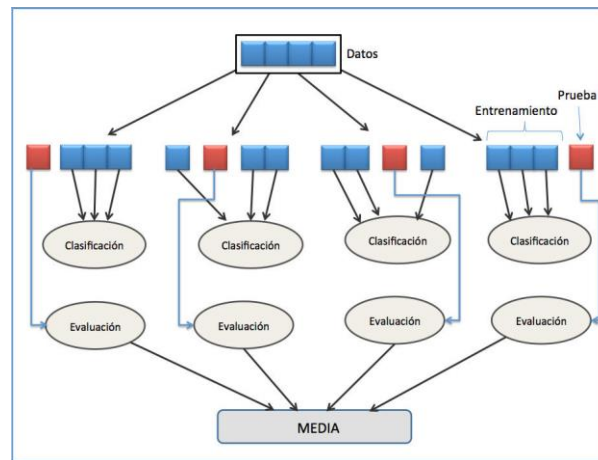


**Ilustración 3:** Tipos de curvas ROC.

Fuente: cbgstat.com

#### 2.4.5.8 Cross-Validation

La validación cruzada se utiliza para evaluar los resultados de un análisis estadístico cuando el conjunto de datos se ha dividido en una muestra de entrenamiento y otra de prueba, de manera de conocer si efectivamente los resultados del análisis son independientes de esta segmentación.



**Ilustración 4:** Cross-Validation representación gráfica.

Fuente: Domenech, J. 2011

#### 2.4.5.9 SMOTE (*Synthetic Minority Oversampling Technique*)

El problema de tener clases desbalanceadas se traduce en un incorrecto desempeño del modelo de predicción, ya que los datos utilizados para su entrenamiento distorsionan los resultados al momento de su entrenamiento.

Una forma de resolver este problema es sobre muestrear (oversampling) los datos de la clase minoritaria. Esto puede lograrse simplemente duplicando los datos de la clase minoritaria en el conjunto de entrenamiento antes de ajustar un modelo de predicción. Esto puede equilibrar la distribución de la clase, pero no proporciona ninguna información adicional al modelo.

Una mejora en la duplicación de datos de la clase minoritaria es “sintetizar” nuevos datos de la clase minoritaria. El enfoque más utilizado para “sintetizar” nuevos datos se denomina *Synthetic Minority Oversampling Technique* (SMOTE) (Chawla, 2002).

SMOTE funciona de la siguiente manera: primero se elige un dato al azar de la clase minoritaria. Luego se encuentran  $k$  de los vecinos más cercanos para ese dato (típicamente  $k=5$ ). Se elige un vecino seleccionado al azar y se crea un dato “sintético” en un punto seleccionado al azar entre los dos datos en el espacio de características.

## CAPÍTULO 3: PLANTEAMIENTO ESTRATÉGICO Y MODELO DE NEGOCIOS

### 3.1 Posicionamiento Estratégico

El posicionamiento estratégico de *MetLife* se identifica bajo la estrategia de “Mejor Producto” según el modelo Delta (Hax & Wilde, 2003), donde se caracteriza por tener numerosos clientes bajo fuertes regulaciones gubernamentales y estandarizaciones propias del rubro. Sumado a lo anterior, la existencia de una alta competencia por participación de mercado y su enfoque a la optimización en la creación de nuevos productos, da como resultado la diferenciación necesaria para que sus clientes prefieran los productos bajo su propuesta de valor. Pero este posicionamiento está sufriendo cambios importantes, donde ahora es el cliente quien se transforma en protagonista de su estrategia.

Esto obliga a la compañía a una redefinición en su estrategia hacia la búsqueda de una solución integral para el cliente, situándose en una posición de “Redefinición en la Experiencia del Cliente”. Esta redefinición se logra a través de una estrecha relación con los clientes y una potenciación de las capacidades internas, para ofrecer una solución integral y completa que comprenda sus necesidades críticas. Bajo esta premisa, la compañía declara que lleva a sus “clientes en el ADN”, como parte de esta nueva estrategia.



**Ilustración 5:** Posicionamiento Estratégico MetLife según Modelo Delta

Fuente: Adaptado de Hax & Wilde, 2003. Elaboración Propia

### 3.2 Balanced Scorecard

Los objetivos estratégicos reflejados en la figura siguiente muestran los principales propósitos de la compañía para cada uno de los cuatro enfoques: financiero, clientes, procesos y personas. Todos ellos incorporan las bases y lineamientos de las actividades llevadas a cabo al interior de la compañía, para el cumplimiento de los objetivos estratégicos establecidos:

Perspectiva	Objetivos
Financiera	<ul style="list-style-type: none"><li>• Sustentabilidad Negocio</li><li>• Optimizar Valor y el Riesgo</li><li>• Productos Rentables</li><li>• Persistencia Primas</li></ul>
Clientes	<ul style="list-style-type: none"><li>• Productos Atractivos.</li><li>• Retención.</li><li>• Satisfacción.</li><li>• Market Share.</li></ul>
Procesos	<ul style="list-style-type: none"><li>• Impulsar la Excelencia Operativa.</li><li>• Eficiencia.</li><li>• De oficina Local a Global.</li></ul>
Personas	<ul style="list-style-type: none"><li>• Retener Mejores Talentos.</li><li>• Alinear Objetivos.</li><li>• Simplificar las relaciones con nuestros socios y clientes.</li></ul>

**Ilustración 6:** BSC objetivos estratégicos MetLife.

Fuente: Elaboración Propia.

La creación y puesta en marcha de productos rentables y atractivos tienen un impacto directo en la perspectiva financiera y de clientes. Si el producto es atractivo para los consumidores, se obtiene una mayor cantidad de ingresos por primas, traduciéndose en un mayor retorno en el mediano/largo plazo.

Conservar la sustentabilidad del negocio por medio de una buena gestión de costos para creación y mantención de productos rentables, impulsar la excelencia operativa y retener a los mejores talentos en *MetLife*, forman parte de la importancia financiera, de procesos y personas respectivamente.

Para todos ellos existe una característica común y de importancia transversal dentro de la estrategia corporativa, que tiene un impacto directo en todas las perspectivas: la persistencia de los clientes. La persistencia es la capacidad de retener a los clientes a lo largo del tiempo, es decir, es el motor de los negocios en *MetLife*. Si ésta se conserva, se logra mantener la sustentabilidad financiera del negocio, ya que muestra la satisfacción de los clientes con la compañía, direcciona a los procesos y la cultura interna.

El cliente es la principal motivación y centro de los objetivos declarados, donde la compañía debe construir una relación producto-cliente que perdure un tiempo razonable para lograr sus objetivos financieros. Pero, la constante fuga de clientes y los bajos esfuerzos actuales de retención afectan fuertemente no sólo en la rentabilidad y sustentabilidad del negocio, sino que también a importantes acuerdos comerciales con socios estratégicos que pueden terminar con el modelo de negocio propuesto.

Este impacto se refleja directamente en las perspectivas financiera y clientes, lo que justifica ahondar en mayor medida en el comportamiento de los actuales clientes, con el fin de determinar y conocer el tipo de éstos que abandonan la compañía y el valor que representan directamente en sus resultados financieros.

### **3.3 Modelo de Negocios**

Este trabajo enfoca sus esfuerzos en la línea de negocios seguros masivos, la cual opera gracias a tres actores principales: empresas (*sponsors*), corredores (*bróker*) y clientes. Esta línea es descrita utilizando el modelo CANVAS (Osterwalder, 2010), donde es posible revisar el esquema visual en Anexo N° 2 de este documento.

- Propuesta de Valor: *Brindar la confianza necesaria para actuar en las necesidades de cobertura de riesgos, por medio de soluciones que permitan proteger lo que al cliente le importa y aprovechar al máximo lo que tiene, a través de respuestas directas y relaciones basadas en la confianza.*

El modelo de negocio de Seguros Masivos puede describirse como sigue:

- Por un lado, las empresas desean mantener a sus clientes ofreciéndoles nuevos y mejores servicios para aumentar sus ganancias, contando además con una numerosa base de clientes quienes se encuentran asociados a un medio de pago o afinidad particular (boletas de servicio, tarjetas de crédito, por mencionar algunos).
- Por el otro lado, *MetLife* desea aumentar su participación de mercado y sus ganancias en el mediano y largo plazo como parte de la estrategia declarada en el BSC.
- Ambas entidades se encuentran enlazadas a través de corredores o *brókers* de negocios, que son responsables de ejecutar los acuerdos comerciales entre ambos, llevándose un porcentaje de comisión por los negocios administrados.

- Con los acuerdos comerciales ya declarados, comienza la creación y distribución de pólizas de seguros a la cartera de clientes de las empresas a través de distintos canales de venta como telemarketing, OTC (*over the counter*) o por medios digitales (páginas web, aplicaciones móviles), llegando a la masividad de clientes con que se dispone, quienes compran los productos y generan la utilidad esperada.
- Esta utilidad se logra bajo márgenes permitidos de persistencia de clientes, donde debe existir un rango aceptable de los mismos, previamente estudiado, para que los ingresos por pago recurrente de primas y, devoluciones de experiencias favorables (DEF), cubran todas las necesidades de mantención, gastos y utilidades de las compañías participantes.
- La estructura de costos de este modelo de negocios se basa en el pago de comisiones por emisión de pólizas, desarrollo de nuevos productos, costos de administración, mantención y siniestralidad de la cartera.

Se puede desprender entonces, que la persistencia de clientes es un factor clave dentro de la configuración de la línea de negocios, ya que permite la sustentabilidad y utilidad esperada por quienes participan de él, transformándose en una métrica de gestión importante para el reflejo de la satisfacción de los clientes con los productos/servicios ofrecidos y, además de la rentabilidad del negocio. Esto tiene sentido siguiendo la estrategia de negocios con que MetLife realiza sus operaciones en la búsqueda de su propósito, por lo que se hace imperativa la búsqueda de una solución de fondo a la constante fuga que los clientes están evidenciando en el tiempo.

## **CAPÍTULO 4: ANÁLISIS DE LA SITUACIÓN ACTUAL**

---

La metodología de *Ingeniería de Negocios* (Barros, 2004) mediante la aplicación de la arquitectura de procesos, permite identificar dentro de la cadena de valor de seguros masivos, la oportunidad de mejora dentro de los procesos y actividades existentes en la compañía, de manera de comprender las dependencias y resultados bajo este esquema metodológico.

En este capítulo se describen los macroprocesos involucrados desde el nivel más general para continuar con los subprocesos particulares, llegando al proceso de “*Análisis y Gestión de Clientes Seguros Masivos*” como determinante del éxito y aplicabilidad del proyecto que se busca rediseñar.

Los macro y subprocesos se presentan en formato IDEF-0, mientras que el proceso que se busca rediseñar se muestra en formato BPMN.

### **4.1 Modelamiento Detallado de Procesos**

La arquitectura de macroprocesos identifica cuatro procesos macro, indicados por niveles (I – IV), que deben estar presentes en todas las organizaciones.

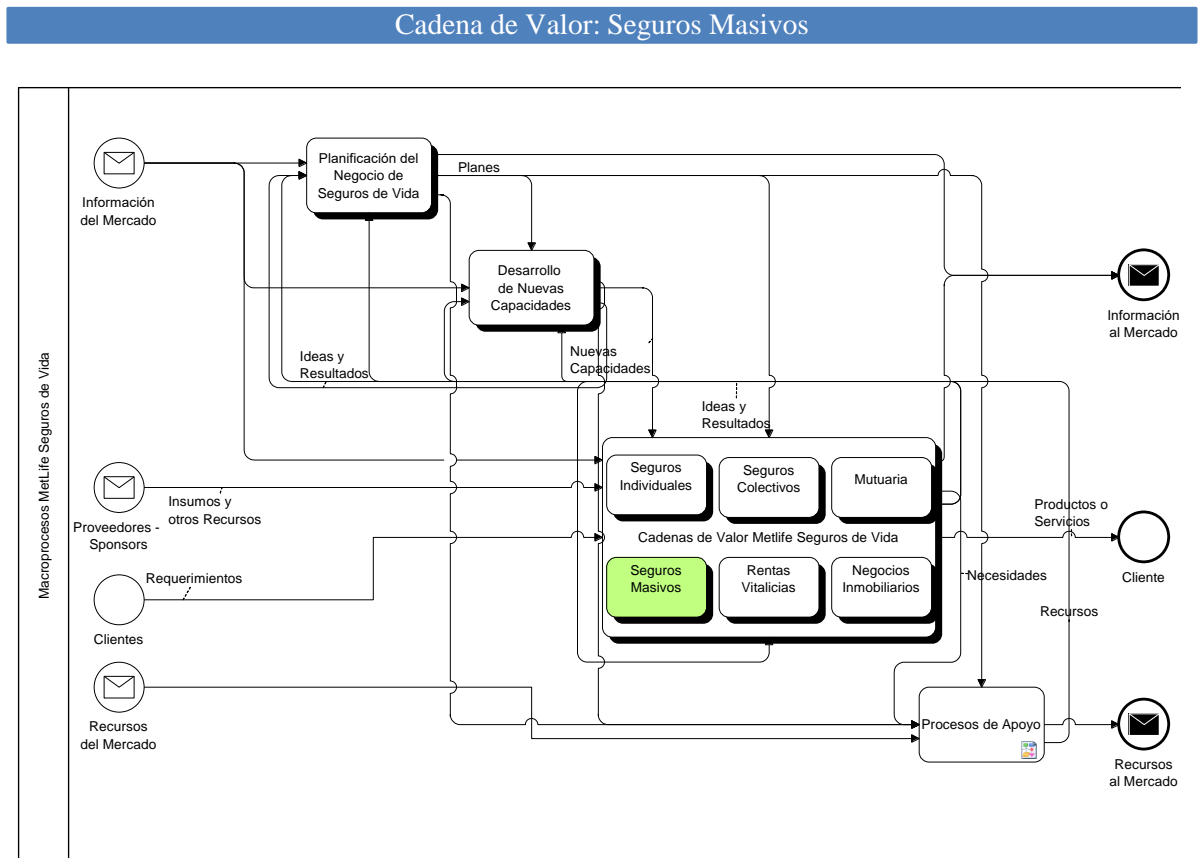
- Cadena de Valor (Nivel I): Contiene las actividades que las empresas deben desarrollar para planificar, producir y entregar al cliente sus productos o servicios.
- Desarrollo de nuevas capacidades (Nivel II): Comprende las actividades relacionadas con el estudio de nuevas capacidades que la empresa debería implementar para ser competitiva (tecnologías o proyectos que inciden en la cadena de valor, por ejemplo).
- Planificación estratégica (Nivel III): Agrupa las actividades relacionadas con la determinación de lineamientos estratégicos de la organización, las cuales se materializan en planes y programas de acción a ser adoptados en las operaciones de esta.
- Gestión de recursos habilitadores (Nivel IV): Abarca las actividades que dan soporte a la ejecución de los otros tres macroprocesos (recursos humanos, insumos, entre otros).

#### **4.1.1 Modelamiento IDEF0**

- Macroprocesos Nivel I:

El primer nivel describe de manera generalizada las interacciones y actividades que se deben desarrollar para la entrega final de los productos o servicios. En MetLife y tal

como se describe en el capítulo I, existen seis cadenas de valor que componen el macroproceso del mismo nombre, las cuales se muestran en la figura siguiente:



**Ilustración 7:** Diagrama de Proceso BPMN de Macroprocesos (Nivel I) para MetLife Seguros de Vida

Fuente: Elaboración Propia (Adaptado de Metodología Ingeniería de Negocios)

Estas cadenas de valor cuentan con sus respectivos planes estratégicos que se determinan según el modelo de negocio, permitiendo así definir las capacidades y objetivos en la construcción del diseño de negocio que cumpla con las especificaciones necesarias para lograrlos. Este trabajo centra sus esfuerzos en la cadena de valor correspondiente a la línea de negocios de *seguros masivos*.

En esta línea corresponde la creación, gestión y comercialización de seguros masivos para clientes de grandes empresas (bancos, clínicas, *retails*, entre otros) y que, en su necesidad de protección, contratan seguros con la compañía a través de los distintos canales de venta habilitados, comprometiendo el pago de primas a través de un medio de afinidad como tarjeta de crédito.

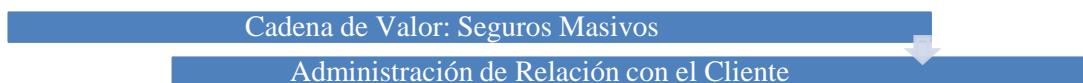


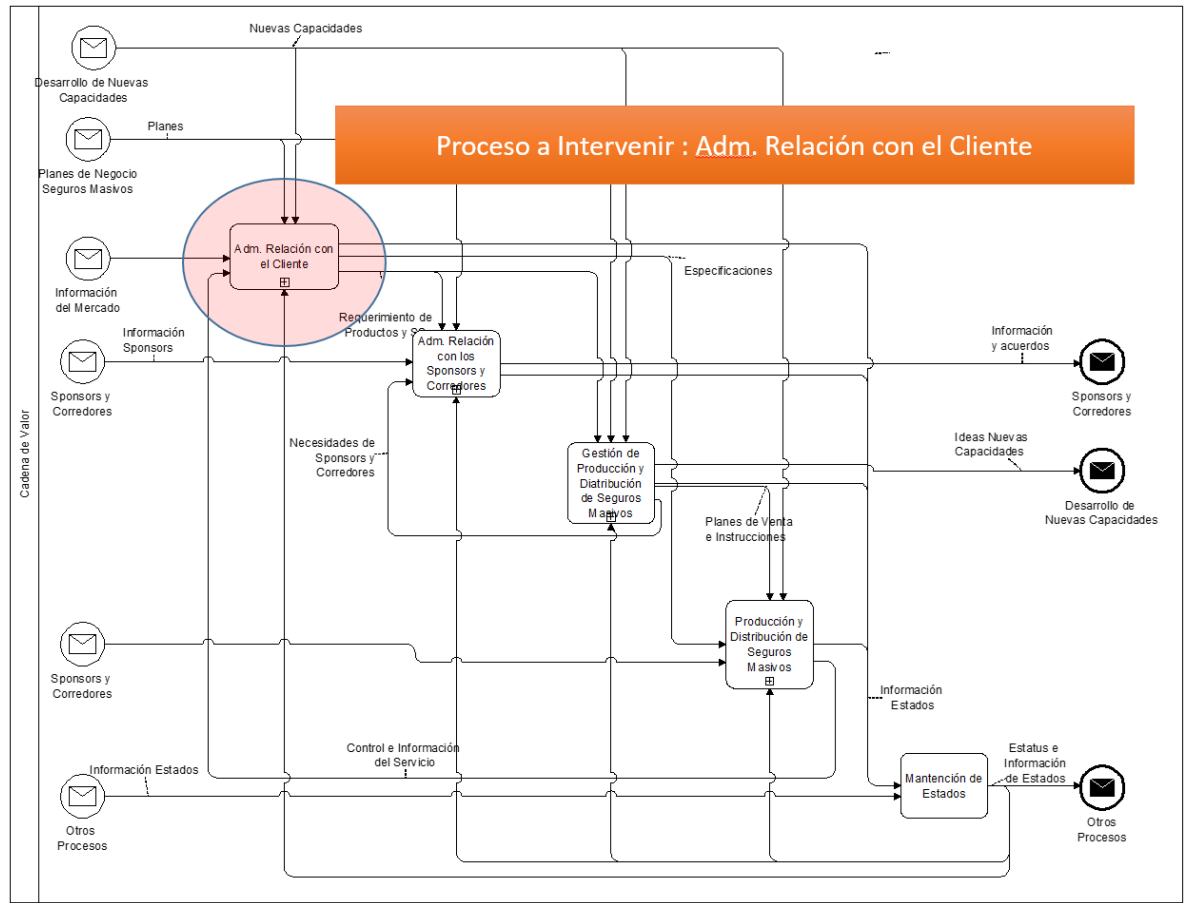
- Macroprocesos Nivel II:

Dentro de la cadena de valor seleccionada existen los procesos relacionados con el estudio de nuevas capacidades y la relación contenida entre ellos, bajo el modelo de arquitectura siguiente:

- *Administración de Relación con el Cliente*: Corresponden a los subprocesos asociados a la relación con los clientes en temas de ventas, planes de negocio, información de mercado, marketing y requerimientos propios.
- *Administración de Relación con los Sponsors y Corredores*: Corresponden a los subprocesos asociados con los sponsors y corredores de seguros, quienes habilitan las bases de clientes para el desarrollo de la línea. Los procesos se relacionan con la especificación de productos, requerimientos, compromisos contractuales - comerciales, seguimiento.
- *Gestión de Producción y Distribución de Seguros Masivos*: Corresponden a los subprocesos asociados a los requerimientos y configuración de los productos a comercializar bajo las plataformas acordadas, para seguros masivos.
- *Producción y Distribución de Seguros Masivos*: Corresponden a los subprocesos asociados a la configuración y suscripción de seguros individualizados para cada *sponsor*. Estos seguros se suscriben y configuran en la compañía pero que, deben ser personalizados según el tipo y requerimiento del *sponsor*.
- *Mantenimiento de Estados*: Corresponden al registro, información y control del estado de las actividades de transformación.

El macroproceso “*Administración de Relación con el Cliente*” es el indicado para continuar el análisis en detalle al siguiente nivel, ya que comprende procesos directamente relacionados con la interacción de clientes con la compañía. Entre ellos, las acciones de marketing junto con diversos planes de negocios e información del mercado, los que pueden ser factores significativos en la decisión de permanencia con los productos para los clientes en *MetLife*.





**Ilustración 8:** Macroprocesos Nivel II, Adm. Relación con el Cliente.

Fuente: Elaboración Propia (Adaptado de Metodología Ingeniería de Negocios)

- Macroprocesos Nivel III:

Desde acá se configuran, en conjunto con los planes estratégicos y capacidad del negocio, los distintos análisis y acciones de mercado que puedan impactar a la compañía para determinar, el tipo de clientes sobre los cuales se debe tomar acciones necesarias para lograr la retención y/o permanencia buscada.

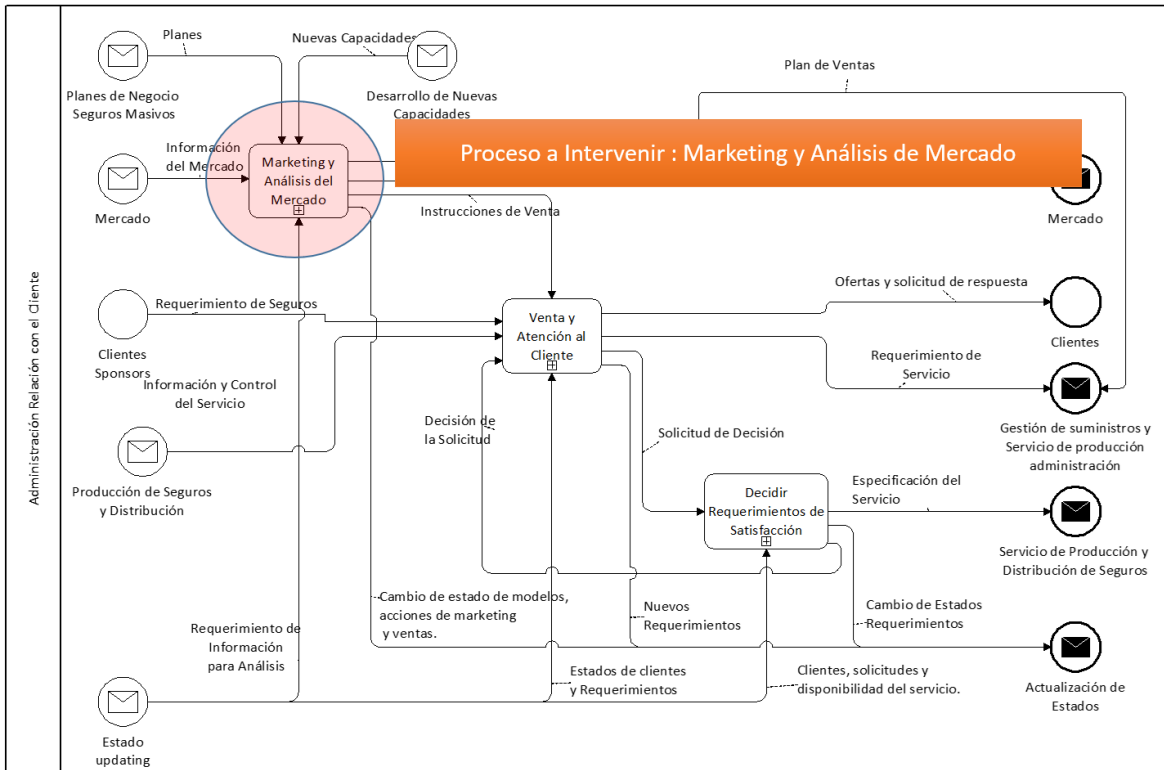
Para ello desde el proceso más general hasta el más específico, se ha identificado el subproceso “Marketing y Análisis de Mercado” a intervenir como el responsable de las acciones que tienen injerencia sobre las decisiones de los clientes y que configuran las instrucciones de venta y atracción de estos hacia la compañía.

La ilustración siguiente muestra los procesos asociados a la cadena de valor seleccionada dentro del proceso de “Administración y relación con cliente”.

## Cadena de Valor: Seguros Masivos

### Administración de Relación con el Cliente

### Marketing y Análisis del Mercado



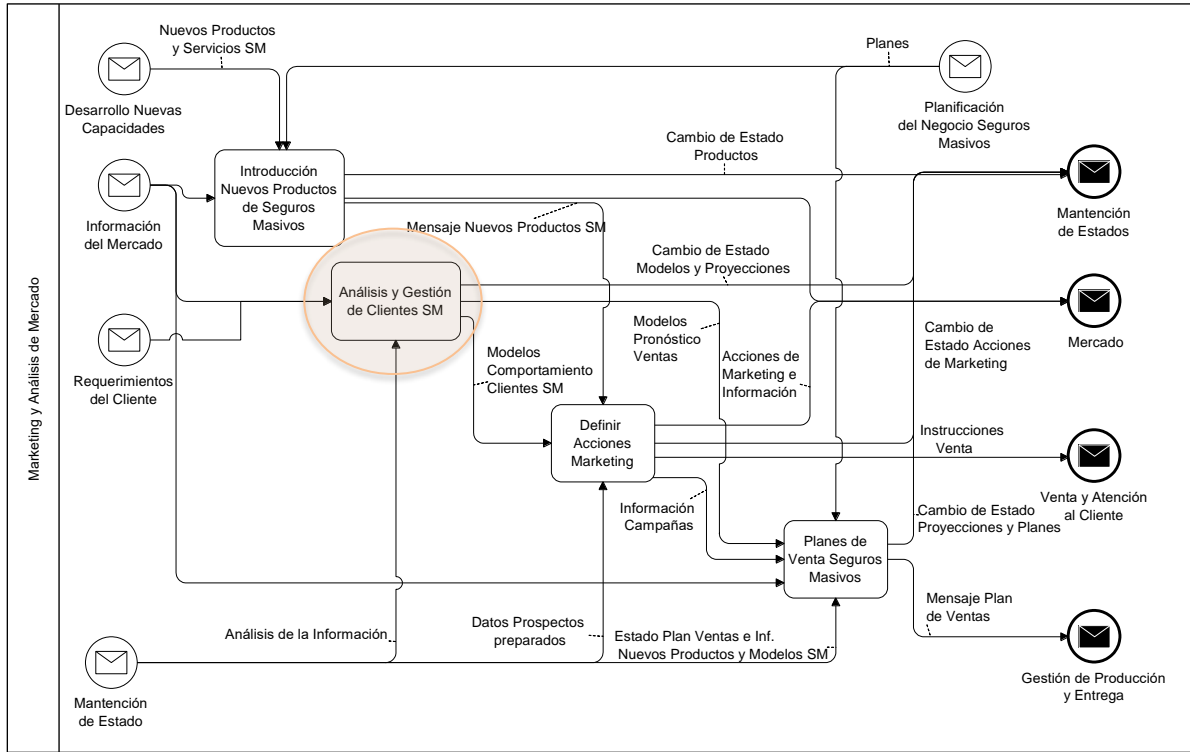
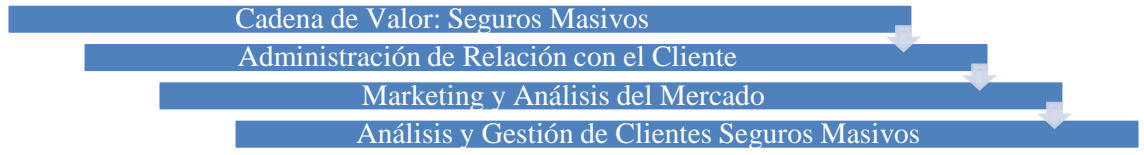
**Ilustración 9:** Macroprocesos Nivel III: Marketing y Análisis del Mercado

Fuente: Elaboración Propia (Adaptado de Metodología Ingeniería de Negocios)

#### ▪ Macroprocesos Nivel IV

Es en este nivel donde se realizan las actividades de planificación y análisis del comportamiento de los clientes respecto a sus productos, ventas y otra información disponible que son de vital ayuda para la toma de decisiones. Sobre todo, para determinar sobre qué clientes es necesario aplicar medidas de retención, que serán primordiales en el accionar de las áreas comerciales.

Se desprende desde acá el subproceso: “*Análisis y Gestión de Clientes de Seguros Masivos*”, donde es posible identificar actividades que tienen como característica principal la interacción directa del personal con clientes bajo condiciones o solicitudes de abandono. En la sección siguiente quedarán reflejadas las actuales falencias en dicho proceso, dadas fundamentalmente por la falta de actividades y acciones de retención, por no contar con información y conocimiento suficiente de los clientes actuales.

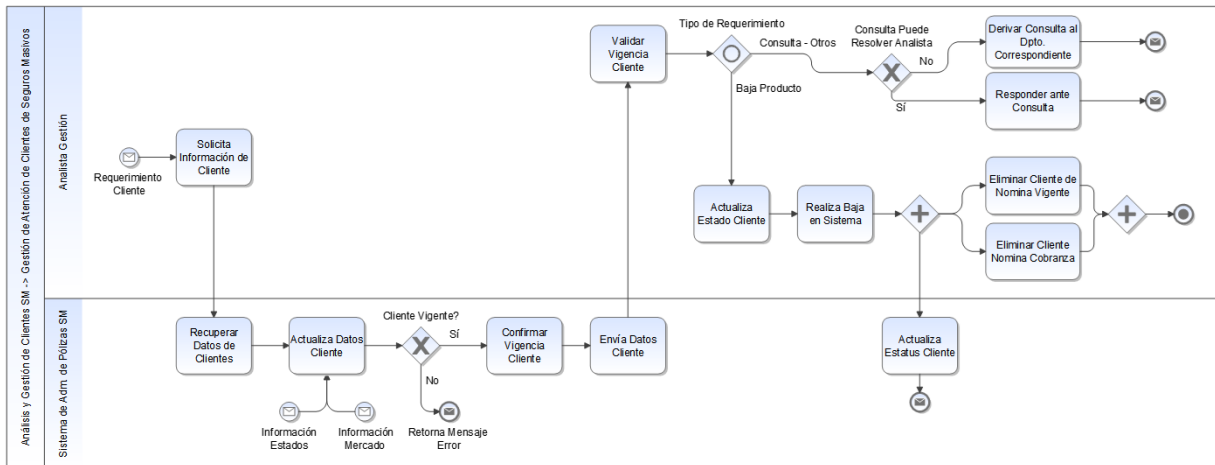


**Ilustración 10:** Macroproceso Nivel IV, Análisis y Gestión del Comportamiento Clientes Seg. Masivos.

Fuente: Elaboración Propia (Adaptado de Metodología Ingeniería de Negocios)

## 4.2 Diagnóstico de la Situación Actual

En el desarrollo de esta metodología se debe llegar al mayor nivel de detalle posible, de manera de identificar el proceso que será objeto de análisis y rediseño en este proyecto. Este proceso tal como se identifica en la sección anterior, corresponde dentro del proceso de “Análisis y Gestión de Clientes de Seguros Masivos”, al subproceso de “Gestión de Atención de Clientes de Seguros Masivos”. En la ilustración siguiente se detalla la situación actual para luego describir en detalle sus actividades comprendidas en él:



**Ilustración 11:** Diagrama actual de subproceso para Gestión de Atención de Clientes de Seguros Masivos.

Fuente: *MetLife*, Elaboración Propia.

1. El analista de Atención de clientes recibe requerimiento del cliente, por medio electrónico (email), presencial o carta.
2. El analista solicita la información por medio de consulta al Sistema de Administración de Pólizas de Seguros Masivos.
3. Este sistema automáticamente recupera los datos de los clientes por medio de consultas y procesos automáticos.
4. Actualizando los datos de clientes por medio de información de estados y del mercado (existiendo otras actividades de apoyo a la actualización de esta información).
5. El Sistema de Administración de Pólizas consulta por medio de procesos de validación si el cliente se encuentra vigente.
  - En caso Afirmativo, el sistema confirma la vigencia del cliente y envía los datos del cliente.
  - En caso Negativo, retorna mensaje de error.
6. El analista valida la vigencia de cliente según los datos que el Sistema le dispone.
7. Luego, según el tipo de requerimiento:
  - Si es requerimiento es de tipo consulta:

- El analista puede enfrentar dos situaciones: que cuente con la información para dar respuesta al requerimiento, donde enviará la respuesta al cliente; en caso contrario, deriva la consulta al departamento de Operaciones correspondiente.
- Si el requerimiento es de tipo “Baja Producto”:
  - *El analista actualiza el Estado del cliente, realiza la baja en el Sistema, y así elimina al cliente de la nómina de clientes vigentes y de la nómina de cobranza, dando fin al proceso.*

Es posible notar a priori la falta de actividades propias de retención de clientes dentro de este proceso, ya que se realiza la gestión de baja (abandono) de los productos en forma inmediata, sin recurrir a estrategias de mantención por falta de herramientas e información que permitan hacerlo, afectando directamente en los resultados de la línea como se detalla en la siguiente sección.

### 4.3 Cuantificación del Problema u Oportunidad

Identificado el origen de la fuga de clientes dentro del proceso de Análisis y Gestión de Clientes SM descrito, es tiempo de cuantificar (de manera aproximada) lo que significa la fuga de éstos en términos de pérdidas monetarias, para la línea de negocios a lo largo del tiempo.

Considerando un precio (prima) promedio de una póliza de seguros en UF 0,235 (aprox. \$6.486) con un número mensual de bajas de clientes en promedio cercano a los 17.000, junto con considerar criterios de siniestralidad y otros costos constantes, se obtiene lo siguiente:

**Tabla 4:** Cuantificación estimada del impacto de fuga clientes en pesos (\$)

Fuente: Elaboración Propia

Cuantificación Fuga Clientes	Promedio de Bajas Mensual #	16.949
	Precio Promedio Póliza UF	0,235
	Perdida Estimada UF Mensual	3.982,96
	Perdida Estimada \$ Mensual	110.128.921
	Perdida Estimada \$ Anual	1.321.547.050

La pérdida anual esperada manteniendo una baja reiterada de 16.949 pólizas de forma mensual, representan para la compañía una merma de más de 1.3MM de pesos, simbolizando la importancia de realizar acciones de retención con urgencia dentro de un proceso que tiene

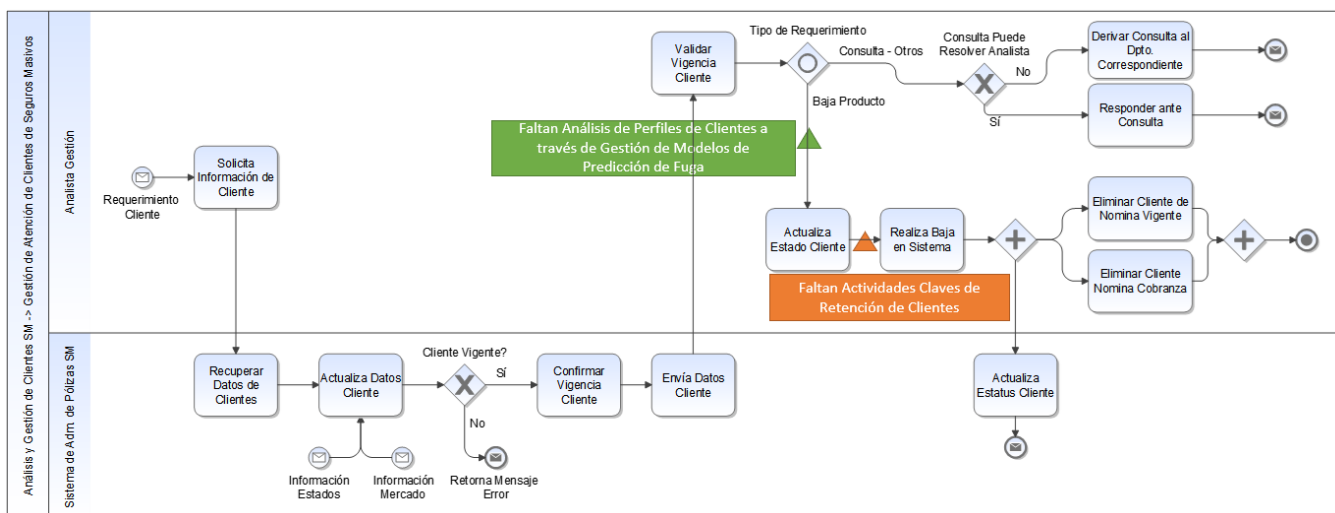
relación directa con los clientes y que puede significar un cambio importante en el cómo ellos ven a la compañía hacerse partícipe de sus preocupaciones, brindándoles alternativas de protección acordes a sus requerimientos.

## CAPÍTULO 5: PROPUESTA DE DISEÑO DE PROCESOS

### 5.1 Direcciones de Cambio

Luego de realizar el análisis y apertura de los macroprocesos descritos en la metodología de Ingeniería de Negocios, la propuesta de cambio impulsada por este trabajo nace con el rediseño del proceso descrito en el capítulo anterior y que corresponde al denominado “*Gestión de Atención de Clientes de Seguros Masivos*”.

En este proceso es posible distinguir la falta de actividades clave para la retención de clientes ante su manifestación de abandono (detallado en Ilustración 12), donde se hace necesario contar con información relevante acerca de sus preferencias y satisfacción con los productos/servicios prestados, que permitan determinar la probabilidad de fuga. Esto último es posible a través de la gestión de modelos de predicción, generando información valiosa a quienes se enfrentan a los clientes en sus demandas por renuncias al producto/servicio, con el fin de proporcionar ofertas de retención por tipo de cliente, con actividades permitan realizar acciones de retención para mantener su preferencia con la compañía.



**Ilustración 12:** Diagrama de Proceso BPMN para Gestión de Atención de Clientes en Seguros Masivos con Identificación de Procesos Clave Faltantes.

Fuente: Elaboración Propia.

De esta forma, es posible aplicar acciones de retención sobre los clientes para mejorar su persistencia dentro de línea de negocios y así aumentar los niveles de rentabilidad y sustentabilidad de la línea, manteniendo las relaciones comerciales y contractuales pactadas con los *sponsors* en el largo plazo.



## 5.2 Arquitectura de Procesos To Be

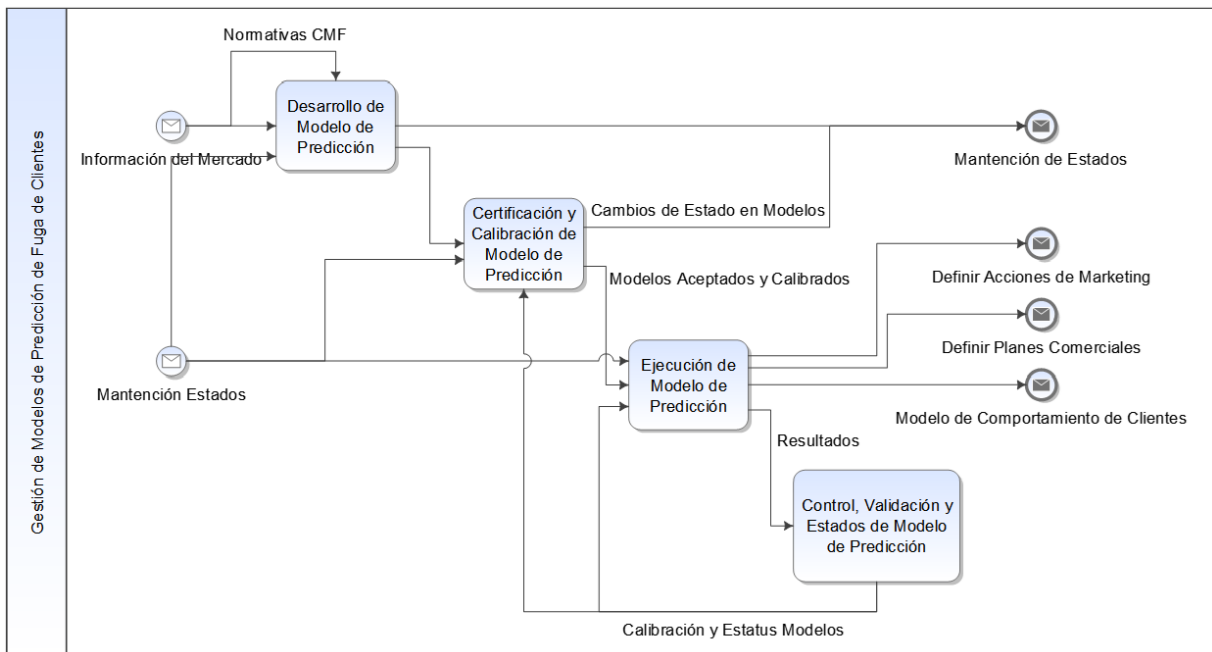
El rediseño del proceso considera la incorporación de nuevas actividades/procesos que involucren la ejecución de modelos de predicción de fuga, los cuales resulten finalmente en recomendaciones de retención. Utilizando las herramientas de minería de datos bajo la metodología CRISP-DM, se propone un nuevo proceso denominado “*Gestión de Modelos de Retención de Clientes*”, que formará parte del proceso descrito anteriormente para la generación, uso y calibración del modelo predictivo.

Este nuevo proceso cuenta con subprocesos y actividades clave descritas a continuación:

- **Desarrollo de Modelos de Predicción;** donde se aplica metodología analítica y reglas de negocio para la elaboración de modelos predictivos que reflejen de mejor manera a los clientes y sus preferencias para con los productos de seguros masivos.
- **Certificación y Calibración de Modelos de Predicción;** durante el cual se valida la construcción de los modelos de retención anteriormente desarrollados, se arrojan resultados y se realizan los ajustes necesarios según la lógica de negocio.
- **Ejecución de Modelos de Predicción;** una vez validado el (los) modelo(s) se incorporan en el proceso impactado para la definición de acciones futuras (comerciales – marketing) para la retención y satisfacción de los clientes.
- **Control, Validación y Estados de Modelos de Predicción;** por último, la etapa de control y estados de modelos permitirán determinar si éstos reflejan los objetivos y resultados esperados, que ayudarán en la calibración y estatus de éstos para su mejora continua.

Todo estos se describen con detalle utilizando la notación IDEF-0 en la sección siguiente.

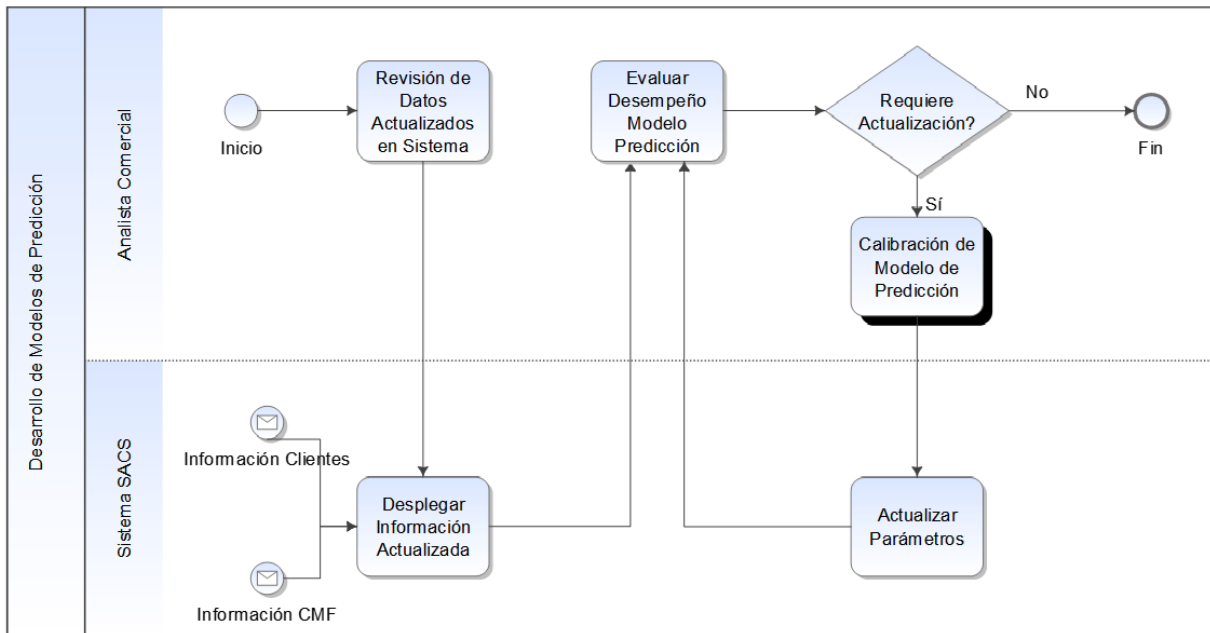
### 5.2.1 Diseño en IDEF0 – Gestión de Modelos de Predicción de Fuga de Clientes



**Ilustración 13:** Nuevo Proceso de Gestión de Modelos de Predicción de Fuga de Clientes.

Fuente: Elaboración Propia

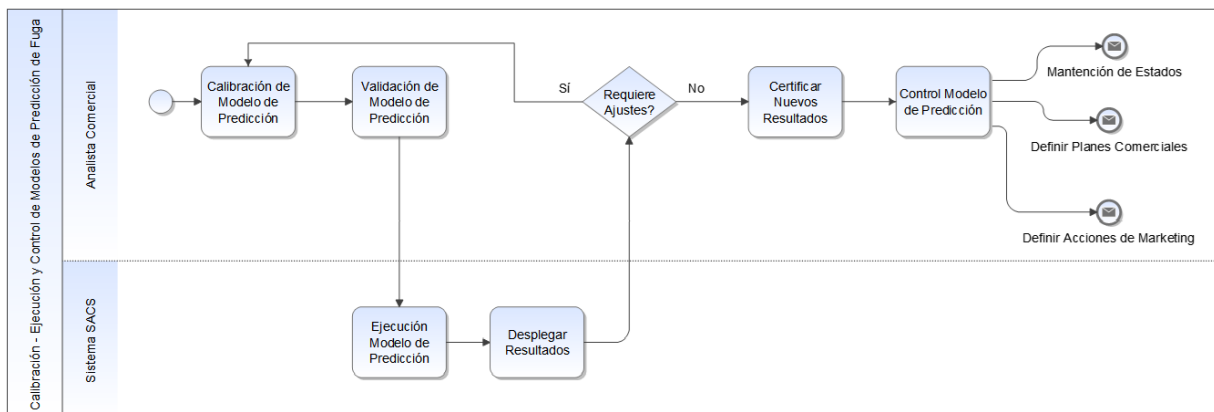
Los subprocesos y actividades contenidas en el nuevo proceso descrito permitirán la optimización de los modelos de predicción, mediante la calibración y actualización de datos de manera continua. Este proceso no forma parte en la actualidad, por lo que será necesario contar con un especialista de datos o en su defecto dentro de la dirección de Operaciones, asignar uno de los analistas a cargo de la gestión comercial de la línea de negocios, que realice las actividades de calibración y control del modelo a medida que se incorporan nuevos datos.



**Ilustración 14:** Actividades contenidas en el “Desarrollo de Modelos de Predicción”

Fuente: Elaboración Propia

El desarrollo de Modelos de Predicción inicia con una revisión detallada de los datos que se encuentran registrados en el Sistema por la persona definida para tal efecto, quien validará y ejecutará el modelo de predicción. Luego el Sistema desplegará los resultados arrojados por el modelo, donde el especialista deberá decidir si es necesario realizar ajustes al modelo. En caso afirmativo, se realiza la calibración del modelo:



**Ilustración 15:** Actividades contenidas en Calibración - Ejecución y Control de Modelos de Predicción de Fuga

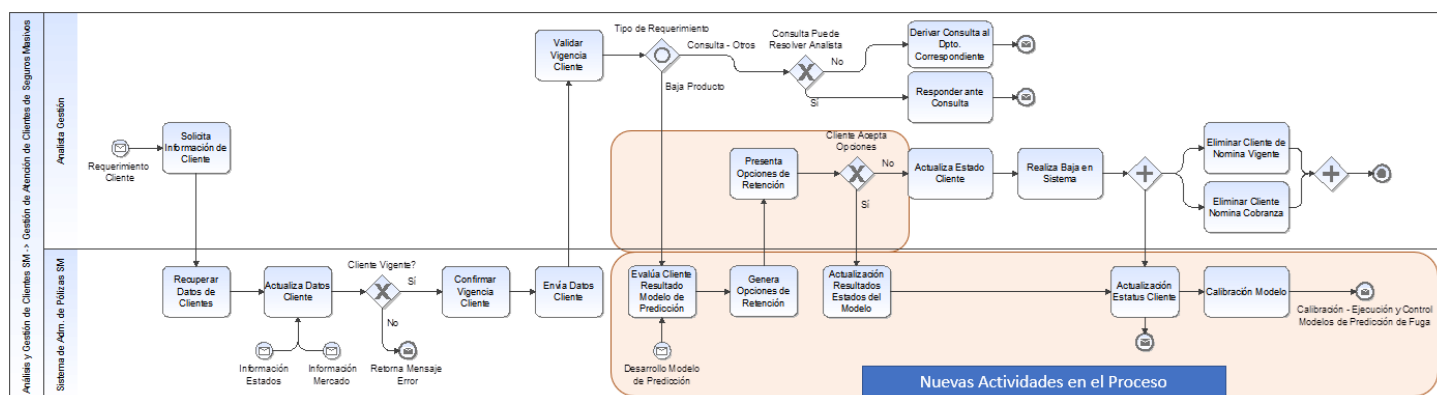
Fuente: Elaboración Propia

La calibración del modelo implica que los resultados obtenidos de la ejecución del modelo no son satisfactorios a los buscados, por lo que se deben validar los datos contenidos en el modelo de predicción y detectar anomalías. Con lo anterior realizado, se ejecuta nuevamente el modelo

en el Sistema que arrojará nuevos resultados. Si es necesario realizar ajustes al modelo se vuelve a recalibrar, en caso negativo, se certifican los resultados obtenidos y se realiza el control y gestión del modelo de predicción, permitiendo la definición (si aplican) de nuevos planes comerciales, acciones de marketing y estados del negocio para ajustar las estrategias de retención de clientes.

## 5.2.2 Diseño en BPMN

La incorporación de estas nuevas actividades y subprocessos dentro de proceso “Análisis y Gestión de Clientes Seguros Masivos”, involucra modificación de las actividades actuales del mismo, para soportar el rediseño y adecuar el flujo a los resultados que se buscan, quedando de la forma siguiente:



**Ilustración 16:** Diagrama de Proceso BPMN para Gestión de Atención de Clientes en Seguros Masivos con nuevas actividades propuestas.

Fuente: Elaboración Propia

El cambio dentro de este proceso ocurre cuando el cliente intenta dar de baja o abandonar el producto que mantiene con la compañía, el cual se describe como sigue:

1. El analista de Atención de clientes recibe requerimiento del cliente, por medio electrónico (email), presencial o carta.
2. El analista solicita la información por medio de consulta al Sistema de Administración de Pólizas de Seguros Masivos.
3. Este sistema automáticamente recupera los datos de los clientes por medio de consultas y procesos automáticos.
4. Actualizando los datos de clientes por medio de información de estados y del mercado.

5. El Sistema de Administración de Pólizas consulta por medio de procesos de validación si el cliente se encuentra vigente.
  - En caso Afirmativo, el sistema confirma la vigencia del cliente y envía los datos del cliente.
  - En caso Negativo, retorna mensaje de error y culmina el proceso.
6. El analista valida la vigencia de cliente según los datos que el Sistema le dispone.
7. Luego, según el tipo de requerimiento:
  - Si el requerimiento es de tipo “Baja Producto”:
    - El analista dispondrá de información del perfil del cliente dado por los resultados entregados en el modelo de predicción (que será descrito en la sección siguiente). Sumado a lo anterior, el analista podrá acceder a un listado de acciones comerciales de retención que deben ser presentadas al cliente.
    - Si el cliente acepta la propuesta dada por el analista comercial, éste último actualiza la base de cliente con la retención exitosa y con la propuesta aceptada para su mantención, permitiendo la calibración del modelo. Luego de esto, se envía aceptación de la oferta comercial ofrecida y se mantiene cliente en estado “Vigente” (retención exitosa).
    - En caso de que el cliente no acepte la propuesta, se actualiza el estado del cliente a “No Vigente” (baja) en el sistema para luego proceder a su eliminación de la nómina de asegurados vigentes y cobranzas, actualizando la base de datos del cliente para calibración del modelo.
  - Si es requerimiento es de tipo consulta:
    - El analista puede enfrentar dos situaciones: que cuente con la información para dar respuesta al requerimiento, enviará la respuesta al cliente; en caso contrario, deriva la consulta al departamento de Operaciones correspondiente.

## CAPÍTULO 6: LÓGICA DE NEGOCIO

---

### 6.1 Preparación y Construcción del Modelo Predictivo

El objetivo de negocio propuesto en el desarrollo de este trabajo incluye el conocimiento de la intención de fuga de clientes en base a la información contenida en la compañía, considerando el comportamiento histórico (pasado) de los clientes y sus características, mediante la utilización del modelo de predicción más adecuado que lo sustente.

#### 6.1.1 Selección de Datos (Set de Datos)

Los datos utilizados son de propiedad de *MetLife* y se encuentran almacenados en diversas fuentes provenientes del sistema de administración de pólizas “SACS”, donde se realizan todas las funciones sobre las pólizas *bancassurance*.

En primer lugar, la diversidad de información que contienen estas bases como datos demográficos, financieros, temporales, personales entre tantos otros, llevan a discutir previamente con los expertos de negocio aquellos que son verdaderamente primordiales durante la configuración y creación de estos productos. Desde acá se desprenden inicialmente 31 variables características que son utilizadas en la construcción de la base de datos (véase Anexo N°3), las cuales son extraídas utilizando lenguaje de consulta SQL bajo una temporalidad comprendida entre los meses de septiembre 2018 y enero 2019 (5 meses), tomando como mes basal el primero de ellos, con un total de 98.310 registros.

##### 6.1.1.1 Exploración del Set Datos (Set de Datos)

Es posible encontrar información relevante asociada a la línea de negocios, realizando una exploración e interpretación del set de datos:

- Los datos contienen información de pólizas de clientes pertenecientes a cinco importantes *sponsors* bancarios: Itaú, BBVA, Scotiabank, Banco de Chile y Banco Security.
- 57.060 registros corresponden a pólizas de Accidentes y 41.250 a pólizas de Salud.
- Más de 67.000 pólizas concentran asegurados mayores a 35 años.
- 62.7% corresponden a asegurados de sexo masculino y 37.3% de sexo femenino.

### 6.1.2 Transformaciones de Datos

La base de datos incluye variables cualitativas que son transformadas en variables *dummy* (variables de tipo binarias que tienen valores 0 y 1, que indica la ausencia o presencia de la cualidad o atributo descrito). De esta forma, es posible incorporar dichas variables dentro de los modelos de predicción sin perder información que puede ser valiosa para la descripción e interpretación posterior de los resultados.

De la misma forma, tanto la edad como el número de meses en que la póliza estuvo vigente, se obtienen de la transformación de la fecha de nacimiento y la diferencia entre fechas de inicio y fin de vigencia. Así mismo, se decide aperturar en intervalos conocidos los valores de edad, prima devengada y capital asegurado, para completar la base con el valor principal (*label*) como input en los modelos predictivos, el cual indica si la póliza se encuentra *Vigente (0) o No Vigente (1)*, (véase Anexo 4).

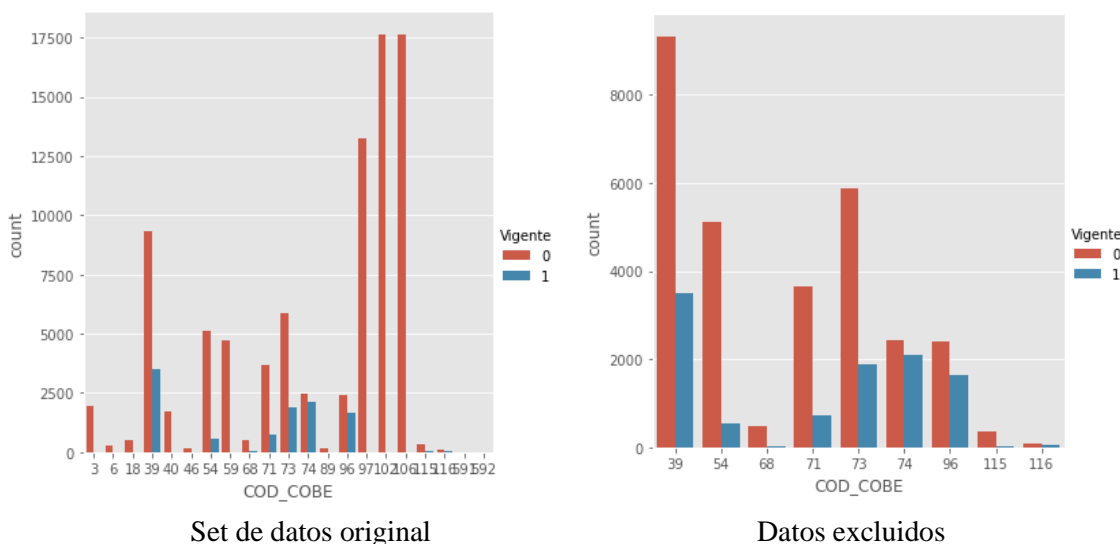
### 6.1.3 Correlación de Datos

Antes de proceder con la aplicación de los modelos predictivos, es necesario conocer si las variables utilizadas son independientes entre sí. Para ello se utiliza la matriz de correlación de datos (Anexo N°5), que indica la existencia de una relación negativa (si uno aumenta el otro disminuye) si el valor es igual a -1 o, por el contrario, si existe una relación positiva (si uno aumenta el otro también lo hace) si el valor es +1. Si el valor es cercano a cero, las variables son independientes entre sí, que es lo que se espera.

- La incorporación de las variables de Prima Devengada, Edades y Coberturas dummy, tendrán una relación negativa entre ellas.
- Existe una relación positiva entre el Riesgo (Cod\_Riesgo) y la Prima Vigente con que el cliente mantiene su cobertura mensualmente.
- Se incorpora la variable de predicción “Vigente” (valor 1 = No Vigente o Fuga), donde es posible observar que la variable Meses Vigente tiene una relación positiva, es decir, que mientras más antiguo sea, mayor es la probabilidad de fuga.
- Así mismo, existe una relación diferenciada entre los tramos de Capitales de los productos, con la probabilidad de fuga: positiva si el Capital Asegurado es < UF 300, lo que dice que aquellos productos que cumplan esta regla tienen mayores probabilidades de fuga; mientras que es negativa en cualquier otro caso.

### 6.1.4 Exclusión de Datos

El set de datos contiene información detallada de distintos tipos de riesgos (coberturas) asociadas a las pólizas de seguros contratadas por los asegurados. Analizando su distribución bajo el criterio de la variable categórica en estudio (vigente), es posible observar que 12 de un total de 21 coberturas registradas en la base, sólo muestran valores de pólizas vigentes en el periodo de estudio considerado, sin ningún registro de fuga de clientes. En el futuro, estos datos pueden traer problemas durante la evaluación de los modelos de predicción al estimar erróneamente fugas sobre coberturas (pólizas) que no lo son, por lo que se excluyen del data set original, reduciendo el universo de datos a 40,277:



**Ilustración 17:** Distribución de Datos por Cobertura y Número de Pólizas Vigentes/No Vigentes.

Fuente: Elaboración Propia

Si bien el número de datos excluidos no es menor, se cuenta con suficiente cantidad para lograr resultados satisfactorios.

### 6.1.5 Set de Entrenamiento de los Datos

El conjunto de datos de entrenamiento se utiliza para el aprendizaje de los modelos, de manera que permitan adaptarse de mejor manera a los parámetros incorporados en la base, para el clasificador de vigencia/no vigencia del cliente (o póliza en este caso). Se divide el conjunto de datos en 2 subconjuntos, donde el primero será para el entrenamiento del modelo. Luego, el segundo set será para evaluar el modelo entrenado. En este caso, se utiliza un 20% de los datos para validar y un 80% para el entrenamiento.



### 6.1.6 Parámetros Utilizados en la Ejecución de los Modelos de Predicción

La siguiente tabla muestra los principales parámetros y configuraciones a considerar durante la ejecución de los modelos de predicción, que permitirán entender más claramente los resultados que se obtengan de cada uno de ellos:

**Tabla 5:** Parámetros Utilizados en la ejecución de los Modelos Predictivos

Fuente: Elaboración Propia

Total de Registros	40,277
<b>Cross Validation</b>	
% Datos Entrenamiento	80%
# Datos Entrenamiento	32,222
% Datos Validación	20%
# Datos Validación	8,055
K-Folds	5

Parámetros Utilizado en la ejecución modelo Regresión Logística		Parámetros Utilizado en la ejecución modelo Árboles de Decisión		Parámetros Utilizado en la ejecución modelo Random Forest	
Nro. De iteraciones	1000	Criterio	Gini Index	Criterio	Gini Index
Tipo	Lineal	Profundidad Máx	17	Profundidad Máx	16

Para todos los modelos de predicción, se realizan iteraciones modificando los parámetros que se indican, llegando a la solución más “óptima” a los valores mostrados en la tabla anterior.

### 6.1.7 Balanceo de Datos

La base de clientes muestra a simple vista que la cantidad de asegurados que se fugan es menor a la cantidad total de ellos, mostrando un claro desbalance entre clases (Clientes Vigentes/Clientes No Vigentes).

**Tabla 6:** Cantidad de Clientes Vigentes/No Vigentes en la base

Fuente: Elaboración Propia

	#	%
<i>Clientes Vigentes</i>	29,747	68.89
<i>Clientes No Vigentes</i>	10,530	31.11
<b>Total</b>	<b>40,277</b>	<b>100</b>

Este desbalance puede tener un impacto significativo en la precisión de algunos modelos predictivos, por lo que se deben evaluar bajo un data set balanceado. Esto se obtiene por medio de la técnica de *oversampling*, que consiste en tomar aleatoriamente observaciones de la clase más baja (No Vigente) y repetirla hasta llegar a un balanceo del 50%.

Descrita la metodología del balanceo de datos SMOTE en la sección 2.4.5.9 de este documento, se procede a utilizar esta técnica de *oversampling* en cada uno de los modelos de predicción seleccionados:

```
print("# Registros usados para entrenamiento: ", len(X_validationRLB))

print("Before OverSampling, counts of label 'No Vigente': {}".format(sum(Y_validationRLB == 1)))
print("Before OverSampling, counts of label 'Vigente': {} \n".format(sum(Y_validationRLB == 0)))

# import SMOTE module from imblearn library
# pip install imblearn (if you don't have imblearn in your system)
from imblearn.over_sampling import SMOTE
os = SMOTE()
X_validation_res, y_validation_res = os.fit_sample(X_validationRLB, Y_validationRLB.ravel())

print("\n After OverSampling, counts of label 'No Vigente': {}".format(sum(y_validation_res == 1)))
print("After OverSampling, counts of label 'Vigente': {}".format(sum(y_validation_res == 0)))

# Registros usados para entrenamiento: 32222
Before OverSampling, counts of label 'No Vigente': 8424
Before OverSampling, counts of label 'Vigente': 23798

After OverSampling, counts of label 'No Vigente': 23798
After OverSampling, counts of label 'Vigente': 23798
```

**Ilustración 18:** Resultados de balanceo de datos aplicando *oversampling*, en el set de entrenamiento.

Fuente: Elaboración Propia

## 6.2 Resultados de Modelos de Predicción

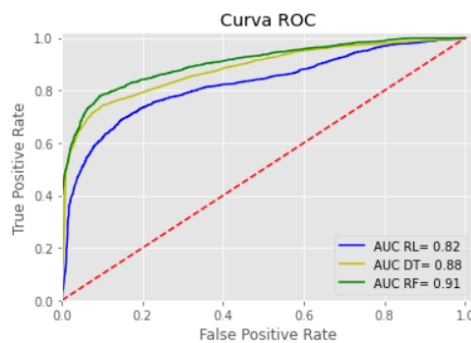
Utilizando el software *Phyton* (versión 3.8.5), se ejecutan los 3 modelos de aprendizaje supervisado de predicción: regresión logística, árboles de decisión y *random forest*. Luego, usando las métricas para medir el desempeño de cada uno de ellos dentro del mismo software y, utilizando el set de datos seleccionado, se obtienen los resultados siguientes:

**Tabla 7:** Resultados y métricas de desempeño de Modelos de Predicción

Fuente: Elaboración Propia

	Regresión Logística	Decision Tree	Random Forest
<i>accuracy</i>	77.54%	87.0%	88.17%
<i>classification_error</i>	22.46%	13.0%	11.83%
<i>AUC</i>	82%	88%	91%
<i>Precisión (Fuga)</i>	74%	78%	80%
<i>Recall (Fuga)</i>	55%	70%	73%
<i>f_measure (f-1 Fuga)</i>	63%	74%	76%
<i>Cross Validation Accuracy</i>	83,56%	85,47%	86,22%

- *Accuracy*: Corresponde al porcentaje de clientes que los modelos indican como fuga y no fuga con respecto al resultado total real. De los modelos evaluados, el de mejor desempeño corresponde al modelo de clasificación *Random Forest*, clasificando un 88.17% de los casos correctamente, aunque el modelo de árbol de decisión arroja también un valor bastante cercano con 87.0%.
- Error de Clasificación: Indica el desempeño del modelo al clasificar correctamente el estado buscado (Vigente / No Vigente o Fuga) luego de realizar el entrenamiento previo. Nuevamente, el mejor desempeño está dado por el modelo de *Random Forest* al presentar el menor valor de error con un 11.83%.
- AUC o Curva ROC: Indica el rendimiento del modelo de clasificación a través de la distribución de las fracciones de verdaderos positivos [TP] y de falsos positivos [FP], el cual se determina con el área bajo la curva:
  - El mejor desempeño, medido a través del área bajo la curva, figura para el modelo de *Random Forest*, seguido muy de cerca por el modelo de *árbol de decisión* y, por último, el modelo de regresión logística.



**Ilustración 19:** Curva ROC de resultados a los modelos predictivos

Fuente: Elaboración Propia

- *Precisión*: Este indicador se refiere al porcentaje de asegurados que el modelo marca como fuga y que efectivamente lo hacen. El modelo que mejor resultado muestra es *Random Forest*.
- *Recall*: Se refiere al porcentaje de asegurados que el modelo marca como fugados con respecto del total real de aquellos en el mismo estado. El modelo con mejor desempeño en esta métrica es *Random Forest*.

- *F-Measure*: Permite comparar el rendimiento combinado de precisión y *recall* entre varias soluciones. Siendo nuevamente el modelo de *Random Forest* con mejor desempeño.
- *Cross Validation Accuracy*: Permite ver la asertividad del modelo luego de utilizar la técnica de *cross validation* en el entrenamiento y validación.

### 6.2.1 Selección del Modelo de Predicción

La selección del modelo de predicción a utilizar está dada por los resultados obtenidos de las métricas de rendimiento y desempeño de cada uno de ellos anteriormente descritos. Se presentan tres modelos de clasificación predictivos, donde el modelo *Random Forest* (RF), muestra resultados aceptables de desempeño.

Es importante realizar un análisis más detallado de las consecuencias sobre las variaciones en los distintos indicadores, según los resultados obtenidos del modelo GBT. Por ejemplo, para RF se tiene un 73% de *recall*, vale decir que, de 100 clientes no vigentes, el modelo acierta en 73 de ellos. Sumado a lo anterior, resulta una precisión del 80%, es decir, para llegar a los 73 clientes no vigentes, el modelo predice que 91 abandonarán la compañía (1,2 veces la cantidad de fugados).

Con lo anterior es posible establecer una cierta relación entre el indicador de precisión y el *recall*, donde mientras más predicciones de fuga de clientes se realicen, aumentan las probabilidades de ser más asertivos en los resultados.

**Tabla 8:** Matriz de Confusión - *Random Forest*

Fuente: Elaboración Propia.

<b>Matriz de Confusión Random Forest</b>	<b>Vigente</b>	<b>No Vigente</b>	<b>Class Precision</b>
<b>Vigente</b>	5.567	382	<b>91%</b>
<b>No Vigente</b>	571	1.535	<b>80%</b>
<b>Class recall</b>	<b>94%</b>	<b>73%</b>	

Adicionalmente, es posible identificar aquellos factores preponderantes en la predicción de abandono dado por el modelo RF, dándonos luces de aquellos indicadores a tomar en cuenta, donde los meses de vigencia actuales en la compañía, pólizas con periodo de gracias o la edad figuran como los de mayor “peso” en la decisión:

**Tabla 9:** 10 Predictores Preponderantes en el Modelo - *Random Forest*

Fuente: Elaboración Propia.

predictor	importancia
MESESVIGENTE	0.197292
PERIODO_GRACIA	0.125401
EDAD	0.118985
COD_SPONSOR	0.071319
BBVA	0.069156
COD_SEXO	0.053630
CAPITAL	0.053244
COD_COBE	0.049741
COD_PLAN	0.040670
Prima Dev <0.1	0.021067

Aun así, la variación de cada una de las métricas mostradas como resultado de la aplicación de los modelos de predicción deben ser comprendidas por los expertos de negocio, con el fin de seleccionar con criterio el modelo más adecuado. El modelo de predicción de fuga de clientes, en el contexto de este trabajo, será utilizado para la generación de acciones de retención en aquellos clientes con mayor probabilidad de fuga, para evitar así el quiebre de acuerdos comerciales y pérdida en la sustentabilidad del negocio.

Para ello es necesario considerar el costo de la acción de retención, el que será sustentado por las actuales responsabilidades del analista comercial encargado. Si este costo es alto, es entonces es más valioso considerar una alta precisión por sobre un buen recall, donde ambas métricas tienen buen desempeño para el modelo RF ya demostrado.

### 6.3 Medidas a Considerar en el Futuro

Como es sabido, la mejor estrategia en lugar de traer nuevos clientes es retener a los que ya se tienen, ya que el costo asociado de traer uno nuevo versus retener a los existentes se cuantifica entre un 5% y 25% más alto (Gallo A., 2014). Es lógico pensar que conlleva un menor costo y esfuerzo retener a un cliente que traer a uno nuevo, donde un servicio y atención diferenciador son dos de las razones principales por las que el cliente se mantiene y no tiene intenciones de irse a la competencia.

Teniendo los resultados del modelo de predicción, es posible visualizar una idea generalizada de aquellos clientes con mayor probabilidad de abandonar los productos con la compañía, pudiendo aplicar medidas proactivas de retención que ayudarían a aumentar la rentabilidad de la línea de negocios y mantener los acuerdos de confianza con sus principales *sponsors*.

Por otro lado, la retención de clientes es una de las actividades que trae consigo mayores beneficios, incluso por sobre la adquisición de nuevos clientes, nuevos canales de venta, nuevas estrategias de promoción de productos, entre otros. Esto obliga a que se ponga especial atención en los motivos que producen el abandono de los clientes de la compañía (véase Anexo 1), para la aplicación de estrategias efectivas de retención que mantengan y beneficien la sustentabilidad del negocio.

Si bien los resultados del modelo nos dan este acercamiento, es necesario realizar continuamente ajustes al mismo para poder obtener mejores resultados en el futuro. Estos ajustes pueden tener relación con la utilización de una mayor cantidad de historia de datos, aumentar el número de árboles a considerar incluyendo distintos niveles de profundidad, consideración de nuevas variables y por supuesto, considerar otros modelos de predicción que puedan mostrar patrones que no es posible ver con los modelos aplicados (considerar, por ejemplo, un modelo de datos no supervisado para clusterización de clientes y comparar con los resultados obtenidos).

#### **6.4 El Valor de los Clientes**

Los recursos destinados para efectuar la retención de clientes son escasos, por lo tanto, es de vital importancia priorizar a aquellos que tienen una alta probabilidad de fuga y que sean más rentables para la compañía. Las estrategias de retención de clientes funcionarán en la medida que éstas estén enfocadas en satisfacer sus necesidades y reflejar la importancia del valor que ellos tienen para la empresa.

Antes de recibir formalmente la solicitud de abandono de productos, la compañía debe ser proactiva en mantener satisfecho a cada uno de ellos bajo distintas estrategias de comunicación, compromiso y flexibilidad. El conocimiento de aquellos clientes con mayor intención de abandono permitirá actuar de manera proactiva, establecer mejores lazos y, sobre todo, crear la confianza, permanencia y satisfacción que se espera con este tipo de productos a lo largo del tiempo.

Pero la selección de contactabilidad de los clientes no es aleatoria. Se deben seleccionar aquellos clientes de mayor valor para la compañía sin impactar los recursos asignados para este propósito. Para ello se hace necesario la incorporación de un valor de calificación de cliente, según el resultado que éstos entregan a la compañía. En las ilustraciones adjuntas en el Anexo N°7, es posible ver la concentración de clientes por monto de prima devengada que actualmente pagan con sus productos, pudiendo establecer que:

- Los clientes que pagan más prima corresponden a los del *sponsor* Banco Security (el 40% de ellos paga primas entre UF 0.75 y UF 1.00).
- La mayor cantidad de clientes que pagan menos prima (< UF 0.1) se concentran en *sponsor* Scotiabank – BBVA.

- Tanto ITAU como Banco de Chile, concentran una cantidad similar de clientes que mantiene pagos de prima entre UF 0.25 y UF 0.5 (sobre 60% para el Banco de Chile y sobre 35% para ITAU).

Con esta información, considerando el volumen de clientes de cada *sponsor*, costo medio de la administración de cartera y el riesgo (considerados dentro de los valores de prima), se calcula el valor de clasificación del cliente considerando además factores de siniestralidad promedio de cada *sponsor* y el valor esperado para los próximos meses en base a los generados anteriormente (temporalidad de 6 meses).

**Tabla 9:** Fragmento de Resultados e Interpretación Modelo de Predicción de Fuga, *Bancassurance*.

Fuente: Elaboración Propia

Demanda Estimada para los próximos 12 Meses												
Meses	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9	Mes 10	Mes 11	Mes 12
BANCO DE CHILE	746	681	855	714	903	834	852	928	759	910	1,070	937
BBVA	7,935	7,247	9,093	7,593	9,609	8,879	9,066	9,870	8,072	9,679	11,388	9,973
ITAU	3,495	3,192	4,005	3,344	4,232	3,911	3,993	4,347	3,555	4,263	5,016	4,392
SCOTIABANK	3,915	3,575	4,486	3,746	4,741	4,381	4,473	4,869	3,983	4,775	5,619	4,920
SECURITY	389	355	445	372	471	435	444	483	395	474	558	489
<b>TOTAL</b>	<b>16,480</b>	<b>15,050</b>	<b>18,884</b>	<b>15,768</b>	<b>19,956</b>	<b>18,440</b>	<b>18,827</b>	<b>20,497</b>	<b>16,764</b>	<b>20,100</b>	<b>23,651</b>	<b>20,711</b>

	Precio Promedio Prima x Sponsor
BANCO DE CHILE	0.298
BBVA	0.162
ITAU	0.411
SCOTIABANK	0.299
SECURITY	0.575

Ingresos Estimados Proximos 12 Meses UF												
Meses	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9	Mes 10	Mes 11	Mes 12
BANCO DE CHILE	222.40	203.10	254.84	212.79	269.31	248.85	254.07	276.61	226.23	271.25	319.17	279.50
BBVA	2,366.35	2,161.02	2,711.54	2,264.12	2,865.47	2,647.79	2,703.36	2,943.15	2,407.13	2,886.15	3,396.04	2,973.88
ITAU	1,042.21	951.78	1,194.24	997.18	1,262.04	1,166.16	1,190.64	1,296.25	1,060.17	1,271.14	1,495.71	1,309.78
SCOTIABANK	1,167.48	1,066.18	1,337.79	1,117.04	1,413.73	1,306.33	1,333.75	1,452.05	1,187.60	1,423.93	1,675.49	1,467.21
SECURITY	115.92	105.86	132.83	110.91	140.37	129.71	132.43	144.18	117.92	141.39	166.37	145.68

Con estos valores parametrizables de ingresos esperados, la probabilidad de fuga ya calculada y la siniestralidad por *sponsor*, se determinará si el cliente a contactar es significativo en base al valor esperado.

1. Se calculará el valor de clasificación de los clientes de toda la cartera vigente.
2. Se seleccionarán aquellos clientes cuyo valor esperado sea superior a un monto definido (parametrizable) por el negocio.
3. Se seleccionarán aquellos clientes que cuenten con una vigencia inferior a los 12 meses (parametrizable).
4. Se debe desplegar en el sistema los clientes que cumplan con dichos requisitos.
5. El analista deberá encargarse de seleccionar los sponsors a contactar en base a su conocimiento y experiencia.

Para la configuración de estos factores, se utilizan las siguientes ponderaciones de siniestralidad según tipo de *sponsor*:

**Tabla 10:** Factores de Siniestralidad Promedio para Cálculo de Valor del Cliente.

Fuente: Elaboración Propia

Sponsor	Factor de Ponderación
ITAU	11,3%
Scotiabank	14.0%
Banco Security	8.5%
BBVA	14.%
Banco de Chile	10,2%

Con esto se determinará un valor de pérdida aproximado que permitirá clasificar al tipo de clientes a contactar, según:

$$\text{Ingreso Esperado Mes Vigente} * \text{Siniestralidad} * \text{Prob Fuga}$$

Por ejemplo, aquellos clientes mayores de 35 años que tienen un Capital Asegurado de UF 1.000 para una cobertura de Muerte Accidental con el banco ITAU (Prob. Fuga = 12.5 %), pagan una prima aprox. de UF 0.26 y están en su mes 6 de vigencia, el cliente se valoriza como:

Valor de Cliente ITAU:  $(\text{Ingreso Esperado } (1.166,16) * \text{Siniestralidad } (11,3\%) * \text{Prob. Fuga } (12.5\%)) = 16.472$

Si la misma configuración corresponde al cliente del Sponsor Scotiabank (Prob. Fuga = 6.4%) se tiene:

Valor Cliente Scotiabank:  $(\text{Ingreso Esperado } (1.306,33) * \text{Siniestralidad } (14\%) * \text{Prob. Fuga } (6.4\%)) = 12.667$

Si bien el cliente del *sponsor* Scotiabank tiene un mayor valor esperado y una baja probabilidad de fuga, la alta siniestralidad de la cartera hace disminuir el valor del cliente en comparación al cliente del *sponsor* ITAU, siendo este último el de mayor valor a contactar.

Por supuesto, estos resultados dependerán del tipo de producto, cobertura asegurada, *sponsor*, prima y probabilidad de fuga dada por el modelo de predicción, desde el punto de vista del producto. El área comercial deberá decidir cuándo y sobre qué cartera de cliente aplicar estrategias de retención, según los lineamientos estratégicos del negocio.

## 6.5 Información Post-Atención de Clientes

Luego de realizar el contacto con los clientes, la información generada como resultado de la aplicación de la estrategia de retención deberá ser almacenada, permitiendo posteriormente la evaluación y ajuste de las variables dentro del modelo predictivo.



Para lograrlo, se presenta la realización de una encuesta de servicio post-atención, la cual contiene:

- Información del Cliente: RUT, Póliza, VIP.
- Información de la Solicitud: Número de solicitud, Fecha de Solicitud.
- Canal de Atención (Presencial, Web, Móvil, Otro) y tiempo de Atención (si aplica).
- Razón de abandono Indicada:
  - Error Suscripción de Producto
  - Mala Gestión de Venta
  - Disconformidad con el Producto
  - Retracción de contratación del Producto
  - Problemas Económicos
  - Otros Motivos
- Propuesta de Retención:
  - Meses de Gracia
  - Rebaja en pago de Prima
  - Incorporación de Nueva Cobertura
  - Otro
- Aceptación de propuesta de Retención.

**Ilustración 20:** Encuesta Post-Atención de Clientes

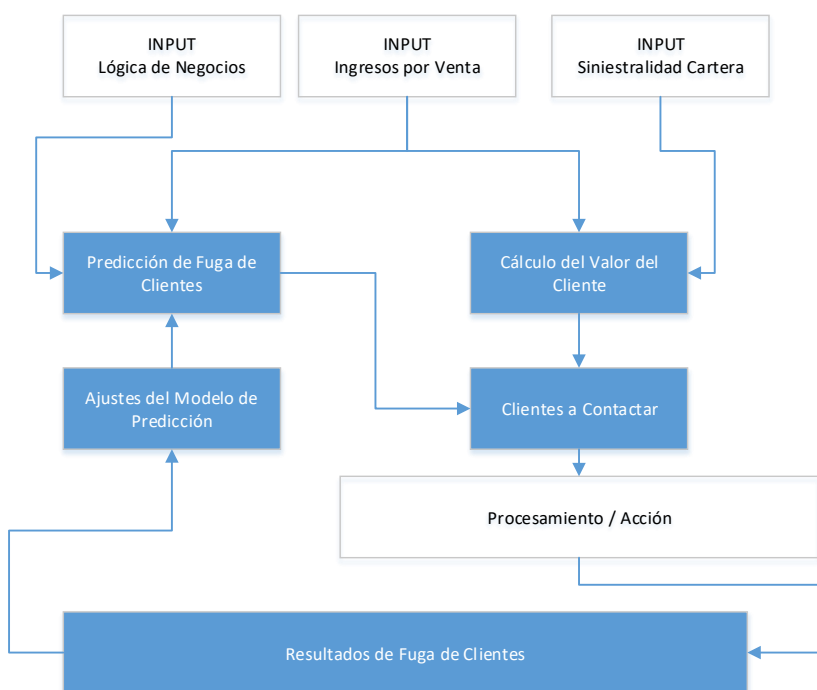
Fuente: Elaboración Propia

Está deberá formar parte de las actividades del analista comercial a cargo de la retención, quien contará con el despliegue de ella dentro del sistema habilitado para dicho efecto.

## CAPÍTULO 7: APOYO TECNOLÓGICO

La mejor forma de describir el proyecto desde el punto de vista tecnológico consiste en la utilización de un lenguaje modelado UML. Esto facilita a quienes son los responsables de desarrollo sobre las plataformas y sistemas, la construcción de los distintos módulos e interacciones entre ambas. Este lenguaje ofrece un estándar para describir gráficamente un sistema, con sus procesos de negocios, expresiones, funciones, esquemas de bases de datos, arquitectura y otros.

El esquema general de solución se muestra en la ilustración siguiente. La idea es que este esquema soporte tanto al proceso rediseñado como a la incorporación de nuevos módulos (destacados en color) en el sistema de administración de pólizas para seguros colectivos SACS, para la ejecución del modelo predictivo y sus ajustes correspondientes.



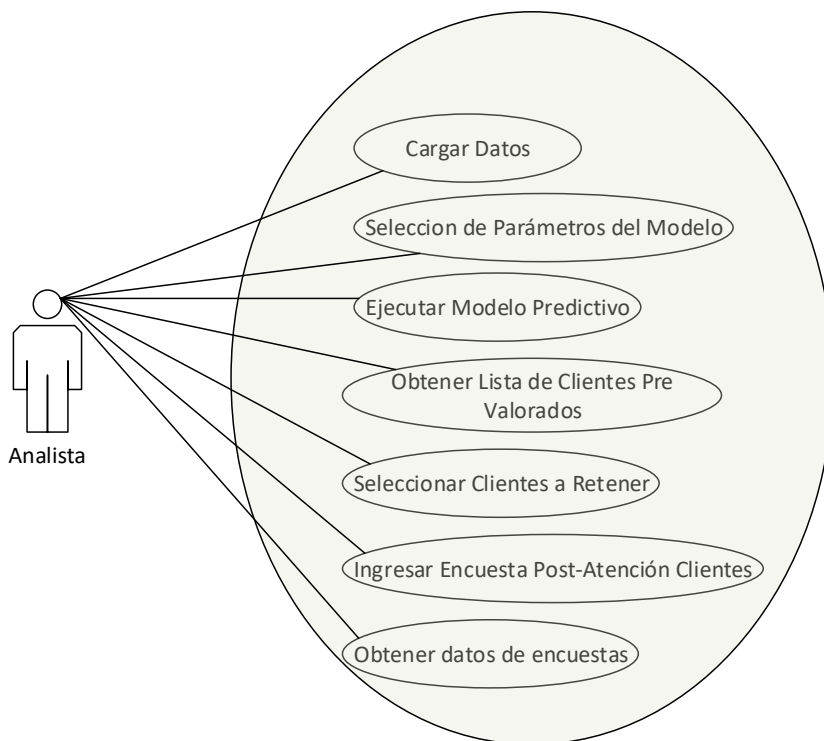
**Ilustración 21:** Esquema General de Solución

Fuente: Elaboración Propia

En este trabajo, se presentan algunos de los diagramas UML más utilizados como: Casos de Uso y Diagrama de Secuencias, para luego dar paso al prototipo propuesto para el desarrollo futuro de los distintos módulos que acá se presentan.

## 7.1 Casos de Uso

Los casos de uso ilustrados a continuación representan una descripción gráfica de las distintas actividades que deben desarrollarse, por medio de un actor, para llevar a cabo una acción o proceso.



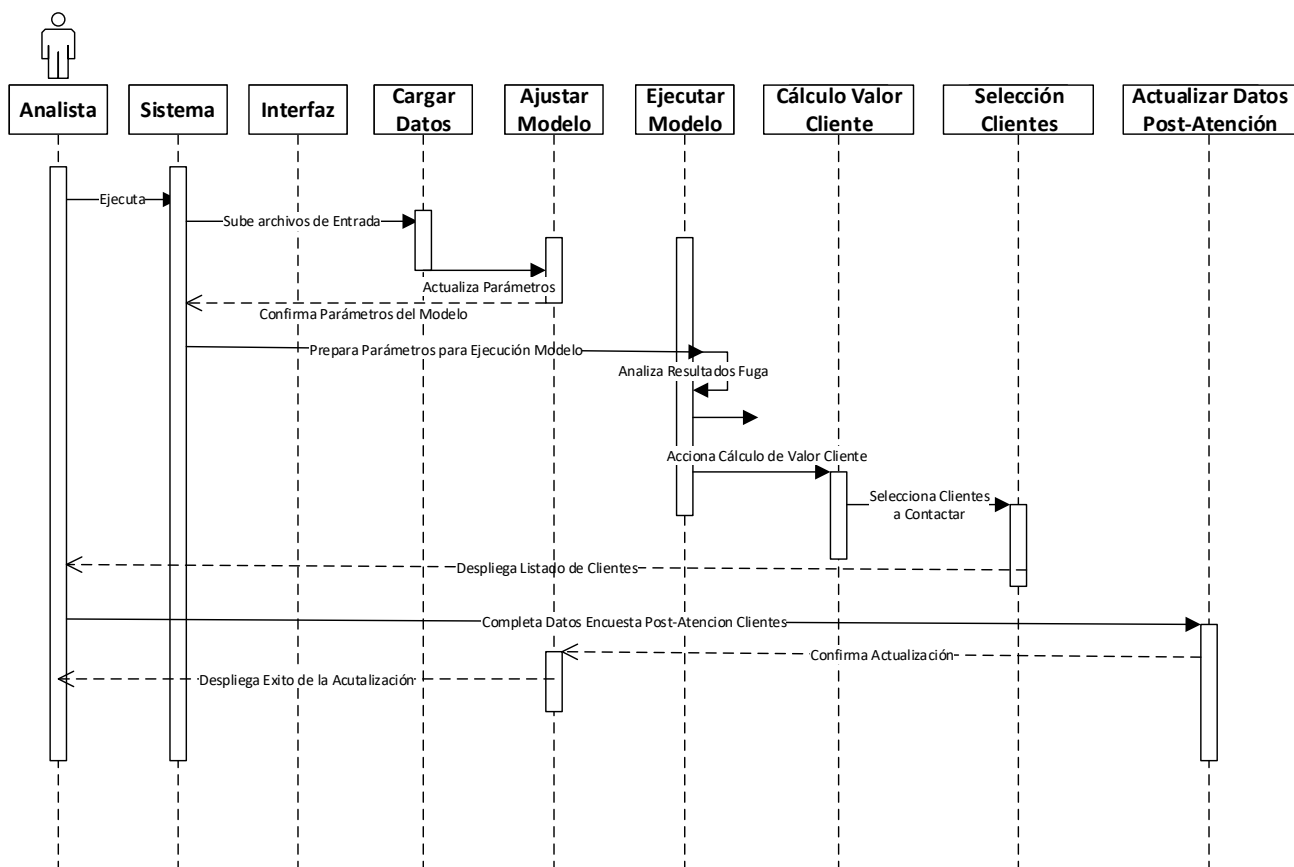
**Ilustración 22:** Diagrama de Casos de Uso

Fuente: Elaboración Propia

## 7.2 Diagrama de Secuencia

El diagrama de secuencia es un diagrama de interacción de un conjunto de objetos de una aplicación a través del tiempo. Muestra cómo operan los procesos entre sí, los módulos o clases que forman parte del programa y las llamadas que se realizan para realizar una tarea determinada.

En primer lugar, se visualiza el diagrama correspondiente a las acciones e interacciones realizadas por el analista, el sistema de administración de pólizas y los distintos módulos nuevos para la ejecución del modelo predictivo, resultados y lista de valoración de clientes



**Ilustración 23:** Diagrama de Secuencias

Fuente: Elaboración Propia

### 7.3 Diseño de Prototipos

Como se ha presentado en las secciones anteriores, a continuación, se ilustran los componentes de los nuevos módulos propuestos a desarrollar dentro del sistema de administración de pólizas SACS. Estos corresponden a:

- Módulo de Parámetros del Modelo Predictivo.
- Módulo de Ejecución y Resultados Modelo Predictivo
- Módulo de Cálculo de Valor de los Clientes.
- Módulo de Ingreso de Datos encuesta Post-Atención

#### 7.3.1 *Módulo de Parámetros del Modelo Predictivo*

Tal como su nombre lo indica, en esta primera pantalla se muestran los parámetros generales previos a la ejecución del modelo predictivo a seleccionar. Para ello se dispone de listas desplegables con las opciones de Modelos de Predicción (que en un futuro puedan incorporarse

otros distintos a RF), selección de base de clientes (Banca, Clínicas, Retails) también pensando en la incorporación de otros tipos de *sponsor* y, por último, el nombre del sponsor a visualizar.

Este módulo permitirá además realizar carga de bases de clientes desde archivo, o consultando directamente la base de datos de la compañía a través de consultas SQL.

Por último, despliega los resultados de las métricas actuales del último modelo ejecutado.

SACS - Módulo de Predicción de Fuga y Valorización de Clientes

Parámetros Modelo de Predicción | Ejecución y Resultados Modelo Predicción | Valorización de Clientes | Ingreso Datos Encuesta Post-Atención Clientes

Periodo: 30/09/2020

Modelo de Predicción: Menu Item

Base de Clientes: Menu Item

Sponsor: Menu Item

Cargar Base desde Archivo | Cargar Base desde Oracle

Clientes No Vigentes: Placeholder

Clientes Vigentes: Placeholder

Métricas Actuales

Precisión: Placeholder

Accuracy: Placeholder

AOC: Placeholder

Descargar Resultados del Modelo

Última Actualización 01/09/2020

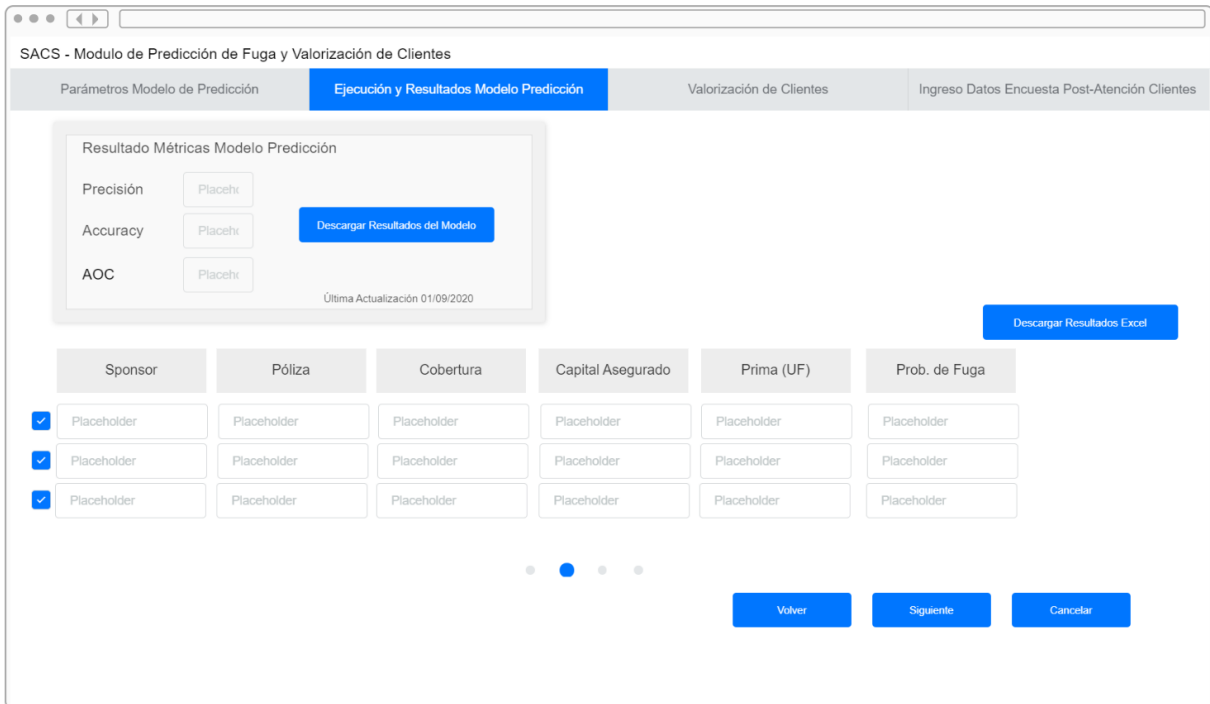
Ejecutar Modelo | Siguiente | Cancelar

**Ilustración 24:** Prototipo Modulo de Parámetros de Modelo de Predicción

Fuente: Elaboración Propia

### 7.3.2 *Módulo de Ejecución y Resultados del modelo predictivo.*

El siguiente módulo despliega los resultados obtenidos luego de ejecutar en la pantalla anterior el modelo de predicción seleccionado. Con ello, será posible descargar el resultado de las métricas obtenidas, así como también la lista de clientes con su probabilidad de fuga calculada.



**Ilustración 25:** Prototipo Módulo Ejecución y Resultados Modelo Predicción

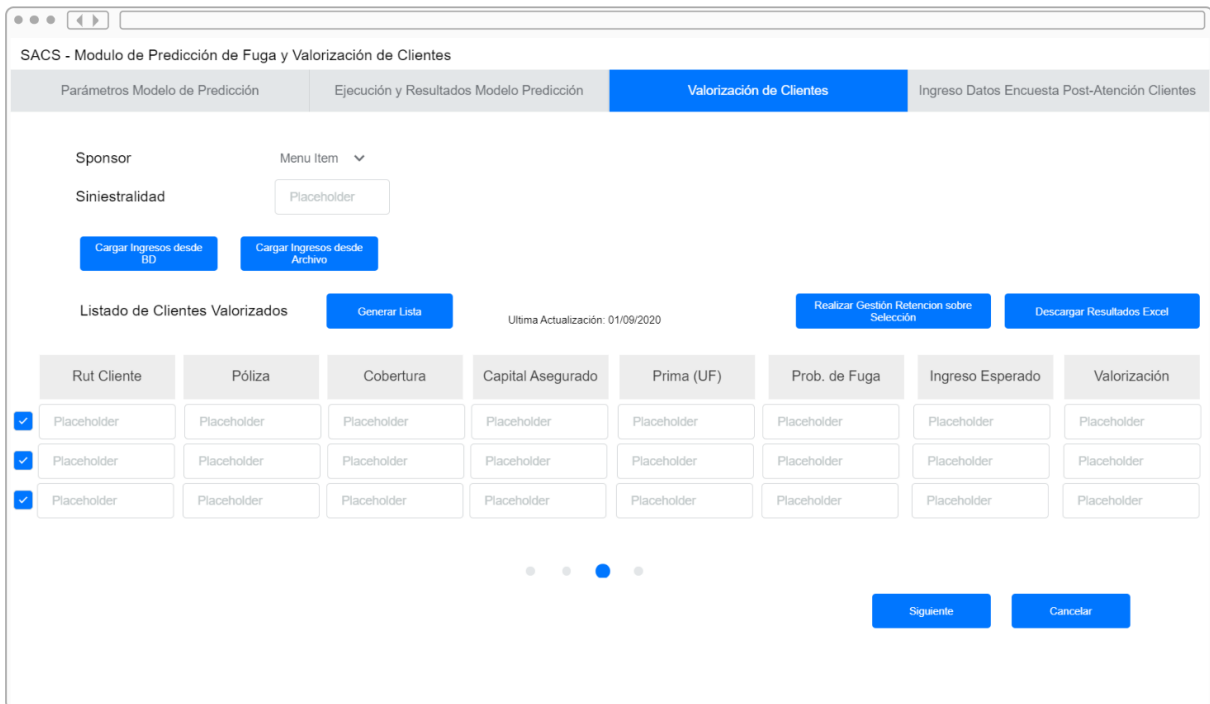
Fuente: Elaboración Propia

### 7.3.3 Módulo de Valorización de Clientes

Teniendo los resultados desde el modelo de predicción ejecutado, se realiza el cálculo de valorización del cliente según la fórmula descrita en la sección 6.5 de este informe. Para ello es necesario seleccionar el *sponsor* a considerar e ingresar el porcentaje de siniestralidad como parámetro obligatorio.

Para que este módulo funcione correctamente, es necesario realizar la carga de los ingresos esperados con los botones de acción desplegados en pantalla.

Con todo esto, se procede al cálculo de la valorización del cliente ejecutando la acción “Generar Lista”, cuyo resultado será una lista desplegable (descargable en Excel) de los distintos resultados de valor obtenidos por cliente. Existe la posibilidad de seleccionar uno o varios desde el listado para continuar con la contactabilidad del cliente y completar la encuesta post-atención del módulo siguiente.



**Ilustración 26:** Prototipo Módulo Valorización de Clientes

Fuente: Elaboración Propia

### 7.3.4 *Módulo de Ingreso de Datos Post-Atención Clientes*

Como se describe en 6.6, el sistema cuenta con un módulo de gestión que permite el ingreso de los datos de los clientes que han sido contactados para evitar su fuga. Con esta información será posible realizar ajustes y posteriores calibraciones al modelo de predicción correspondiente.

**SACS - Modulo de Predicción de Fuga y Valorización de Clientes**

Parámetros Modelo de Predicción | Ejecución y Resultados Modelo Predicción | Valorización de Clientes | **Ingreso Datos Encuesta Post-Atención Clientes**

Rut Cliente: Placeholder | VIP:  SI  NO

Póliza: Placeholder | N° Solicitud: Placeholder

Fecha Gestion: Placeholder | Tiempo Atención: Placeholder

Canal de Atención:  Presencial  Web  Movil  Otro

**Alternativa Retencion**

Meses de Gracia

Rebaja Prima Monto: Placeholder

Nueva Cobertura Cual: Placeholder

Elimina Cobertura Cual: Placeholder

Modifica Cap. Asegurado Monto: Placeholder

**Acepta Propuesta**

SI  NO

**Razón Abandono**

Mala Gestión de Venta

Error de Suscripción de Producto

Disconformidad con el Producto

Retracción contratación

Problemas Económicos

Otros

Guardar Cancelar

**Ilustración 27:** Prototipo Modulo de Ingreso de Datos Encuesta Post-Atención de Clientes

Fuente: Elaboración Propia



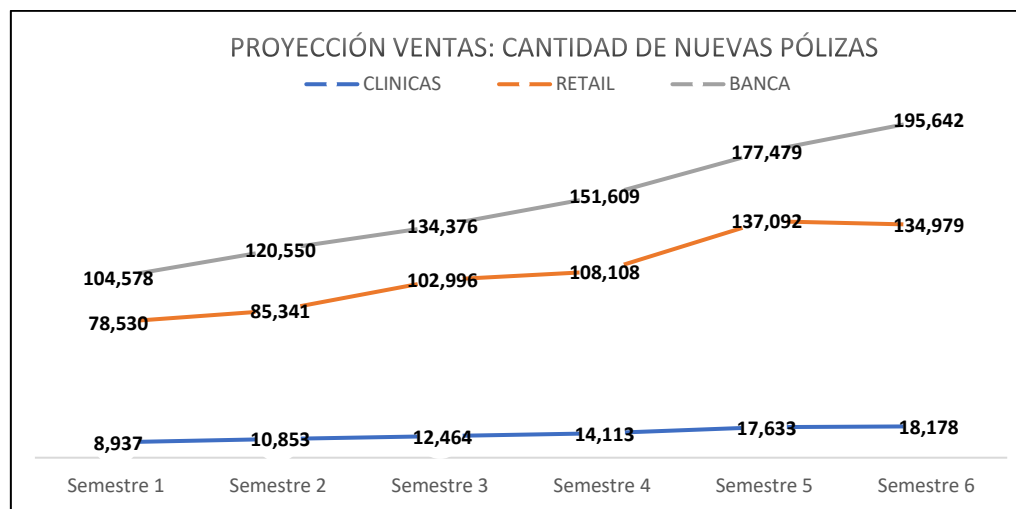
## CAPÍTULO 8: EVALUACIÓN ECONÓMICA DEL PROYECTO

Para poder evaluar la factibilidad económica de realizar el trabajo propuesto, es necesario incorporar dentro del análisis de solución la evaluación económica del mismo. Para ello se considera la situación actual de la línea de negocios versus una situación de mejora, que incluye una digitalización de actuales procesos en la Administración y Ventas, para finalizar con la ejecución del proyecto y su evaluación frente a análisis de sensibilidad.

### 8.1 Supuestos Utilizados

Durante la evaluación económica es necesario considerar algunos supuestos para la completitud y validez respectiva de los resultados, los cuales nacen a partir de la experiencia y consulta de los diversos responsables dentro del control y gestión de la línea:

- Se construye flujo de caja en un horizonte de 3 años evaluado en forma semestral ya que, al tratarse de un proyecto tecnológico, tanto su implementación como su desarrollo se puede realizar en un tiempo reducido, permitiendo así visualizar rápidamente los resultados.
- La moneda original de evaluación es la Unidad de Fomento (UF) que para efectos prácticos se transforma a miles de pesos al último día del mes de junio 2019 (UF valor en pesos: \$27.903).
- Se considera una proyección de nuevas ventas de pólizas (ingresos) durante los próximos 36 meses por tipo de clúster para la construcción de los ingresos por ventas, siguiendo los lineamientos comerciales y manteniendo el stock vigente de pólizas al inicio de la evaluación:



**Gráfico 4:** Proyección Semestral de Pólizas por Clúster para Mercado Masivo en MetLife.

Fuente: Elaboración Propia

- Se calcula un promedio de precios de los distintos productos (pólizas) para cada uno de los clúster y sponsors que forman parte de la línea considerando sólo *bancassurance*, los cuales serán parte de la construcción de los ingresos por venta:

**Tabla 11:** Prima Promedio Mensual en pesos por Sponsor y Clúster Banca.

Fuente: Elaboración Propia

Productos	Precio Prima Mensual (UF) Promedio	Precio \$
Banco de Chile	0.291	\$ 8,110
BBVA	0.140	\$ 3,919
Scotiabank	0.483	\$ 13,472
Security	0.617	\$ 17,212
Otros <i>Runoff</i>	0.223	\$ 6,220

- Se considera una estimación porcentual promedio de los Gastos de Administración y Ventas y de Siniestralidad por clúster que se aplica sobre el ingreso por primas de seguro:

**Tabla 12:** Promedio Porcentual de Siniestralidad y Costos de Adm. y Ventas por Clúster Banca.

Fuente: Elaboración Propia

Siniestralidad	14%
Costos de Adm y Ventas	19.33%

- Los costos variables consideran además del punto anterior, un 12% de provisión que corresponden a las Reservas Técnicas de Seguros + Incobrables.
- El cálculo de depreciación para los sistemas de Software (SACS) no es incorporado dentro del flujo de caja, pero tienen un costo de mantención que debe considerarse dentro de los costos fijos de la evaluación, así como también un costo promedio de operación de la línea estimado luego del resultado anual obtenido en el año 2018:

**Tabla 13:** Costo de mantención de Software más Costos operacionales Promedio en M\$.

Fuente: Elaboración Propia

Costo Mantención Software Línea DM	SACS
Valor UF x Hora	1.30
Cantidad de Horas Mes	40
Total UF Mes	52.0

Total \$ Mes	1,450,972
Total \$ Semestral	8,705,830
Costos Op. Promedio M\$	- 61,263

- La Tasa de Interés Compañía corresponde a un 27%, la cual es revisada constantemente por el regulador local.
- Valor Mercado Activo Fijo, Capital de trabajo, Préstamo y amortización no aplican para el desarrollo de la evaluación económica.
- Tratándose además de un proyecto con financiamiento directo desde la línea de negocios (proyecto puro), permitirá realizar el cálculo de la tasa de descuento utilizando la metodología de CAPM (Método Capital Asset Pricing Model) por simplicidad y conocimiento de las variables que lo componen:

**Método Capital Asset Pricing Model CAPM**

CAPM = Tasa Libre de Riesgo + (Rendimiento del Mercado - Tasa Libre Riesgo) X Beta + Ajuste Riesgo Pais

**CAPM = 4,41% + (9,03% - 4,14%) \* 1,11 + 2,98%**

**CAPM = 12,55%**

Tasa Libre de Riesgo	Este dato es actualizado anualmente y se basa en la tasa libre de riesgo del bono del Banco Central de Chile (BC-10) a 8 años.	<b>4,14%</b>
Beta (2)	Dado que en Chile no existe un mercado lo suficientemente grande y activo para una adecuada determinación del BETA se usan betas de EE.UU.	<b>1,11</b>
Rendimiento del Mercado	Tasa Anual del Promedio simple de la variación del indicador DOW JONES a 5 años	<b>9,03%</b>
Ajuste Riesgo País a Ene 18 (1)		<b>298</b>

**Ilustración 28:** Cálculo de la Tasa de Interés utilizando método CAPM.

Fuente: Elaboración Propia

## 8.2 Evaluación Económica (Situación Actual – Mejorada – Implementación del Proyecto).

La construcción del flujo de caja muestra que la evaluación de la situación actual arroja resultados positivos con un VAN igual a M\$ 7.843.578 como resultado al final del horizonte de evaluación mientras que, la situación actual mejorada, que comprende una mejora de un 8% en los Costos de Adm. y Ventas debido a la mayor digitalización y mejora en los actuales procesos de venta, refleja como resultado un VAN igual a M\$ 9.785.235 (esto es M\$1.941.657 superior con sólo la aplicación de mejora).

La implementación de la propuesta tecnológica, bajo los desarrollos y soportes necesarios para incorporar analítica de clientes dentro del sistema de Administración de Pólizas SACS, implica un resultado en VAN de M\$12.473.421, con una Tasa Interna de Retorno (TIR) del 4.9% que

en comparación con la situación actual mejorada representa un incremento adicional de más de M\$2.688.186 en rentabilidad para el negocio, siendo una alternativa altamente factible y adecuada de realizar para obtener mejores y mayores beneficios a los ya actuales.

**Tabla 14:** Resumen resultados en M\$ de los escenarios económicos propuestos.

Fuente: Elaboración Propia

<b>7,843,578</b>	<b>9,785,235</b>	<b>12,473,421</b>	<b>2,688,186</b>
		<b>15.97%</b>	<b>4.93%</b>
Situación Actual	Mejora Situación Actual	Implementación Proyecto	Delta

### 8.3 Flujo de Caja

El detalle de las cifras en el flujo de caja para la implementación del proyecto se muestra a continuación:

**Tabla 15:** Evaluación Económica Implementación del Proyecto.

Fuente: Elaboración Propia.

Concepto / Horizonte Tiempo	\$	Año 1		Año 2		Año 3		
		Semestre 0	Semestre 1	Semestre 2	Semestre 3	Semestre 4	Semestre 5	Semestre 6
<b>Ingresos por Venta</b>			4,336,406	5,690,581	7,090,291	8,637,093	10,487,030	12,339,326
Banca			4,336,406	5,690,581	7,090,291	8,637,093	10,487,030	12,339,326
Banco de Chile			515,357	686,837	868,016	1,070,776	1,313,513	1,561,720
BBVA			427,442	562,633	702,797	857,832	1,043,015	1,228,853
Scotiabank			2,216,016	2,872,265	3,537,456	4,264,212	5,132,853	5,985,522
Security			1,145,670	1,525,633	1,926,585	2,374,928	2,911,552	3,459,580
Otros Runoff			31,921	43,213	55,437	69,346	86,097	103,652
<b>Costo Variable Total</b>		- 1,801,449	- 2,148,794	- 2,507,820	- 2,904,575	- 3,379,084	- 3,854,198	
Siniestralidad		- 591,919	- 776,764	- 967,825	- 1,178,963	- 1,431,480	- 1,684,318	
Siniestralidad Banca		- 591,919	- 776,764	- 967,825	- 1,178,963	- 1,431,480	- 1,684,318	
Administración y Ventas		- 1,209,529	- 1,372,030	- 1,539,995	- 1,725,612	- 1,947,604	- 2,169,880	
CV Adm y Venta Banca		- 689,160	- 689,160	- 689,160	- 689,160	- 689,160	- 689,160	
Reservas		- 520,369	- 682,870	- 850,835	- 1,036,451	- 1,258,444	- 1,480,719	
<b>Costo Fijo Total</b>	- 69,969	- 463,371	- 585,247	- 711,221	- 850,433	- 1,016,927	- 1,183,634	
<b>UAI</b>	- 69,969	2,071,587	2,956,540	3,871,251	4,882,085	6,091,019	7,301,495	
Impuesto 27%	- 18,892	- 559,328	- 798,266	- 1,045,238	- 1,318,163	- 1,644,575	- 1,971,404	
<b>UDI</b>	- 88,861	1,512,259	2,158,274	2,826,013	3,563,922	4,446,444	5,330,091	
<b>FCO</b>	- 88,861	1,512,259	2,158,274	2,826,013	3,563,922	4,446,444	5,330,091	
Inversión Activo Fijo	- 8,371	-	-	-	-	-	-	
<b>FCC</b>	- 8,371	-	-	-	-	-	-	
<b>Flujo Caja Privado</b>	- 97,232	1,512,259	2,158,274	2,826,013	3,563,922	4,446,444	5,330,091	

### 8.4 Análisis de Sensibilidad

Luego de obtener los resultados para cada uno de los escenarios propuestos, es necesario realizar un análisis de sensibilidad que permita reflejar situaciones de mercado, externalidades o

modificaciones a la actual operación de negocio durante el transcurso de la evaluación, que puedan cambiar los resultados y criterio de elección del proyecto respectivo.

Para ello, se utilizan dos parámetros fundamentales a evaluar bajo situación desfavorable y favorable: Nivel de Ventas y Siniestralidad.

- **Nivel de Ventas:** El negocio de seguros masivos se sustenta en la cantidad y masividad de clientes dentro de su cartera, por lo que las variaciones en los niveles de venta afectan directamente a los ingresos obtenidos por la línea de negocio.
- **Siniestralidad:** Aplicación efectiva de la cobertura de seguro contratada, que obliga a la compañía de seguros a realizar el desembolso económico pactado. Con esto presente, mientras mayor sea el número de personas siniestradas, mayor será el impacto económico desfavorable en los resultados.

Se consideran los siguientes escenarios para los parámetros antes descritos:

Parámetros Que Considerar	Peor Escenario (PE)	Mejor Escenario (ME)
Nivel de Ventas	- 15%	+ 15%
Siniestralidad	+ 10%	- 10%

**Ilustración 29:** Descripción de Escenarios para Análisis de Sensibilidad.

Fuente: Elaboración Propia

Dando como resultados las siguientes cifras en M(\$).

Resultados (M\$)					
VAN	7,701,568	7,745,705	7,843,578	10,614,614	11,188,395
	Siniestralidad PE	Ventas PE	Escenario Normal	Siniestralidad ME	Ventas ME
Delta Vs Esc. Normal	142,010	97,873	-	2,771,036	3,344,817

**Ilustración 30:** Evaluación Económica ante análisis de Sensibilidad.

Fuente: Elaboración Propia

## CAPÍTULO 9: CONCLUSIONES

---

El mercado de seguros ha ido transformando su orientación original desde los productos hacia sus clientes, donde *MetLife* busca una relación duradera, cercana y satisfactoria con ellos, plasmando en su estrategia de negocios lo importante que son sus clientes. La constante fuga de sus clientes en el mercado de seguros masivos, en específico para aquellos asociados a la banca, es detonante para la utilización de nuevos mecanismos que permitan aumentar su permanencia con la compañía y disminuir su intención de abandono.

La inexistencia de procesos de retención de clientes dentro de la línea de negocios obliga a incorporar nuevas actividades y rediseñar el actual proceso de “Análisis y Gestión de Clientes de Seguros Masivos”, utilizando la metodología de negocios. Esta metodología impulsada a través del estudio del Máster permite abordar e identificar de una manera estructurada y metódica, dentro de un escenario de macroprocesos, los procesos/actividades críticas que dificultan el logro de los objetivos estratégicos declarados por la compañía. Lo que sin duda se transforma en una herramienta transcendental en la generación de nuevos recursos y proyectos orientados directamente en el sentido estratégico que se manifiesta desde el origen.

Dentro de estos nuevos recursos, es posible incorporar inteligencia en la detección de aquellos clientes con mayor probabilidad de abandono. Esta inteligencia se traduce en la ejecución de tres modelos de predicción de datos, utilizando información histórica contenida en las bases de la compañía en un período de 5 meses, sobre un total de más 40.000 registros. Todos ellos tratados y depurados utilizando el *framework* CRISP-DM, altamente utilizado para la obtención, depuración, evaluación y control de los datos. Con esta herramienta, es posible obtener información relevante de la cartera de clientes como, por ejemplo, más del 60% de las pólizas corresponden al sexo masculino o que la cobertura de riesgo más común corresponde a la de accidentes.

Se evaluaron tres modelos de predicción de fuga: regresión logística, árboles de decisión y *random forest (RF)*. Todos ellos ampliamente utilizados y legitimados en la literatura por sus buenos desempeños/resultados. Estos modelos se trabajan dentro de un ambiente de desarrollo *Python*, donde la programación y evaluación de cada uno se logra por medio de la investigación e iteraciones de los resultados.

En este trabajo, el modelo *RF* demuestra un mejor desempeño a partir de los resultados, configuraciones y ajustes, con una precisión cercana al 80%. Este modelo indica que dos grandes sponsors concentran la mayor cantidad de clientes propensos a la fuga, donde la edad de los asegurados, el número de meses vigente en la compañía, la forma de pago de la prima el capital, cobertura y prima pagada, son los principales factores en la determinación de abandono.

Esto no basta para determinar el éxito/fracaso de la permanencia de los clientes, para ello es necesario evaluarlos y clasificarlos en términos de valor, creando para ello un método de valorización de los clientes que considera el valor esperado de los ingresos futuros, siniestralidad y los resultados del modelo de predicción RF. Todo ello bajo una configuración personalizada por *sponsor* y que arroja resultados de acuerdo al “peso” y siniestralidad de cada uno. Con estos resultados es posible elaborar estrategias de retención diferenciadas, pudiendo determinar a priori aquellos clientes que tienen un mayor valor para la compañía y enfocar sus esfuerzos en su retención y permanencia en el tiempo.

El desarrollo de este trabajo tiene foco principalmente en la detección de aquellos clientes con alta probabilidad de fuga y un alto costo según el valor que ellos entregan a la compañía. El alcance se reduce en realizar contacto con aquellos que presentan mayor probabilidad de fuga y mayor valor para la compañía, registrar los motivos razones de fuga y ver qué soluciones pueden aplicarse, las que pueden ir desde la aplicación de nuevas tarifas hasta la creación de productos diferenciados sin perder el costo real del producto/servicio ofrecido.

Lo anterior se sustenta bajo apoyo tecnológico, donde se propone la creación de nuevos módulos que soportan estas nuevas actividades para realizar ajustes y ejecutar el modelo de predicción seleccionado. Todo esto utilizando el actual sistema de administración de pólizas de seguros masivos SACS, conocido por los expertos de negocio y de quienes se enfrentan día a día con los clientes para su constante actualización. Esta propuesta incluye diagramación UML de las actividades necesarias y sus interacciones con los distintos módulos o actores que son clave para la obtención de resultados de valor de los clientes.

La evaluación económica para llevar a cabo este proyecto indica que es altamente rentable la inversión inicial para la obtención de los resultados en el corto plazo, con una TIR del 4.9%, los cuales se evalúan bajo distintos escenarios posibles de ventas y siniestralidad aplicando niveles de sensibilidad propias del negocio. Estos dos factores demuestran ser claves en la rentabilidad y éxito del trabajo en su totalidad. Todo esto bajo un escenario cortoplacista (3 años) dada la naturaleza del negocio y los ingresos/egresos de primas, los cuales configuran finalmente los resultados obtenidos.

Entendiendo la importancia estratégica detrás de la ejecución de este proyecto, donde la persistencia de los clientes es el factor clave y transversal para la rentabilidad del negocio, no es posible implementarlo de manera explícita, pero no quiere decir que no se debe realizar. Los recursos originalmente destinados para su realización se ven mermados debido a los altos costos de negocio asumidos producto de la contingencia social ocurrida en el país en octubre del año 2019 y para el año 2020 la pandemia de COVID19, lo que obliga a repensar las prioridades del negocio. Lo anterior imposibilita la opción de incorporar los distintos procesos y actividades de retención por la falta del apoyo tecnológico (módulos propuestos dentro del sistema de

administración de pólizas), ejecución del modelo de predicción de fuga, cálculo del valor del cliente y sobre todo la rentabilidad final obtenida por su implementación.

Esta importancia estratégica demostrada en términos financieros y de clientes justifica la inversión que la compañía debe realizar en recursos (tecnológicos, humanos), procesos y gestión de cambio al interior de la compañía. Cada cliente debe sentir que se están considerando sus necesidades y atenciones, bajo nuevas formas de hacer negocio. Lamentablemente, la imposibilidad de implementar los hallazgos y beneficios de este trabajo temporalmente (por decisión de la compañía), da paso a la digitalización y mejora de otros procesos y actividades del día a día consideradas también críticas, en este periodo excepcional y complejo que a nivel de la industria en general se está viviendo.

Finalmente, este trabajo tiene variados aspectos de mejora a considerar en el futuro en términos de selección de datos y ajustes a los modelos de predicción. Si bien el foco principal es el conocimiento de aquellos clientes con alta probabilidad de fuga y mayor valor para la compañía, sería de gran ayuda contar con un sistema integrado de recomendación de estrategias de retención para las demás líneas de negocio presentes. Así como también, la aplicación de modelos predictivos diferenciados para cada una de ellas, permitiendo la personalización y caracterización de los productos que la componen.

Durante la ejecución del proyecto se enfrentaron diversos obstáculos que hicieron muy difícil su realización, asumiendo la realidad a nivel compañía con respecto a las actividades, responsables de negocio y barreras existentes. La constante reinversión, levantamiento de procesos y rediseño de otros, mostraron claramente algunas oportunidades de mejora, como reutilización de la información almacenada o creación de nuevos roles o puestos de trabajo (especialistas de datos) que formarán parte de futuras investigaciones al respecto.



## CAPÍTULO 10: BIBLIOGRAFÍA

---

1. Alba, J. W. (1983). The Effects of Product Knowledge on the Comprehension, Retention, and Evaluation of Product Information. *ACR North American Advances*, NA-10. Recuperado de <http://acrwebsite.org/volumes/6184/volumes/v10/NA-10>.
2. Athanassopoulos, A: *Customer Satisfaction cues to Support Market Segmentation and Explain Switching behavior*. *Journal of Business Research*, 47:197, 2000.
3. Barrientos, F., & Ríos, S. A. (2013). *Aplicación de Minería de Datos para Predecir Fuga de Clientes en la Industria de las Telecomunicaciones*. 36.
4. Barros, Óscar (2004). *Ingeniería de Negocios: Diseño Integrado de Servicios, sus Procesos y Apoyo TI*. 375.
5. Biau, G. (s. f.). Analysis of a Random Forests Model. 33.
6. Breiman, Leo (2001). *Random Forests*. *Machine Learning* 45, 5–32 .
7. Chapman, P. et al, 2000. CRISP-DM 1.0 - Step-by-step data mining guide.
8. Chawla, N. et al, 2002. “SMOTE: *Synthetic Minority Over-sampling Technique*.”
9. Chen, Z.-Y., Fan, Z.-P. and Sun, M. (2012) ‘*Decision support A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data*’, *European Journal of Operational Research*, 223, pp. 461–472.
10. C. Chu, G. Xu, J. Brownlow and B. Fu, “*Deployment of churn prediction model in financial services industry*,” *2016 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC), Durham, NC, 2016*, pp. 1-2.
11. Coussement, K. and Van den Poel, D. (2008) ‘*Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques*’, *Expert Systems with Applications*, 34(1), pp. 313–327.
12. Domenech, J. (2011) <https://commons.wikimedia.org/w/index.php?curid=17617674>
13. Dupouy, C (2014). *Aplicación de árboles de decisión para la estimación del escenario económico y la estimación de movimiento la tasa de interés en Chile*.
14. Gallo, A (2014). *The value of Keeping the Right Customers*.
15. Goldie, C., Klüppelberg, C.: Subexponential distributions. In Adler, R., Feldman, R., Taqqu, M., eds.: *A practical guide to heavy tails: Statistical techniques for analysing heavy tails*. Birkhauser (1997).
16. Guelman, L. (2012). *Gradient boosting trees for auto insurance loss cost modeling and prediction*. *Expert Systems with Applications*, 39(3), 3659–3667.
17. Günther, C. C. et al. (2014) ‘*Modelling and predicting customer churn from an insurance company*’, *Scandinavian Actuarial Journal*, 2014(1), pp. 58–71.
18. Hax, A. C., & Wilde, D. L. (2003). *El Proyecto Delta*. Editorial Norma.
19. Lang, S., Kragler, P., Haybach, G., Fahrmeir, L.: Bayesian space-time analysis of health insurance data. In Schwaiger, M., O. Opitz, eds.: *Exploratory Data Analysis in Empirical Research*. Springer (2002)
20. López, R (2005). “*Análisis de la eficacia de los modelos de árboles de decisión para la gestión de políticas promocionales centradas en la segmentación de clientes*”.











21. Martinez J. (2008). Precisión, Recall F1, Accuracy en Clasificación. <https://iartificial.net/precision-recall-f1-accuracy-en-clasificacion/>
22. Martínez, Z (2007). "*Predicción de crisis empresariales en seguros no vida, mediante arboles de decisión y reglas de clasificación*".
23. Osterwalder, A. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. New Jersey: John Wiley & Sons.
24. Qian, Z., Jiang, W. and Tsui, K. -L. (2006) "*Churn detection via customer profile modeling*", International Journal of Production Research, Vol. 44, No. 14, pp. 2913–2933.
25. Raymond Mikulak, R. M. (1996). *The basics of FMEA*. Boca Raton, FL, USA: CRC Press.
26. Reinartz, W. J. and Kumar, V. (2003) 'The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration', Journal of Marketing, 67(1), pp. 77–99.
27. Sahar F. Sabbeh (2018) '*Machine-Learning Techniques for Customer Retention: A Comparative Study*', International Journal of Advanced Computer Science and Applications, 9(2).
28. Stucki, Oskar (2019). "*Predicting the customer churn with machine learning methods - CASE: private insurance customer data*"
29. Staudt, M., Kietz, J.U., Reimer, U.: A data mining support environment and its application to insurance data. In: Procs. KDD. (1998) insurance mining.
30. Tamaddoni, A., Stakhovych, S. and Ewing, M. (2016) '*Comparing Churn Prediction Techniques and Assessing Their Performance: A Contingent Perspective*', Journal of Service Research, 19(2), pp. 123–141.
31. Verbeke W, Martens D, Mues C, Baesens B (2011) *Building comprehensible customer churn prediction models with advanced rule induction techniques*. Expert Syst Appl 38:2354–2364
32. Wirth, R. (2000). *CRISP-DM: Towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 29–39.

## CAPÍTULO 11: ANEXOS

**Anexo 1:** Número de Casos y % de Abandono por Motivo registrado en el periodo 201712 a 201808.  
Elaboración Propia

		# Casos	% del Total
Renuncias Voluntarias	Error suscripción	152	0,0%
	Mala gestión venta	6.320	0,8%
	Disconformidad con el producto	49.961	6,3%
	Retractación	2.697	0,3%
	Problemas económicos	13.838	1,7%
	Otras renuncias	79.571	10,0%
Renuncias por No Pago	Mora	134.251	16,8%
	Problemas con el medio de pago	1.213	0,2%
	Cuenta/ tarjeta sin fondos	4.014	0,5%
	Cobro automático anulado por el asegurado	1.756	0,2%
	Mandato no instalado	5.714	0,7%
	Otros motivos	497.421	62,4%

**Anexo 2:** Modelo de negocios de Seguros Masivos según modelo CANVAS. Elaboración Propia

<p><b>Partners Clave</b></p> <ul style="list-style-type: none"> <li>Sponsors</li> </ul>    <ul style="list-style-type: none"> <li>Corredores</li> </ul>	<p><b>Actividades Clave</b></p> <ul style="list-style-type: none"> <li>Desarrollo Productos Rentables</li> <li>Marketing Inter Empresas</li> <li>Campañas</li> </ul>  <p><b>Recursos Clave</b></p> <ul style="list-style-type: none"> <li>Medio de Pago afin</li> <li>Bases de Clientes Sponsors</li> </ul> 	<p><b>Propuesta de Valor</b></p> <p>Brindar la confianza necesaria para actuar en sus necesidades de <b>cobertura de riesgos</b>, por medio de soluciones que permiten proteger lo que al cliente le importa y aprovechar al máximo lo que tiene, a través de respuestas directas y relaciones basadas en la confianza.</p>	<p><b>Relación con los Clientes</b></p>  <ul style="list-style-type: none"> <li>Póliza de Seguros</li> <li>Cond. General</li> <li>Cond. Particular</li> </ul> <p><b>Canales de comunicación y distribución</b></p>   	<p><b>Segmento de Clientes</b></p> <p><b>Orientado a la Masividad</b>, dependerán del Sponsor:</p> <ul style="list-style-type: none"> <li>Cliente Bancario</li> <li>Cliente Retail</li> <li>Cliente Clínicas de Salud</li> </ul> 
<p><b>Estructura de Costes</b></p> <ul style="list-style-type: none"> <li>Comisiones</li> <li>Desarrollo de Productos</li> <li>Costos de Administración, Mantención</li> <li><b>Siniestralidad</b></li> </ul>		<p><b>Ingresos</b></p> <ul style="list-style-type: none"> <li>Pago de Primas.</li> <li><b>Persistencia de Clientes.</b></li> <li>Devolución por Experiencias Favorables</li> </ul>		

**Anexo 3:** Tabla de valores brutos de datos (total). Elaboración Propia.

#	Nombre Variable Origen	Ejemplo Tipo Dato	Descripción
1	PERIODO	201809	Indica el periodo de registro del dato (Formato yyyyymm)
2	COD_LOB	4	Código que indica Tipo de Negocio al cual pertenece el dato (4 = Masivos)
3	COD_CANAL	3	Código del canal de negocios
4	DSC_CANAL	Otros	Descripción del Canal de Negocios
5	COD_SPONSOR	1	Código que indica el sponsor de la póliza
6	DSC_SPONSOR	BBVA	Descripción del sponsor
7	COD_PROD	14	Código que indica el producto asociado a la póliza
8	COD_PLAN	13	Código que indica el plan de la póliza
9	COD_POLIZA	340011399	Código que indica el número póliza
10	COD_RELACION	2	Código que indica relación del registro con el asegurado ppal.
11	COD_SEXO	0	Código que indica el sexo (0 =Masculino / 1 =Femenino)
12	COD_COBE	39	Código que indica la cobertura asegurada
13	DSC_COBE	Enfermedades Graves	Descripción de la cobertura asegurada
14	PERIODO_GRACIA	2	Numero de meses con periodo de gracia (si aplica)
15	CAPITAL	600	Capital Asegurado
16	PRIMA_DEVENGADA	0.4116	Prima Devengada
17	PRIMA_VIGENTE	0.4116	Prima Vigente de pago
18	COD_SUN	2710	Código contable SUN
19	DSC_SUN	TMK BBVA SALUD	Descripción de código contable SUN
20	COD_CANAL_VTA	2	Código de Canal de Venta
21	DSC_CANAL_VTA	TMK	Descripción Canal de Venta
22	COD_COBEPROPHET	2	Código de Cobertura del Modelo de Suscripción
23	DSC_COBEPROPHET	ENFERMEDADES GRAVES	Descripción de cobertura del modelo de suscripción
24	COD_RIESGO	2	Código que indica el riesgo
25	DSC_RIESGO	SALUD	Descripción del Riesgo (cobertura de la póliza)
26	DATA	INFORCE_COL	Indica el tipo de dato de la póliza en términos de vigencia
27	MOTIVO_BAJA	Morosidad en Cobranza	Descripción del motivo de abandono de la póliza
28	FECHANACIMIENTO	12/08/1983	Fecha Nacimiento del asegurado
29	FECHAVIGENCIA	01/09/2018	Fecha Vigencia Inicio de la póliza
30	FECHAVENCIMIENTO	01/09/2021	Fecha Fin de Vigencia de la póliza
31	FECHABAJA	15/10/2019	Fecha de baja del asegurado

**Anexo 4: Transformaciones y variables Dummy a Utilizar. Elaboración Propia.**

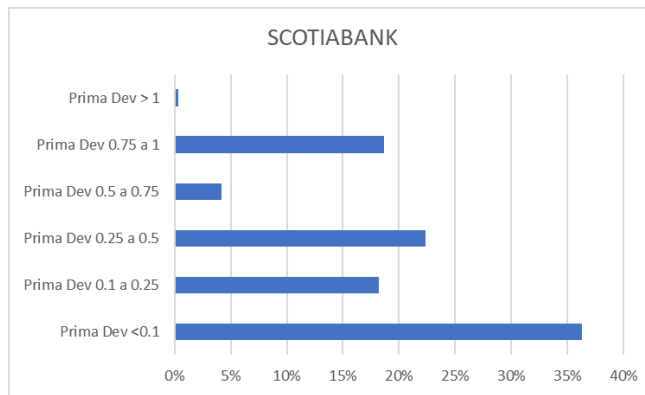
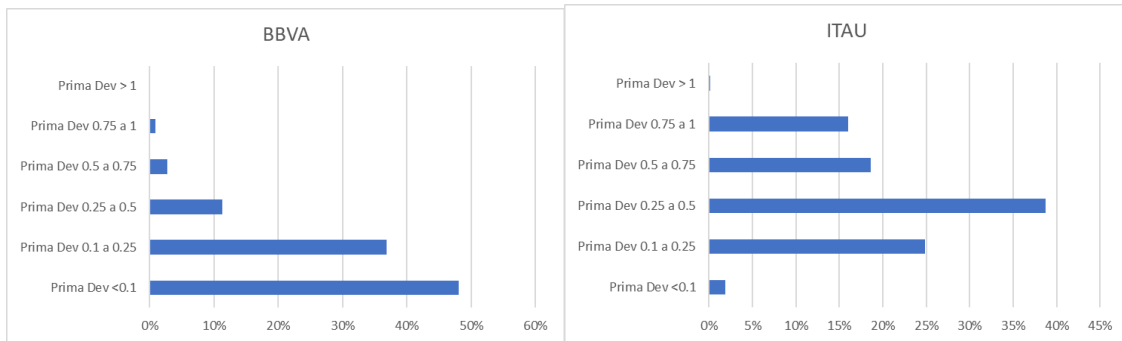
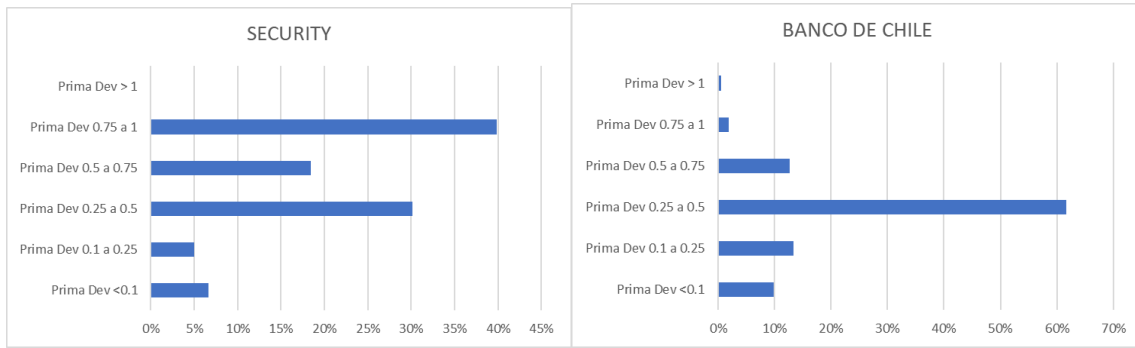
#	Nombre Variable Origen	Tipo	Ejemplo Tipo Dato	Descripción
46	EDAD	<i>Transformación</i>	35	Indica la Edad del asegurado edad
47	MESESVIGENTE	<i>Transformación</i>	33	Numero de Meses con que la póliza se mantiene vigente
48	Vigente	<i>Label</i>	0	Indica si la póliza se encuentra Vigente (0) o No Vigente (1)
49	BBVA	<i>Dummy</i>	1	Variable Dummy indica si la póliza pertenece al sponsor BBVA
50	BChile	<i>Dummy</i>	0	Variable Dummy indica si la póliza pertenece al sponsor Bchile
51	ITAU	<i>Dummy</i>	0	Variable Dummy indica si la póliza pertenece al sponsor Itaú
52	SCOTIABANK	<i>Dummy</i>	0	Variable Dummy indica si la póliza pertenece al sponsor Scotiabank
53	SECURITY	<i>Dummy</i>	0	Variable Dummy indica si la póliza pertenece al sponsor Security
54	Catastrofico	<i>Dummy</i>	0	Variable Dummy indica si la póliza contiene cobertura Catastrófica
55	EnfGraves	<i>Dummy</i>	0	Variable Dummy indica si la póliza contiene cobertura Enfermedades Graves
56	Fractura	<i>Dummy</i>	0	Variable Dummy indica si la póliza contiene cobertura Fractura
57	MuerteAcc	<i>Dummy</i>	0	Variable Dummy indica si la póliza contiene cobertura Muerte Accidental
58	Oncologico	<i>Dummy</i>	0	Variable Dummy indica si la póliza contiene cobertura Oncológica
59	PTD	<i>Dummy</i>	0	Variable Dummy indica si la póliza contiene cobertura PTD
60	RDH	<i>Dummy</i>	0	Variable Dummy indica si la póliza contiene cobertura RDH
61	Accidente	<i>Dummy</i>	0	Variable Dummy indica si la póliza pertenece al ramo FECU de Accidente
62	Salud	<i>Dummy</i>	1	Variable Dummy indica si la póliza pertenece al ramo FECU de Salud
63	Vida	<i>Dummy</i>	0	Variable Dummy indica si la póliza pertenece al ramo FECU de Vida
64	Edad <18	<i>Dummy</i>	0	Variable Dummy indica si la edad del asegurado es menor a 18 años
65	Edad Entre 18 y 24	<i>Dummy</i>	0	Variable Dummy indica si la edad del asegurado es entre 18 años a 24 años
66	Edad Entre 24 y 35	<i>Dummy</i>	0	Variable Dummy indica si la edad del asegurado es entre 24 años a 35 años
67	Edad Entre 35 y 45	<i>Dummy</i>	1	Variable Dummy indica si la edad del asegurado es entre 35 años a 45 años
68	Edad > 45	<i>Dummy</i>	0	Variable Dummy indica si la edad del asegurado es mayor a 45 años
69	Prima Dev <0.1	<i>Dummy</i>	0	Variable Dummy indica si la prima devengada es menor a UF 0.1

70	Prima Dev 0.1 a 0.25	<i>Dummy</i>	0	Variable Dummy indica si la prima devengada es entre UF 0.1 y UF 0.25
71	Prima Dev 0.25 a 0.5	<i>Dummy</i>	1	Variable Dummy indica si la prima devengada es entre UF 0.25 y UF 0.5
72	Prima Dev 0.5 a 0.75	<i>Dummy</i>	0	Variable Dummy indica si la prima devengada es entre UF 0.5 y UF 0.75
73	Prima Dev 0.75 a 1	<i>Dummy</i>	0	Variable Dummy indica si la prima devengada es entre UF 0.75 y UF 1.0
74	Prima Dev > 1	<i>Dummy</i>	0	Variable Dummy indica si la prima devengada es mayor a UF 1.0
75	CapAseg < 300	<i>Dummy</i>	0	Variable Dummy indica si el capital asegurado es menor a UF 300
76	CapAseg 300 a 500	<i>Dummy</i>	0	Variable Dummy indica si el capital asegurado es entre UF 300 a UF 500
77	CapAseg 500 a 700	<i>Dummy</i>	1	Variable Dummy indica si el capital asegurado es entre UF 500 a UF 700
78	CapAseg 700 a 1000	<i>Dummy</i>	0	Variable Dummy indica si el capital asegurado es entre UF 700 a UF 1000
79	CapAseg 1000 a 1500	<i>Dummy</i>	0	Variable Dummy indica si el capital asegurado es entre UF 1000 a UF 1500
80	CapAseg 1500 a 2000	<i>Dummy</i>	0	Variable Dummy indica si el capital asegurado es entre UF 1500 a UF 2000
81	CapAseg 2000 a 2500	<i>Dummy</i>	0	Variable Dummy indica si el capital asegurado es entre UF 2000 a UF 2500
82	CapAseg > 3000	<i>Dummy</i>	0	Variable Dummy indica si el capital asegurado es mayor a UF 2500

Anexo 5: Matriz de Correlación de Datos

Accidente	CapSeg < 300	CapSeg 300 a 500	CapSeg 500 a 700	CapSeg > 1500	CapAs eg	CAPIT	COD_CANAL	COD_COBE	COD_PROP	COD_HET	COD_PROD	COD_LOB	COD_PLAN	COD_PROD	COD_RIESG	COD_SEXO	COD_SOR	COD_AL	CodRif	esgo	>45	Edad Entre 18 y 24	Edad Entre 24 y 35	Edad Entre 35 y 45	Edad Entre 45 y 55	ITAU	SVIGE N	MESE N	ORIGE N	PERO DO_G A_VIG	PRIM RACIA	PRIM ENTE	PRIM <0.1	PRIM >0.1	PRIM 0.1 a 0.25	PRIM 0.25 a 0.5	PRIM 0.5 a 1	PRIM >1	SCOTI ABAN	Vigen range	
Accidente	1.0	0.6	-0.2	-0.3	-0.3	-0.2	-0.3	0.5	0.8	-0.3	0.3	-0.5	-1.0	0.0	-0.3	1.0	-1.0	-0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	-0.4	0.3	0.0	-0.3	-0.6	0.4	0.2	-0.3	-0.4	-1.0	0.0	0.1	0.0	0.1		
CapSeg < 300	0.6	1.0	-0.5	-0.3	-0.4	-0.2	-0.5	0.5	0.6	-0.5	0.5	-0.6	-0.6	0.0	-0.4	0.6	-0.6	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	-0.2	0.6	0.2	-0.2	-0.5	0.3	0.3	-0.4	-0.3	-0.4	-0.3	-0.6	-0.2	0.2		
CapSeg 300 a 500	-0.2	-0.5	1.0	-0.1	-0.2	0.1	0.2	-0.2	-0.3	0.2	-0.1	0.2	0.0	0.4	-0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	-0.3	0.0	0.2	0.5	-0.3	-0.1	0.0	0.6	0.2	0.2	0.2	-0.1			
CapSeg 500 a 700	-0.3	-0.3	-0.1	1.0	0.0	0.3	-0.2	-0.1	0.3	-0.4	0.3	0.3	0.0	0.3	0.0	-0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	-0.2	-0.3	-0.1	0.0	0.0	-0.2	0.2	-0.1	0.3	0.0	-0.2				
CAPITAL	-0.3	-0.4	-0.2	-0.1	1.0	0.0	0.2	-0.2	-0.3	0.2	-0.3	0.3	0.0	0.2	-0.3	0.3	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	-0.2	0.1	0.1	0.1	-0.1	-0.1	0.4	-0.1	0.3	0.0	-0.1				
COD_CANAL	-0.2	-0.2	0.1	0.0	0.0	1.0	0.2	-0.1	-0.2	0.2	-0.2	0.1	0.2	0.0	0.2	-0.2	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.2	0.1	0.0	0.1	0.0	-0.1	0.0	-0.1	0.0	0.1	0.2	0.2	0.0		
COD_COBE	-0.3	-0.5	0.2	-0.2	0.3	0.2	1.0	-0.5	-0.4	1.0	-0.8	0.9	0.3	0.0	0.5	-0.3	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.4	-0.8	-0.9	-0.4	0.1	0.0	-0.2	0.1	0.0	-0.2	0.1	0.1	0.3	0.7	-0.4	
COD_COBEPROPHET	0.8	0.6	-0.3	-0.1	-0.3	-0.2	-0.4	0.6	1.0	-0.4	0.5	-0.7	-0.8	0.0	-0.6	0.8	-0.8	-0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.3	0.0	-0.4	0.3	0.2	-0.4	-0.2	-0.5	-0.4	0.2	-0.5	-0.4	0.2	
COD_LOB	-0.3	-0.5	0.2	0.3	0.2	0.2	1.0	-0.5	-0.4	1.0	-0.8	0.9	0.3	0.0	0.5	-0.3	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.4	-0.8	-0.9	-0.4	0.1	0.0	-0.2	0.1	0.1	0.3	0.7	-0.4			
COD_PLAN	0.3	0.5	-0.1	-0.4	-0.3	0.2	0.8	0.4	0.5	-0.8	1.0	-0.8	-0.3	0.3	-0.5	1.0	0.0	-0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.5	0.0	0.0	0.0	0.2	0.0	0.2	0.0	-0.2	0.0	-0.3	-0.5	0.3	
COD_PROD	-0.5	-0.6	0.2	0.3	0.3	0.1	0.9	-0.5	-0.7	0.9	-0.8	1.0	0.5	0.0	0.6	-0.5	0.5	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.7	-0.6	0.0	0.3	-0.1	0.3	-0.1	0.2	0.2	0.2	0.5	0.6	-0.4		
COD_RIESGO	-1.0	-0.6	0.2	0.3	0.3	0.2	0.3	-0.5	-0.8	0.3	-0.3	0.5	1.0	0.0	0.3	-1.0	1.0	0.0	-0.1	0.0	-0.1	0.0	-0.1	0.0	0.0	0.4	-0.3	0.0	0.3	0.6	-0.4	-0.2	0.3	0.4	1.0	0.0	-0.1				
COD_SEXO	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
COD_SPONSOR	-0.3	-0.4	0.4	0.0	0.2	0.2	0.5	-0.5	-0.6	0.5	-0.6	0.6	0.3	0.0	1.0	-0.3	0.3	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.3	-0.5	-0.2	0.3	0.4	-0.2	0.2	0.3	0.3	0.7	-0.1				
CodRiesgo ACC	1.0	0.6	-0.2	-0.3	-0.3	-0.2	0.3	0.5	0.8	-0.3	0.3	-0.5	-1.0	0.0	-0.3	1.0	-1.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.4	0.3	0.0	-0.3	-0.6	0.4	0.2	-0.3	-0.4	-1.0	0.0	-0.1			
CodRiesgo SAL	-1.0	-0.6	0.2	0.3	0.3	0.2	0.3	-0.5	-0.8	0.3	-0.3	0.5	1.0	0.0	0.3	-1.0	1.0	0.0	-0.1	0.0	-0.1	0.0	-0.1	0.0	0.0	0.4	-0.3	0.0	0.3	0.6	-0.4	-0.2	0.3	0.4	1.0	0.0	-0.1				
Edad > 45	-0.1	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	0.1	-0.1	0.1	-0.1	0.1	0.1	0.0	-0.1	0.1	-0.1	-0.5	-0.6	0.1	-0.2	0.0	0.1	-0.2	0.0	0.1	-0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0		
Edad Entre 18 y 24	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Edad Entre 24 y 35	0.1	0.1	0.0	0.0	-0.1	0.0	0.0	0.1	0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	
Edad Entre 35 y 45	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ITAU	-0.4	-0.2	0.2	-0.1	0.1	0.0	-0.4	0.0	-0.4	0.0	0.0	0.4	0.0	0.3	-0.4	0.4	0.4	0.1	-0.1	-0.1	0.0	0.0	0.0	0.0	0.0	1.0	0.1	0.8	1.0	0.3	-0.3	0.0	0.2	0.1	0.4	-0.3	0.2				
MESESVIGENTE	0.3	0.6	-0.3	-0.2	-0.2	-0.2	-0.8	0.7	0.5	-0.8	0.7	-0.7	-0.3	0.0	-0.5	0.3	-0.3	-0.2	-0.1	0.2	0.0	0.1	0.2	0.0	0.1	1.0	0.6	0.1	-0.2	0.1	-0.3	-0.1	0.3	-0.1	-0.3	-0.1	-0.3	-0.1	-0.5	0.3	
ORIGEN	0.0	0.2	0.0	-0.3	-0.1	-0.1	-0.9	0.3	0.0	-0.9	0.5	-0.6	0.0	0.0	-0.2	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.6	1.0	0.8	0.1	-0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PERIODO_GRACIA	-0.3	-0.2	0.2	-0.1	0.1	0.0	-0.4	0.0	-0.4	0.0	0.0	0.3	0.0	0.3	0.0	0.3	0.3	0.1	-0.1	-0.1	0.0	-0.1	-0.1	0.0	1.0	0.1	0.8	1.0	0.3	0.0	0.2	0.1	0.3	-0.3	0.0	0.2	0.1	0.3	-0.3	0.2	
PRIMA_VIGENTE	-0.6	-0.5	0.5	0.0	0.1	0.1	-0.4	-0.5	0.1	0.0	0.3	0.6	0.0	0.4	-0.6	0.6	0.1	-0.1	-0.1	0.0	0.3	-0.2	0.1	0.3	1.0	0.3	-0.2	0.1	0.3	1.0	-0.7	-0.2	0.2	0.7	0.6	0.1	0.1	0.3	-0.1		
Prima Dev <0.1	0.4	0.3	-0.3	0.0	-0.1	0.0	0.0	0.3	0.4	0.0	0.0	-0.1	-0.4	0.0	-0.2	0.4	-0.4	-0.1	0.1	0.0	0.0	0.0	0.0	0.0	-0.3	0.1	-0.2	-0.3	-0.7	1.0	-0.4	-0.4	-0.2	-0.4	-0.4	-0.2	-0.4	0.0	0.0		
Prima Dev 0.1 a 0.25	0.2	0.3	-0.1	-0.2	-0.1	-0.1	-0.2	0.2	0.2	-0.2	0.0	-0.2	0.0	-0.2	0.2	0.0	-0.2	0.2	-0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.0	-0.2	-0.4	1.0	-0.3	-0.2	-0.2	-0.2	-0.2	-0.2	-0.1	-0.1	
Prima Dev 0.25 a 0.5	-0.3	-0.4	0.0	0.2	0.4	0.0	0.1	-0.4	-0.3	0.1	-0.2	0.2	0.3	0.0	-0.3	0.3	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	-0.3	0.0	0.2	0.2	0.2	-0.4	-0.3	-1.0	-0.2	0.3	0.0	0.0	0.0	0.0	
Prima Dev 0.75 a 1	-0.4	-0.3	0.6	-0.1	0.1	0.1	-0.2	-0.3	0.1	0.0	0.2	0.4	0.0	0.3	-0.4	0.4	0.0	0.0	0.0	0.1	-0.1	0.0	0.0	0.0	0.0	0.1	-0.1	0.0	0.1	0.7	-0.2	-0.2	1.0	0.4	0.2	0.0	0.4	0.2	-0.1		
Salud	-1.0	-0.6	0.2	0.3	0.3	0.2	0.3	-0.5	-0.8	0.3	0.5	1.0	0.0	0.3	-1.0	1.0	0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	-0.3	0.0	0.3	0.6	-0.4	-0.2	0.3	0.4	1.0	0.0	0.4	1.0	0.0	-0.1	
SCOTIABANK	0.0	-0.2	0.2	0.0	0.0	0.2	0.7	-0.4	-0.3	0.7	-0.5	0.6	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.3	-0.5	-0.6	-0.3	0.1	0.0	-0.1	0.0	0.2	0.0	0.2	0.0	1.0	-0.2		
Vigente = range1	0.1	0.2	-0.1	-0.2	-0.1	0.0	-0.4	0.2	0.1	-0.4	0.3	-0.4	-0.1	0.0	-0.1	0.1	-0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.4	0.2	-0.1	0.0	0.1	0.0	-0.1	0.0	-0.1	-0.1	-0.2	1.0		

**Anexo 7: Concentración de Clientes por Sponsor y Montos de Prima Devengada.**





## Anexo 8: Utilización de SMOTE para balanceo de clases en Phytton.

```
print("# Registros usados para entrenamiento: ", len(X_validationRLB))

print("Before OverSampling, counts of label 'No Vigente': {}".format(sum(Y_validationRLB == 1)))
print("Before OverSampling, counts of label 'Vigente': {}".format(sum(Y_validationRLB == 0)))

# import SMOTE module from imblearn library
# pip install imblearn (if you don't have imblearn in your system)
from imblearn.over_sampling import SMOTE
os = SMOTE()
X_validation_res, y_validation_res = os.fit_sample(X_validationRLB, Y_validationRLB.ravel())

print("\n After OverSampling, counts of label 'No Vigente': {}".format(sum(y_validation_res == 1)))
print("After OverSampling, counts of label 'Vigente': {}".format(sum(y_validation_res == 0)))

# Registros usados para entrenamiento: 32222
Before OverSampling, counts of label 'No Vigente': 8424
Before OverSampling, counts of label 'Vigente': 23798

After OverSampling, counts of label 'No Vigente': 23798
After OverSampling, counts of label 'Vigente': 23798
```

## Anexo 9: Utilización de Modelo Regresión Logística Balanceado en Phytton.

```
# Regresion Logistica Balanceado
mRLB = LogisticRegression(random_state=0, solver='liblinear', max_iter=100)

#Train the classifier.
mRLB.fit(os_data_XRLB, os_data_yRLB) # Entrenamiento del Modelo
clf = mRLB.fit(os_data_XRLB, os_data_yRLB)
pred_yRLB = mRLB.predict(X_trainRLB)
mostrar_resultados(Y_trainRLB, pred_yRLB)

RLB = round(mRLB.score(X_trainRLB, Y_trainRLB) * 100, 2)
print("Precision del modelo RLB:",RLB, "%")

#Error Clasificacion (Cuadrático Medio MSE)
mse = (np.square(Y_trainRLB - pred_yRLB)).mean()
print("Error Clasificacion (MSE): ",mse *100, "%")

# Cross Validation usando KFold = 5
scores = cross_val_score(clf, X_trainRLB, Y_trainRLB, cv=5, scoring="accuracy")
print("Metricas cross_validation: \n", scores)
print("Media de cross_validation: \n", scores.mean())
```

	precision	recall	f1-score	support
0	0.90	0.79	0.84	5949
1	0.55	0.74	0.63	2106
accuracy			0.78	8055
macro avg	0.72	0.76	0.74	8055
weighted avg	0.81	0.78	0.78	8055

Precision del modelo RLB: 77.54 %

Error Clasificacion (MSE): 22.45810055865922 %

Metricas cross\_validation:

[0.84357542 0.83985102 0.82805711 0.83302297 0.8336437 ]

Media de cross\_validation:

0.8356300434512726

## Anexo 10: Utilización de Modelo Árbol de Decisión Balanceado en Phytón.

```
# Arbol de Decision Balanceado
mDTB = DecisionTreeClassifier(#random_state=0,
                             criterion = 'gini',
                             #min_samples_split = 20,
                             max_depth = 15)
                             #min_samples_Leaf = 5) #,

#Train the classifier.
mDTB.fit(os_data_XRLB, os_data_yRLB)
clfDT = mDTB.fit(os_data_XRLB, os_data_yRLB)
pred_yDTB = mDTB.predict(X_trainDTB)
mostrar_resultados(Y_trainDTB, pred_yDTB)

DTB = round(mDTB.score(X_trainDTB, Y_trainDTB) * 100, 2)
print("Precision del modelo DTB:",DTB, "%")

#Error Clasificacion (Cuadrático Medio MSE)
mse = (np.square(Y_trainDTB - pred_yDTB)).mean()
print("Error Clasificacion (MSE): ",mse *100, "%")

###
# Cross Validation usando KFold = 5
scoresDT = cross_val_score(clfDT, X_trainDTB, Y_trainDTB, cv=5, scoring="accuracy")
print("Metricas cross_validation: \n", scoresDT)
print("Media de cross_validation: \n", scoresDT.mean())
```

	precision	recall	f1-score	support
0	0.90	0.93	0.91	5949
1	0.78	0.70	0.74	2106
accuracy			0.87	8055
macro avg	0.84	0.82	0.83	8055
weighted avg	0.87	0.87	0.87	8055

Precision del modelo DTB: 87.0 %  
Error Clasificacion (MSE): 12.998137802607076 %  
Metricas cross\_validation:  
[0.86468032 0.87212911 0.83923029 0.84233395 0.85536934]  
Media de cross\_validation:  
0.8547486033519555

## Anexo 11: Utilización de Modelo Random Forest Balanceado en Phytón.

```
#Random Forest Balanceado
mRFB = RandomForestClassifier(criterion = 'gini',max_depth = 16)
mRFB.fit(os_data_XRLB, os_data_yRLB)
clfRF = mRFB.fit(os_data_XRLB, os_data_yRLB)
pred_yRFB = mRFB.predict(X_trainRF)
mostrar_resultados(Y_trainRFB, pred_yRFB)

RFB = round(mRFB.score(X_trainRFB, Y_trainRFB) * 100, 2)
print("Precision del modelo RFB:",RFB, "%")

#Error Clasificacion (Cuadrático Medio MSE)
mse = (np.square(Y_trainRFB - pred_yRFB)).mean()
print("Error Clasificacion (MSE): ",mse *100, "%")

# Cross Validation usando KFold = 5
scoresRF = cross_val_score(clfRF, X_trainRFB, Y_trainRFB, cv=5, scoring="accuracy")
print("Metricas cross_validation: \n", scoresRF)
print("Media de cross_validation: \n", scoresRF.mean())
```

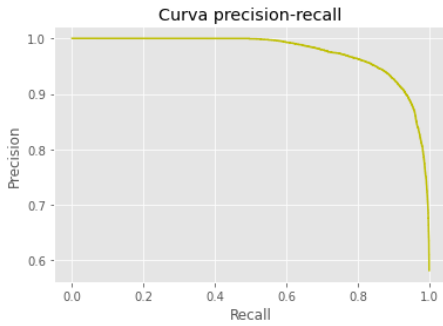
	precision	recall	f1-score	support
0	0.91	0.94	0.92	5949
1	0.80	0.73	0.76	2106
accuracy			0.88	8055
macro avg	0.85	0.83	0.84	8055
weighted avg	0.88	0.88	0.88	8055

Precision del modelo RFB: 88.17 %  
Error Clasificacion (MSE): 11.831160769708255 %  
Metricas cross\_validation:  
[0.86902545 0.87026691 0.85350714 0.86405959 0.85412787]  
Media de cross\_validation:  
0.8621973929236499

## Anexo 12: Curva de Precisión Modelo Random Forest Balanceado en Python.

```
# Curva de Precision
from sklearn.metrics import precision_recall_curve
precision, recall, _ = precision_recall_curve(os_data_yRFB, mRFB.predict_proba(os_data_XRFB)[:,:1])
plt.step(recall, precision, color='y')
plt.title('Curva precision-recall')
plt.xlabel('Recall')
plt.ylabel('Precision')
```

Text(0, 0.5, 'Precision')



## Anexo 13: Flujo de Caja situación actual. Elaboración Propia.

Concepto / Horizonte Tiempo	§	Año 1		Año 2		Año 3	
		Semestre 0	Semestre 1	Semestre 2	Semestre 3	Semestre 4	Semestre 5
Ingresos por Venta	-	3,084,004	4,246,901	5,481,030	6,821,462	8,361,587	9,996,367
Banca	-	3,084,004	4,246,901	5,481,030	6,821,462	8,361,587	9,996,367
Banco de Chile		369,181	518,336	680,189	858,862	1,065,439	1,288,258
BBVA		304,478	420,889	544,796	679,569	834,334	998,816
Scotiabank		1,566,908	2,124,021	2,703,392	3,323,189	4,031,258	4,771,190
Security		820,414	1,150,701	1,508,650	1,903,398	2,359,562	2,851,100
Otros Runoff		23,022	32,955	44,003	56,445	70,994	87,004
<b>Costo Variable Total</b>	-	<b>1,397,979</b>	<b>1,925,120</b>	<b>2,484,551</b>	<b>3,092,169</b>	<b>3,790,307</b>	<b>4,531,353</b>
Siniestralidad	-	431,761	594,566	767,344	955,005	1,170,622	1,399,491
Siniestralidad Banca	-	431,761	594,566	767,344	955,005	1,170,622	1,399,491
Administración y Ventas	-	966,218	1,330,554	1,717,207	2,137,164	2,619,685	3,131,862
CV Adm y Venta Banca	-	596,138	820,926	1,059,483	1,318,589	1,616,295	1,932,298
Reservas	-	370,080	509,628	657,724	818,575	1,003,390	1,199,564
<b>Costo Fijo Total</b>	-	<b>350,655</b>	<b>455,315</b>	<b>566,387</b>	<b>687,026</b>	<b>825,637</b>	<b>972,767</b>
<b>UAI</b>	-	<b>1,335,370</b>	<b>1,866,465</b>	<b>2,430,092</b>	<b>3,042,267</b>	<b>3,745,642</b>	<b>4,492,247</b>
Impuesto 27%	-	360,550	503,946	656,125	821,412	1,011,323	1,212,907
<b>UDI</b>	-	<b>974,820</b>	<b>1,362,520</b>	<b>1,773,967</b>	<b>2,220,855</b>	<b>2,734,319</b>	<b>3,279,340</b>
<b>FCO</b>	-	<b>974,820</b>	<b>1,362,520</b>	<b>1,773,967</b>	<b>2,220,855</b>	<b>2,734,319</b>	<b>3,279,340</b>
<b>FCC</b>	-	-	-	-	-	-	-
<b>Flujo Caja Privado</b>	-	<b>974,820</b>	<b>1,362,520</b>	<b>1,773,967</b>	<b>2,220,855</b>	<b>2,734,319</b>	<b>3,279,340</b>

**Anexo 14: Flujo de Caja situación actual mejorada. Elaboración Propia.**

Concepto / Horizonte Tiempo	§	Año 1		Año 2		Año 3	
		Semestre 0	Semestre 1	Semestre 2	Semestre 3	Semestre 4	Semestre 5
Ingresos por Venta							
Banca		3,084,004	4,246,901	5,481,030	6,821,462	8,361,587	9,996,367
Banco de Chile		369,181	518,336	680,189	858,862	1,065,439	1,288,258
BBVA		304,478	420,889	544,796	679,569	834,334	998,816
Scotiabank		1,566,908	2,124,021	2,703,392	3,323,189	4,031,258	4,771,190
Security		820,414	1,150,701	1,508,650	1,903,398	2,359,562	2,851,100
Otros Runoff		23,022	32,955	44,003	56,445	70,994	87,004
<b>Costo Variable Total</b>	-	1,291,964	1,594,317	1,915,191	2,263,703	2,664,136	3,089,179
Siniestralidad	-	431,761	594,566	767,344	955,005	1,170,622	1,399,491
Siniestralidad Banca	-	431,761	594,566	767,344	955,005	1,170,622	1,399,491
Administración y Ventas	-	860,204	999,751	1,147,847	1,308,699	1,493,514	1,689,687
CV Adm y Venta Banca	-	490,123	490,123	490,123	490,123	490,123	490,123
Reservas	-	370,080	509,628	657,724	818,575	1,003,390	1,199,564
<b>Costo Fijo Total</b>	-	350,655	455,315	566,387	687,026	825,637	972,767
<b>UAI</b>	-	<b>1,441,385</b>	<b>2,197,268</b>	<b>2,999,452</b>	<b>3,870,733</b>	<b>4,871,814</b>	<b>5,934,421</b>
Impuesto 27%	-	389,174	593,262	809,852	1,045,098	1,315,390	1,602,294
<b>UDI</b>	-	<b>1,052,211</b>	<b>1,604,006</b>	<b>2,189,600</b>	<b>2,825,635</b>	<b>3,556,424</b>	<b>4,332,127</b>
<b>FCO</b>	-	<b>1,052,211</b>	<b>1,604,006</b>	<b>2,189,600</b>	<b>2,825,635</b>	<b>3,556,424</b>	<b>4,332,127</b>
<b>FCC</b>	-	-	-	-	-	-	-
<b>Flujo Caja Privado</b>	-	<b>1,052,211</b>	<b>1,604,006</b>	<b>2,189,600</b>	<b>2,825,635</b>	<b>3,556,424</b>	<b>4,332,127</b>