



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**DISEÑO DE UN MODELO PREDICTIVO PARA EL AUMENTO DE PÓLIZAS
PRINCIPALES EN UNA COMPAÑÍA DE SEGUROS**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

FELIPE IGNACIO GARCÍA BACCHIEGA

PROFESOR GUÍA:

ALEJANDRA PUENTE CHANDÍA

MIEMBROS DE LA COMISIÓN:

PABLO MARÍN VICUÑA

ERICK MÉNDEZ GUZMÁN

SANTIAGO DE CHILE

2017

**RESUMEN DE LA MEMORIA PARA
OPTAR AL TÍTULO DE:** Ingeniero Civil
Industrial
POR: Felipe Ignacio García Bacchiega
FECHA: 13/11/2017
PROFESOR GUÍA: Alejandra Puentes Chandía

DISEÑO DE UN MODELO PREDICTIVO PARA EL AUMENTO DE PÓLIZAS PRINCIPALES EN UNA COMPAÑÍA DE SEGUROS

Seguros Falabella es una Corredora de Seguros Grandes Tiendas perteneciente al grupo Falabella. Tiene participación en algunos países de Latinoamérica, en particular en Chile, donde se realiza este trabajo de memoria.

El objetivo general de esta memoria consiste en diseñar un modelo predictivo basado en las características individuales de los clientes, que entregue la propensión de compra para los seis productos principales de la compañía, con el fin de incrementar el stock de estas pólizas.

Se realizan modelos de propensión para todos los ramos principales de seguros, correspondientes a automotrices, vida, vida con bonificación, salud, hogar y transaccionales. En particular se utilizan dos metodologías distintas de clasificación binaria: Árbol de Decisión y Logit Binario.

Dada la naturaleza de los datos y la diferencia de información para los clientes con y sin la tarjeta de crédito del holding, es necesario calcular modelos independientes para ambos tipos de clientes. Así, se generan 12 modelos distintos tanto para Árbol de Decisión como también para Logit Binario.

La empresa cuenta con un total de 3 millones de registros de clientes entre los periodos 2015 y 2016. Debido a la metodología elegida, sólo se utiliza un 40% de la base resultante para obtener resultados consistentes y en un tiempo de ejecución razonable.

En base a los resultados, se elige el método Logit Binario puesto que tuvo un mejor desempeño en las métricas más significativas para el negocio. Éste, en promedio, logró recuperar un 60% de las ventas generadas con sólo un 20% de la base de datos, en comparación con el 53,3% del árbol de decisión.

Para elegir el producto más adecuado para cada cliente, se propone un modelo de “Next Best Offer”. Se realiza una simulación de esta metodología versus la actual, obteniendo un beneficio estimado 11% mayor y un aumento de 22% en pólizas vendidas.

Para validar este método, se plantean dos experimentos de igualdad de proporciones. El primero busca comprobar que la elección mayor decil de propensión genera más ventas cuando existe un único máximo decil. El segundo busca verificar que la ponderación por el beneficio de la compañía genera más ventas cuando existe más de un producto que compartan el máximo decil.

Finalmente, para estudios futuros, se propone la realización de un modelo que determine, para cada cliente, el canal más propenso para la contratación de un producto. Además, se plantea una idea de mejora del método “Next Best Offer” mediante la optimización según una ponderación tanto del decil de propensión como también del beneficio esperado para la compañía.

AGRADECIMIENTOS

Quisiera agradecerle a mi familia por todo el apoyo que me han brindado desde el comienzo de mi carrera. Ha sido un proceso largo, sin embargo ustedes siempre han estado ahí para apoyarme y aconsejarme cuando lo necesité.

Darle las gracias a mis amigos, quienes siempre han estado ahí para llenar con risas, alegrías y buenos momentos este proceso. Gracias a los de la U, el Harem y mis amigos más cercanos por todos los recuerdos, anécdotas e historias que me llevo conmigo. Ha sido por lejos lo más valioso que he podido sacar de esta etapa.

Gracias la Rosi, que me apoyó durante todo este tiempo, dándome ánimo y energía para lograr terminar esta memoria. Gracias linda por tu apoyo y paciencia, aun cuando estaba mañoso y cansado.

Agradecerle al equipo de Business Intelligence de Seguros Falabella por toda la paciencia, consejos y apoyo que me brindaron en este proceso. Gracias por hacer más ameno el trabajo en la oficina y por las risas en los almuerzos.

Finalmente darle las gracias a los miembros de mi comisión por toda la ayuda brindada para poder entregar la memoria: a Erick Méndez por su buena disposición y rápidas correcciones; a Pablo Marín por su visión más crítica que me permitió ser más prolijo y enfocado en los detalles; y a Alejandra Puente por su constante apoyo en la gestión de la titulación anticipada y por siempre darse el tiempo para reunirse conmigo a pesar de estar siempre con muchas cosas.

Ahora puedo decirles, mamá y papá: ¡Misión Cumplida!

TABLA DE CONTENIDO

1. ANTECEDENTES GENERALES	1
1.1 Industria Corredora Seguros Grandes Tiendas	1
1.2 Seguros Falabella	1
2. DESCRIPCIÓN Y JUSTIFICACIÓN DEL PROYECTO	2
2.1 IDENTIFICACIÓN DEL PROBLEMA Y OPORTUNIDAD	2
3. OBJETIVOS.....	4
3.1 Objetivo general	4
3.2 Objetivos específicos	4
4. ALCANCES	4
5. RESULTADOS ESPERADOS	5
6. MARCO TEÓRICO	6
6.1 Proceso de extracción y transformación de datos	6
6.2 Modelos de propensión	6
6.2.1 Modelo Logit Binario	7
6.2.2 Modelo Árbol de Decisión	8
6.3 Next Best Offer	8
6.4 Métricas de comparación	8
7. METODOLOGÍA	10
7.1 Descripción de fuentes de datos	10
7.2 Tratamiento de datos	10
7.3 Desarrollo de modelos de propensión	10
7.4 Evaluación.....	11
7.4.1 Desempeño modelos de propensión.....	11
7.5 Propuestas de acción	11
7.6 Propuesta de diseño experimental.....	11
8. DESARROLLO.....	11
8.1 Descripción de los datos	11
8.1.1 Tenencia de seguros	11
8.1.2 Gastos tarjeta CMR.....	12
8.1.3 Datos demográficos	12
8.2 Tratamiento de datos	13
8.2.1 Selección.....	13

8.2.2	Pre-procesamiento	14
8.2.3	Transformación.....	14
8.2.4	Descripción de datos	15
8.3	Desarrollo de modelos de propensión	22
8.3.1	Modelo Logit Binario	22
8.3.2	Modelo Árbol de Decisión.....	27
8.4	Evaluación de modelos.....	31
8.4.1	Desempeño modelos de propensión.....	31
8.5	Propuestas de acción	37
8.6	Propuesta de diseño experimental.....	38
8.7	Evaluación económica	40
9.	CONCLUSIONES	42
9.1	Conclusiones del proyecto.....	42
9.2	Limitaciones y trabajos futuros	43
10.	BIBLIOGRAFÍA	45
11.	ANEXOS	47
11.1	Anexo A.....	47
11.2	Anexo B.....	47
11.3	Anexo C	49
11.4	Anexo D.....	51
11.5	Anexo E.....	75
11.6	Anexo F	75

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Aumento porcentual de primas intermediadas por compañía	2
Ilustración 2: Evolución fuga y cross-selling seguro Automotriz.....	3
Ilustración 3: Esquema proceso ETL.....	6
Ilustración 4: Distribución base entrenamiento y testeó	7
Ilustración 5: Gráfico de método de creación de Matriz de Confusión	9
Ilustración 6: Distribución de tenencia según ramo	15
Ilustración 7: Distribución de clientes por edades	16
Ilustración 8: Porcentaje de contrataciones de seguros de vida según edad clientes con CMR.....	17
Ilustración 9: Porcentaje de contrataciones de seguros de vida según edad clientes no CMR.....	17
Ilustración 10: Distribución de clientes por género	18
Ilustración 11: Distribución de clientes por GSE.....	18
Ilustración 12: Distribución de clientes por Estado Civil	19
Ilustración 13: Distribución de clientes por habitación	19
Ilustración 14: Distribución tenencia vehicular.....	20
Ilustración 15: Contratación de seguros automotriz según cantidad de vehículos para clientes con CMR.....	20
Ilustración 16: Contratación de seguros automotriz según cantidad de vehículos para clientes con CMR.....	21
Ilustración 17: Distribución de clientes según monto de gastos por rubros	22
Ilustración 18: Extracto perfil más propenso modelo auto con CMR.....	23
Ilustración 19: Extracto perfil más propenso modelo vida con CMR	24
Ilustración 20: Extracto perfil más propenso modelo vida con bonificación con CMR ...	25
Ilustración 21: Extracto perfil más propenso modelo salud con CMR.....	25
Ilustración 22: Extracto perfil más propenso modelo hogar con CMR	26
Ilustración 23: Extracto perfil más propenso modelo transaccional con CMR.....	27
Ilustración 24: Gráfico Árbol de Decisión modelo auto con CMR.....	28
Ilustración 25: Gráfico Árbol de Decisión modelo vida con CMR	28
Ilustración 26: Gráfico Árbol de Decisión modelo vida con bonificación con CMR.....	29
Ilustración 27: Gráfico Árbol de Decisión modelo salud con CMR.....	29
Ilustración 28: Gráfico Árbol de Decisión modelo hogar con CMR.....	30
Ilustración 29: Gráfico Árbol de Decisión modelo transaccionales con CMR.....	30
Ilustración 30: Probabilidad media por producto modelo Logit Binario y Árbol de Decisión sin CMR.....	31
Ilustración 31: Desempeño modelos según matrices de confusión sin CMR	31
Ilustración 32: Diferencia en métricas de desempeño Logit – Arbol sin CMR	32
Ilustración 33: Curva de ganancia modelo auto sin CMR	33
Ilustración 34: Desempeño mejor quintil sin CMR modelo Logit y Arbol	33
Ilustración 35: Probabilidad media por producto modelo Logit Binario y Árbol de Decisión con CMR.....	34
Ilustración 36: Desempeño modelos según matrices de confusión con CMR.....	34
Ilustración 37: Diferencia métricas de desempeño modelo Logit - Árbol con CMR	34
Ilustración 38: Curva de ganancia modelo auto con CMR.....	35
Ilustración 39: Desempeño mejor quintil con CMR modelo Logit y Arbol	35
Ilustración 40: Ejemplo máximo decil	38
Ilustración 41: Descripción experimento máximo decil.....	39

Ilustración 42: Descripción experimento máximo beneficio esperado	39
Ilustración 43: Ventas simuladas según modelo propuesto.....	40

1. ANTECEDENTES GENERALES

1.1 Industria Corredora Seguros Grandes Tiendas

Las corredoras de seguros son entidades jurídicas o naturales que actúan como intermediarios entre las compañías de seguros y los clientes que se quieren asegurar, pudiendo ser personas naturales o empresas.

Los principales actores en este mercado son las compañías de seguros y los clientes. Los primeros son quienes ofrecen distintas pólizas para cubrir algún riesgo, mientras que los segundos contratan estos seguros que son ofrecidos por dichas aseguradoras.

Las corredoras entran entre estos dos actores, siendo el nexo entre las compañías de seguros y los clientes. Estas entidades, que son independientes a las aseguradoras, ofrecen a sus clientes ciertas coberturas de acuerdo a sus requerimientos y en base a la oferta de seguros que tienen [1]. De esta manera, las corredoras obtienen una comisión por intermediación, que corresponde a la diferencia entre el precio que le cobra la corredora al asegurado y el precio al que la compañía de seguros ofrece el producto a la corredora.

Si bien, las corredoras de seguros ofrecen a los clientes los distintos productos que las compañías aseguradoras les permiten proporcionar, también brindan a dichos clientes un servicio de asesoría, ya sea en la contratación de este seguro como en el proceso de ejecución de la póliza. Por esta razón, la compañía brinda tanto productos como servicios a sus clientes al momento de contratar un seguro.

Este mercado está regulado por la Superintendencia de Valores y Seguros, entidad que supervisa el mercado, exigiendo a los corredores que estén registrados y entreguen reportes de manera periódica, entre otras solicitudes. Es esta misma entidad quien mantiene un estudio sobre la tendencia de esta industria, la cual ha sido positiva desde el año 2013, destacándose el año 2016 con un crecimiento de un 18,03% en la prima intermediada [2], vendiendo un total de 311.758 millones de pesos.

1.2 Seguros Falabella

Seguros Falabella cuenta con el prestigio, tamaño y participación de mercado de la marca asociada al holding Falabella, dándole un poder de marca importante en comparación con sus competidores. Lo anterior les permite negociar mejores precios con las compañías aseguradoras que su competencia.

Otra de sus ventajas consiste en la instalación de una agencia de Seguros Falabella en las tiendas de retail, permitiéndoles aprovechar el flujo de personas que concurren a estos locales. Esto se traduce en un menor costo de adquisición de clientes por parte de la compañía

En el sector de corredoras de seguros grandes tiendas, Seguros Falabella tiene una participación de mercado cercana a un 60% por prima intermediada grandes tiendas, donde es líder de mercado tanto en seguros generales como en seguros de vida, segmentos donde tiene un 76% y 35% de participación respectivamente [2].

Los montos asociados a las primas intermediadas por la compañía han ido en aumento entre los años 2010 y 2016, desde \$99.556 millones de pesos hasta llegar a los \$186.738 millones respectivamente (Anexo Figura 1) [2]. Sin embargo, al analizar el porcentaje que representan dichos montos frente al mercado, la compañía muestra un crecimiento sostenido entre los años 2010 y 2014 llegando a este último con un 64%, mientras que en los dos años posteriores ha perdido participación hasta llegar a un 59,9% (Anexo Figura 2). [2]

Por esta razón, si bien la organización se encuentra en una etapa de crecimiento con respecto al monto de primas intermediadas, en el tema de participación de mercado se encuentra en un leve decrecimiento.

2. DESCRIPCIÓN Y JUSTIFICACIÓN DEL PROYECTO

2.1 IDENTIFICACIÓN DEL PROBLEMA Y OPORTUNIDAD

En los pasados años, Seguros Falabella no ha mostrado un aumento suficiente en la venta de sus productos para mantener la cuota de mercado de los años anteriores. Dicha cuota ha ido disminuyendo desde 2014, teniendo una participación de 64%, 63% y 59,9% en 2014, 2015 y 2016, respectivamente [2].

Con el fin de determinar las posibles causas de esta disminución en la cuota de mercado y en la cantidad de ventas, se comparó el aumento porcentual en el monto de primas intermediadas por cada compañía en relación al año anterior, la cual se muestra en la ilustración 1.

	dic-16	dic-15	dic-14	dic-13	dic-12	dic-11
Corredora de Seguros Tricot Ltda	14%	51%	105%	424%	-	-
Líder Corredores de Seguros y Gestión Financiera S.A.	46%	7%	3%	1%	-11%	15%
Corona Corredora de Seguros Ltda.	48%	26%	1012%	-	-	-
Seguros Falabella Corredores Ltda.	12%	7%	17%	6%	4%	21%
ABCDIN Corredores de Seguros Ltda.	1%	25%	15%	6%	-17%	12%
La Polar Corredores de Seguros Ltda.	42%	37%	50%	24%	-46%	-44%
Agencias Briner Corredores de Seguros S.A.	-1%	-10%	-6%	3%	-12%	2%
Total	18%	9%	16%	6%	-4%	11%

*Ilustración 1: Aumento porcentual de primas intermediadas por compañía
Fuente: Elaboración propia en base a Informe Corredoras de Seguros Grandes Tiendas [2]*

En ella es posible apreciar que 4 de las 6 empresas que compiten contra Seguros Falabella han mostrado un aumento porcentual mayor al obtenido por ésta última. Se puede observar que las corredoras La Polar, Líder y Corona tuvieron un aumento mayor a un 40% en el último año, mientras que Seguros Falabella sólo aumentó un 12% en relación al año anterior.

En base a esto, podemos concluir que el mercado de las Corredoras de Seguros Grandes Tiendas sigue activo, dado que aún se generan aumentos en los porcentajes de montos por primas intermediadas.

Por otro lado, analizaremos el comportamiento de la fuga y el cross-selling de la compañía. Éste último consiste en ofrecer a los clientes actuales otros productos que aún no tienen. En la ilustración 2 podemos observar la evolución de la fuga y el cross-selling, ambas medidas en cantidad de pólizas, representada por los gráficos de líneas en el eje

principal. Además, la diferencia neta entre estos dos valores en el eje secundario, muestra que ha habido varios periodos donde existe más fugas que ventas por cross-selling.

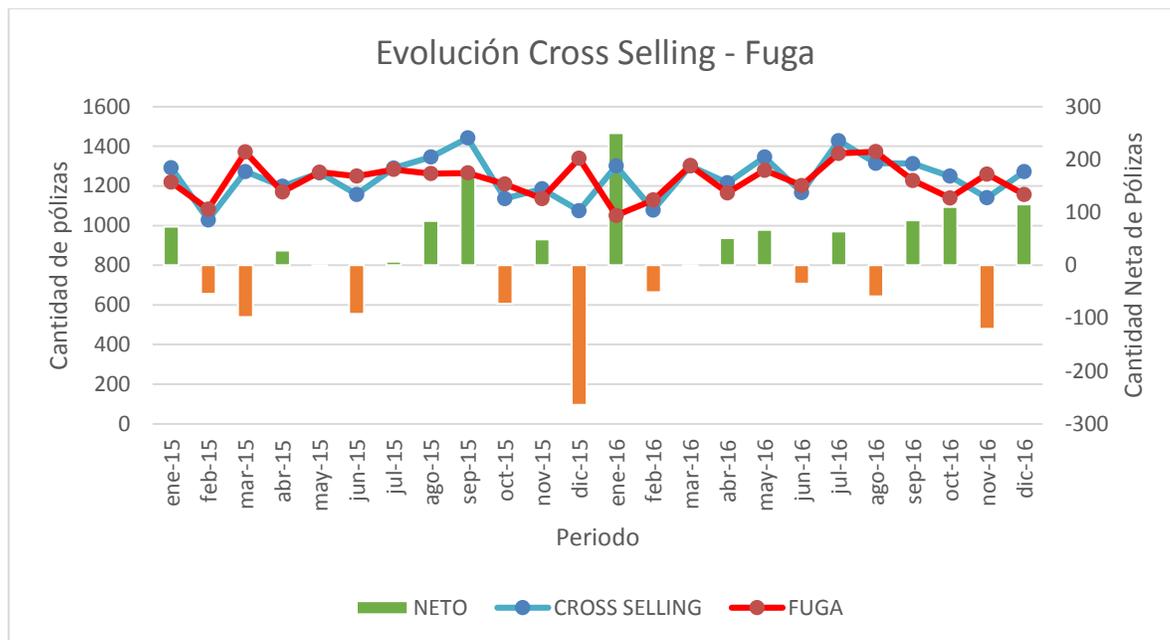


Ilustración 2: Evolución fuga y cross-selling seguro Automotriz
Fuente: Elaboración Propia

Lo anterior se debe a la pequeña diferencia existente entre el cross-selling y la fuga. Si consideramos el promedio de los valores netos, obtenemos un pequeño incremento de 12 pólizas.

Es de suma importancia que Seguros Falabella pueda mejorar en esta situación para competir y mantener su posición como líder del mercado. Por lo tanto, es indispensable para la empresa que aumente el stock de pólizas, dado que no atraer nuevos ingresos, entendiendo estos como la venta de productos y no como un aumento en la comisión cobrada, tiene un efecto negativo en la participación de mercado. Por esta razón resulta interesante generar un aumento en el stock de pólizas principales mediante el cross-selling.

Actualmente la compañía utiliza un método de selección de clientes más propensos para la contratación de un seguro automotriz basado en un modelo de regresión logística. Este modelo sólo considera a aquellos clientes que cuentan con tarjeta de crédito de la tienda de retail del mismo holding Falabella, llamada CMR Falabella. Es con este modelo de propensión con el cual generan una base de datos que se le envía posteriormente a call-center con el fin de realizar las ofertas de productos para generar cross-selling.

De esta manera, resulta útil para la compañía poder realizar ofertas sobre todos los productos principales que tienen. Así, no solamente se podrá realizar un cross-selling de seguros automotriz, sino que se podrá ofrecer a cada cliente el producto más propenso de adquisición.

El presente trabajo pretende entregar las herramientas necesarias para generar un aumento en el stock de pólizas principales de la compañía. Para esto se plantearán dos modelos de propensión para luego determinar, dado del desempeño de estos, cuál es el modelo que cumple de mejor manera el objetivo principal de determinar el producto al que cada cliente está más propenso de elegir.

Estos modelos permitirán aprovechar la oportunidad entregada por la amplia variedad de datos sobre los clientes que tiene la compañía con el fin de dar una mejor oferta a sus clientes. Con esto se busca generar un mayor interés y atracción por parte del cliente sobre los productos ofrecidos, aumentando la posibilidad que éste los acepte. De esta manera se busca también aumentar el cross-selling de la compañía y, por ende, la participación de mercado.

3. OBJETIVOS

3.1 Objetivo general

Diseñar un modelo predictivo, basado en las características individuales de cada cliente, que entregue la probabilidad de propensión de compra para los productos elegidos con el fin de incrementar el stock de pólizas principales de Seguros Falabella.

3.2 Objetivos específicos

Los objetivos específicos de este proyecto son:

1. Generar un levantamiento de los datos a utilizar en el modelo y realizar un análisis descriptivo de éstos, con el fin de determinar las características a utilizar.
2. Seleccionar las características que tengan relevancia con el modelo predictivo en cuestión e identificar el nivel de agregación del modelo.
3. Desarrollar modelos que permitan determinar el producto al cual cada cliente está más propenso a contratar.
4. Evaluar la capacidad predictiva de cada modelo, tanto con una visión analítica enfocada en el desempeño de cada uno, como otra comercial enfocada en el valor para la compañía.
5. Plantear un diseño experimental que tenga como finalidad poder determinar la efectividad del modelo Next Best Offer.

4. ALCANCES

Este proyecto se encuentra limitado a los clientes actuales de Seguros Falabella, quienes tienen como mínimo un producto vigente contratado en la empresa, por lo que quedan excluidos todos aquellos clientes que se han fugado de la compañía y quienes no forman parte de ella.

Este trabajo de memoria consta de tres partes principales:

1. La primera consiste en un análisis de los datos y la ejecución de los distintos modelos de propensión en base a estos datos. El alcance de esta primera parte consiste en crear un modelo de propensión de venta por canal, para los 6 productos más importantes para la compañía, los cuales son seguros de auto, vida, vida con bonificación, salud, hogar y transacciones.
2. La segunda parte corresponde a la evaluación de los modelos y creación de un modelo “Next Best Offer” para seleccionar el mejor producto para un cliente. Para ello se compararán los modelos según el desempeño en el mejor quintil calculado por éstos y su desempeño según métricas asociadas a la matriz de confusión. Posteriormente se evaluará algún método para elegir el mejor producto para el cliente en base a los resultados del modelo de propensión que más se adecúe al negocio.
3. Finalmente, la tercera parte de este proyecto consiste en el planteamiento de un diseño experimental que permita comprobar que el modelo “Next Best Offer” propuesto en la parte anterior, obtiene mejores resultados que un modelo aleatorio. Para ello se deberá obtener como resultado que existe una diferencia entre las proporciones de los grupos de control y experimental. Este diseño experimental quedará propuesto para futuras investigaciones, sin embargo, se detallará los distintos grupos de control y experimentales.

5. RESULTADOS ESPERADOS

El proyecto pretende entregar cuatro resultados importantes:

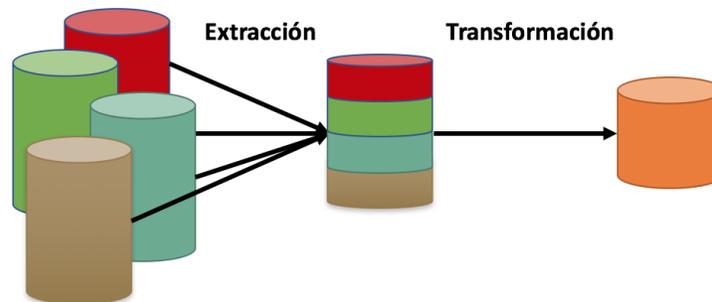
1. El primero consiste en la selección del conjunto de características más importantes de los clientes para determinar la propensión a cada producto. Este resultado es útil para la compañía dado que permitirá obtener un perfil de aquellas características más importantes para la propensión. Además, ayudará a determinar cuáles son las variables de mayor impacto para los distintos productos, con el fin de mantenerlas actualizados y sin errores.
2. El segundo consiste en generar los distintos modelos para obtener la propensión que tiene cada cliente para cada uno de los productos.
3. El tercero, corresponde a encontrar los métodos más efectivos para la evaluación de modelos de propensión realizados, con el fin de elegir el modelo de mejor desempeño y que más se adecúe a las necesidades de la compañía.
4. Finalmente, el cuarto resultado consiste en encontrar la mejor forma de comparar el producto por el cuál un cliente está más propenso, obteniendo una manera de poder evaluar las distintas propensiones obtenidas por cada modelo y para cada

uno de sus ramos. Lo anterior se debe a que cada probabilidad por producto, proviene de un modelo distinto, por lo que es necesario encontrar la manera de poder comparar estos valores entre sí.

6. MARCO TEÓRICO

6.1 Proceso de extracción y transformación de datos

El proceso de extracción y transformación de datos consiste en obtener información de una base de datos desde diversas fuentes y realizar transformaciones sobre estos con el fin de asegurar su calidad y consistencia, como se muestra en la ilustración 3. De esta manera, se generará una base de datos limpia y lista para poder trabajar con dichos datos, con el objetivo de no tener errores en los campos ni valores vacíos.



*Ilustración 3: Esquema proceso ETL
Fuente: Elaboración propia*

La extracción de datos se realizará mediante consultas en SQL directamente de los datos de la compañía. Dentro de este proceso se encuentra la fusión de las distintas tablas de datos en una sola. Así, se tendrá toda la información en una misma tabla de datos, la cual tendrá registros únicos para cada cliente.

La transformación es un proceso muy importante debido a que permitirá detectar anomalías en los datos. Además, es con este proceso que se corregirán, completarán datos vacíos, se eliminarán entradas duplicadas y estandarizarán los datos según codificación, formato o unidades de medida [9]. De esta manera se evitarán los errores en la realización de cálculos o uso de modelos con la base de datos, de manera de obtener resultados que representen correctamente la realidad de lo estudiado.

Para el desarrollo de este trabajo de título se utilizará el software “R Studio”, con el cual se podrán trabajar los datos, permitiendo eliminar aquellas entradas de valores mal ingresados y facilitar el trabajo de *outliers*. Estos últimos son datos que presentan valores atípicos en alguna variable.

6.2 Modelos de propensión

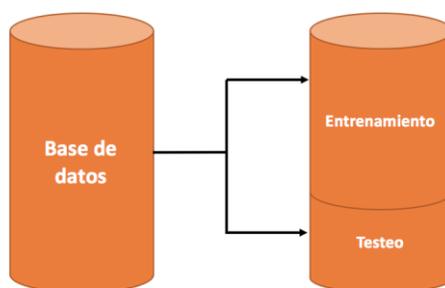
Con los datos seleccionados, pre-procesados y transformados, se generará un modelo de propensión para cada ramo. Este modelo utilizará las características personales como variables independientes para predecir el comportamiento de la

variable dependiente que, en este caso, corresponde a la propensión de contratación de un producto en particular. Para el desarrollo de los modelos de propensión se utiliza el software “R Studio”, software de código abierto utilizado en estudios computacionales estadísticos.

Los modelos de propensión utilizan estadística para predecir alguna variable en particular, basándose en los datos entregados al modelo. Para ello se les suele entregar un porcentaje de datos designado para el entrenamiento de los modelos y otro para el testeo de este.

Un problema que surge del proceso anterior corresponde al “over fitting”, el cual ocurre cuando el modelo se sobre ajusta a los datos de entrenamiento. Si sucede esto, al probar el modelo con la data de testeo, se podrá observar que este no predice de manera correcta a la variable dependiente. Esto se debe a que el modelo asume que la base de testeo se comportará de manera idéntica a la de entrenamiento, lo cual ocurre a causa de la desproporción existente entre las dos bases.

Con el fin de evitar este problema, se utilizará un 70% para la base de datos para entrenamiento y un 30% para la de testeo, como se puede observar en la ilustración 4.



*Ilustración 4: Distribución base entrenamiento y testeo
Fuente: Elaboración propia*

Estos modelos predictivos se pueden separar en dos grandes grupos: supervisado y sin supervisión. Los primeros reciben un set de datos etiquetados como datos de entrenamiento y realizan predicciones en base a factores no observables, mientras que los segundos reciben datos sin etiquetar [4]. Esta etiqueta corresponde a especificarle al modelo que tipo de variable se le está entregando, siendo numérica, binaria, texto, entre otros.

En particular para este proyecto se utilizarán modelos supervisados dado que se le entregarán datos con etiquetas a los distintos tipos de variables que se tratarán. Además de la característica anterior, deben cumplir con la característica de ser de clasificación binaria, por lo que los modelos a utilizar son: “Logit Binario” y “Árbol de Decisión”.

6.2.1 Modelo Logit Binario

Con el fin de estimar la propensión de cada cliente a la contratación de algún producto, se utilizará un modelo Logit Binario, que entrega la probabilidad de contratación del producto en cuestión. Así, cada probabilidad se calcula de la siguiente manera:

$$E[y] = \frac{\exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}{1 + \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}$$

Donde el término dentro de la exponencial corresponde a la componente determinística asociada a la utilidad para cada uno de los clientes según sus n atributos. En ella cada β corresponde a los coeficientes de cada variable explicativa x_i [5].

6.2.2 Modelo Árbol de Decisión

El modelo de Árbol de Decisión corresponde a utilizar como base un árbol de clasificación en términos de los datos que se tienen para una variable dependiente. Así, se busca realizar un método de clasificación para determinar la propensión de cada cliente dependiendo de cada una de las características de éstos. De esta manera se van generando nodos de acuerdo a las características más significativas, dado que el modelo mismo elimina aquellas variables que no lo son [6].

El modelo determinará en base a las características del cliente en qué nodo final del Árbol de Decisión se encuentra. Es así como el modelo calcula la probabilidad de propensión de cada individuo según la posición final del árbol en la que se encuentre cada cliente.

6.3 Next Best Offer

Un método que busca generar mayor cantidad de ventas hacia sus clientes consiste en ofrecerle el mejor producto, el cual se ajusta a sus características. Este método se llama “Next Best Offer” y consiste en ofrecerle un único producto a cada cliente.

Lo anterior surge debido a la intención de ofrecer a cada cliente el producto más adecuado según sus características. Además, dado que si se ofrecen todos los productos disponibles a cada cliente, sin importar si se adecúa o no a sus necesidades, se podría generar un rechazo por parte de éstos. Es por esta razón que mediante este método se busca limitar la cantidad de productos ofertados a cada cliente y elegir el más apropiado para estos.

En particular para este proyecto de título, el modelo de “Next Best Offer” busca ofrecerle a cada cliente aquel producto por el que estaría más propenso a contratar. Esto se realizará en base a la información que se tiene del cliente, utilizando el resultado de un modelo de propensión para determinar el producto a elegir por un cliente [8].

De esta manera se evitar la sobreexposición de ofertas a los clientes eligiendo un único producto, que a su vez maximiza la propensión de compra.

6.4 Métricas de comparación

Posterior a la realización de los distintos modelos de propensión mediante la base de entrenamiento, es necesario realizar una comparación entre estos con el fin de obtener el modelo que más se adapte a los requerimientos del problema. Para ello se necesitará crear una matriz de confusión, donde se determinarán los verdaderos positivos (TP), los verdaderos negativos (TN), los falsos negativos (FN) y los falsos positivos (FP). Para esto, se evaluarán mediante la base de testeo los distintos modelos,

generando así, los valores previamente descritos, tal como se muestra en la ilustración 5 [9].

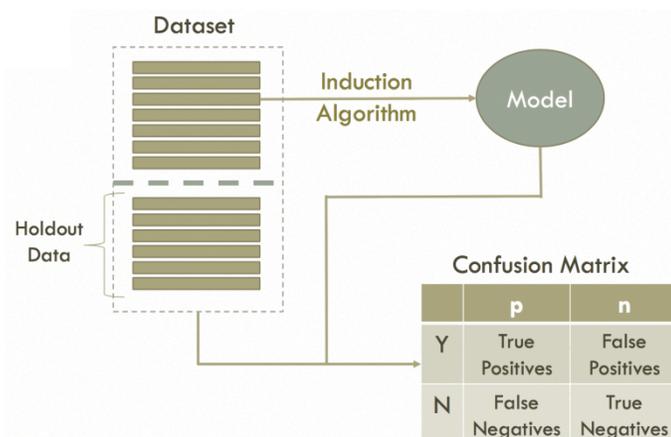


Ilustración 5: Gráfico de método de creación de Matriz de Confusión
Fuente: Apuntes de Big Data, Capítulo 2, Lecture 11: A key analytical framework.

A partir de esta matriz de confusión se pueden obtener métricas como la “Presición”, “Accuracy” y “Sensitivity” [9].

$$\begin{aligned}
 \text{Presición} &= \frac{TP}{TP + FP} \\
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\
 \text{Sensitivity} &= \frac{TP}{TP + FN}
 \end{aligned}$$

La precisión permite ver la dispersión de los datos obtenidos, mientras que el accuracy mide a qué tan efectivo es mi modelo para predecir mis casos reales.

Finalmente, la sensibilidad mide la proporción de verdaderos positivos a los cuales el modelo es capaz de identificar de manera correcta. Es un indicador que permite controlar la cantidad de falsos negativos que el modelo genera.

Otra de manera para comparar el desempeño del modelo, consiste en ordenar de mayor a menor la data según la probabilidad calculada por el modelo. Luego se selecciona el mejor quintil de los valores y se calcula el porcentaje de casos positivos que captura el modelo.

La gran diferencia que radica en estos dos métodos de comparación corresponde a que, en la matriz de confusión, se debe fijar un porcentaje mínimo de la probabilidad calculada por sobre el cual asumir como casos. Lo anterior corresponde a discretizar la probabilidad generada a partir de este valor. Mientras que para el segundo método, este valor queda libre y se calcula el desempeño real del modelo sin fijar un valor de probabilidad.

Otra diferencia del segundo método, es que el uso de matriz de confusión es muy útil frente a bases de datos más balanceadas. En este caso la base no se encuentra balanceada, por lo que las métricas de desempeño pueden resultar engañosas.

Así, cuando una base se encuentra desbalanceada con mayor cantidad de casos negativos, el modelo asume más falsos negativos. En caso contrario, de estar balanceada con más casos positivos, el modelo tiende a asumir mayor cantidad de falsos positivos.

Mientras que, en la segunda metodología de comparación, sólo se mide el desempeño real del modelo según la probabilidad calculada de manera continua y no discreta.

7. METODOLOGÍA

7.1 Descripción de fuentes de datos

Para poder entender el trabajo de datos a realizar más adelante, es necesario tener conocimiento de los datos que se tienen. Es por esto que se realizará un proceso de descripción, donde se entrará a explicar en qué consisten los datos que la compañía tiene de acuerdo a las distintas fuentes de información y que variables serán rescatadas para el desarrollo de esta memoria.

7.2 Tratamiento de datos

Dado que gran parte de los puntos de esta memoria dependen de los datos utilizados, es importante realizar un proceso de selección, pre-procesamiento, transformación y descripción de los mismos. Es en base a los datos que se desarrollarán los modelos de propensión, razón por la cual el trabajo realizado en ellos debe ser muy consistente para que represente realmente lo que se busca obtener.

La selección de datos consiste en la extracción de los datos a utilizar desde las bases de datos de la compañía y la elección de aquellos datos que se utilizarán para los modelos de propensión.

El pre-procesamiento de los datos pretende limpiar la data de aquellos mal ingresados, *missing values* y *outliers*. Esto permitirá tener una mejor calidad de los datos con los cuales trabajar, dado que también este proceso contempla el trato de registros mal ingresados.

El último paso de trabajo de datos en si corresponde a la transformación de estos mismos. Este proceso contempla la categorización de variables que permitirá manejar de mejor manera los datos *outliers* que pueden generar problemas con los modelos. Además, la categorización permite tener una muestra más homogénea para ciertas variables como es el caso de los montos de gastos de los clientes.

7.3 Desarrollo de modelos de propensión

El desarrollo de los modelos de propensión será de manera independiente tanto para cada uno de los tipos de modelos como para cada producto. De esta manera se contará con 2 modelos distintos para cada rama, donde cada uno de ellos determinará la propensión de dicho cliente en la contratación del producto en sí.

En esta parte se desarrollarán los dos modelos elegidos, comenzando por el modelo Logit Binario y luego el Árbol de Decisión. En cada uno de esos puntos se

detallará tanto la fórmula utilizada como el desempeño de cada uno de ellos. Además, se especificarán aquellas variables más significativas para el modelo y aquellos perfiles de clientes más propensos a contratar según los resultados obtenidos para cada modelo.

7.4 Evaluación

7.4.1 Desempeño modelos de propensión

Otra de las partes fundamentales de esta memoria consiste en la manera en la cual se seleccionará aquel modelo a utilizar. En esta parte se contrastará el desempeño de los modelos obtenidos para cada uno de los productos según las métricas detalladas previamente.

Lo anterior permitirá determinar el modelo de mejor desempeño. Sin embargo, los resultados obtenidos por estas métricas deberán ser contrastados con los requerimientos del negocio, con el fin de obtener un modelo que se ajuste a las necesidades de la compañía.

7.5 Propuestas de acción

Luego de la comparación de los modelos, se presentarán las propuestas de acción para la compañía que hagan uso de los resultados obtenidos por los modelos, con el fin de determinar la “Next Best Offer”. De esta manera se espera entregar toda la información y las herramientas necesarias para que la empresa pueda tomar la decisión que más se alinee al plan de marketing y ventas de la compañía.

7.6 Propuesta de diseño experimental

Se planteará un diseño experimental que permita determinar la efectividad del modelo “Next Best Offer” propuesto. Este permitirá a la empresa determinar si efectivamente este modelo es el más adecuado para la compañía.

8. DESARROLLO

8.1 Descripción de los datos

Seguros Falabella cuenta con tres fuentes de datos que se utilizarán para el desarrollo de esta memoria. Estas son tenencia de seguros, gastos realizados con la tarjeta CMR y datos demográficos de los clientes. En anexos 5 se encuentra un esquema entidad relación simplificado de la base de datos utilizada.

8.1.1 Tenencia de seguros

Esta primera tabla contiene la tenencia de seguros por cada uno de los clientes de la compañía. Aquí se encuentra la información asociada a la tenencia tanto inicial como del periodo anterior. Esta base de datos cuenta con más de 3 millones de registros, y sus entradas son binarias para cada tenencia de seguro.

Resulta interesante calcular la proporción en la tenencia de los seguros principales, las cuales son:

- 45% son seguros automotrices.
- 18% son seguros de vida
- 15% son seguros de vida con bonificación
- 9% son seguros de hogar
- 7% son seguros de salud
- 6% son seguros transaccionales

Se puede observar que los seguros automotrices representan el mayor porcentaje de la tenencia de seguros principales. Con esto se puede observar la importancia de este producto para la compañía.

Además de la tenencia de seguros, existen dos etiquetas asociadas a la tenencia de a lo menos un seguro principal y la otra de a lo menos un seguro no principal. Esta etiqueta será útil para el ver el impacto de la tenencia de algún producto, ya sea principal o no, en el modelo de propensión.

8.1.2 Gastos tarjeta CMR

La segunda base de datos a utilizar contiene los montos de los gastos realizados por cada cliente en el último año, segmentado por rubros. Dentro de estos se encuentran los rubros automotriz, combustible, comunicación, educación, empresa, entretención, farmacias, mejoramiento del hogar, pago web, recaudación, restaurant, salud, supermercados, tiendas por departamento, transporte, viajes viviendas y otros rubros. En Anexos Figura 3 se encuentra un glosario con ejemplo de los gastos pertenecientes a cada rubro.

Además de lo anterior, la base de datos cuenta con marcas de segmento de valor de cliente, segmento uso de la tarjeta CMR, la cantidad de meses en los que se usó la tarjeta CMR en el último año, qué tan seguido compra en Falabella, la cantidad de visitas a Falabella en el último año, cantidad de transacciones realizadas con la tarjeta CMR y el gasto total con la tarjeta CMR en el último año. Los datos anteriores corresponden a datos internos de CMR Falabella que permiten caracterizar a un cliente, entregando una mejor idea de cómo es el cliente según su uso de CMR.

Esta segunda base de datos, que contiene los datos presentados, cuenta con 2,9 millones de registros únicos para cada cliente. Cabe destacar que sólo aquellos clientes con la tarjeta CMR cuentan con los datos de esta base, por lo que aquellos clientes sin la tarjeta no poseen esta información.

8.1.3 Datos demográficos

Esta última tabla cuenta con los datos demográficos de cada cliente, base de datos que contiene 3,2 millones de registros. En ella se tiene la información del rango de edad, sexo, grupo socioeconómico, estado civil, tipo de habitación donde vive, actividad laboral y si es empleado vigente del holding.

Además de lo anterior, de esta tabla se puede obtener una marca de tenencia de tarjeta CMR, la cual se obtiene de analizar el estado actual de la cuenta. Así, si este estado es “cuenta abierta” se asocia a la tenencia de la tarjeta.

Aquellos clientes que tienen la tarjeta cuentan con información extra en esta misma tabla. Esta información consiste en la descripción de la cuenta, según su tipo de morosidad; la antigüedad de la cuenta; el tipo de cliente, pudiendo ser normal, premium o elite; el tipo de tarjeta con la que cuenta el cliente; el tipo de envío de la facturación de la tarjeta; el rango del cupo; y finalmente el hábito de pago.

Por otro lado, se tienen datos de la tenencia de vehículos de los clientes, donde se cuenta con la cantidad de vehículos que tiene dicho cliente y el año de fabricación de éste.

8.2 Tratamiento de datos

Como se mencionó en la sección anterior, es necesario realizar un trabajo de los datos entregados por la compañía. Si bien estos se encuentran estandarizados y sin errores para la mayor parte de los campos, existen ciertas inconsistencias que será necesario trabajar para no tener errores en los datos que se utilizarán para los modelos.

8.2.1 Selección

La selección de los datos se realizó a través de tres fuentes de datos de la compañía como se mencionó previamente. Para realizar el cruce entre los datos se comenzó desde la tabla de tenencia, ya que esta es la base de clientes activos que tiene la empresa. Esa tabla se juntó, mediante el ID de cliente, con las otras 2 tablas. Así se obtuvo una gran base donde cada registro es único y contiene toda la información disponible por cliente. De esta manera, se generó una base de datos de 3,2 millones de registros.

Dentro de los datos existen registros asociados a la tenencia de los distintos seguros para el periodo anterior y el actual. Estos registros cumplen con una de las siguientes cuatro condiciones para cada uno de los productos:

1. Aquellos que en el periodo anterior ni en el actual tienen el producto, a los que llamaremos “0-0”.
2. Aquellos que en el periodo anterior no tienen el producto y en el periodo actual si, a los que llamaremos “0-1”.
3. Aquellos que tanto en el periodo anterior como en el periodo actual tienen el producto, llamados “1-1”.
4. Finalmente aquellos que en el periodo anterior si tienen el producto y en el actual no, llamados “1-0”.

Dada las condiciones anteriores, el modelo propuesto sólo debe considerar aquellos datos que cumplen la condición 1 o 2. Ambos datos corresponden a aquellos clientes que se mantuvieron en la compañía. Por un lado los primeros no contrataron producto en el periodo siguiente, mientras que los segundos si lo hicieron.

Es importante destacar lo anterior, dado que el modelo pretende predecir la venta cross-selling de productos únicamente para aquellos clientes que no tenían el producto inicial. Por esta razón, no es pertinente utilizar a aquellos clientes que cumplen la condición 3 o 4, dado que no es razonable entrenar el modelo con aquellos que mantuvieron su producto o los clientes que se ya no poseen dicho producto.

8.2.2 Pre-procesamiento

Al analizar la data, se pudo observar que la cantidad de datos mal ingresados y missing values no era significativa dado que la cantidad de éstos fue de un 0,015% de la base de datos. El primer problema que se encontró en los datos, es que existen 4 registros de clientes con CMR cuya facturación se encuentra sin información, lo cual no puede ser dado que todo cliente de CMR debe tener una manera de hacer llegar la facturación, por lo que se eliminaron esos registros.

Otros dos problemas relacionados con los datos vacíos es que existen 5 registros que no tienen información asociada a su género y 8 registros que no tienen rango de edad, a pesar de ser clientes CMR. Se prosiguió a eliminar esos registros dado que todo registro de CMR debe tener el rango de la edad y el género del cliente.

Finalmente, el trato de datos mal ingresados realizado mostró que existen 298 clientes cuyo rango de edad está entre los 1 a 17 años, lo cual no se puede dar debido a que no se venden productos a menores de 18 años.

El trato de outliers se dejó para el proceso de transformación de los datos, dado que estos datos serán categorizados, con lo cual el problema asociado al uso de dichos outliers no se ve afectado.

8.2.3 Transformación

El proceso de transformación de datos permite evitar que existan outliers o que existan un conjunto muy pequeño de datos que presenten una cierta característica. Así, los datos de gastos de la tarjeta CMR serán categorizados para evitar *outliers* asociados a aquellos clientes que tienen gastos muy elevados en comparación al gasto normal.

En particular, para este trabajo de memoria se utilizaron únicamente variables binarias. Lo anterior se debe a que las variables categóricas se pueden representar como un conjunto de variables binarias. Así, la no inclusión de variables continuas permite comparar los coeficientes asociados al modelo entre ellos. Esto permite un mejor análisis de las variables más impactantes para cada modelo.

Por esta razón, cada gasto se categorizó dividiendo el rango de datos en tres segmentos de igual tamaño, sin considerar los valores iguales a cero. Así, se llamó a cada categoría: “gasto bajo la media”, “gasto en la media” y “gasto por sobre la media”, mientras que a aquellos datos donde el gasto era cero, se les llamó “no gasta”.

Se prosiguió a discretizar la variable asociada a la cantidad de transacciones realizadas con la tarjeta CMR. Para ello se utilizó la categoría de “0 transacciones”, “1 transacción”, “2 transacciones” y “3 o más transacciones”. El mismo proceso fue realizado con la variable de recency, que corresponde a qué tan reciente un cliente compró en Falabella. Los valores definidos para asignar a las distintas categorías son:

- No informado: aquellos registros de los que no se tiene información.
- No compra: clientes que tienen un recency de cero.
- Reciente: aquellos clientes que tienen un recency menor o igual a 12 meses.
- No reciente: aquellos clientes que tienen un recency mayor a 12 meses.

Finalmente, se categorizó la variable de visitas a Falabella, que contiene la cantidad de veces que se ha comprado en alguna de las tiendas en los últimos 12 meses. Para esta categorización nuevamente se eligieron 5 categorías distintas:

- No informado: aquellos registros de los que no se tiene información.
- No visita: no ha visitado una sucursal en el último año.
- Baja visita: aquellos clientes que han visitado una sucursal menos de 10 veces en el año.
- Media visita: aquellos clientes que han visitado una sucursal más de 10 pero menos de 20 veces en un año.
- Alta visita: Aquellos clientes que van más de 20 veces en un año.

Con esto, todas las variables utilizadas se encuentran discretizadas y no existen outliers en la base de datos a utilizar.

8.2.4 Descripción de datos

Luego del proceso de selección, pre-procesamiento y transformación de datos se redujo el número de registros en 315 registros, que corresponde a un 0,01% de los datos. Adicional a esto, se seleccionó una muestra de un 40% de los datos con el fin de poder realizar los modelos con una cantidad de datos lo suficientemente robusta y que a su vez permitieran una ejecución más rápida de los modelos debido a la menor cantidad de registros, dejando un total de 1,2 millones de registros.

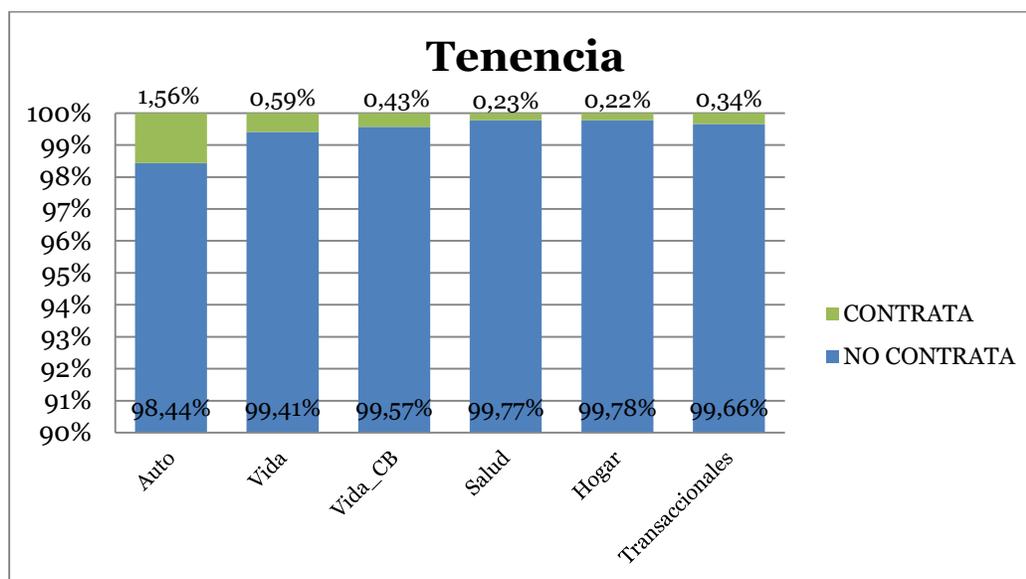


Ilustración 6: Distribución de tenencia según ramo

Fuente: Elaboración propia

Se comenzará con la descripción de los datos de tenencia, dado que estos son los valores principales para el modelo. Se puede observar en los datos que la cantidad de casos positivos para el modelo es muy baja, sobre todo para los ramos de salud y hogar, tal como se muestra en la ilustración 6.

Así se puede observar que será necesario balancear la data de entrenamiento para poder aumentar la cantidad de casos positivos en el modelo. El método a utilizar para el balanceo será un método de *over-sampling* que consiste en añadir casos positivos a la base. Esto se realizará hasta que el número de casos positivos sea mayor a un 50% de la base de entrenamiento, el cual se realizará duplicando los registros positivos con el fin de no perder las proporciones de las características de los casos positivos.

Para la descripción de los datos que vienen a continuación, es importante distinguir aquellos clientes con y sin tarjeta CMR, dado que la tenencia de CMR indica la existencia de más datos asociados a cada cliente.

Analizando los rangos de edad, se puede observar en la ilustración 7 que la distribución de edades para aquellos clientes con y sin CMR, es similar tanto en porcentaje como en distribución.

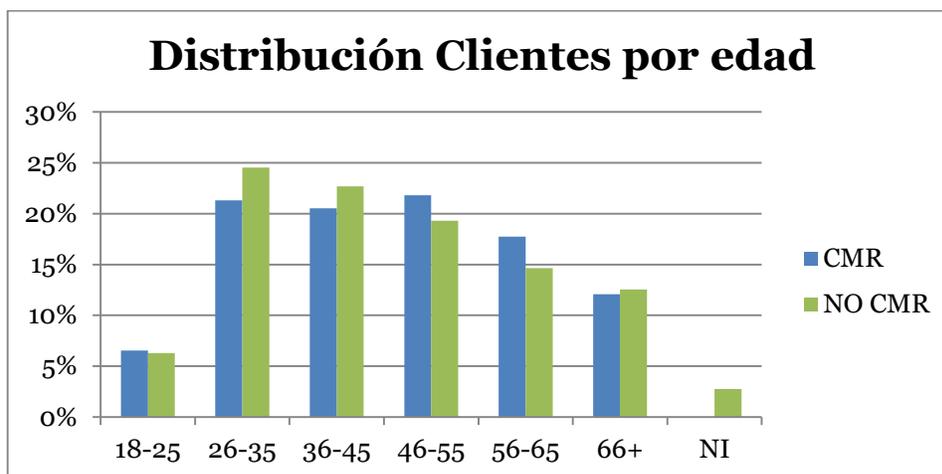
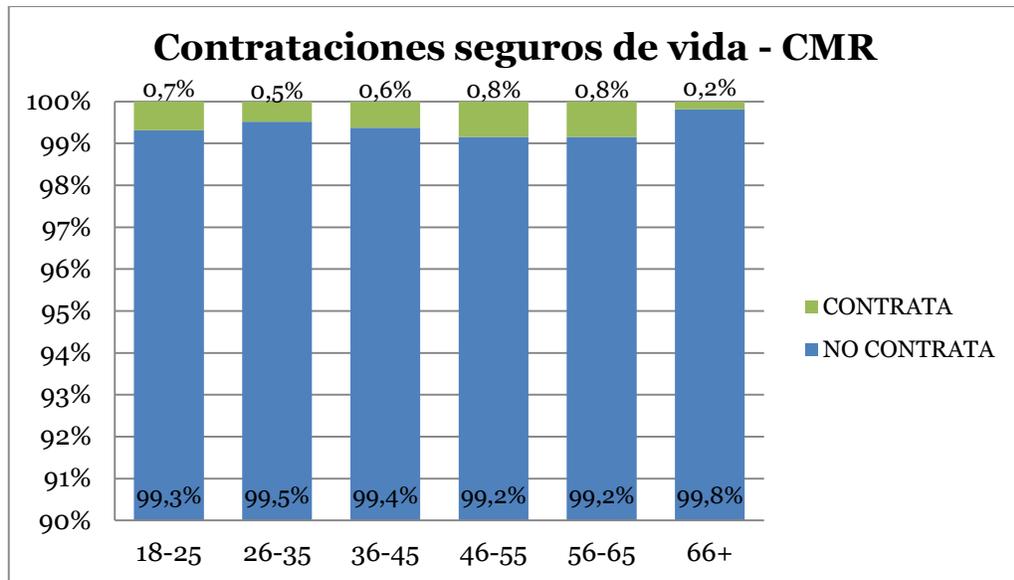


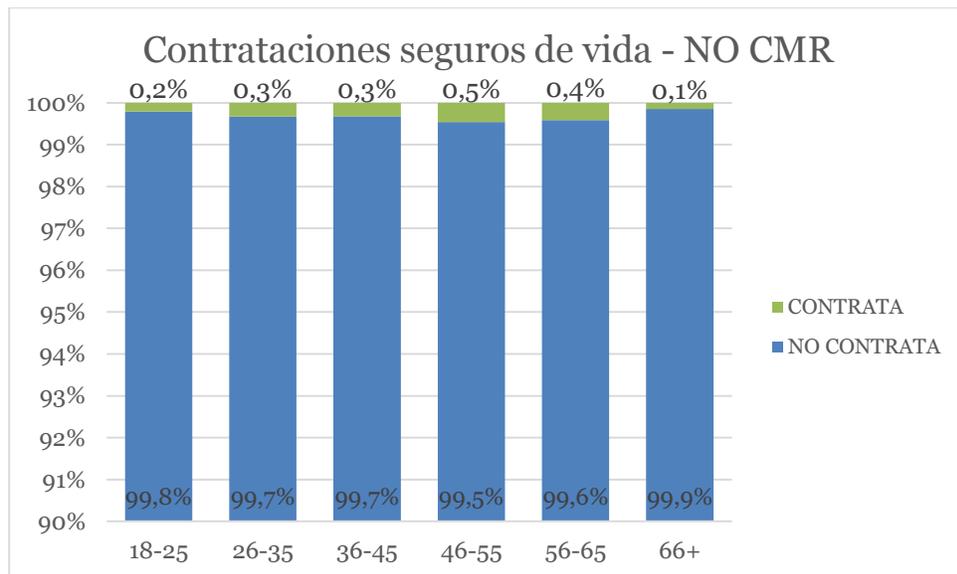
Ilustración 7: Distribución de clientes por edades
Fuente: Elaboración propia

Es interesante analizar la contratación de seguros de vida asociados a la edad. Para el siguiente análisis se incluyeron tanto seguros de vida como seguro de vida con bonificación y se separó según tenencia de CMR. En la ilustración 8 se puede observar que la mayor cantidad de contrataciones de seguros de vida se genera en los segmentos entre los 46 y 65 años para clientes con la tarjeta CMR.



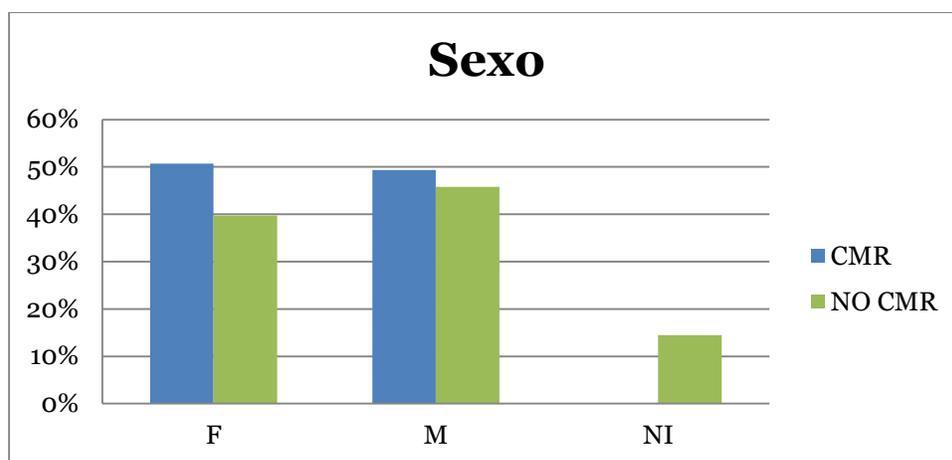
*Ilustración 8: Porcentaje de contrataciones de seguros de vida según edad clientes con CMR
Fuente: Elaboración propia*

De igual manera se realizó el mismo análisis para clientes sin la tarjeta. En estos clientes se puede observar que es mayor el porcentaje de contratación en el periodo entre los 46-55 años, como se observa en la ilustración 9.



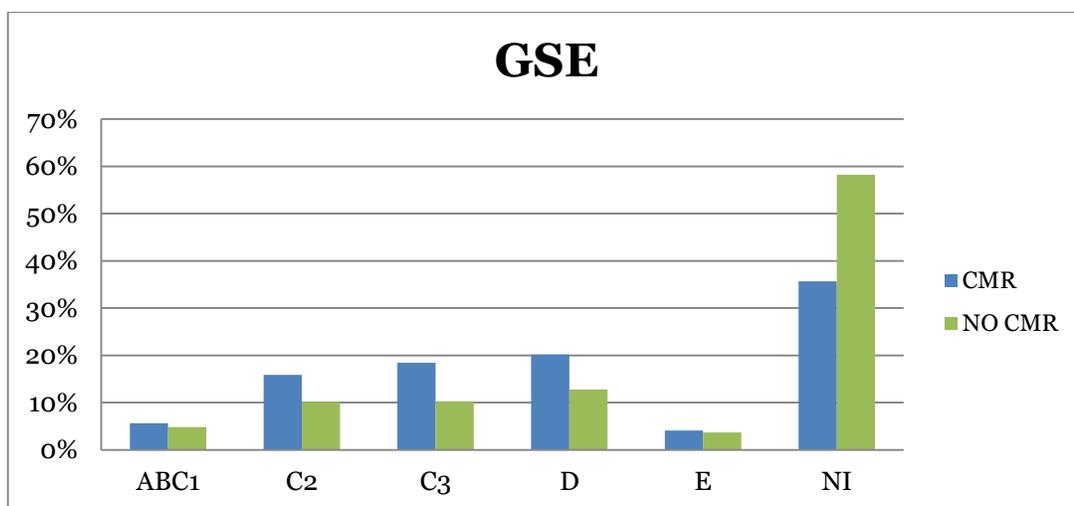
*Ilustración 9: Porcentaje de contrataciones de seguros de vida según edad clientes no CMR
Fuente: Elaboración propia*

Por otro lado, al analizar el género de los clientes en la base de datos, se puede observar en la ilustración 10 que la proporción de hombres y mujeres es prácticamente el 50% en aquellos clientes con la tarjeta CMR, mientras que en aquellos clientes sin la tarjeta la proporción de mujeres es un poco menor.



*Ilustración 10: Distribución de clientes por género
Fuente: Elaboración propia*

Se puede observar en la ilustración 11 que el grupo socioeconómico de mayor porcentaje, sin considerar a aquellos clientes sin información, corresponde al grupo D tanto para clientes con y sin tarjeta CMR. Se puede observar que la cantidad de datos no informados es alta, sin embargo, esta marca entrega información relevante sobre el nivel económico de cada cliente.



*Ilustración 11: Distribución de clientes por GSE
Fuente: Elaboración propia*

También se tiene información respecto al estado civil de los clientes. Se puede apreciar en la ilustración 12 que la mayor cantidad de clientes, sin contar a aquellos clientes sin información, son casados o solteros.

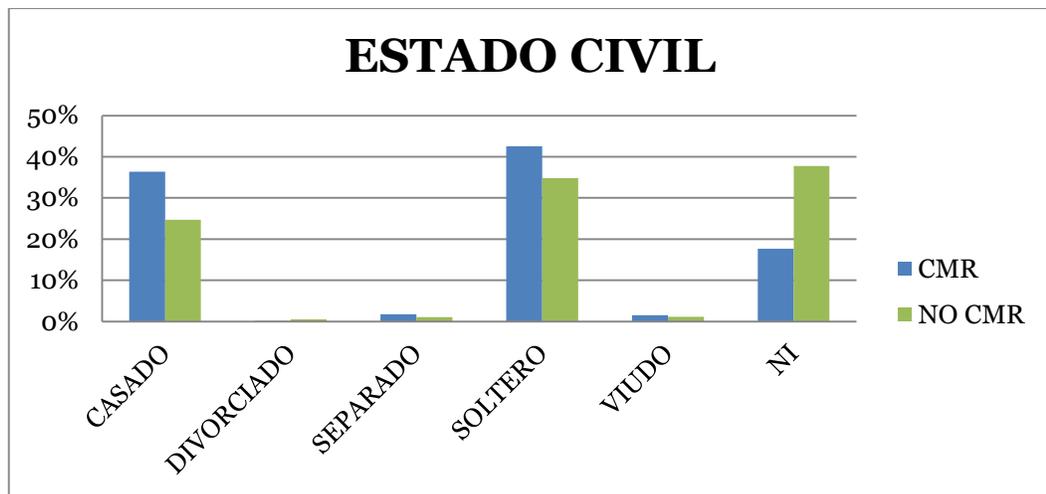


Ilustración 12: Distribución de clientes por Estado Civil
Fuente: Elaboración Propia

Otra característica demográfica corresponde al tipo de habitación en la cual vive el cliente. Se puede observar, en la ilustración 13, la distribución de los clientes, donde la mayor proporción tanto de clientes con y sin CMR corresponden a aquellos que tienen casa propia o viven con los padres.

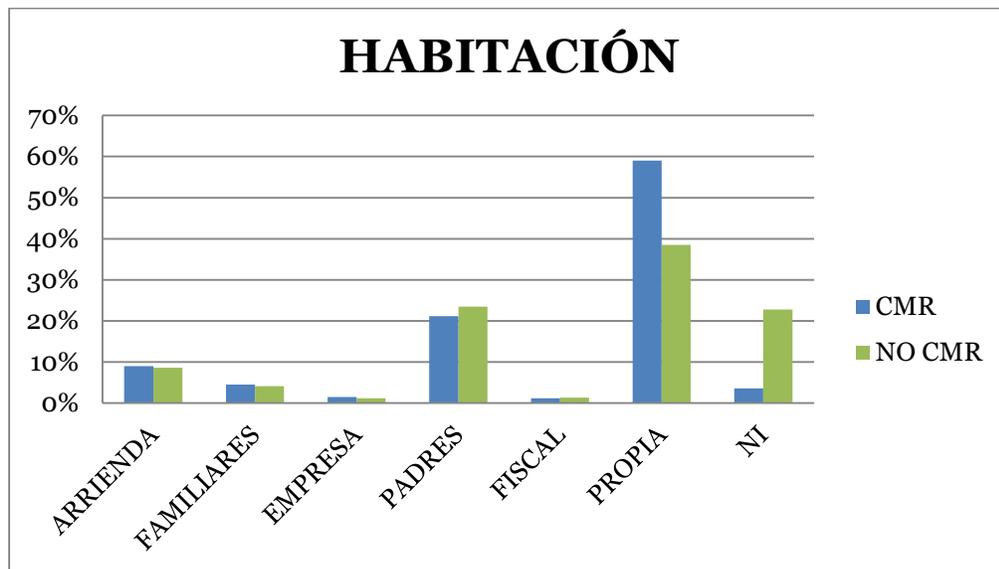
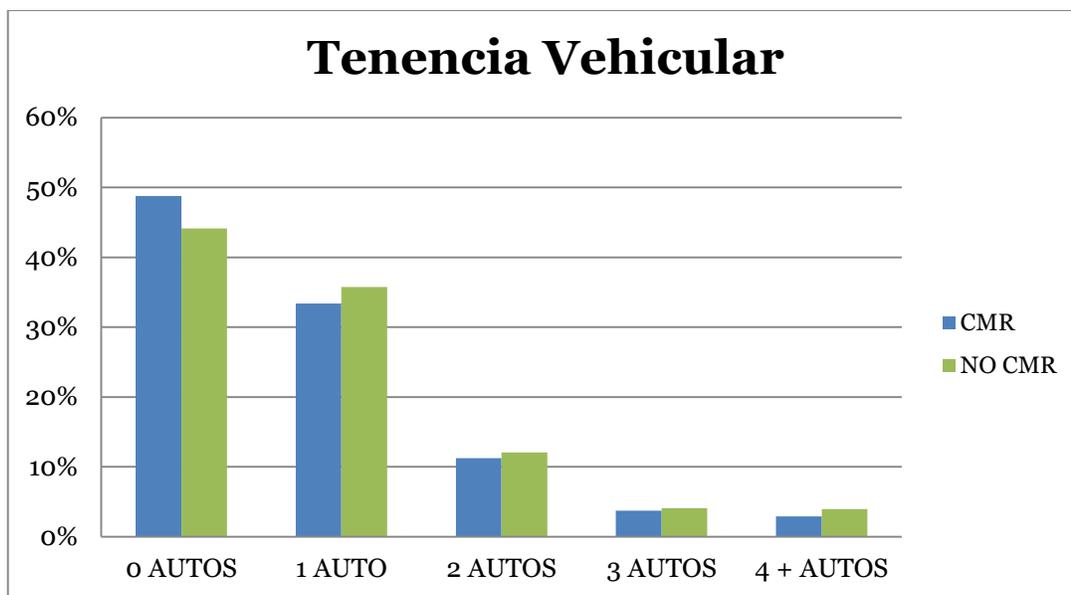


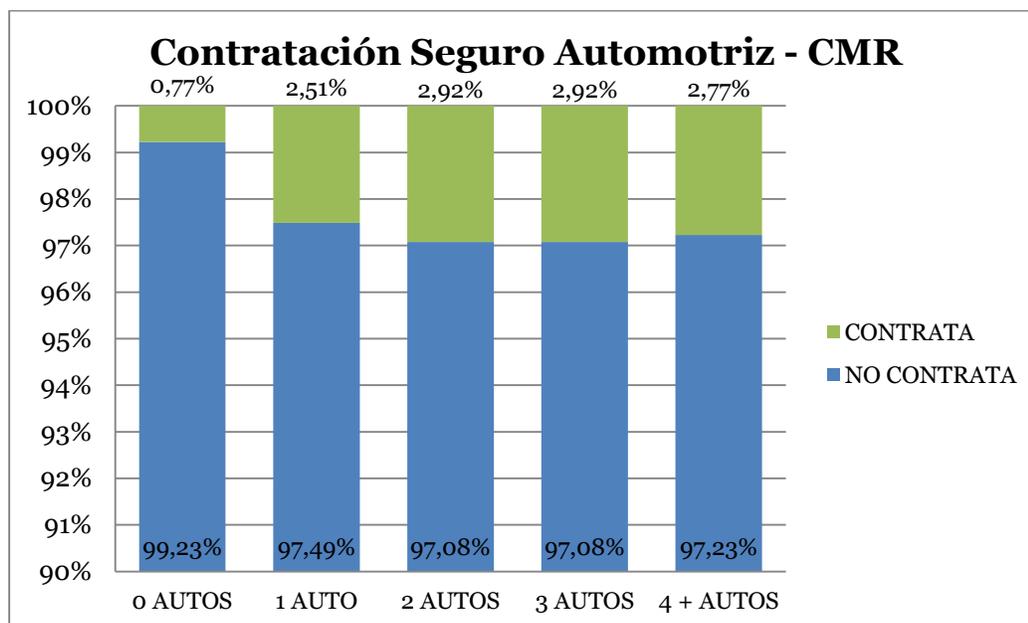
Ilustración 13: Distribución de clientes por habitación
Fuente: Elaboración Propia

Es útil analizar la cantidad de vehículos que tienen los clientes, separando nuevamente aquellos con y sin CMR. Se puede observar en la ilustración 14 la distribución de clientes según la cantidad de vehículos que poseen.



*Ilustración 14: Distribución tenencia vehicular
Fuente: Elaboración propia*

Es interesante analizar a aquellos clientes que, según la cantidad vehicular que tienen, contratan un seguro vehicular en cada periodo. Se puede observar de la ilustración 15 que, a mayor cantidad de vehículos, más probable es la contratación de un seguro automotriz.



*Ilustración 15: Contratación de seguros automotriz según cantidad de vehículos para clientes con CMR
Fuente: Elaboración propia*

De igual manera se analizó a los clientes sin CMR, obteniendo el mismo segmento de mayor compra y un comportamiento similar a aquellos clientes con CMR, como se muestra en la ilustración 16.

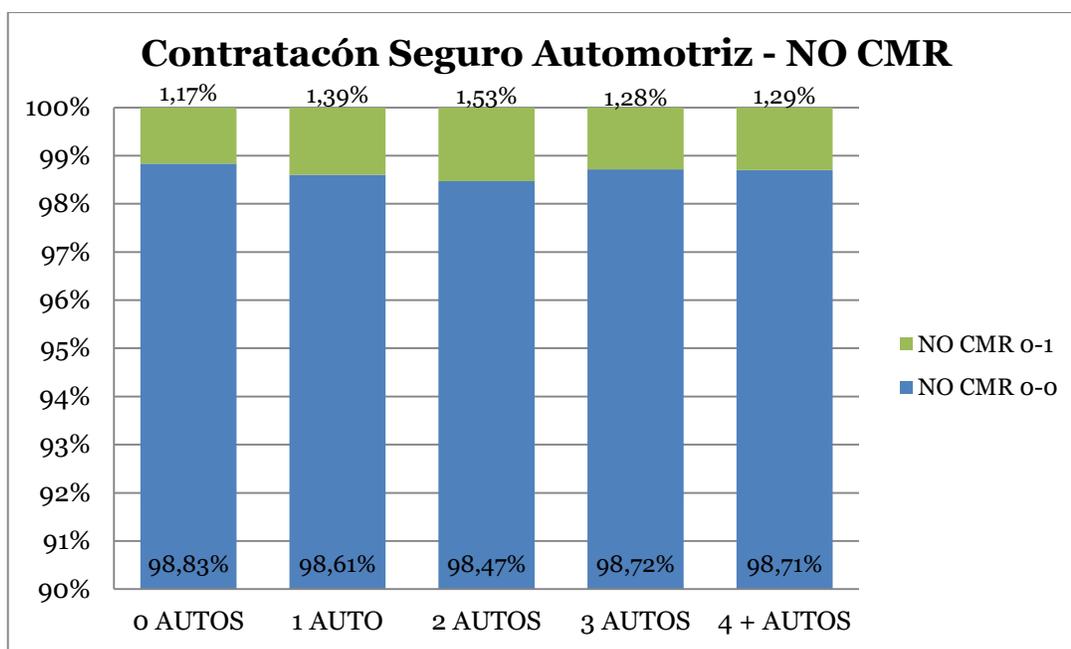


Ilustración 16: Contratación de seguros automotriz según cantidad de vehículos para clientes con CMR
Fuente: Elaboración propia

El último dato demográfico corresponde a si el cliente es empleado actual del holding. Lo anterior es importante mencionarlo dado que puede existir una influencia de contratación de seguros debido a la confianza que le podría brindar a algún cliente, empleado del holding, el contratar otro seguro dentro de la compañía. La cantidad de clientes con CMR empleados de la compañía es un 1% del total de clientes, mientras que los sin la tarjeta es un 0,5%.

Los siguientes datos corresponden a aquellos clientes que tienen tarjeta CMR, dado que para aquellos que no tienen, no cuentan con estos datos. El primero de ellos corresponde a los gastos realizados con la tarjeta según los rubros. Estos datos se categorizaron según los terciles al cual pertenecían. De esta manera se generaron 3 categorías de acuerdo al nivel de gasto.

Para el nivel de mayor gasto se llamó “gasto sobre la media”, aquellos en el segundo tercil se llamó “gastos en la media” y aquellos con el menor gasto se llamó “gastos bajo la media”. Además de lo anterior, se categorizó de manera independiente a aquellos clientes que no gastaron, llamándolos “no gasta”.

La categoría “no gasta” es aquella que tiene la mayor cantidad de clientes, donde para el gasto de tiendas por departamento posee un 50% de los datos, hasta llegar incluso a un 97% en gastos en los rubros tanto de educación como empresa.

Dejando de lado la categoría asociada al no gasto, podemos observar en la ilustración 17 que los rubros donde más se gasta con la tarjeta CMR corresponden a los gastos realizados en supermercados y en tiendas por departamento, lo cual se puede deber a las ofertas que generan tanto el supermercado Tottus como Falabella Retail utilizando la tarjeta CMR.

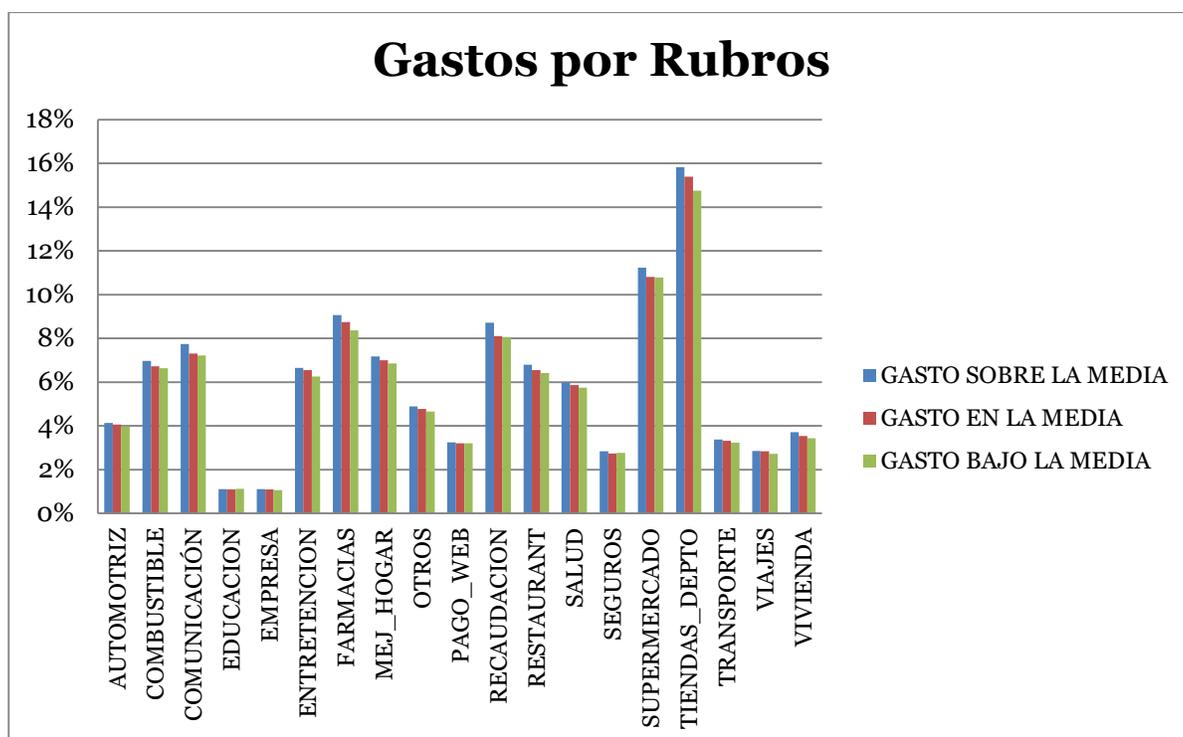


Ilustración 17: Distribución de clientes según monto de gastos por rubros
Fuente: Elaboración propia

8.3 Desarrollo de modelos de propensión

Habiendo realizado la selección, pre-procesamiento y transformación de los datos, se desarrollarán los modelos de propensión para cada uno de los canales. Se realizarán modelos independientes para cada uno de los productos que se tienen, tanto para clientes con y sin tarjeta CMR. Estos últimos tienen que estar diferenciados debido a que los primeros cuentan con más datos que los segundos, por lo que para cada uno de los 6 productos se realizarán dos modelos.

8.3.1 Modelo Logit Binario

Para el desarrollo del modelo Logit Binario se utilizó un proceso llamado Backward elimination, que consiste en añadir todas las variables al modelo e ir eliminando aquellas variables no significativas del modelo. En este caso en particular, se utilizó un filtro de una significancia al 90%, por lo que aquellas variables que tienen una significancia mayor o igual al 90% se mantuvieron, mientras que aquellas menores se eliminaron.

Como cada producto tiene modelos independientes, las variables explicativas usadas en cada uno de ellos variarán según el ramo y la tenencia de CMR. Sin embargo, para los modelos sin la tarjeta del holding, las variables independientes son siempre las mismas dado que, para todos los modelos, estas variables son significativas.

Así, la variable dependiente y_i asociada a la contratación del producto i queda determinada por una ecuación lineal, donde el coeficiente β_0 se pueden interpretar como la aversión al riesgo innata de los clientes. En cambio, los β_j capturan el impacto de la

variable j en la contratación del producto i . De esta manera, la ecuación para la variable independiente es como sigue:

$$y_i = \beta_0 + \beta_1 OTROS_ANT + \beta_2 PRINCIPAL_ANT + \beta_3 RANGO_EDAD + \beta_4 SEXO + \beta_5 GSE + \beta_6 ESTADO_CIVIL + \beta_7 HABITACION + \beta_8 ACTIVIDAD + \beta_9 EMPLEADO_VIGENTE$$

Sin embargo, para los modelos asociados a los clientes con CMR, las variables significativas van cambiando según el producto al cuál se le calcula el modelo de propensión. Esto tiene sentido dado que, para algún producto en particular, es más importante una variable que otra.

Es importante recalcar que todas las ecuaciones son lineales y de variables binarias o categóricas, al igual que las de clientes sin tarjeta. Las variables utilizadas en cada modelo se pueden observar en la tabla de anexo 4.

Dado que los coeficientes de los modelos provienen de variables binarias, se puede obtener aquellas de mayor impacto para cada uno de los productos. Con ello se puede verificar que los modelos generados tienen relación con el negocio dado que dichas variables hacen sentido.

En particular se verificarán los coeficientes de los modelos con la tarjeta CMR dado que la explicación de los modelos sin esta resulta trivial. Las tablas completas con los coeficientes obtenidos se encuentran en las tablas de anexo D.

Como se observa en la ilustración 18, correspondiente a un extracto de los 5 coeficientes de mayor impacto en la propensión. En ella se encuentra la variable, su valor asociado, el coeficiente entregado por el modelo y el porcentaje de la base de datos que tiene esa variable. Se observa que la cantidad de vehículos y el año de este son una de las características más relevantes. Una posible interpretación de lo anterior puede ser que un cliente con auto nuevo tiende a contratar un seguro en caso de accidentes por tener aversión a este riesgo.

De igual manera, la cantidad de vehículos es una característica de alto impacto que se esperaba obtener según la lógica, cuyo valor más importante corresponde a la tenencia de 2 vehículos. Lo anterior se puede interpretar dado que, mientras más vehículos, mayor posibilidad de que un cliente tenga a lo menos un seguro de este tipo.

VARIABLE	VALOR	COEF	PORCENTAJE
HABITO_PAGO	PÉSIMO	1,83	0,009 %
ANIO_VEHIC	NUEVOS	1,46	4,91 %
N_VEHICULOS	2 AUTOS	0,87	11,25 %
RANGO_CUPO_BI	1.250.0001 - 2.000.000	0,31	0,14 %
ACTIVIDAD	OFICIAL FF.AA.	0,29	0,31 %

*Ilustración 18: Extracto perfil más propenso modelo auto con CMR
Fuente: Elaboración propia*

Por otra parte, la variable de hábito pago pésimo es una variable poco esperada según la razón lógica del negocio. Sin embargo, lo anterior hace sentido dado que, en

promedio, lo seguros automotrices son los de mayor costo para los clientes, por lo que el mal hábito de pago puede deberse a los altos montos gastados con su tarjeta CMR. Lo anterior cobra aún más sentido al observar que la variable asociada al rango de cupo de tarjeta es de lo más altos.

Finalmente la actividad oficial FF.AA se puede deber a los beneficios a los cuales los oficiales tienen acceso, los cuales hacen relación a la compra de vehículos con descuentos únicos. Esto genera impacto dado que, al tener descuentos en la compra de vehículos, es más probable que tengan vehículos nuevos.

Se puede observar que el porcentaje de participación de estas variables es bajo, razón por la cual aquellos clientes que cumplen dichas características tienen una alta propensión para el seguro automotriz.

En el caso de los seguros de vida y vida con bonificación, tienen resultados muy distintos entre sí. Como se muestra en la ilustración 19, para la propensión a un seguro de vida, dentro de las variables que más influyen se encuentran la actividad y características asociadas a la tarjeta CMR.

La actividad rentista considera a aquellos clientes cuyos ingresos dependen de las rentas generadas por propiedades. Esta actividad es riesgosa dado que no asegura una renta constante, razón por la cual puede resultar de interés contar con un seguro de vida para cubrir a los familiares en caso de cualquier fatalidad.

VARIABLE	VALOR	COEF	PORCENTAJE
ACTIVIDAD	RENTISTA	0,71	0,26 %
ANTIGÜEDAD_CM	1-2 ANIOS	0,68	9,28 %
MES_COMPRA_CM	MUY ALTO USO	0,62	40 %
TIPO_TARJETA	VISA PLATINUM	0,55	5,24 %
SEGMENTO_USO_CM	MEDIA FRECUENCIA	0,52	38,93 %

*Ilustración 19: Extracto perfil más propenso modelo vida con CMR
Fuente: Elaboración Propia*

Por otro lado, las características asociadas a la tarjeta CMR se pueden explicar debido a que estas generaron confianza en el holding por parte del cliente, dado que todas las variables en cuestión muestran que el cliente tiene antigüedad, usa de manera recurrente su tarjeta y además cuenta con una tarjeta de alto valor.

Al igual que para el modelo automotriz, algunos de las variables tienen un bajo porcentaje de clientes que tienen dicha característica, lo cual puede deberse a que dichas variables son muy significativas para la propensión del modelo. Podemos observar que el mes de compra y el segmento de uso de la CMR son variables de alto impacto, donde además un alto porcentaje de clientes tienen dicha característica. Con lo cual se puede ver el impacto del uso de la tarjeta en la propensión de seguros de vida.

Al analizar el perfil de clientes más propensos para el seguro de vida con bonificación, podemos observar en la ilustración 20, que un rango elevado de edad tiene un alto impacto en la propensión, además de un 17,45% de clientes con dicha

característica. Esto concuerda con la realidad dado que, a mayor edad, mayor probabilidad de sufrir algún accidente de salud.

Lo anterior, sumado a que los clientes más propensos no visitan sucursales del retail Falabella genera dos posibles interpretaciones. La primera corresponde a que la contratación de este tipo de seguros es generada en su mayor parte por call-center, lo cual concuerda con que los clientes no asisten a las sucursales. La segunda se podría deber a que el segmento con mayor edad son clientes menos activos, por lo que visitan menos las tiendas de retail. Esta variable también tiene un alto porcentaje de clientes con dicha característica, con un 16,42%

VARIABLE	VALOR	COEF	PORCENTAJE
VISITA_FAL	NO VISITA	3,78	16,42 %
HABITO_PAGO	MALO	0,70	0,83 %
RANGO_EDAD	56 - 65	0,61	17,45 %
TIPO_CLIENTE	CLIENTE PREMIUM	0,58	11,47 %
ESTADO_CIVIL	VIUDO	0,49	1,56 %

*Ilustración 20: Extracto perfil más propenso modelo vida con bonificación con CMR
Fuente: Elaboración propia*

Tal como ocurrió con los seguros automotrices, el hábito de pago tiene relación con el valor de los seguros de vida con bonificación. Éstos ocupan el segundo lugar como los seguros de mayor precio para el cliente, los cuales son un 50% en promedio más barato que los seguros automotrices.

De igual manera que en seguros de vida, que el tipo de cliente sea Premium puede ser interpretado como la confianza que tiene el cliente frente al holding.

Finalmente, un acontecimiento como el fallecimiento de un conyugue, puede incentivar la contratación de un seguro de vida, razón por la cual un estado civil viudo tiene sentido en este modelo.

Para el caso de salud, se puede observar en la ilustración 21 que nuevamente es muy alto el impacto de la actividad rentista, asociado al riesgo de esta actividad dada la poca constancia de los flujos de ingresos.

El alto rango de edad concuerda con el cual las personas comienzan a tener más problemas de salud, razón por la cual hace sentido que esta variable tenga importancia para este modelo, al igual que en el modelo de vida con bonificación.

VARIABLE	VALOR	COEF	PORCENTAJE
ACTIVIDAD	RENTISTA	0,96	0,26 %
RANGO_EDAD	46 - 55	0,86	21,7 %
ESTADO_CIVIL	VIUDO	0,53	1,54 %
ANTIGUEDAD	1-2 ANIOS	0,52	9,1 %
TIPO_TARJETA	VISA PLATINUM	0,46	5,3 %

*Ilustración 21: Extracto perfil más propenso modelo salud con CMR
Fuente: Elaboración propia*

Nuevamente el estado civil viudo tiene un alto impacto en la propensión, en este caso para un seguro de salud. Lo anterior nuevamente hace sentido debido a un temor producto de una experiencia pasada con un conyugue.

Finalmente, las variables asociadas a la antigüedad y el tipo de tarjeta se pueden interpretar nuevamente como la confianza que tiene el cliente con el holding, lo cual puede afectar positivamente la contratación de un seguro de salud en la compañía.

Se puede observar en la ilustración 22 que la característica más importante consiste en el rango de edad por sobre los 66 años, lo cual se puede deber a la aversión al riesgo que una persona de ese rango de edad presenta.

Podemos observar que nuevamente es significativa una actividad cuyos flujos de ingreso de capital no son constantes y dependen de las ventas generadas, como lo son para un comisionista.

VARIABLE	VALOR	COEF	PORCENTAJE
RANGO_EDAD	66 +	1,44	11,79 %
ACTIVIDAD	COMISIONISTA	0,74	0,34 %
HABITO_PAGO	MALO	0,71	0,85 %
ANIO_VEHIC	NUEVOS	0,65	5,62 %
HABITACION	CASA PROPIA	0,61	58,74 %

*Ilustración 22: Extracto perfil más propenso modelo hogar con CMR
Fuente: Elaboración propia*

El hábito de pago malo puede deberse al que los seguros de hogares se encuentran en promedio en el tercer lugar de los seguros principales más caros, por lo que se puede interpretar que el hábito de pago es malo producto del alto valor que éste podría tener.

Por otra parte, es razonable encontrar que otra de las variables importantes consiste en la tenencia de una casa propia, dado que hace poco sentido el asegurar la casa de algún tercero. Esta variable la posee el 58,74%, por lo cual podemos determinar que es muy común en los clientes propensos, por lo cual tiene aún más sentido la interpretación anterior.

Los seguros transaccionales consideran complementos de otros seguros principales y seguros extras. Por ejemplo, un seguro complementario a uno automotriz consiste en asegurar el vehículo por dentro, lo cual protege al cliente en caso de robo y daños realizados en el interior de este.

Por otro lado, un seguro extra consiste en un seguro que no pertenece a ninguno de los productos principales, como lo son los seguros de mascotas.

Para estos seguros, el hábito de pago pésimo y la descripción de la cuenta como mora leve son de alto impacto, como se ve en la ilustración 23. Lo anterior se puede justificar porque el precio de estos seguros es bajo, dado que su valor promedio es un 10% del precio de los automotrices, por lo cual los clientes pueden considerar su contratación dado su bajo valor.

Sumando a lo anterior, los clientes con mora leve, que considera 1 a 2 días de retraso en sus pagos de la tarjeta, pueden estar en esta situación dado que olvidan realizar los pagos. Por lo cual pueden pertenecer a este segmento aquellos clientes olvidadizos que dejan bolsos u objetos de valor en el interior de sus vehículos, por lo cual les resulta útil contratar un seguro complementario que proteja sus objetos de valor en los automóviles.

VARIABLE	VALOR	COEF	PORCENTAJE
HABITO_PAGO	PESIMO	1,17	0,009 %
DESC_SITUACTA	MORA LEVE	0,93	9,74 %
ESTADO_CIVIL	DIVORCIADO	0,8	0,17 %
ACTIVIDAD	COMISIONISTA	0,73	0,34 %
ANTIGÜEDAD_CMR	0 – 6 MESES	0,60	4,54 %

*Ilustración 23: Extracto perfil más propenso modelo transaccional con CMR
Fuente: Elaboración propia*

Lo anterior sumado a que la actividad como comisionista es también de alto impacto, se puede concluir que estos seguros son preferidos por clientes cuyos estados financieros con su tarjeta CMR es malo.

Estos tipos de seguros son preferidos por aquellos clientes nuevos, cuya antigüedad es menor a 6 meses. Lo anterior puede explicarse debido a que para la contratación de estos seguros no es necesario que exista una confianza de parte del cliente hacia el holding como era el caso de los seguros de vida, dado que estos no tienen un carácter tan personal como los ya mencionados.

En base a lo anterior, se puede interpretar que los modelos explican de manera correcta el negocio dado que sus variables hacen sentido. Por esta razón es pertinente su uso para el cálculo de propensión de los clientes de la compañía para cada uno de los productos principales.

8.3.2 Modelo Árbol de Decisión

A fin de comparar el desempeño del modelo Logit Binario se utilizará un modelo Árbol de Decisión. Para que este cumpla con su finalidad, se utilizarán las mismas variables que en el Logit para cada producto, además del balanceo de los datos considerando un 50% de casos positivos en la variable dependiente.

El método utilizado es de clasificación y, tanto la profundidad como la poda, están controlados según reglas de detención estadísticas definidas en el código del paquete “rpart” de R. Los demás parámetros están definidos por defecto en el mismo código [10].

De igual manera que para el modelo Logit, se analizarán si los modelos generados tienen sentido para el negocio y sólo se realizará el análisis para los modelos con la tarjeta CMR, dado que para los modelos sin esta es trivial.

Para el modelo de auto, podemos observar que los primeros filtros hacen sentido al modelo, dado que tienen relación a la tenencia y año del vehículo de los clientes, como se observa en la ilustración 24.

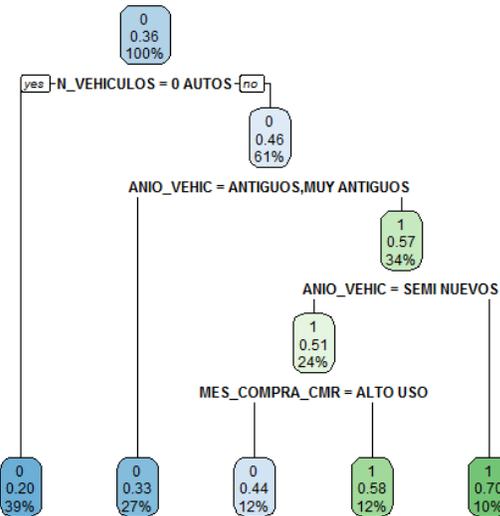


Ilustración 24: Gráfico Árbol de Decisión modelo auto con CMR
Fuente: Elaboración propia

Para el modelo de vida, se puede observar en la ilustración 25, que la tenencia de algún seguro principal es uno de los primeros filtros que utiliza el modelo. Así, aquellos clientes con un seguro principal tienen una propensión de un 68% de contratación de un seguro de vida. Lo cual se puede interpretar como la confianza que tiene un cliente en la compañía.

Nuevamente podemos observar que el segundo filtro que se utiliza es el uso de CMR, que nuevamente se puede interpretar como la confianza en el holding. Lo mismo ocurre con la antigüedad de un cliente, donde se observa que una antigüedad entre 1 a 5 años es el último filtro que realiza el modelo para determinar una propensión mayor a un 50%.

Por otro lado, la edad determinante para la contratación de un seguro de vida es para aquellos clientes menores a 66 años, dado que estos terminan con una propensión igual o mayor a 47%. En cambio, aquellos clientes con 66 años o más, terminan con una propensión de un 20%.

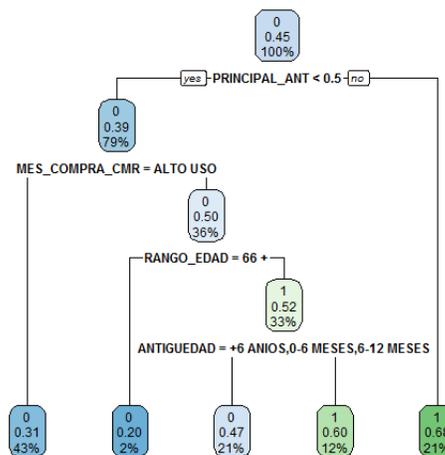


Ilustración 25: Gráfico Árbol de Decisión modelo vida con CMR
Fuente: Elaboración propia

Para el caso del modelo de vida con bonificación, al igual que en el seguro de vida, la tenencia de otro seguro principal es el primer filtro, como se puede observar en la ilustración 26. Nuevamente se asocia a la confianza del cliente en la compañía.

Finalmente, y al igual que en el seguro de vida, una edad mayor a 66 años es filtro importante, dado que para aquellos clientes que cumplen con dicha clasificación, su propensión es un 24%, mientras que aquellos que la cumplen tienen una propensión de un 76%.

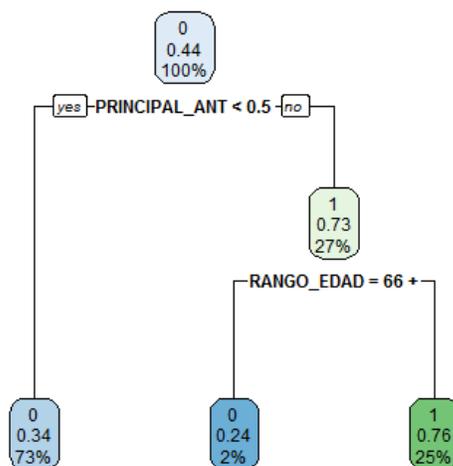


Ilustración 26: Gráfico Árbol de Decisión modelo vida con bonificación con CMR
Fuente: Elaboración propia

Al igual que para los modelos con vida, para los modelo de salud y hogar el primer y único filtro que realiza el Árbol de Decisión corresponde a la tenencia de otro seguro principal, como se puede observar en las ilustraciones 27 y 28 respectivamente. Con lo anterior podemos concluir que la confianza dese el cliente hacia la compañía es determinante para el cálculo de la propensión de este modelo.

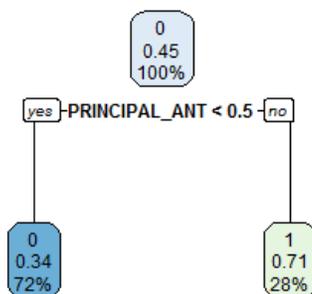


Ilustración 27: Gráfico Árbol de Decisión modelo salud con CMR
Fuente: Elaboración propia

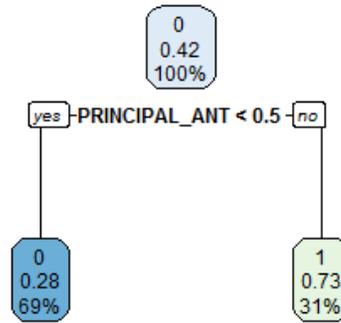


Ilustración 28: Gráfico Árbol de Decisión modelo hogar con CMR
Fuente: Elaboración propia

Por último, para el modelo de seguros transaccionales, el modelo Árbol de Decisión realiza un primer filtro asociado al gasto con la tarjeta CMR, por lo que a aquellos clientes que tienen un gasto bajo la media, tienen una propensión de 21%, como se puede observar en la ilustración 29.

Podemos observar que la tenencia de un seguro principal es un indicador importante en el filtro, dado que aquellos que tienen un seguro principal además de gastar con su CMR, tienen una propensión de un 61%.

Podemos observar además, que los siguientes filtros están relacionados con la confianza en el holding, la antigüedad del cliente según la tenencia de tarjeta CMR, la tenencia de otros seguros en la compañía, el tipo de tarjeta que poseen y el segmento de seguros Falabella al que pertenecen.

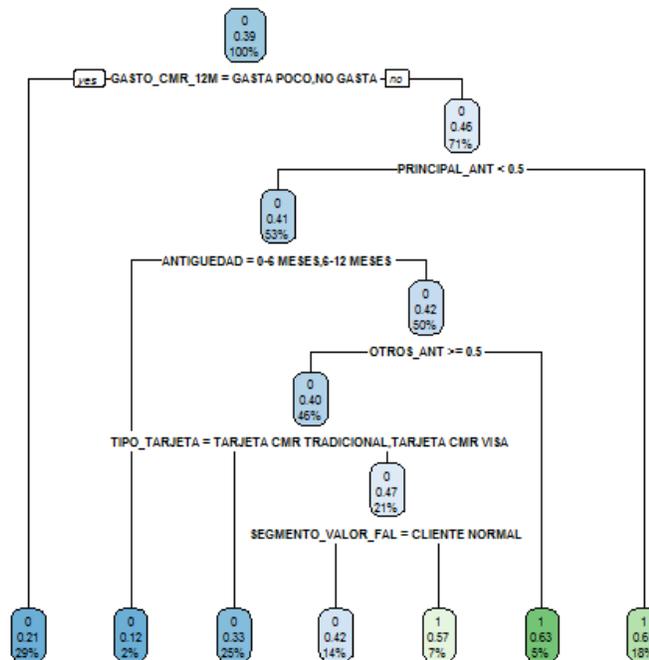


Ilustración 29: Gráfico Árbol de Decisión modelo transaccionales con CMR
Fuente: Elaboración propia

En base a lo anterior, podemos establecer que los modelos obtenidos mediante el Árbol de Decisión representan de buena manera el negocio. Con esto, podemos concluir que su uso se encuentra justificado.

8.4 Evaluación de modelos

Con el fin de obtener resultados robustos, se ejecutaron los modelos 5 veces para comparar los desempeños obtenidos. En estas ejecuciones se cambiaron tanto la base de testeo como la de entrenamiento dado que ambas son elegidas de manera aleatoria. Las variaciones de los desempeños del mejor quintil entre estas 5 ejecuciones fueron menor a un 3% en cada uno de los modelos, con lo cual se puede concluir que son modelos robustos.

8.4.1 Desempeño modelos de propensión

Como se mencionó previamente, se evaluarán los modelos según su desempeño en la matriz de confusión y en su desempeño según el mejor quintil de probabilidad entregado por el modelo. De esta manera se podrá determinar aquel modelo que se adapte más a los requerimientos de la compañía según el indicador más adecuado. El análisis comenzará con los modelos sin la tarjeta CMR y luego los modelos con la tarjeta.

Dado que la matriz de confusión utiliza un valor mínimo de probabilidad para asumir como positivos los valores de la variable dependiente calculada para la generación de cada matriz de confusión, se utilizó la probabilidad media obtenida por cada modelo.

	AUTO	VIDA	VIDA_CB	SALUD	HOGAR	TRANSACCIONALES
LOGIT	32%	24%	31%	31%	25%	39%
ÁRBOL	38%	27%	35%	38%	30%	35%

*Ilustración 30: Probabilidad media por producto modelo Logit Binario y Árbol de Decisión sin CMR
Fuente: Elaboración propia*

Para analizar el desempeño de los modelos sin CMR es indispensable conocer las probabilidades medias calculada por los modelos para los productos según cada modelo, lo cual se muestra en la ilustración 30. De esta ilustración se puede observar que el comportamiento de los modelos es similar para cada producto, donde la mayor diferencia según la probabilidad media calculada por estos corresponde a un 7%.

MODELO	Métrica	AUTO	VIDA	VIDA_CB	SALUD	HOGAR	TRANSACC.
LOGIT	Accuracy	61,8%	75,5%	51,8%	55,7%	67,4%	58,1%
	Sensibilidad	83,4%	84,5%	85,8%	83,9%	80,7%	77,3%
	Precisión	2,8%	2,0%	0,3%	0,2%	0,2%	0,2%
ARBOL	Accuracy	73,1%	78,0%	68,1%	68,0%	79,5%	79,9%
	Sensibilidad	65,7%	74,9%	67,8%	63,2%	73,0%	51,0%
	Precisión	3,2%	1,8%	0,3%	0,1%	0,3%	0,3%

*Ilustración 31: Desempeño modelos según matrices de confusión sin CMR
Fuente: Elaboración propia*

Con dichas probabilidades se calcularon los desempeños de las métricas según la matriz de confusión obtenida. De los resultados obtenidos, se calcularon las diferencias en las métricas restando el desempeño de los modelos Logit versus los obtenidos por los modelos Árbol de Decisión como se observa en la ilustración 31.

Con lo anterior se pudo observar, tal como muestra la ilustración 32, que la diferencia en precisión es muy cercana a 0, lo que indica que ambos modelos son similares en precisión. Sin embargo, la diferencia significativa se encuentra al analizar la sensibilidad y el accuracy de los modelos. Para la sensibilidad se puede observar que el modelo Logit supera hasta en un 17% al modelo de Árbol de Decisión, mientras que para el accuracy el primero se ve superado por el segundo en hasta en un 18%. Podemos observar que para el producto hogar ambos modelos tienen un desempeño similar.

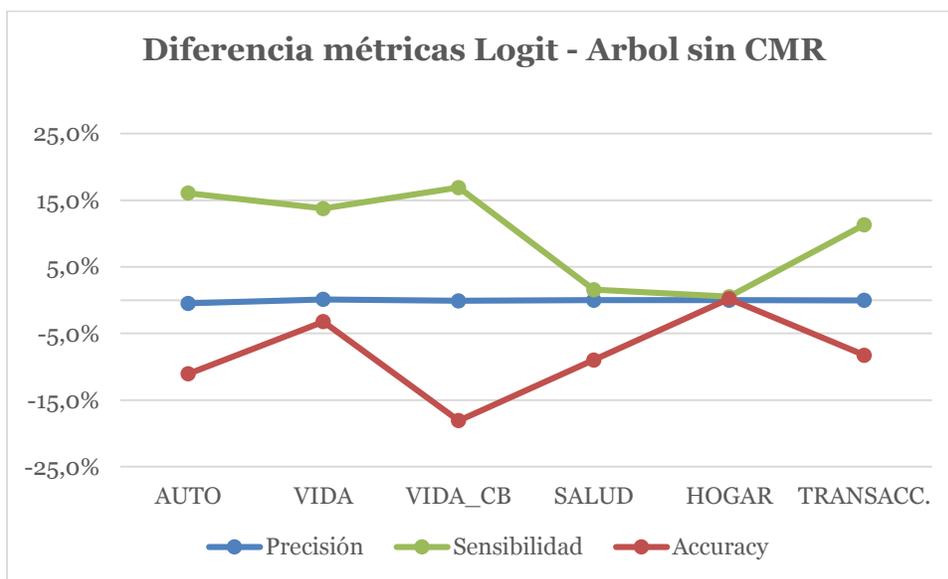


Ilustración 32: Diferencia en métricas de desempeño Logit – Arbol sin CMR
Fuente: Elaboración propia

Además del análisis anterior, se pueden comparar el desempeño de los modelos para cada producto según las curvas de ganancia. En la ilustración 33 se puede observar que para el modelo auto sin CMR, el modelo logit obtiene un mejor desempeño que el modelo árbol de decisión en todos los deciles. En anexos F se encuentran las curvas de ganancia para todos los productos.

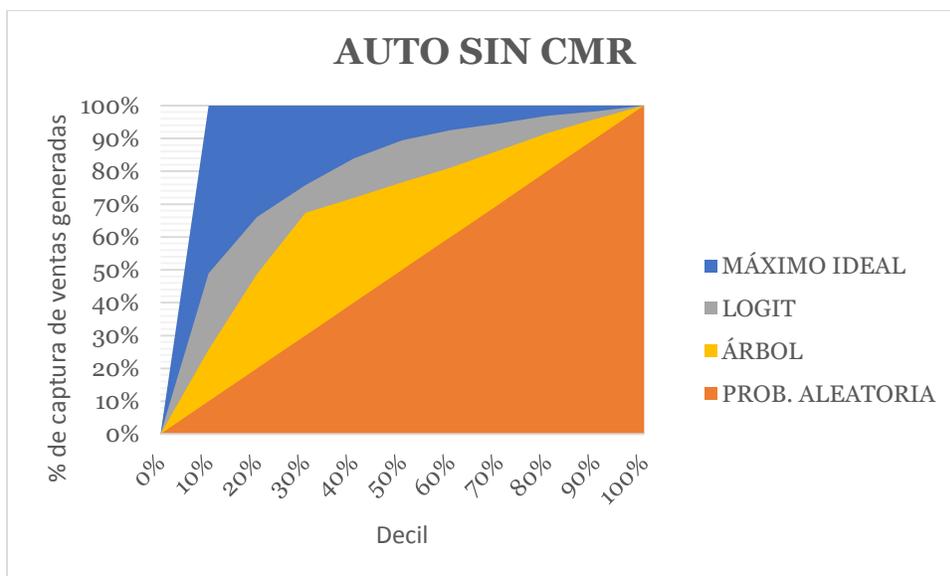


Ilustración 33: Curva de ganancia modelo auto sin CMR
Fuente: Elaboración propia

Por otra parte, al analizar el desempeño según el mejor quintil, se observa que el modelo Logit es superior al Árbol de Decisión, tal como se observa en la ilustración 34, en la cual están graficados los desempeños de ambos modelos para cada producto. Se observa que la mayor diferencia corresponde a un 17,3% en el modelo de auto

Esta diferencia llega hasta un 17,3% en el modelo Logit de auto, lo cual es un claro indicio del mejor desempeño del modelo. El beneficio generado por este modelo es superior en todos los productos, sin embargo la menor diferencia se encuentra en salud, donde alcanza un 1,1%.

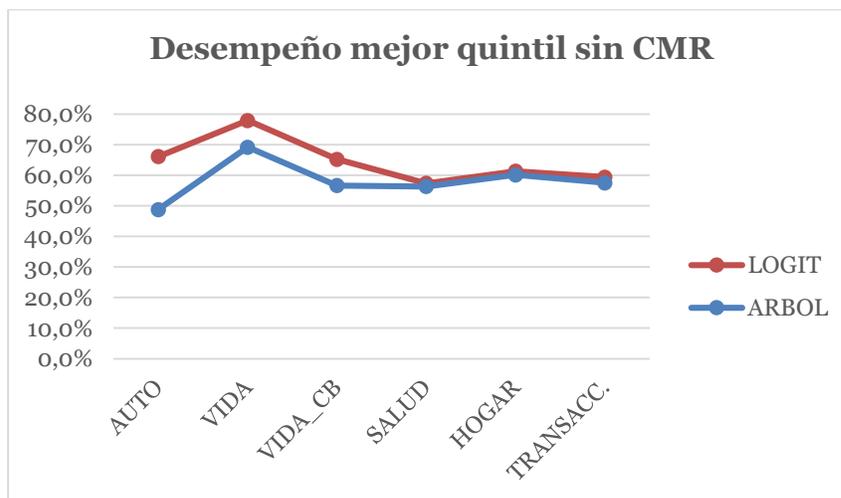


Ilustración 34: Desempeño mejor quintil sin CMR modelo Logit y Arbol
Fuente: Elaboración propia

Ahora es el caso de analizar los modelos asociados a la tenencia de la tarjeta CMR. De igual manera que en el caso sin la tarjeta, se utilizó la probabilidad media calculada por cada modelo para generar las matrices de confusión en cada caso. Dichas probabilidades pueden observarse en la ilustración 35.

Al igual que en el caso de los clientes sin CMR, la diferencia existente entre las probabilidades medias de los modelos son muy pequeñas, donde la máxima diferencia corresponde a un 11%.

	AUTO	VIDA	VIDA_CB	SALUD	HOGAR	TRANSACCIONALES
LOGIT	29%	37%	33%	35%	30%	31%
ARBOL	32%	42%	39%	40%	35%	42%

Ilustración 35: Probabilidad media por producto modelo Logit Binario y Árbol de Decisión con CMR
Fuente: Elaboración propia

Analizando las métricas de desempeño de la ilustración 36, nuevamente se comparó la diferencia en el desempeño de las métricas restando el desempeño del modelo Logit menos el modelo Árbol de Decisión.

MODELO	Métrica	AUTO	VIDA	VIDA_CB	SALUD	HOGAR	TRANSACC.
LOGIT	Accuracy	63,8%	52,4%	55,6%	56,2%	61,5%	53,6%
	Sensibilidad	79,8%	80,3%	83,4%	79,7%	94,9%	82,2%
	Precisión	3,7%	1,0%	1,1%	0,6%	0,6%	0,9%
ARBOL	Accuracy	68,4%	58,6%	88,6%	84,9%	85,4%	67,6%
	Sensibilidad	71,7%	69,3%	44,2%	44,6%	91,8%	63,3%
	Precisión	4,0%	1,1%	2,3%	0,9%	1,1%	0,9%

Ilustración 36: Desempeño modelos según matrices de confusión con CMR
Fuente: Elaboración propia

Se puede observar en la ilustración 37 que el desempeño de los modelos es similar al caso sin CMR. En esta ilustración se puede observar que la precisión nuevamente es cercano a 0, el modelo Logit nuevamente es superior en sensibilidad hasta en un 40,2% en Vida con Bonificación y el modelo árbol es superior en accuracy para todos los productos.

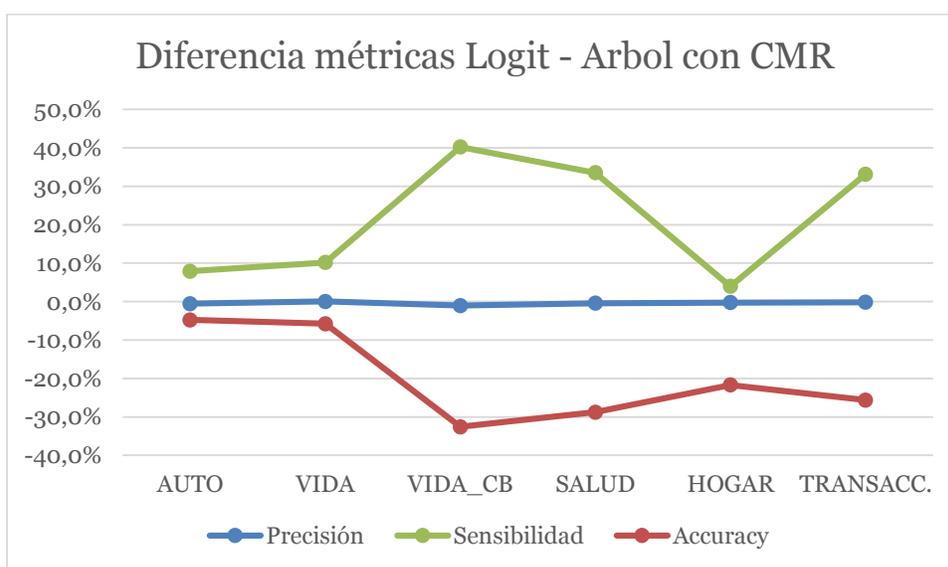


Ilustración 37: Diferencia métricas de desempeño modelo Logit - Árbol con CMR
Fuente: Elaboración propia

De igual manera que para los modelos sin CMR, se puede comprar el desempeño de los modelos mediante las curvas de ganancia de estos. En la ilustración 38 se puede observar que el modelo logit para el producto auto sin CMR obtiene mejor desempeño que el modelo árbol de decisión para cada uno de los deciles. En anexos F se encuentran las curvas de ganancia para todos los productos.

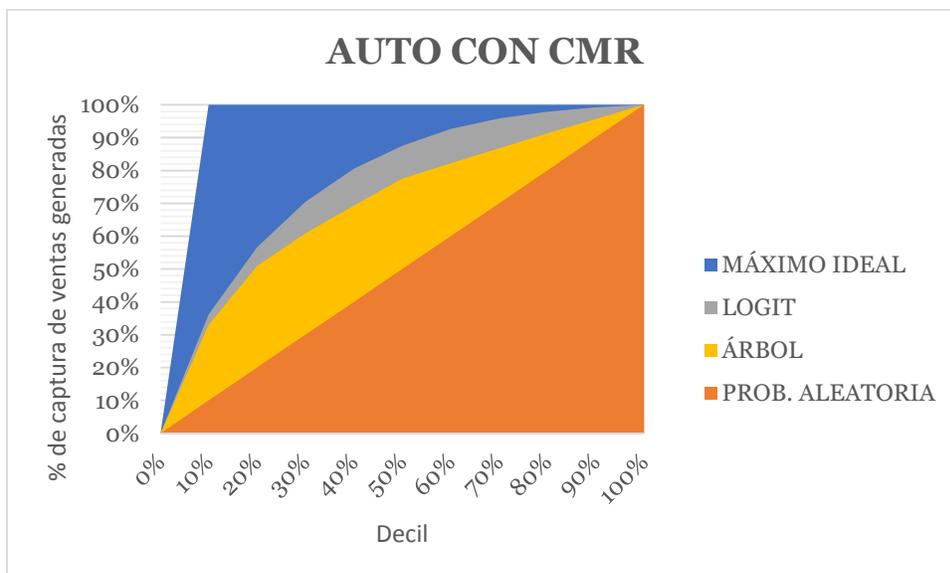


Ilustración 38: Curva de ganancia modelo auto con CMR
Fuente: Elaboración propia

El siguiente paso es analizar su desempeño en términos del mejor quintil. Para este análisis se graficó nuevamente el desempeño de ambos modelos, donde en la ilustración 39 se puede observar que el modelo Logit es superior al Árbol de Decisión en todos los productos, con una diferencia máxima de un 12,7% en vida con bonificación.

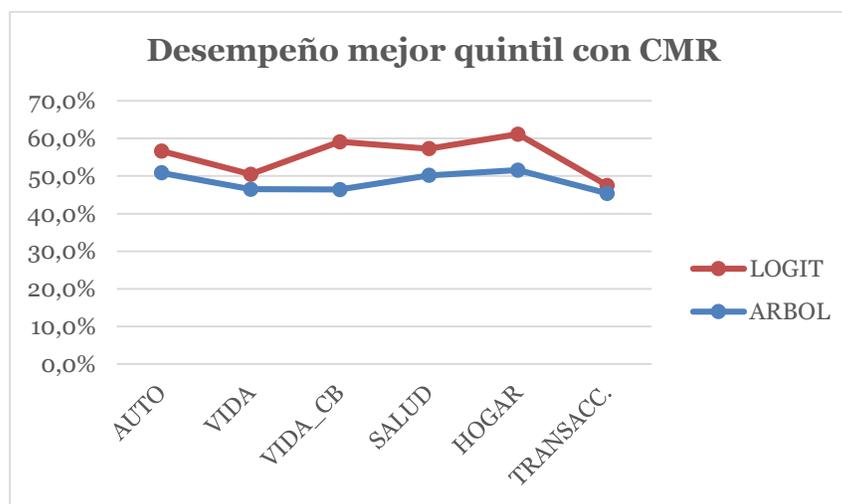


Ilustración 39: Desempeño mejor quintil con CMR modelo Logit y Arbol
Fuente: Elaboración propia

Dados los resultados obtenidos previamente, no podemos concluir de manera simple qué modelo es mejor dado que tanto para los clientes con y sin tarjeta CMR el

Logit es mejor en sensibilidad y desempeño del mejor quintil, mientras que el Árbol de Decisión es mejor en accuracy.

Para poder comparar ambos modelos es necesario entender el negocio y, en particular, en qué consiste el email marketing. Éste corresponde a enviar correos a una base de datos de clientes con el fin de incentivar la contratación de algún producto por parte de éste. Dado que se selecciona un pequeño porcentaje de la base de datos, el objetivo es elegir aquellos clientes más propensos a la contratación con el fin de maximizar la esperanza de contratación de un producto.

Si bien estas campañas tienen un costo de ejecución muy bajo dado que los procesos se ejecutan de manera automática, el costo asociado al proceso debe ser considerado como aquellos clientes propensos a los que no se les envían campañas. Por lo tanto, el costo asociado a un falso positivo es mucho menor a un falso negativo, dado que el último considera la pérdida de una posible contratación.

Siguiendo esta línea, la medida de accuracy no es adecuada para este caso ya que considera la proporción de valores que el modelo predice de manera correcta, ya sean positivos como negativos, versus el total de predicciones. Lo anterior, sumado a la baja cantidad de casos positivos en los datos, generaría una mala medición de desempeño.

Un ejemplo de lo anterior ocurriría si un modelo predice todos los valores como negativos. Dada la desproporción de los datos, un modelo que clasifique todos los casos como negativos tendría un accuracy muy alto dada la baja cantidad de casos que se clasificarían como falsos negativos versus la alta cantidad de aciertos asociados al verdadero negativo. Sin embargo, este resultado no resulta útil para el negocio dado que la intención del modelo es capturar la mayor cantidad de verdaderos positivos y la menor cantidad de falsos negativos.

Por otra parte, si se considera un modelo que reduce la cantidad de falsos negativos pero aumenta la cantidad de falsos positivos, éste tendrá un menor accuracy. Sin embargo, este resultado es más útil para el negocio dado que predice de mejor manera a los clientes propensos al cross-selling al disminuir los falsos negativos.

Es por esto que la medida de accuracy no es un indicador consistente con los requerimientos del negocio y, en particular, a la finalidad para la cual se utilizará este modelo.

Por lo tanto, el modelo debe considerar aquella métrica que permita comprar la cantidad de casos positivos reales predichos por sobre los falsos negativos. Ésta métrica corresponde a la sensibilidad, que considera la proporción de valores positivos predichos de manera correcta por el modelo sobre el total de valores positivos reales. Ella es independiente de la proporción de casos positivos que se tenga en los datos, lo cual la convierte en una métrica robusta.

Por último, y dado que las campañas de email marketing no son enviadas a todos los clientes, resulta pertinente el uso del mejor quintil de desempeño de los modelos. De esta manera se selecciona a aquellos clientes más propensos, con el fin de maximizar las ventas generadas por cross-selling. Así, la métrica de desempeño del mejor quintil que considera la proporción de casos positivos capturados por el modelo por sobre el total de

casos positivos de la base de testeo, resulta útil para medir el desempeño del modelo en el porcentaje al cuál se le enviará la campaña.

Para ello se debe seleccionar al mejor 20% de los clientes según la probabilidad de propensión calculada por el modelo. Luego, se debe calcular el porcentaje de casos positivos capturados en ese 20% versus el total de positivos de la base de testeo. De esta forma, la métrica se puede comparar con un aleatorio, donde este obtendría un 20% de los valores positivos, por lo que el desempeño obtenido por los modelos se debe comparar con dicho valor. Dicho desempeño se puede observar en las curvas ganancia de anexos F.

Considerando todas las métricas comentadas previamente, el mejor modelo según su desempeño, la naturaleza del negocio y finalidad de este, corresponde al modelo Logit Binario. Esto debido a que este permite determinar en mejor manera que el Árbol de Decisión la cantidad de positivos reales y minimizar la cantidad de falsos negativos. Este es un buen indicador para el negocio dado que así se minimizan la cantidad de clientes a los que, siendo propenso a la contratación, no se les ofrece un producto.

8.5 Propuestas de acción

Como ya se determinó que el modelo Logit Binario tiene un mejor desempeño y además se ajusta a las necesidades de la compañía, el siguiente paso es evaluar el modelo de acuerdo al método “Next Best Offer” cuyo objetivo es seleccionar el mejor producto según la propensión de un cliente.

Lo anterior indica que será necesario comparar la propensión entregada por los modelos entre los distintos productos. Dado que los modelos son independientes, una manera de comparar los valores de las propensiones consiste en analizar el decil al que pertenece cada valor. De esta manera se elegirá aquella propensión que tenga el mejor decil de todos.

Por lo tanto es necesario que el output generado por el modelo incluya tanto la propensión de un producto como el decil al cual pertenece dicho valor. Así, se contará con dos valores para cada producto.

Lo anterior generará dos casos particulares para cada cliente:

1. Existe un único producto cuyo valor de decil es el más alto.
2. Existe más de un producto cuyos valores de decil corresponden al valor más alto.

Para el primer caso, dado que es un único el producto que tiene el decil más alto, se elegirá dicho producto como el “Next Best Offer” para el cliente. Para el segundo caso, como pueden existir dos o más de un producto con mismo decil, se ponderará la propensión por el valor de la prima promedio de cada producto. Se elegirá aquel que tenga mayor valor.

A modo de ejemplo, en la ilustración 40 se observan ambos casos. Para el cliente 1 existe un único máximo decil, a quién se le ofrece el producto Vida. Por otro lado, para el cliente 2 existe más de un valor máximo de decil, por lo que se elige el ramo auto ya que es el de mayor beneficio esperado.

ID	Prop. Auto	Decil Auto	Benef. Auto	Prop.Vida	Decil Vida	Benef. Vida
1	0,74	8	18,5	0,65	9	13
2	0,68	7	17	0,50	7	10

Ilustración 40: Ejemplo máximo decil
Fuente: Elaboración propia

Lo anterior es un método que permite, por un lado, elegir el producto por el cuál un cliente se encuentra más propenso y, además, pondera el beneficio generado para la compañía. Este método permite elegir el mejor producto para el cliente y en caso de que exista más de uno, se elegirá aquél que le entregue mayor beneficio a la compañía.

En base a lo anterior, se propone ofrecer a cada cliente el producto en base al proceso de selección previamente descrito, dado que en una primera instancia maximiza la propensión para cliente, mientras que en una segunda maximiza el beneficio para la compañía en caso de que la primera no permita determinar una única opción.

8.6 Propuesta de diseño experimental

La realización de un experimento, cuya finalidad sea comprobar que el método de “Next Best Offer” propuesto es una buena manera de elegir el producto, resulta interesante y a la vez útil para la compañía. Para verificar lo anterior, será necesario realizar dos experimentos independientes. Dichos experimentos se compararán con sus respectivos grupos de control mediante un test de igualdad de proporciones.

Estos experimentos únicamente se verificarán para los productos de auto, vida, vida con bonificación, salud y hogar, dado que son los únicos productos por los que se hace campaña de call-center y email marketing.

El tamaño para los grupos experimentales como para el grupo de control de cada experimento corresponde a un tercio del tamaño de muestra, correspondiente a 1.437 clientes. Dicho valor se calculó para que el experimento sea significativo con una confianza de 95% y un error de un 1%. De esta manera, el tamaño total de cada experimento corresponde a 4.311 clientes.

El primer experimento se realizará a aquellos clientes que tienen un único producto cuyo valor de decil es el más alto. Las características de este experimento se detallan en la ilustración 41.

Experimento máximo decil

<i>Objetivo</i>	Probar que la elección del máximo decil genera más ventas.
<i>Grupo experimental 1</i>	Se le ofrecerá el producto de mejor decil. (1.437 clientes)
<i>Grupo experimental 2</i>	Se le ofrecerá el producto de peor decil. (1.437 clientes)
<i>Grupo de control</i>	Ofrecer el producto de manera aleatoria. (1.437 clientes)
<i>Hipótesis nula</i>	Proporción de clientes que contratan un seguro es igual para el grupo control como para cada grupo experimental de manera

	independiente.
<i>Hipótesis alternativa</i>	Proporción de clientes que contratan seguros es distinta para el grupo de control y cada grupo experimental de manera independiente.
<i>Total</i>	4.311 clientes

Ilustración 41: Descripción experimento máximo decil
Fuente: Elaboración propia

El segundo experimento es para aquellos clientes que cuentan con más de un producto cuyos valores de decil corresponden al valor más alto. Las características de este experimento se detallan en la ilustración 42.

Experimento máximo beneficio esperado

<i>Objetivo</i>	Probar que la elección del máximo decil genera más ventas
<i>Grupo experimental 1</i>	Se le ofrecerá el producto de mejor decil. (1.437 clientes)
<i>Grupo experimental 2</i>	Se le ofrecerá el producto de peor decil. (1.437 clientes)
<i>Grupo de control</i>	Ofrecer el producto de manera aleatoria. (1.437 clientes)
<i>Hipótesis nula</i>	Proporción de clientes que contratan un seguro es igual para el grupo control como para cada grupo experimental de manera independiente.
<i>Hipótesis alternativa</i>	Proporción de clientes que contratan seguros es distinta para el grupo de control y cada grupo experimental de manera independiente.
<i>Total</i>	4.311 clientes

Ilustración 42: Descripción experimento máximo beneficio esperado
Fuente: Elaboración propia

Dado lo anterior, estos experimentos se deberán realizar considerando únicamente los 5 productos mencionados previamente, diferenciando aquellos modelos con y sin la tarjeta CMR. Así, se realizarán 10 experimentos distintos.

Con el fin de evitar un posible sesgo en los experimentos debido a diferencias existentes en ciertas épocas del año, semana o día en el que se realizarán los distintos experimentos, se propone realizar los experimentos de manera simultánea. Así coincidirán tanto los tiempos como los plazos de ejecución.

Además, es importante que la proporción de características se mantenga tanto en los grupos de control como en los grupos experimentales. Para lograr lo anterior se propone utilizar un resamplio aleatorio para luego comparar las características entre los distintos grupos, con el fin de no sesgar al experimento. Asegurando que sean representativos y similares.

Con este diseño experimental propuesto se espera obtener que las proporciones entre los grupos experimentales y los grupos de control sean distintas. En particular, tanto los grupos experimentales tanto del mejor decil como para el mayor beneficio esperado, se espera que su proporción de contratación de productos sea mayor a la de su

grupo de control respectivo. En cambio, para el peor decil y el menor valor esperado, se espera que sus proporciones sean menores al grupo de control.

Con estos resultados se podrá comprobar que el método de “Next Best Offer” presentado en este proyecto de título tiene un impacto positivo significativo en la venta de pólizas principales para Seguros Falabella.

8.7 Evaluación económica

Con el fin de obtener el beneficio económico que tiene para la compañía el trabajo de título, se realizó una evaluación económica mediante una simulación, contrastando el beneficio esperado calculado con el modelo actual existente versus el modelo propuesto de “Next Best Offer”.

Para ello se tomaron únicamente los ramos a los cuales se les realiza campaña de email marketing: automotriz, vida, vida con bonificación, salud y hogar. En particular, se empleó el mismo grupo de testeo utilizado para evaluar los modelos, que considera el periodo entre los años 2015 a 2016 y tiene un total de 660.000 clientes aproximadamente.

De esta manera, se calculó la propensión de cada cliente para los cinco productos mencionados previamente y se utilizó el método “Next Best Offer” propuesto para determinar el producto a ofrecer a cada cliente. Se eligieron únicamente aquellos productos cuyo desempeño del modelo se encontró por sobre el decil 8.

Este resultado se comparó con la venta en el periodo final, por lo que si el producto elegido por el método propuesto fue efectivamente contratado por el cliente, se contabilizó como venta.

Este resultado se contrastó con el método actual que tiene la compañía. Éste método consiste en un modelo de propensión de seguro automotriz, donde se ofrece dicho producto a aquellos clientes que tienen una propensión por sobre el decil 8.

PRODUCTO	CANTIDAD	PORCENTAJE
AUTOMOTRIZ	7.076	74,5%
VIDA	718	7,6%
VIDA CON BONIFICACIÓN	1.107	11,7%
SALUD	265	2,8%
HOGAR	326	3,4%
Total		9.492

*Ilustración 43: Ventas simuladas según modelo propuesto
Fuente: Elaboración propia*

En términos de pólizas principales, la simulación permitió determinar que el modelo actual generaría una venta de 7.781 seguros automotrices, mientras que la presentada en este trabajo de título sería de 9.492. En la ilustración 43 se encuentra la cantidad de pólizas vendidas según rubro, donde se observa que 2.416 de ellas fueron en un rubro distinto al automotriz.

Si bien se vendieron más pólizas de los distintos ramos, a excepción del automotriz, es importante comprobar que el impacto económico generado sea positivo. Lo anterior dado que el ramo automotriz es el que genera mayor comisión por venta.

Para medir el impacto económico se ponderó, para cada ramo, la cantidad vendida por la comisión promedio que se obtiene en cada uno. Así se estimó que el modelo actual generaría un beneficio de 739 millones de pesos, mientras que el modelo propuesto generaría 824 millones de pesos.

Con el modelo propuesto se generaron menor cantidad de ventas en el ramo automotriz, sin embargo, el aumento en las ventas de los otros ramos generó un ingreso correspondiente a un 18,4% del beneficio total. Con esto se puede concluir que el método propuesto es capaz de generar un ingreso mayor que el actual producto de las ventas generadas por los demás ramos.

9. CONCLUSIONES

9.1 Conclusiones del proyecto

Este trabajo de título tenía como objetivo principal diseñar un modelo predictivo basado en las características de los clientes con el fin de conocer la propensión de compra de estos para los seis productos principales de Seguros Falabella. Para lograr esto se realizaron múltiples modelos de propensión de dos tipos distintos: Árbol de Decisión y Logit Binario.

A partir de las métricas de comparación se obtuvo como resultado que el modelo Logit Binario tiene mejor desempeño que el modelo Árbol de Decisión para las métricas de comparación utilizadas.

Con la primera métrica correspondiente a la sensibilidad, el modelo Logit Binario fue superior al modelo Árbol de Decisión con una métrica de hasta un 40% superior para el seguro de Vida con bonificación para aquellos clientes con la tarjeta de crédito de la compañía.

Para la segunda métrica, correspondiente al desempeño del modelo en su mejor quintil, el modelo Logit Binario nuevamente fue superior al Árbol de Decisión en hasta un 17,3% en el modelo de seguro automotriz para clientes sin la tarjeta de crédito.

Por último, en la única métrica de desempeño en la cual el modelo Árbol de Decisión superó al Logit Binario corresponde al accuracy. En esta obtuvo un desempeño de hasta un 32,6% superior para el modelo de Vida con bonificación para clientes con la tarjeta de crédito.

Sin embargo, dada la baja proporción de valores positivos en la variable dependiente, el accuracy no se ajusta al negocio. Esto se debe a que los falsos negativos tienen un costo mayor a un falso positivo dado que la intención es capturar la mayor cantidad de casos positivos posibles. Por esta razón, esta métrica no es tan significativa para el negocio y para este problema en particular.

En base a estos resultados, se recomienda a la empresa el uso de un método de “Next Best Offer” que permita comparar las propensiones para los distintos productos. Se propone la elección del producto con el mejor decil de desempeño en caso de que este sea único. En caso contrario, se propone la ponderación del beneficio de la compañía por la propensión de cada producto, eligiendo finalmente el que entregue mayor valor esperado.

Como se pudo observar durante el desarrollo de este trabajo de título, tanto los clientes con la tarjeta de crédito del retail Falabella como los que no la poseen, tienen características demográficas y comportamientos similares. Sin embargo, es necesario hacer una diferenciación entre estos clientes, dado que se tiene más información para aquellos con dicha tarjeta. Es esta información extra la que lleva a la necesidad de realizar modelos distintos para estos dos tipos de clientes, dado que esta información entrega un aporte extra en el desempeño de los modelos.

Para cada uno de los modelos generados, las variables de mayor impacto tienen sentido con el negocio, por lo cual se puede interpretar que los modelos representan de

manera correcta el problema en cuestión. Así, tanto el modelo Logit Binario como el Árbol de Decisión permiten predecir la variable dependiente de manera correcta.

En particular, en el modelo Logit se observó que, dentro de las características más importantes al momento de contratar un seguro de vida, vida con bonificación o salud, se encuentran las variables que tienen más relación con la confianza en el holding. Lo anterior es información relevante, dado que para ofrecer dichos productos y tener buena recepción, el cliente debe tener una buena impresión y confianza en el holding.

Se concluye además que, en base a la simulación realizada, el método de “Next Best Offer” propuesto permitiría generar un aumento de 22% en la cantidad de pólizas principales en comparación con el método utilizado actualmente por la compañía. Además, en esta misma simulación se obtuvo un beneficio económico 11,5% mayor al método actual de la compañía.

De esta manera se puede concluir que este trabajo de título cumple el objetivo general de generar un aumento en las pólizas principales de la compañía. Concluyendo así, que los modelos obtenidos son un aporte tanto en el cálculo de la propensión de un cliente como en el beneficio económico generado para la compañía.

Es así como los modelos de propensión presentados y el método de “Next Best Offer” proporcionado, entregan una herramienta de gestión para el envío de campañas de email marketing según el mejor producto para cada cliente.

9.2 Limitaciones y trabajos futuros

Una de las dificultades de este trabajo de título consistió en la baja cantidad de casos positivos para la variable dependiente. Con lo cual, en un comienzo se obtuvieron valores muy bajos de probabilidades generadas por el modelo. Lo anterior llevó a la realización de un balanceo de la base de entrenamiento mediante un *over-sampling*, hasta obtener la misma proporción de casos positivos como negativos. Así se obtuvo un incremento en los valores de estas probabilidades generadas.

En particular se utilizó este método debido a que un *under-sampling* puede llevar a la eliminación de alguna combinación de casos negativos muy importante. Por esto, un *over-sampling* de los casos positivos es una mejor opción porque permite mantener la distribución de características de estos casos.

Por otro lado, dada la gran cantidad de canales por los cuales un cliente puede ser contactado, se propone realizar un modelo predictivo multinomial para cada producto, cuyo fin sea estimar por qué canal es más propenso un cliente de contratar dicho producto.

Además del trabajo futuro anterior, se propone generar un método de optimización del modelo “Next Best Offer”. Para esto se propone la creación de un método de cálculo de beneficio que pondere tanto el decil como el beneficio generado para la compañía mediante un polinomio. De esta manera se podrá ponderar de mejor manera tanto la preferencia para el cliente i como el beneficio de la compañía para cada producto j .

Este polinomio puede ser del estilo:

$$\text{Next Best Offer}_{i,j} = \alpha(\text{Propensión}_{i,j} * \text{Decil}_{i,j}) + (1 - \alpha)(\text{Propensión}_{i,j} * \text{Beneficio}_j)$$

Donde este α puede ser obtenido mediante la prueba de distintos valores con un diseño experimental. Es esperable que este método permita encontrar un “Next Best Offer” que maximice tanto el beneficio para el cliente como para la compañía.

Por otra parte, dado que este tema de memoria sólo abarca la propensión de cada cliente a los distintos ramos, se propone crear algún modelo de propensión para determinar el producto que elegiría cada cliente dada la propensión a cierto ramo. Esto permitirá ofrecer a cada cliente el producto más adecuado de acuerdo al ramo de mayor propensión.

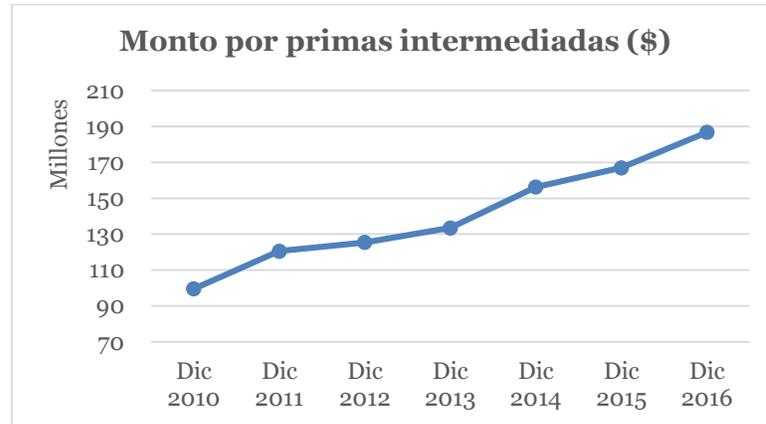
10. BIBLIOGRAFÍA

- [1] Superintendencia de Valores y Seguros (2017). Portal de Educación Financiera. [en línea]. <http://www.svs.cl/educa/602/w3-article-1518.html> [consulta: 20 septiembre de 2017]
- [2] Superintendencia de Valores y Seguros (2016). Informe Corredoras de Seguros Grandes Tiendas. [en línea]. http://www.svs.cl/portal/principal/605/articles-23300_recurso_1.pdf [consulta: 30 abril 2017]
- [3] Superintendencia de Valores y Seguros (2017). Identificación Seguros Falabella Corredores Limitada. [en línea]. http://www.svs.cl/institucional/mercados/entidad.php?auth=&send=&mercado=S&rut=77099010&grupo=&tipoentidad=CSJUR&vig=VI&row=AABaKOAAWA_AAAfjAAn&control=svs&pestanian=1 [consulta: 30 abril 2017]
- [4] Mohri, M., Rostamizadeh, A. y Talwalkar, A. (2012). Foundations of Machine Learning. Boston, Estados Unidos. Massachusetts Institute of Technology. pp. 1-9.
- [5] Mitchell, C. (1992). Logistic regression analysis. Department of Measurement, Statistics & Evaluation. Washington, Estados Unidos.
- [6] Rokach, Lior; Maimon, O. (2008). Data mining with decision trees: theory and applications. World Scientific Pub Co Inc
- [7] Breiman, L. Random Forest. (2001). Kluwer Academic Publishers. Manufactured in The Netherlands. California, Estados Unidos. Statistics Department, University of California.
- [8] Davenport, T., D'Almeida, L. y D'Unger, J. (2010). Know What Your Customers Want Before They Do. Boston, Estados Unidos. Harvard Business Review. pp 5-15.
- [9] Vildoso, F. y Velásquez, J. (2017). Apuntes de Big Data, Capítulo 2, Lecture 11: A key analytical framework. Santiago, Chile. Escuela de Ingeniería y Ciencias, Universidad de Chile. pp. 1-41.
- [10] Therneau, T. M., Atkinson, B., y Ripley, M. B. (2010). The rpart package.

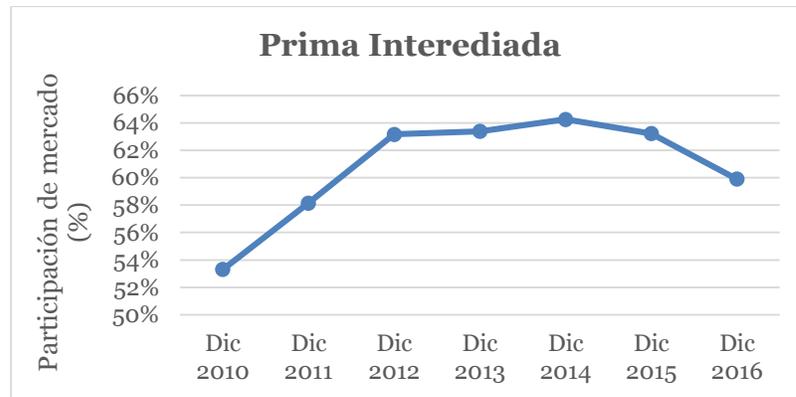
11. ANEXOS

11.1 Anexo A

Antecedentes Generales:



Anexo 1: Monto de primas intermediadas en Seguros Generales y Seguros de Vida por Seguros Falabella
Fuente: Elaboración Propia en base a Informe Corredoras de Seguros Grandes Tiendas [2]



Anexo 2: Porcentaje de prima intermediaria Seguros Falabella en relación al mercado
Fuente: Elaboración propia en base a Informe Corredoras de Seguros Grandes Tiendas [2]

11.2 Anexo B

Ejemplo de rubros Gastos

RUBRO GASTO	CATEGORÍA
AUTOMOTRIZ	Accesorios automotriz
	Arriendo automotriz
	Servicio técnico automotriz
	Automotora
	Autolavados
VIVIENDA	Construcción y propiedades
	Funerarias y cementerios
	Servicios básicos

	Inmobiliarias y corredoras de propiedades
TIENDAS POR DEPARTAMENTO Y COMERCIOS	Grandes Tiendas Retail menor Muebles y artículos electrónicos Farmacias Accesorios, joyerías y regalos Maletería Librerías y editoriales Artículos deportivos Bebés y niños Outdoor
SUPERMERCADOS Y TIENDAS DE ALIMENTOS	Supermercados Minimarkets, carnicerías y verdulerías Panaderías, pastelerías y chocolaterías Tiendas de mascotas Botillerías y licores
MEJORAMIENTO DEL HOGAR	Muebles y decoración Ferretería y materiales de construcción Electro hogar Computación
PAGO WEB	Servicios financieros e inmobiliarias Publicidad
VIAJES	Hoteles y hostales viajes aéreos Agencias de viajes
COMBUSTIBLE	Combustible
SALUD	Clínicas y hospitales Isapres/ Fonasa Ópticas Dentistas
FARMACIAS	Farmacias
RECAUDACIÓN	Inversiones Donaciones Impuestos y gobierno
COMUNICACIÓN	TV Cable

	Telefonía Hogar Internet Celulares y accesorios
RESTAURANT	Comida rápida Heladerías Restaurantes Pubs y cervecerías Cafetería Casinos y discoteca
EDUCACIÓN	Universidades Colegios Preuniversitarios Escuelas de Conducción
ENTRETENCIÓN	Casinos y juegos de azar Cines y espectáculos Peluquerías y salones de belleza Gimnasios y deportes
SEGUROS	Seguros generales
TRANSPORTES	Transporte local Peajes Taxis y similares
OTROS	Ropa

Anexo 3: Glosario ejemplificado de Rubro Gasto
Fuente: CMR Falabella

11.3 Anexo C

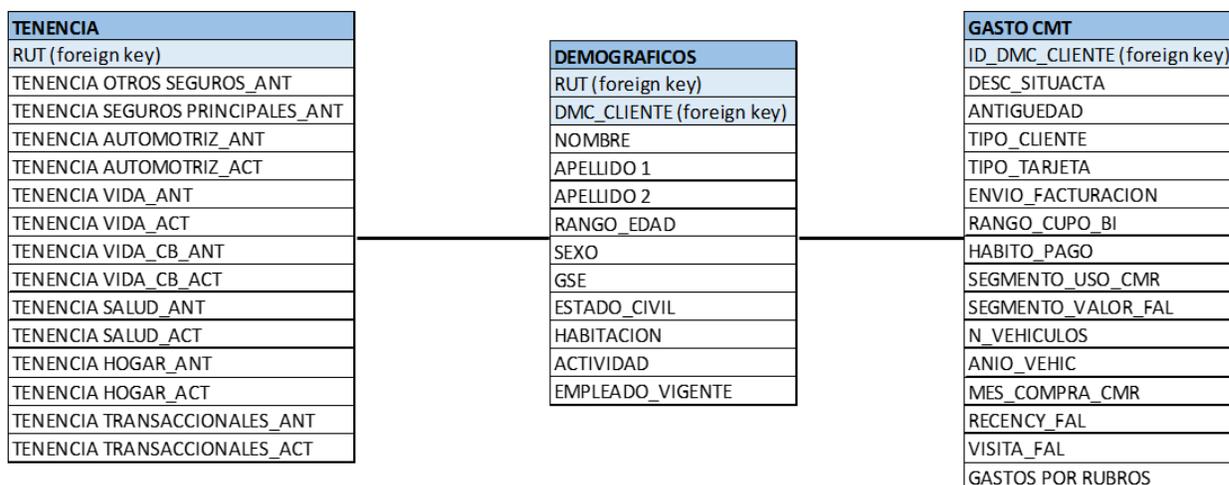
VARIABLES por modelo con CMR

VARIABLE	AUTO	VIDA	VIDA_CB	SALUD	HOGAR	TRANSACC.
OTROS_ANT	X	X	X		X	X
PRINCIPAL_ANT	X	X	X	X	X	X
RANGO_EDAD	X	X	X	X	X	X
SEXO	X	X	X	X	X	X
GSE	X	X	X	X	X	X
ESTADO_CIVIL	X	X	X	X	X	X
HABITACION	X	X	X	X	X	X
ACTIVIDAD	X	X	X	X	X	X
EMPLEADO_VIGENTE	X	X	X	X	X	X
DESC_SITUACTA		X		X	X	X
ANTIGUEDAD	X	X	X	X	X	X
TIPO_CLIENTE	X	X	X	X	X	X

TIPO_TARJETA	X	X	X	X	X	X
ENVIO_FACTURACION	X		X	X	X	X
RANGO_CUPO_BI	X	X	X	X	X	X
HABITO_PAGO	X		X	X	X	X
SEGMENTO_USO_CMV	X	X	X	X	X	X
SEGMENTO_VALOR_FAL	X	X	X	X	X	X
N_VEHICULOS	X	X	X	X	X	X
ANIO_VEHIC	X	X	X	X	X	X
MES_COMPRA_CMV	X	X	X	X	X	X
REGENCY_FAL		X	X	X	X	X
VISITA_FAL	X	X		X	X	
GASTO_CMV_12M	X	X	X	X	X	X
GASTO_RUBRO_AUTOMOTRIZ_12M	X	X		X	X	X
GASTO_RUBRO_COMBUSTIBLE_12M	X	X		X	X	X
GASTO_RUBRO_COMUNICACION_12M		X	X	X	X	X
GASTO_RUBRO_EDUCACION_12M		X	X		X	
GASTO_RUBRO_EMPRESA_12M		X	X	X		X
GASTO_RUBRO_ENTRETENCION_12M		X		X	X	X
GASTO_RUBRO_FARMACIAS_12M	X	X	X	X	X	X
GASTO_RUBRO_MEJ_HOGAR_12M			X		X	X
GASTO_RUBRO_OTROS_12M	X	X	X	X	X	X
GASTO_RUBRO_PAGO_WEB_12M	X		X	X	X	X
GASTO_RUBRO_RECAUDACION_12M	X	X	X	X	X	X
GASTO_RUBRO_RESTAURANT_12M		X	X	X	X	X
GASTO_RUBRO_SALUD_12M	X	X	X		X	X
GASTO_RUBRO_SPM_12M	X	X	X	X	X	X
GASTO_RUBRO_TIENDAS_DEPTO_12M	X	X			X	X
GASTO_RUBRO_TRANSPORTE_12M	X	X	X		X	X
GASTO_RUBRO_VIAJES_12M	X	X	X	X	X	X
GASTO_RUBRO_VIVIENDA_12M	X	X	X	X	X	X

Anexo 4: Variables por modelo con CMV

Fuente: Elaboración propia



Anexo 5: Modelo entidad relación bases de datos
Fuente: Elaboración propia

11.4 Anexo D

Coefficientes modelo automotriz

VariableName	Estimate
ACTIVIDADCOMISIONISTA	0,49
ACTIVIDADDUENA DE CASA	-0,71
ACTIVIDADEMPL. DOMESTICA	-0,33
ACTIVIDADEMPL. TECNICO	-0,15
ACTIVIDADEMPRESARIO	0,33
ACTIVIDADEST. UNIVER.	0,36
ACTIVIDADJUBILADO/MONTEPIADO	-0,32
ACTIVIDADOBREROS, AUXI.	-0,70
ACTIVIDADOFICIAL FF.AA.	0,88
ACTIVIDADPERS. DE SEGURIDAD	-0,80
ACTIVIDADPERSONAL FF. AA.	-0,08
ACTIVIDADPROF. UNIV, (INDP.)	0,56
ACTIVIDADPROF. UNIV. (EMP.)	0,38
ACTIVIDADRENTISTA	0,14
ACTIVIDADTRABAJ. INDEPENDIENT	-0,30
EMPLEADO_VIGENTE	0,38
ESTADO_CIVILDIVORCIADO	-0,19
ESTADO_CIVILSEPARADO	-0,20
ESTADO_CIVILSOLTERO	-0,06
ESTADO_CIVILVIUDO	-0,08
GSEC2	-0,07
GSEC3	-0,43
GSED	-0,77
GSEE	-0,89

HABITACIONCASA DE FAMILIARES	-0,16
HABITACIONCASA DE LA EMPRESA	-0,37
HABITACIONCASA DE LOS PADRES	-0,04
HABITACIONCASA FISCAL	-0,29
HABITACIONCASA PROPIA	-0,03
OTROS_ANT	-1,75
PRINCIPAL_ANT	0,41
RANGO_EDAD26 - 35	0,89
RANGO_EDAD36 - 45	0,96
RANGO_EDAD46 - 55	0,78
RANGO_EDAD56 - 65	0,62
RANGO_EDAD66 +	0,13
SEXOM	0,02

*Anexo 6: Coeficientes modelo auto sin CMR
Fuente: Elaboración propia*

VariableName	Estimate
ACTIVIDADCOMISIONISTA	-0,07
ACTIVIDADDUEÑA DE CASA	-0,35
ACTIVIDADEMPL. DOMESTICA	-0,10
ACTIVIDADEMPL. TECNICO	-0,09
ACTIVIDADEMPRESARIO	-0,14
ACTIVIDADEST. UNIVER.	-0,24
ACTIVIDADJUBILADO/MONTEPIADO	-0,24
ACTIVIDADNO INFORMADO	0,48
ACTIVIDADOBRREROS, AUXI.	-0,40
ACTIVIDADOFICIAL FF.AA.	0,29
ACTIVIDADPERS. DE SEGURIDAD	-0,16
ACTIVIDADPERSONAL FF. AA.	-0,04
ACTIVIDADPROF. UNIV, (INDP.)	-0,01
ACTIVIDADPROF. UNIV. (EMP.)	0,11
ACTIVIDADRENTISTA	0,28
ACTIVIDADTRABAJ. INDEPENDIENT	-0,15
ANIO_VEHICMUY ANTIGUOS	-0,28
ANIO_VEHICNUEVOS	1,46
ANIO_VEHICSEMI NUEVOS	0,52
ANTIGUEDAD0-6 MESES	-1,24
ANTIGUEDAD1-2 ANIOS	0,10
ANTIGUEDAD3-5 ANIOS	0,12
ANTIGUEDAD6-12 MESES	-1,71
EMPLEADO_VIGENTE	0,04
ENVIO_FACTURACIONHOGAR	-0,20
ENVIO_FACTURACIONOFICINA	-0,31

ESTADO_CIVILDIVORCIADO	0,07
ESTADO_CIVILSEPARADO	0,06
ESTADO_CIVILSOLTERO	-0,06
ESTADO_CIVILVIUDO	-0,10
GASTO_CMR_12MGASTA NORMAL	-0,09
GASTO_CMR_12MGASTA POCO	-0,16
GASTO_CMR_12MNO GASTA	-0,12
GASTO_RUBRO_AUTOMOTRIZ_12MGASTA NORMAL	-0,09
GASTO_RUBRO_AUTOMOTRIZ_12MGASTA POCO	-0,11
GASTO_RUBRO_AUTOMOTRIZ_12MNO GASTA	-0,20
GASTO_RUBRO_COMBUSTIBLE_12MGASTA NORMAL	0,11
GASTO_RUBRO_COMBUSTIBLE_12MGASTA POCO	0,05
GASTO_RUBRO_COMBUSTIBLE_12MNO GASTA	-0,10
GASTO_RUBRO_FARMACIAS_12MGASTA NORMAL	0,10
GASTO_RUBRO_FARMACIAS_12MGASTA POCO	0,08
GASTO_RUBRO_FARMACIAS_12MNO GASTA	0,15
GASTO_RUBRO_OTROS_12MGASTA NORMAL	-0,02
GASTO_RUBRO_OTROS_12MGASTA POCO	0,06
GASTO_RUBRO_OTROS_12MNO GASTA	-0,03
GASTO_RUBRO_PAGO_WEB_12MGASTA NORMAL	-0,06
GASTO_RUBRO_PAGO_WEB_12MGASTA POCO	-0,17
GASTO_RUBRO_PAGO_WEB_12MNO GASTA	-0,41
GASTO_RUBRO_RECAUDACION_12MGASTA NORMAL	-0,08
GASTO_RUBRO_RECAUDACION_12MGASTA POCO	-0,09
GASTO_RUBRO_RECAUDACION_12MNO GASTA	-0,19
GASTO_RUBRO_SALUD_12MGASTA NORMAL	-0,05
GASTO_RUBRO_SALUD_12MGASTA POCO	-0,11
GASTO_RUBRO_SALUD_12MNO GASTA	-0,19
GASTO_RUBRO_SPM_12MGASTA NORMAL	0,13
GASTO_RUBRO_SPM_12MGASTA POCO	0,16
GASTO_RUBRO_SPM_12MNO GASTA	0,27
GASTO_RUBRO_TIENDAS_DEPTO_12MGASTA NORMAL	0,02
GASTO_RUBRO_TIENDAS_DEPTO_12MGASTA POCO	0,06
GASTO_RUBRO_TIENDAS_DEPTO_12MNO GASTA	0,13
GASTO_RUBRO_TRANSPORTE_12MGASTA NORMAL	-0,18
GASTO_RUBRO_TRANSPORTE_12MGASTA POCO	0,06
GASTO_RUBRO_TRANSPORTE_12MNO GASTA	-0,23
GASTO_RUBRO_VIAJES_12MGASTA NORMAL	-0,11
GASTO_RUBRO_VIAJES_12MGASTA POCO	0,07
GASTO_RUBRO_VIAJES_12MNO GASTA	-0,04
GASTO_RUBRO_VIVIENDA_12MGASTA NORMAL	0,04
GASTO_RUBRO_VIVIENDA_12MGASTA POCO	-0,15

GASTO_RUBRO_VIVIENDA_12MNO GASTA	-0,01
GSEC2	0,02
GSEC3	0,00
GSED	-0,05
GSEE	0,09
HABITACIONCASA DE FAMILIARES	-0,11
HABITACIONCASA DE LA EMPRESA	-0,17
HABITACIONCASA DE LOS PADRES	0,01
HABITACIONCASA FISCAL	-0,06
HABITACIONCASA PROPIA	0,04
HABITO_PAGOMALO	0,01
HABITO_PAGOPESIMO	1,83
MES_COMPRA_CMERMUY ALTO USO	0,08
N_VEHICULOS1 AUTOS	0,45
N_VEHICULOS2 AUTOS	0,87
N_VEHICULOS3 AUTOS	0,85
N_VEHICULOS4 o + AUTOS	0,77
OTROS_ANT	-0,78
PRINCIPAL_ANT	0,17
RANGO_CUPO_BI1.250.0001 - 2.000.000	0,31
RANGO_CUPO_BI2.000.0001 - +	0,29
RANGO_CUPO_BI501.000 - 1.250.0000	0,21
RANGO_EDAD26 - 35	0,23
RANGO_EDAD36 - 45	0,21
RANGO_EDAD46 - 55	0,22
RANGO_EDAD56 - 65	0,27
RANGO_EDAD66 +	0,20
SEGMENTO_USO_CMREBAJA FRECUENCIA	0,06
SEGMENTO_USO_CMRFUGADO	-0,07
SEGMENTO_USO_CMRIINACTIVO	-0,08
SEGMENTO_USO_CMREMEDIA FRECUENCIA	0,12
SEGMENTO_USO_CMRNUEVOS	-0,54
SEGMENTO_USO_CMRPOR FUGARSE	0,07
SEGMENTO_VALOR_FALCLIENTE NORMAL	-0,11
SEXOM	0,01
TIPO_CLIENTECLIENTE NORMAL	-0,10
TIPO_CLIENTECLIENTE PREMIUM	0,16
TIPO_TARJETATARJETA CMR TRADICIONAL	-0,28
TIPO_TARJETATARJETA CMR VISA	-0,14
TIPO_TARJETAVISA PLATINUM	0,02
TIPO_TARJETAVISA SIGNATURE	0,04
VISITA_FALMEDIA VISITA	0,01
VISITA_FALNO VISITA	-0,58

Anexo 7: Coeficientes modelo auto con CMR
Fuente: Elaboración propia

Coeficientes modelo vida

VariableName	Estimate
OTROS_ANT	-2,42
PRINCIPAL_ANT	-0,08
RANGO_EDAD26 - 35	0,05
RANGO_EDAD36 - 45	-0,10
RANGO_EDAD46 - 55	0,34
RANGO_EDAD56 - 65	0,32
RANGO_EDAD66 +	-0,09
SEXOM	0,30
GSEC2	-0,23
GSEC3	-0,21
GSED	-0,45
GSEE	-0,30
ESTADO_CIVILDIVORCIADO	-0,56
ESTADO_CIVILSEPARADO	-0,03
ESTADO_CIVILSOLTERO	-0,16
ESTADO_CIVILVIUDO	-0,49
HABITACIONCASA DE FAMILIARES	0,28
HABITACIONCASA DE LA EMPRESA	0,21
HABITACIONCASA DE LOS PADRES	-0,02
HABITACIONCASA FISCAL	0,14
HABITACIONCASA PROPIA	-0,07
ACTIVIDADCOMISIONISTA	-0,28
ACTIVIDADDUEÑA DE CASA	-0,26
ACTIVIDADEMPL. DOMESTICA	0,91
ACTIVIDADEMPL. TECNICO	0,32
ACTIVIDADEMPRESARIO	0,17
ACTIVIDADEST. UNIVER.	-2,35
ACTIVIDADJUBILADO/MONTEPIADO	-0,07
ACTIVIDADOBREROS, AUXI.	0,24
ACTIVIDADOFICIAL FF.AA.	-0,45
ACTIVIDADPERS. DE SEGURIDAD	0,29
ACTIVIDADPERSONAL FF. AA.	-0,44
ACTIVIDADPROF. UNIV, (INDP.)	-0,12
ACTIVIDADPROF. UNIV. (EMP.)	-0,36
ACTIVIDADRENTISTA	0,02
ACTIVIDADTRABAJ. INDEPENDIENT	0,17
EMPLEADO_VIGENTE	-0,56

Anexo 8: Coeficientes modelo vida sin CMR
Fuente: Elaboración propia

VariableName	Estimate
ACTIVIDADCOMISIONISTA	-0,07
ACTIVIDADDUEÑA DE CASA	-0,46
ACTIVIDADEMPL. DOMESTICA	0,20
ACTIVIDADEMPL. TECNICO	0,20
ACTIVIDADEMPRESARIO	0,11
ACTIVIDADEST. UNIVER.	-1,40
ACTIVIDADJUBILADO/MONTEPIADO	-0,32
ACTIVIDADOBREROS, AUXI.	0,17
ACTIVIDADOFICIAL FF.AA.	-0,94
ACTIVIDADPERS. DE SEGURIDAD	0,21
ACTIVIDADPERSONAL FF. AA.	-0,65
ACTIVIDADPROF. UNIV, (INDP.)	-0,02
ACTIVIDADPROF. UNIV. (EMP.)	-0,17
ACTIVIDADRENTISTA	0,71
ACTIVIDADTRABAJ. INDEPENDIENT	0,20
ANIO_VEHICMUY ANTIGUOS	0,14
ANIO_VEHICNUEVOS	-0,09
ANIO_VEHICSEMI NUEVOS	-0,23
ANTIGUEDAD0-6 MESES	-1,97
ANTIGUEDAD1-2 ANIOS	0,68
ANTIGUEDAD3-5 ANIOS	0,11
ANTIGUEDAD6-12 MESES	-1,41
DESC_SITUACTANO NORMAL	-0,64
DESC_SITUACTANORMAL	-1,09
EMPLEADO_VIGENTE	-0,20
ESTADO_CIVILDIVORCIADO	0,20
ESTADO_CIVILSEPARADO	0,01
ESTADO_CIVILSOLTERO	-0,07
ESTADO_CIVILVIUDO	0,02
GASTO_CMV_12MGASTA NORMAL	0,03
GASTO_CMV_12MGASTA POCO	-0,09
GASTO_CMV_12MNO GASTA	-0,23
GASTO_RUBRO_AUTOMOTRIZ_12MGASTA NORMAL	-0,14
GASTO_RUBRO_AUTOMOTRIZ_12MGASTA POCO	-0,23
GASTO_RUBRO_AUTOMOTRIZ_12MNO GASTA	-0,06
GASTO_RUBRO_COMBUSTIBLE_12MGASTA NORMAL	0,09
GASTO_RUBRO_COMBUSTIBLE_12MGASTA POCO	-0,01
GASTO_RUBRO_COMBUSTIBLE_12MNO GASTA	0,01
GASTO_RUBRO_COMUNICACION_12MGASTA NORMAL	-0,04

GASTO_RUBRO_COMUNICACION_12MGASTA POCO	-0,19
GASTO_RUBRO_COMUNICACION_12MNO GASTA	-0,24
GASTO_RUBRO_EDUCACION_12MGASTA NORMAL	-0,02
GASTO_RUBRO_EDUCACION_12MGASTA POCO	-0,04
GASTO_RUBRO_EDUCACION_12MNO GASTA	0,18
GASTO_RUBRO_EMPRESA_12MGASTA NORMAL	0,05
GASTO_RUBRO_EMPRESA_12MGASTA POCO	-0,24
GASTO_RUBRO_EMPRESA_12MNO GASTA	-0,18
GASTO_RUBRO_ENTRETENCION_12MGASTA NORMAL	-0,11
GASTO_RUBRO_ENTRETENCION_12MGASTA POCO	-0,08
GASTO_RUBRO_ENTRETENCION_12MNO GASTA	0,03
GASTO_RUBRO_FARMACIAS_12MGASTA NORMAL	-0,01
GASTO_RUBRO_FARMACIAS_12MGASTA POCO	-0,05
GASTO_RUBRO_FARMACIAS_12MNO GASTA	-0,12
GASTO_RUBRO_OTROS_12MGASTA NORMAL	-0,09
GASTO_RUBRO_OTROS_12MGASTA POCO	-0,18
GASTO_RUBRO_OTROS_12MNO GASTA	-0,05
GASTO_RUBRO_RECAUDACION_12MGASTA NORMAL	0,07
GASTO_RUBRO_RECAUDACION_12MGASTA POCO	0,04
GASTO_RUBRO_RECAUDACION_12MNO GASTA	-0,23
GASTO_RUBRO_RESTAURANT_12MGASTA NORMAL	0,23
GASTO_RUBRO_RESTAURANT_12MGASTA POCO	0,19
GASTO_RUBRO_RESTAURANT_12MNO GASTA	0,20
GASTO_RUBRO_SALUD_12MGASTA NORMAL	0,05
GASTO_RUBRO_SALUD_12MGASTA POCO	0,13
GASTO_RUBRO_SALUD_12MNO GASTA	0,12
GASTO_RUBRO_SPM_12MGASTA NORMAL	-0,13
GASTO_RUBRO_SPM_12MGASTA POCO	-0,23
GASTO_RUBRO_SPM_12MNO GASTA	-0,19
GASTO_RUBRO_TIENDAS_DEPTO_12MGASTA NORMAL	0,14
GASTO_RUBRO_TIENDAS_DEPTO_12MGASTA POCO	0,17
GASTO_RUBRO_TIENDAS_DEPTO_12MNO GASTA	0,07
GASTO_RUBRO_TRANSPORTE_12MGASTA NORMAL	0,00
GASTO_RUBRO_TRANSPORTE_12MGASTA POCO	0,08
GASTO_RUBRO_TRANSPORTE_12MNO GASTA	0,02
GASTO_RUBRO_VIAJES_12MGASTA NORMAL	0,16
GASTO_RUBRO_VIAJES_12MGASTA POCO	0,08
GASTO_RUBRO_VIAJES_12MNO GASTA	0,00
GASTO_RUBRO_VIVIENDA_12MGASTA NORMAL	0,18
GASTO_RUBRO_VIVIENDA_12MGASTA POCO	0,13
GASTO_RUBRO_VIVIENDA_12MNO GASTA	0,14
GSEC2	-0,19

GSEC3	0,01
GSED	0,10
GSEE	0,03
HABITACIONCASA DE FAMILIARES	0,19
HABITACIONCASA DE LA EMPRESA	0,12
HABITACIONCASA DE LOS PADRES	-0,12
HABITACIONCASA FISCAL	0,01
HABITACIONCASA PROPIA	-0,11
MES_COMPRA_CMERMUY ALTO USO	0,62
N_VEHICULOS1 AUTOS	-0,19
N_VEHICULOS2 AUTOS	-0,04
N_VEHICULOS3 AUTOS	-0,07
N_VEHICULOS4 o + AUTOS	-0,03
OTROS_ANT	-0,28
PRINCIPAL_ANT	0,19
RANGO_CUPO_BI1.250.0001 - 2.000.000	0,14
RANGO_CUPO_BI2.000.0001 - +	0,02
RANGO_CUPO_BI501.000 - 1.250.0000	0,17
RANGO_EDAD26 - 35	-0,55
RANGO_EDAD36 - 45	-0,32
RANGO_EDAD46 - 55	0,01
RANGO_EDAD56 - 65	0,06
RANGO_EDAD66 +	-0,83
RECENCY_FALNO RECIENTE	0,01
RECENCY_FALRECIENTE	0,12
SEGMENTO_USO_CMREBAJA FRECUENCIA	0,18
SEGMENTO_USO_CMRFUGADO	0,09
SEGMENTO_USO_CMRIINACTIVO	-0,04
SEGMENTO_USO_CMRMEDIA FRECUENCIA	0,52
SEGMENTO_USO_CMRNUEVOS	-0,07
SEGMENTO_USO_CMRPOR FUGARSE	0,13
SEGMENTO_VALOR_FALCLIENTE NORMAL	0,00
SEXOM	0,04
TIPO_CLIENTECLIENTE NORMAL	-0,31
TIPO_CLIENTECLIENTE PREMIUM	-0,25
TIPO_TARJETATARJETA CMR TRADICIONAL	-0,06
TIPO_TARJETATARJETA CMR VISA	0,02
TIPO_TARJETAVISA PLATINUM	0,55
TIPO_TARJETAVISA SIGNATURE	0,00
TRX_CMRI_12M1 TRX	0,26
TRX_CMRI_12M2 TRX	0,40
TRX_CMRI_12M3 o + TRX	0,46
VISITA_FALMEDIA VISITA	-0,26

Anexo 9: Coeficientes modelo vida con CMR
Fuente: Elaboración propia

Coeficientes modelo vida con bonificación

VariableName	Estimate
ACTIVIDADCOMISIONISTA	-0,06
ACTIVIDADDUEÑA DE CASA	-0,87
ACTIVIDADEMPL. DOMESTICA	0,52
ACTIVIDADEMPL. TECNICO	-0,10
ACTIVIDADEMPRESARIO	-0,17
ACTIVIDADEST. UNIVER.	-1,20
ACTIVIDADJUBILADO/MONTEPIADO	-1,10
ACTIVIDADOBREROS, AUXI.	-0,07
ACTIVIDADOFICIAL FF.AA.	-16,77
ACTIVIDADPERS. DE SEGURIDAD	0,22
ACTIVIDADPERSONAL FF. AA.	-0,60
ACTIVIDADPROF. UNIV, (INDP.)	-0,45
ACTIVIDADPROF. UNIV. (EMP.)	-0,71
ACTIVIDADRENTISTA	-1,04
ACTIVIDADTRABAJ. INDEPENDIENT	-0,17
EMPLEADO_VIGENTE	1,51
ESTADO_CIVILDIVORCIADO	0,77
ESTADO_CIVILSEPARADO	0,91
ESTADO_CIVILSOLTERO	-0,05
ESTADO_CIVILVIUDO	-0,04
GSEC2	0,06
GSEC3	0,02
GSED	-0,24
GSEE	-0,12
HABITACIONCASA DE FAMILIARES	-0,30
HABITACIONCASA DE LA EMPRESA	-0,04
HABITACIONCASA DE LOS PADRES	-0,01
HABITACIONCASA FISCAL	-0,06
HABITACIONCASA PROPIA	0,16
OTROS_ANT	-0,92
PRINCIPAL_ANT	1,70
RANGO_EDAD26 - 35	-0,23
RANGO_EDAD36 - 45	-0,53
RANGO_EDAD46 - 55	-0,27
RANGO_EDAD56 - 65	-0,21
RANGO_EDAD66 +	-2,35
SEXOM	-0,15

Anexo 10: Coeficientes modelo vida con bonificación sin CMR
Fuente: Elaboración propia

VariableName	Estimate
ACTIVIDADCOMISIONISTA	0,03
ACTIVIDADDUENÑA DE CASA	-0,54
ACTIVIDADEMPL. DOMESTICA	0,18
ACTIVIDADEMPL. TECNICO	0,00
ACTIVIDADEMPRESARIO	0,02
ACTIVIDADEST. UNIVER.	-1,58
ACTIVIDADJUBILADO/MONTEPIADO	-0,45
ACTIVIDADOBREROS, AUXI.	0,11
ACTIVIDADOFICIAL FF.AA.	-1,29
ACTIVIDADPERS. DE SEGURIDAD	0,33
ACTIVIDADPERSONAL FF. AA.	-0,66
ACTIVIDADPROF. UNIV, (INDP.)	0,04
ACTIVIDADPROF. UNIV. (EMP.)	-0,31
ACTIVIDADRENTISTA	-0,48
ACTIVIDADTRABAJ. INDEPENDIENT	0,03
ANIO_VEHICMUY ANTIGUOS	0,02
ANIO_VEHICNUEVOS	0,02
ANIO_VEHICSEMI NUEVOS	-0,11
ANTIGUEDAD0-6 MESES	-0,65
ANTIGUEDAD1-2 ANIOS	0,40
ANTIGUEDAD3-5 ANIOS	0,04
ANTIGUEDAD6-12 MESES	0,11
EMPLEADO_VIGENTE	0,20
ENVIO_FACTURACIONHOGAR	-0,04
ENVIO_FACTURACIONOFICINA	-0,22
ESTADO_CIVILDIVORCIADO	-0,77
ESTADO_CIVILSEPARADO	0,02
ESTADO_CIVILSOLTERO	0,08
ESTADO_CIVILVIUDO	0,49
GASTO_CMV_12MGASTA NORMAL	-0,10
GASTO_CMV_12MGASTA POCO	-0,26
GASTO_CMV_12MNO GASTA	-0,30
GASTO_RUBRO_COMUNICACION_12MGASTA NORMAL	-0,02
GASTO_RUBRO_COMUNICACION_12MGASTA POCO	-0,23
GASTO_RUBRO_COMUNICACION_12MNO GASTA	-0,33
GASTO_RUBRO_EDUCACION_12MGASTA NORMAL	-0,33
GASTO_RUBRO_EDUCACION_12MGASTA POCO	-0,12
GASTO_RUBRO_EDUCACION_12MNO GASTA	-0,13
GASTO_RUBRO_EMPRESA_12MGASTA NORMAL	-0,11

GASTO_RUBRO_EMPRESA_12MGASTA POCO	-0,28
GASTO_RUBRO_EMPRESA_12MNO GASTA	-0,05
GASTO_RUBRO_FARMACIAS_12MGASTA NORMAL	0,02
GASTO_RUBRO_FARMACIAS_12MGASTA POCO	-0,04
GASTO_RUBRO_FARMACIAS_12MNO GASTA	-0,10
GASTO_RUBRO_MEJ_HOGAR_12MGASTA NORMAL	-0,09
GASTO_RUBRO_MEJ_HOGAR_12MGASTA POCO	-0,20
GASTO_RUBRO_MEJ_HOGAR_12MNO GASTA	-0,12
GASTO_RUBRO_OTROS_12MGASTA NORMAL	0,03
GASTO_RUBRO_OTROS_12MGASTA POCO	-0,03
GASTO_RUBRO_OTROS_12MNO GASTA	-0,02
GASTO_RUBRO_PAGO_WEB_12MGASTA NORMAL	-0,05
GASTO_RUBRO_PAGO_WEB_12MGASTA POCO	-0,24
GASTO_RUBRO_PAGO_WEB_12MNO GASTA	-0,06
GASTO_RUBRO_RECAUDACION_12MGASTA NORMAL	-0,13
GASTO_RUBRO_RECAUDACION_12MGASTA POCO	-0,04
GASTO_RUBRO_RECAUDACION_12MNO GASTA	-0,43
GASTO_RUBRO_RESTAURANT_12MGASTA NORMAL	0,07
GASTO_RUBRO_RESTAURANT_12MGASTA POCO	0,24
GASTO_RUBRO_RESTAURANT_12MNO GASTA	0,06
GASTO_RUBRO_SALUD_12MGASTA NORMAL	-0,21
GASTO_RUBRO_SALUD_12MGASTA POCO	-0,04
GASTO_RUBRO_SALUD_12MNO GASTA	0,01
GASTO_RUBRO_SPM_12MGASTA NORMAL	0,08
GASTO_RUBRO_SPM_12MGASTA POCO	0,02
GASTO_RUBRO_SPM_12MNO GASTA	0,00
GASTO_RUBRO_TRANSPORTE_12MGASTA NORMAL	-0,13
GASTO_RUBRO_TRANSPORTE_12MGASTA POCO	-0,06
GASTO_RUBRO_TRANSPORTE_12MNO GASTA	-0,09
GASTO_RUBRO_VIAJES_12MGASTA NORMAL	-0,15
GASTO_RUBRO_VIAJES_12MGASTA POCO	-0,02
GASTO_RUBRO_VIAJES_12MNO GASTA	0,04
GASTO_RUBRO_VIVIENDA_12MGASTA NORMAL	0,02
GASTO_RUBRO_VIVIENDA_12MGASTA POCO	-0,09
GASTO_RUBRO_VIVIENDA_12MNO GASTA	-0,07
GSEC2	0,13
GSEC3	0,20
GSED	0,22
GSEE	0,26
HABITACIONCASA DE FAMILIARES	-0,05
HABITACIONCASA DE LA EMPRESA	0,26
HABITACIONCASA DE LOS PADRES	-0,08
HABITACIONCASA FISCAL	-0,09

HABITACIONCASA PROPIA	-0,07
HABITO_PAGOMALO	0,70
HABITO_PAGOPEσιμο	-9,09
MES_COMPRA_CMRYMUY ALTO USO	0,17
N_VEHICULOS1 AUTOS	-0,05
N_VEHICULOS2 AUTOS	0,08
N_VEHICULOS3 AUTOS	0,01
N_VEHICULOS4 o + AUTOS	0,26
OTROS_ANT	-0,05
PRINCIPAL_ANT	0,37
RANGO_CUPO_BI1.250.0001 - 2.000.000	-0,01
RANGO_CUPO_BI2.000.0001 - +	-0,09
RANGO_CUPO_BI501.000 - 1.250.0000	0,15
RANGO_EDAD26 - 35	-0,17
RANGO_EDAD36 - 45	0,14
RANGO_EDAD46 - 55	0,48
RANGO_EDAD56 - 65	0,61
RANGO_EDAD66 +	-1,62
SEGMENTO_USO_CMRYBAJA FRECUENCIA	0,11
SEGMENTO_USO_CMRYFUGADO	-0,17
SEGMENTO_USO_CMRYINACTIVO	-0,49
SEGMENTO_USO_CMRYMEDIA FRECUENCIA	0,15
SEGMENTO_USO_CMRYNUEVOS	-0,17
SEGMENTO_USO_CMRYPOR FUGARSE	0,10
SEGMENTO_VALOR_FALCLIENTE NORMAL	-0,22
SEXOM	0,18
TIPO_CLIENTECLIENTE NORMAL	0,08
TIPO_CLIENTECLIENTE PREMIUM	0,58
TIPO_TARJETATARJETA CMRY TRADICIONAL	-0,33
TIPO_TARJETATARJETA CMRY VISA	-0,35
TIPO_TARJETAVISA PLATINUM	0,18
TIPO_TARJETAVISA SIGNATURE	0,14
VISITA_FALMEDIA VISITA	0,00
VISITA_FALNO VISITA	3,78
VISITA_FALPOCA VISITA	0,01

*Anexo 11: Coeficientes modelo vida con bonificación con CMR
Fuente: Elaboración propia*

Coeficientes modelo salud

VariableName	Estimate
ACTIVIDADCOMISIONISTA	1,37
ACTIVIDADDUEÑA DE CASA	-0,84
ACTIVIDADEMPL. DOMESTICA	-16,67

ACTIVIDADEMPL. TECNICO	0,15
ACTIVIDADEMPRESARIO	0,64
ACTIVIDADEST. UNIVER.	0,08
ACTIVIDADJUBILADO/MONTEPIADO	0,39
ACTIVIDADOBREROS, AUXI.	-0,39
ACTIVIDADOFICIAL FF.AA.	-17,12
ACTIVIDADPERS. DE SEGURIDAD	-16,95
ACTIVIDADPERSONAL FF. AA.	0,09
ACTIVIDADPROF. UNIV, (INDP.)	0,42
ACTIVIDADPROF. UNIV. (EMP.)	0,04
ACTIVIDADRENTISTA	-17,06
ACTIVIDADTRABAJ. INDEPENDIENT	-0,15
EMPLEADO_VIGENTE	1,68
ESTADO_CIVILDIVORCIADO	-18,65
ESTADO_CIVILSEPARADO	-0,14
ESTADO_CIVILSOLTERO	-0,51
ESTADO_CIVILVIUDO	0,25
GSEC2	-0,15
GSEC3	-0,83
GSED	-0,93
GSEE	-17,31
HABITACIONCASA DE FAMILIARES	-0,05
HABITACIONCASA DE LA EMPRESA	-17,11
HABITACIONCASA DE LOS PADRES	0,05
HABITACIONCASA FISCAL	-0,79
HABITACIONCASA PROPIA	-0,17
OTROS_ANT	-1,22
PRINCIPAL_ANT	1,45
RANGO_EDAD26 - 35	-0,30
RANGO_EDAD36 - 45	-0,36
RANGO_EDAD46 - 55	-0,14
RANGO_EDAD56 - 65	-0,33
RANGO_EDAD66 +	-1,96
SEXOM	-0,40

*Anexo 12: Coeficientes modelo salud sin CMR
Fuente: Elaboración propia*

VariableName	Estimate
ACTIVIDADCOMISIONISTA	0,05
ACTIVIDADDUEÑA DE CASA	-1,31
ACTIVIDADEMPL. DOMESTICA	-0,50
ACTIVIDADEMPL. TECNICO	-0,02
ACTIVIDADEMPRESARIO	0,00

ACTIVIDADEST. UNIVER.	-0,78
ACTIVIDADJUBILADO/MONTEPIADO	-0,34
ACTIVIDADOBREROS, AUXI.	0,05
ACTIVIDADOFICIAL FF.AA.	-0,41
ACTIVIDADPERS. DE SEGURIDAD	0,12
ACTIVIDADPERSONAL FF. AA.	-0,51
ACTIVIDADPROF. UNIV, (INDP.)	0,19
ACTIVIDADPROF. UNIV. (EMP.)	-0,06
ACTIVIDADRENTISTA	0,96
ACTIVIDADTRABAJ. INDEPENDIENT	0,20
ANIO_VEHICMUY ANTIGUOS	0,11
ANIO_VEHICNUEVOS	-0,13
ANIO_VEHICSEMI NUEVOS	-0,31
ANTIGUEDAD0-6 MESES	0,12
ANTIGUEDAD1-2 ANIOS	0,52
ANTIGUEDAD3-5 ANIOS	0,10
ANTIGUEDAD6-12 MESES	0,42
DESC_SITUACTANO NORMAL	0,22
DESC_SITUACTANORMAL	-0,18
EMPLEADO_VIGENTE	0,17
ENVIO_FACTURACIONHOGAR	-0,30
ENVIO_FACTURACIONOFICINA	-0,05
ESTADO_CIVILDIVORCIADO	0,23
ESTADO_CIVILSEPARADO	0,11
ESTADO_CIVILSOLTERO	0,06
ESTADO_CIVILVIUDO	0,53
GASTO_CM_12MGASTA NORMAL	-0,13
GASTO_CM_12MGASTA POCO	-0,61
GASTO_CM_12MNO GASTA	-0,45
GASTO_RUBRO_AUTOMOTRIZ_12MGASTA NORMAL	0,02
GASTO_RUBRO_AUTOMOTRIZ_12MGASTA POCO	0,04
GASTO_RUBRO_AUTOMOTRIZ_12MNO GASTA	0,13
GASTO_RUBRO_COMBUSTIBLE_12MGASTA NORMAL	0,13
GASTO_RUBRO_COMBUSTIBLE_12MGASTA POCO	0,04
GASTO_RUBRO_COMBUSTIBLE_12MNO GASTA	0,01
GASTO_RUBRO_COMUNICACION_12MGASTA NORMAL	-0,06
GASTO_RUBRO_COMUNICACION_12MGASTA POCO	-0,34
GASTO_RUBRO_COMUNICACION_12MNO GASTA	-0,37
GASTO_RUBRO_EMPRESA_12MGASTA NORMAL	-0,20
GASTO_RUBRO_EMPRESA_12MGASTA POCO	-0,45
GASTO_RUBRO_EMPRESA_12MNO GASTA	-0,35
GASTO_RUBRO_ENTRETENCION_12MGASTA NORMAL	0,26

GASTO_RUBRO_ENTRETENCION_12MGASTA POCO	0,11
GASTO_RUBRO_ENTRETENCION_12MNO GASTA	0,26
GASTO_RUBRO_FARMACIAS_12MGASTA NORMAL	-0,17
GASTO_RUBRO_FARMACIAS_12MGASTA POCO	-0,09
GASTO_RUBRO_FARMACIAS_12MNO GASTA	-0,15
GASTO_RUBRO_OTROS_12MGASTA NORMAL	0,08
GASTO_RUBRO_OTROS_12MGASTA POCO	0,01
GASTO_RUBRO_OTROS_12MNO GASTA	-0,01
GASTO_RUBRO_PAGO_WEB_12MGASTA NORMAL	-0,24
GASTO_RUBRO_PAGO_WEB_12MGASTA POCO	-0,06
GASTO_RUBRO_PAGO_WEB_12MNO GASTA	-0,01
GASTO_RUBRO_RECAUDACION_12MGASTA NORMAL	-0,17
GASTO_RUBRO_RECAUDACION_12MGASTA POCO	0,01
GASTO_RUBRO_RECAUDACION_12MNO GASTA	-0,35
GASTO_RUBRO_RESTAURANT_12MGASTA NORMAL	0,04
GASTO_RUBRO_RESTAURANT_12MGASTA POCO	0,10
GASTO_RUBRO_RESTAURANT_12MNO GASTA	0,22
GASTO_RUBRO_SPM_12MGASTA NORMAL	-0,09
GASTO_RUBRO_SPM_12MGASTA POCO	-0,10
GASTO_RUBRO_SPM_12MNO GASTA	-0,08
GASTO_RUBRO_VIAJES_12MGASTA NORMAL	-0,14
GASTO_RUBRO_VIAJES_12MGASTA POCO	-0,06
GASTO_RUBRO_VIAJES_12MNO GASTA	-0,07
GASTO_RUBRO_VIVIENDA_12MGASTA NORMAL	0,12
GASTO_RUBRO_VIVIENDA_12MGASTA POCO	-0,12
GASTO_RUBRO_VIVIENDA_12MNO GASTA	-0,05
GSEC2	0,05
GSEC3	0,14
GSED	0,17
GSEE	-0,30
HABITACIONCASA DE FAMILIARES	0,28
HABITACIONCASA DE LA EMPRESA	0,11
HABITACIONCASA DE LOS PADRES	0,03
HABITACIONCASA FISCAL	-0,12
HABITACIONCASA PROPIA	-0,03
HABITO_PAGOMALO	0,00
HABITO_PAGOPESIMO	-9,82
MES_COMPRA_CMARMUY ALTO USO	0,10
N_VEHICULOS1 AUTOS	-0,09
N_VEHICULOS2 AUTOS	0,02
N_VEHICULOS3 AUTOS	-0,01
N_VEHICULOS4 o + AUTOS	0,11
PRINCIPAL_ANT	0,38

RANGO_CUPO_BI1.250.0001 - 2.000.000	-0,01
RANGO_CUPO_BI2.000.0001 - +	-0,11
RANGO_CUPO_BI501.000 - 1.250.0000	-0,05
RANGO_EDAD26 - 35	0,12
RANGO_EDAD36 - 45	0,37
RANGO_EDAD46 - 55	0,86
RANGO_EDAD56 - 65	0,44
RANGO_EDAD66 +	0,31
RECENCY_FALNO RECIENTE	-0,29
RECENCY_FALRECIENTE	-0,25
SEGMENTO_USO_CMRTAJA FRECUENCIA	0,04
SEGMENTO_USO_CMRFUGADO	0,40
SEGMENTO_USO_CMRIACTIVO	-0,76
SEGMENTO_USO_CMRTEDIA FRECUENCIA	0,02
SEGMENTO_USO_CMRTUEVOS	-0,16
SEGMENTO_USO_CMRTOR FUGARSE	0,05
SEGMENTO_VALOR_FALCLIENTE NORMAL	0,01
SEXOM	-0,07
TIPO_CLIENTECLIENTE NORMAL	-0,83
TIPO_CLIENTECLIENTE PREMIUM	-0,49
TIPO_TARJETATARJETA CMR TRADICIONAL	0,12
TIPO_TARJETATARJETA CMR VISA	0,25
TIPO_TARJETAVISA PLATINUM	0,46
TIPO_TARJETAVISA SIGNATURE	0,08
VISITA_FALMEDIA VISITA	0,20
VISITA_FALPOCA VISITA	0,09

*Anexo 13: Coeficientes modelo salud con CMR
Fuente: Elaboración propia*

Coeficientes modelo hogar

VariableName	Estimate
ACTIVIDADCOMISIONISTA	1,41
ACTIVIDADDUEÑA DE CASA	-0,40
ACTIVIDADEMPL. DOMESTICA	0,96
ACTIVIDADEMPL. TECNICO	-0,41
ACTIVIDADEMPRESARIO	0,10
ACTIVIDADEST. UNIVER.	-0,03
ACTIVIDADJUBILADO/MONTEPIADO	0,32
ACTIVIDADOBRREROS, AUXI.	-0,40
ACTIVIDADOFICIAL FF.AA.	0,82
ACTIVIDADPERS. DE SEGURIDAD	-0,27

ACTIVIDADPERSONAL FF. AA.	0,22
ACTIVIDADPROF. UNIV, (INDP.)	0,08
ACTIVIDADPROF. UNIV. (EMP.)	0,17
ACTIVIDADRENTISTA	-0,13
ACTIVIDADTRABAJ. INDEPENDIENT	-0,07
EMPLEADO_VIGENTE	2,17
ESTADO_CIVILDIVORCIADO	-0,48
ESTADO_CIVILSEPARADO	-0,87
ESTADO_CIVILSOLTERO	-0,05
ESTADO_CIVILVIUDO	0,46
GSEC2	-0,32
GSEC3	-0,63
GSED	-1,19
GSEE	-1,71
HABITACIONCASA DE FAMILIARES	-0,17
HABITACIONCASA DE LA EMPRESA	-0,41
HABITACIONCASA DE LOS PADRES	-0,11
HABITACIONCASA FISCAL	-0,14
HABITACIONCASA PROPIA	-0,05
OTROS_ANT	-1,48
PRINCIPAL_ANT	1,28
RANGO_EDAD26 - 35	0,78
RANGO_EDAD36 - 45	0,78
RANGO_EDAD46 - 55	0,97
RANGO_EDAD56 - 65	1,07
RANGO_EDAD66 +	1,13
SEXOM	-0,20

*Anexo 14: Coeficientes modelo hogar sin CMR
Fuente: Elaboración propia*

VariableName	Estimate
ACTIVIDADCOMISIONISTA	0,74
ACTIVIDADDUEÑA DE CASA	-0,43
ACTIVIDADEMPL. DOMESTICA	0,19
ACTIVIDADEMPL. TECNICO	-0,03
ACTIVIDADEMPRESARIO	-0,54
ACTIVIDADEST. UNIVER.	-0,36
ACTIVIDADJUBILADO/MONTEPIADO	0,03
ACTIVIDADOBRREROS, AUXI.	0,10
ACTIVIDADOFICIAL FF.AA.	0,03
ACTIVIDADPERS. DE SEGURIDAD	-0,46
ACTIVIDADPERSONAL FF. AA.	0,37
ACTIVIDADPROF. UNIV, (INDP.)	-0,09

ACTIVIDADPROF. UNIV. (EMP.)	-0,17
ACTIVIDADRENTISTA	0,66
ACTIVIDADTRABAJ. INDEPENDIENT	0,07
ANIO_VEHICMUY ANTIGUOS	0,11
ANIO_VEHICNUEVOS	0,65
ANIO_VEHICSEMI NUEVOS	0,16
ANTIGUEDAD0-6 MESES	-0,70
ANTIGUEDAD1-2 ANIOS	0,00
ANTIGUEDAD3-5 ANIOS	-0,12
ANTIGUEDAD6-12 MESES	-0,82
DESC_SITUACTANO NORMAL	-1,45
DESC_SITUACTANORMAL	-1,65
EMPLEADO_VIGENTE	0,51
ENVIO_FACTURACIONHOGAR	0,03
ENVIO_FACTURACIONOFICINA	-0,33
ESTADO_CIVILDIVORCIADO	0,14
ESTADO_CIVILSEPARADO	-0,15
ESTADO_CIVILSOLTERO	0,04
ESTADO_CIVILVIUDO	-0,10
GASTO_CM_12MGASTA NORMAL	0,08
GASTO_CM_12MGASTA POCO	-0,10
GASTO_CM_12MNO GASTA	-0,52
GASTO_RUBRO_AUTOMOTRIZ_12MGASTA NORMAL	-0,10
GASTO_RUBRO_AUTOMOTRIZ_12MGASTA POCO	-0,01
GASTO_RUBRO_AUTOMOTRIZ_12MNO GASTA	-0,12
GASTO_RUBRO_COMBUSTIBLE_12MGASTA NORMAL	0,07
GASTO_RUBRO_COMBUSTIBLE_12MGASTA POCO	0,26
GASTO_RUBRO_COMBUSTIBLE_12MNO GASTA	0,24
GASTO_RUBRO_COMUNICACION_12MGASTA NORMAL	-0,07
GASTO_RUBRO_COMUNICACION_12MGASTA POCO	-0,17
GASTO_RUBRO_COMUNICACION_12MNO GASTA	-0,13
GASTO_RUBRO_EDUCACION_12MGASTA NORMAL	-0,25
GASTO_RUBRO_EDUCACION_12MGASTA POCO	-0,17
GASTO_RUBRO_EDUCACION_12MNO GASTA	0,00
GASTO_RUBRO_ENTRETENCION_12MGASTA NORMAL	-0,06
GASTO_RUBRO_ENTRETENCION_12MGASTA POCO	0,22
GASTO_RUBRO_ENTRETENCION_12MNO GASTA	0,19
GASTO_RUBRO_FARMACIAS_12MGASTA NORMAL	-0,22
GASTO_RUBRO_FARMACIAS_12MGASTA POCO	-0,01
GASTO_RUBRO_FARMACIAS_12MNO GASTA	-0,26
GASTO_RUBRO_MEJ_HOGAR_12MGASTA NORMAL	-0,22
GASTO_RUBRO_MEJ_HOGAR_12MGASTA POCO	-0,31

GASTO_RUBRO_MEJ_HOGAR_12MNO GASTA	-0,30
GASTO_RUBRO_OTROS_12MGASTA NORMAL	-0,26
GASTO_RUBRO_OTROS_12MGASTA POCO	-0,18
GASTO_RUBRO_OTROS_12MNO GASTA	-0,04
GASTO_RUBRO_PAGO_WEB_12MGASTA NORMAL	0,02
GASTO_RUBRO_PAGO_WEB_12MGASTA POCO	-0,09
GASTO_RUBRO_PAGO_WEB_12MNO GASTA	-0,18
GASTO_RUBRO_RECAUDACION_12MGASTA NORMAL	-0,15
GASTO_RUBRO_RECAUDACION_12MGASTA POCO	0,02
GASTO_RUBRO_RECAUDACION_12MNO GASTA	-0,18
GASTO_RUBRO_RESTAURANT_12MGASTA NORMAL	-0,01
GASTO_RUBRO_RESTAURANT_12MGASTA POCO	0,00
GASTO_RUBRO_RESTAURANT_12MNO GASTA	0,06
GASTO_RUBRO_SALUD_12MGASTA NORMAL	0,10
GASTO_RUBRO_SALUD_12MGASTA POCO	-0,03
GASTO_RUBRO_SALUD_12MNO GASTA	0,02
GASTO_RUBRO_SPM_12MGASTA NORMAL	-0,04
GASTO_RUBRO_SPM_12MGASTA POCO	0,05
GASTO_RUBRO_SPM_12MNO GASTA	0,04
GASTO_RUBRO_TIENDAS_DEPTO_12MGASTA NORMAL	0,12
GASTO_RUBRO_TIENDAS_DEPTO_12MGASTA POCO	0,14
GASTO_RUBRO_TIENDAS_DEPTO_12MNO GASTA	0,06
GASTO_RUBRO_TRANSPORTE_12MGASTA NORMAL	0,43
GASTO_RUBRO_TRANSPORTE_12MGASTA POCO	0,17
GASTO_RUBRO_TRANSPORTE_12MNO GASTA	0,22
GASTO_RUBRO_VIAJES_12MGASTA NORMAL	0,15
GASTO_RUBRO_VIAJES_12MGASTA POCO	0,13
GASTO_RUBRO_VIAJES_12MNO GASTA	0,00
GASTO_RUBRO_VIVIENDA_12MGASTA NORMAL	-0,08
GASTO_RUBRO_VIVIENDA_12MGASTA POCO	0,14
GASTO_RUBRO_VIVIENDA_12MNO GASTA	-0,08
GSEC2	0,09
GSEC3	0,00
GSED	0,10
GSEE	0,00
HABITACIONCASA DE FAMILIARES	0,03
HABITACIONCASA DE LA EMPRESA	-0,51
HABITACIONCASA DE LOS PADRES	0,10
HABITACIONCASA FISCAL	0,47
HABITACIONCASA PROPIA	0,61
HABITO_PAGOMALO	0,71
HABITO_PAGOPESIMO	-15,31

MES_COMPRA_CMERMUY ALTO USO	0,20
N_VEHICULOS1 AUTOS	-0,07
N_VEHICULOS2 AUTOS	-0,10
N_VEHICULOS3 AUTOS	-0,07
N_VEHICULOS4 o + AUTOS	0,11
OTROS_ANT	-0,56
PRINCIPAL_ANT	0,35
RANGO_CUPO_BI1.250.0001 - 2.000.000	0,22
RANGO_CUPO_BI2.000.0001 - +	0,22
RANGO_CUPO_BI501.000 - 1.250.0000	0,02
RANGO_EDAD26 - 35	0,37
RANGO_EDAD36 - 45	0,60
RANGO_EDAD46 - 55	0,91
RANGO_EDAD56 - 65	1,31
RANGO_EDAD66 +	1,44
RECENCY_FALNO RECIENTE	-0,61
RECENCY_FALRECIENTE	-0,67
SEGMENTO_USO_CMRTAJA FRECUENCIA	-0,16
SEGMENTO_USO_CMRFUGADO	-0,20
SEGMENTO_USO_CMRIINACTIVO	0,04
SEGMENTO_USO_CMRMEDIA FRECUENCIA	-0,03
SEGMENTO_USO_CMRNUEVOS	-14,24
SEGMENTO_USO_CMRPOR FUGARSE	-0,17
SEGMENTO_VALOR_FALCLIENTE NORMAL	-0,07
SEXOM	0,05
TIPO_CLIENTECLIENTE NORMAL	-0,46
TIPO_CLIENTECLIENTE PREMIUM	-0,27
TIPO_TARJETATARJETA CMR TRADICIONAL	-0,50
TIPO_TARJETATARJETA CMR VISA	-0,21
TIPO_TARJETAVISA PLATINUM	0,12
TIPO_TARJETAVISA SIGNATURE	-0,18
VISITA_FALMEDIA VISITA	-0,08
VISITA_FALPOCA VISITA	-0,38

*Anexo 15: Coeficientes modelo hogar con CMR
Fuente: Elaboración propia*

Coeficientes modelo transaccionales

VariableName	Estimate
ACTIVIDADCOMISIONISTA	0,37
ACTIVIDADDUEÑA DE CASA	-0,04

ACTIVIDADEMPL. DOMESTICA	0,87
ACTIVIDADEMPL. TECNICO	0,19
ACTIVIDADEMPRESARIO	-0,97
ACTIVIDADEST. UNIVER.	-0,28
ACTIVIDADJUBILADO/MONTEPIADO	-0,09
ACTIVIDADOBREROS, AUXI.	0,32
ACTIVIDADOFICIAL FF.AA.	1,13
ACTIVIDADPERS. DE SEGURIDAD	0,48
ACTIVIDADPERSONAL FF. AA.	-0,03
ACTIVIDADPROF. UNIV, (INDP.)	0,73
ACTIVIDADPROF. UNIV. (EMP.)	-0,05
ACTIVIDADRENTISTA	-16,46
ACTIVIDADTRABAJ. INDEPENDIENT	0,12
EMPLEADO_VIGENTE	0,10
ESTADO_CIVILDIVORCIADO	0,75
ESTADO_CIVILSEPARADO	0,61
ESTADO_CIVILSOLTERO	0,08
ESTADO_CIVILVIUDO	-0,04
GSEC2	-0,33
GSEC3	-0,50
GSED	-0,54
GSEE	-1,12
HABITACIONCASA DE FAMILIARES	-0,42
HABITACIONCASA DE LA EMPRESA	0,36
HABITACIONCASA DE LOS PADRES	-0,33
HABITACIONCASA FISCAL	-0,37
HABITACIONCASA PROPIA	-0,06
OTROS_ANT	-0,84
PRINCIPAL_ANT	1,42
RANGO_EDAD26 - 35	-0,36
RANGO_EDAD36 - 45	-0,49
RANGO_EDAD46 - 55	-0,52
RANGO_EDAD56 - 65	-0,41
RANGO_EDAD66 +	-0,61
SEXOM	-0,27

*Anexo 16: Coeficientes modelo transaccionales sin CMR
Fuente: Elaboración propia*

VariableName	Estimate
ACTIVIDADCOMISIONISTA	0,73
ACTIVIDADDUEÑA DE CASA	0,04
ACTIVIDADEMPL. DOMESTICA	0,49

ACTIVIDADEMPL. TECNICO	0,09
ACTIVIDADEMPRESARIO	0,04
ACTIVIDADEST. UNIVER.	-0,69
ACTIVIDADJUBILADO/MONTEPIADO	0,05
ACTIVIDADOBREROS, AUXI.	0,15
ACTIVIDADOFICIAL FF.AA.	-0,17
ACTIVIDADPERS. DE SEGURIDAD	0,04
ACTIVIDADPERSONAL FF. AA.	-0,11
ACTIVIDADPROF. UNIV, (INDP.)	-0,14
ACTIVIDADPROF. UNIV. (EMP.)	-0,25
ACTIVIDADRENTISTA	-0,59
ACTIVIDADTRABAJ. INDEPENDIENT	-0,03
ANIO_VEHICMUY ANTIGUOS	0,16
ANIO_VEHICNUEVOS	0,42
ANIO_VEHICSEMI NUEVOS	0,13
ANTIGUEDAD0-6 MESES	0,60
ANTIGUEDAD1-2 ANIOS	0,20
ANTIGUEDAD3-5 ANIOS	0,10
ANTIGUEDAD6-12 MESES	-1,92
DESC_SITUACTANO NORMAL	0,93
DESC_SITUACTANORMAL	-1,39
EMPLEADO_VIGENTE	0,30
ENVIO_FACTURACIONHOGAR	0,07
ENVIO_FACTURACIONOFICINA	0,42
ESTADO_CIVILDIVORCIADO	0,80
ESTADO_CIVILSEPARADO	0,13
ESTADO_CIVILSOLTERO	0,05
ESTADO_CIVILVIUDO	-0,13
GASTO_CM_12MGASTA NORMAL	-0,25
GASTO_CM_12MGASTA POCO	-0,58
GASTO_CM_12MNO GASTA	-0,30
GASTO_RUBRO_AUTOMOTRIZ_12MGASTA NORMAL	-0,14
GASTO_RUBRO_AUTOMOTRIZ_12MGASTA POCO	-0,08
GASTO_RUBRO_AUTOMOTRIZ_12MNO GASTA	-0,14
GASTO_RUBRO_COMBUSTIBLE_12MGASTA NORMAL	0,06
GASTO_RUBRO_COMBUSTIBLE_12MGASTA POCO	0,07
GASTO_RUBRO_COMBUSTIBLE_12MNO GASTA	0,02
GASTO_RUBRO_COMUNICACION_12MGASTA NORMAL	-0,02
GASTO_RUBRO_COMUNICACION_12MGASTA POCO	-0,18
GASTO_RUBRO_COMUNICACION_12MNO GASTA	-0,15
GASTO_RUBRO_EMPRESA_12MGASTA NORMAL	-0,03
GASTO_RUBRO_EMPRESA_12MGASTA POCO	0,10

GASTO_RUBRO_EMPRESA_12MNO GASTA	0,01
GASTO_RUBRO_ENTRETENCION_12MGASTA NORMAL	-0,10
GASTO_RUBRO_ENTRETENCION_12MGASTA POCO	-0,02
GASTO_RUBRO_ENTRETENCION_12MNO GASTA	-0,09
GASTO_RUBRO_FARMACIAS_12MGASTA NORMAL	0,17
GASTO_RUBRO_FARMACIAS_12MGASTA POCO	0,07
GASTO_RUBRO_FARMACIAS_12MNO GASTA	0,11
GASTO_RUBRO_MEJ_HOGAR_12MGASTA NORMAL	-0,05
GASTO_RUBRO_MEJ_HOGAR_12MGASTA POCO	-0,06
GASTO_RUBRO_MEJ_HOGAR_12MNO GASTA	0,06
GASTO_RUBRO_OTROS_12MGASTA NORMAL	-0,09
GASTO_RUBRO_OTROS_12MGASTA POCO	-0,12
GASTO_RUBRO_OTROS_12MNO GASTA	-0,04
GASTO_RUBRO_PAGO_WEB_12MGASTA NORMAL	0,01
GASTO_RUBRO_PAGO_WEB_12MGASTA POCO	0,17
GASTO_RUBRO_PAGO_WEB_12MNO GASTA	0,06
GASTO_RUBRO_RECAUDACION_12MGASTA NORMAL	0,12
GASTO_RUBRO_RECAUDACION_12MGASTA POCO	0,17
GASTO_RUBRO_RECAUDACION_12MNO GASTA	0,01
GASTO_RUBRO_RESTAURANT_12MGASTA NORMAL	0,02
GASTO_RUBRO_RESTAURANT_12MGASTA POCO	-0,07
GASTO_RUBRO_RESTAURANT_12MNO GASTA	-0,14
GASTO_RUBRO_SALUD_12MGASTA NORMAL	0,01
GASTO_RUBRO_SALUD_12MGASTA POCO	-0,02
GASTO_RUBRO_SALUD_12MNO GASTA	0,11
GASTO_RUBRO_SPM_12MGASTA NORMAL	0,15
GASTO_RUBRO_SPM_12MGASTA POCO	0,08
GASTO_RUBRO_SPM_12MNO GASTA	0,11
GASTO_RUBRO_TIENDAS_DEPTO_12MGASTA NORMAL	0,04
GASTO_RUBRO_TIENDAS_DEPTO_12MGASTA POCO	0,12
GASTO_RUBRO_TIENDAS_DEPTO_12MNO GASTA	0,10
GASTO_RUBRO_TRANSPORTE_12MGASTA NORMAL	-0,01
GASTO_RUBRO_TRANSPORTE_12MGASTA POCO	0,01
GASTO_RUBRO_TRANSPORTE_12MNO GASTA	-0,01
GASTO_RUBRO_VIAJES_12MGASTA NORMAL	-0,16
GASTO_RUBRO_VIAJES_12MGASTA POCO	-0,22
GASTO_RUBRO_VIAJES_12MNO GASTA	0,05
GASTO_RUBRO_VIVIENDA_12MGASTA NORMAL	0,20
GASTO_RUBRO_VIVIENDA_12MGASTA POCO	0,27
GASTO_RUBRO_VIVIENDA_12MNO GASTA	0,17
GSEC2	-0,13
GSEC3	-0,05

GSED	0,05
GSEE	-0,11
HABITACIONCASA DE FAMILIARES	0,17
HABITACIONCASA DE LA EMPRESA	0,43
HABITACIONCASA DE LOS PADRES	0,21
HABITACIONCASA FISCAL	0,30
HABITACIONCASA PROPIA	0,17
HABITO_PAGOMALO	0,20
HABITO_PAGOPESIMO	1,17
MES_COMPRA_CMRMUY ALTO USO	0,03
N_VEHICULOS1 AUTOS	0,04
N_VEHICULOS2 AUTOS	0,06
N_VEHICULOS3 AUTOS	0,05
N_VEHICULOS4 o + AUTOS	0,27
OTROS_ANT	-0,49
PRINCIPAL_ANT	0,21
RANGO_CUPO_BI1.250.0001 - 2.000.000	0,07
RANGO_CUPO_BI2.000.0001 - +	-0,01
RANGO_CUPO_BI501.000 - 1.250.0000	0,09
RANGO_EDAD26 - 35	-0,24
RANGO_EDAD36 - 45	-0,19
RANGO_EDAD46 - 55	0,07
RANGO_EDAD56 - 65	0,35
RANGO_EDAD66 +	0,46
REGENCY_FALNO RECIENTE	-0,02
REGENCY_FALRECIENTE	0,15
SEGMENTO_USO_CM RB AJA FRECUENCIA	0,12
SEGMENTO_USO_CM RFUGADO	-0,25
SEGMENTO_USO_CM RINACTIVO	-1,12
SEGMENTO_USO_CM RMEDIA FRECUENCIA	0,15
SEGMENTO_USO_CM RNUEVOS	0,55
SEGMENTO_USO_CM RPOR FUGARSE	-0,12
SEGMENTO_VALOR_FALCLIENTE NORMAL	-0,38
SEXOM	-0,02
TIPO_CLIENTECLIENTE NORMAL	-0,09
TIPO_CLIENTECLIENTE PREMIUM	-0,01
TIPO_TARJETATARJETA CMR TRADICIONAL	-0,36
TIPO_TARJETATARJETA CMR VISA	-0,43
TIPO_TARJETAVISA PLATINUM	0,08
TIPO_TARJETAVISA SIGNATURE	0,23

Anexo 17: Coeficientes modelo transaccionales con CMR
Fuente: Elaboración propia

11.5 Anexo E

Métricas de desempeño modelos

MODELO	Métrica	AUTO	VIDA	VIDA_CB	SALUD	HOGAR	TRANSACC.
LOGIT	Accuracy	61,8%	75,5%	51,8%	55,7%	67,4%	58,1%
	Sensibilidad	83,4%	84,5%	85,8%	83,9%	80,7%	77,3%
	Precisión	2,8%	2,0%	0,3%	0,2%	0,2%	0,2%
ARBOL	Accuracy	73,1%	78,0%	68,1%	68,0%	79,5%	79,9%
	Sensibilidad	65,7%	74,9%	67,8%	63,2%	73,0%	51,0%
	Precisión	3,2%	1,8%	0,3%	0,1%	0,3%	0,3%

Anexo 18: Desempeño modelos según matrices de confusión sin CMR

Fuente: Elaboración propia

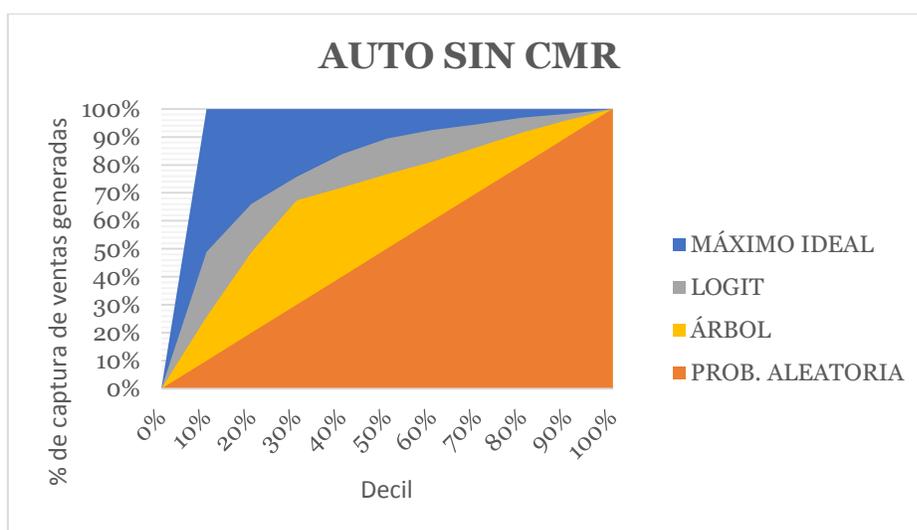
MODELO	Métrica	AUTO	VIDA	VIDA_CB	SALUD	HOGAR	TRANSACC.
LOGIT	Accuracy	63,8%	52,4%	55,6%	56,2%	61,5%	53,6%
	Sensibilidad	79,8%	80,3%	83,4%	79,7%	94,9%	82,2%
	Precisión	3,7%	1,0%	1,1%	0,6%	0,6%	0,9%
ARBOL	Accuracy	68,4%	58,6%	88,6%	84,9%	85,4%	67,6%
	Sensibilidad	71,7%	69,3%	44,2%	44,6%	91,8%	63,3%
	Precisión	4,0%	1,1%	2,3%	0,9%	1,1%	0,9%

Anexo 19: Desempeño modelos según matrices de confusión con CMR

Fuente: Elaboración propia

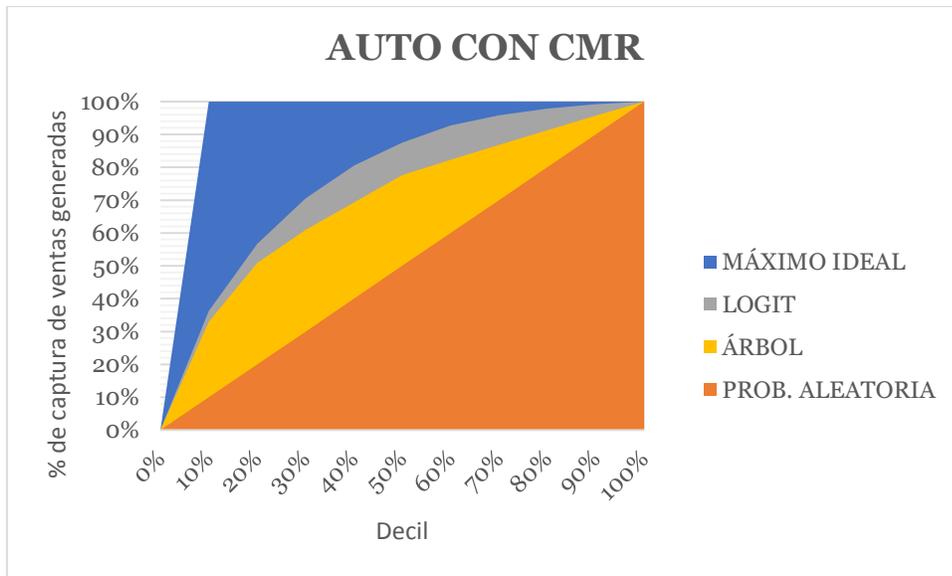
11.6 Anexo F

Curvas ganancia de modelos

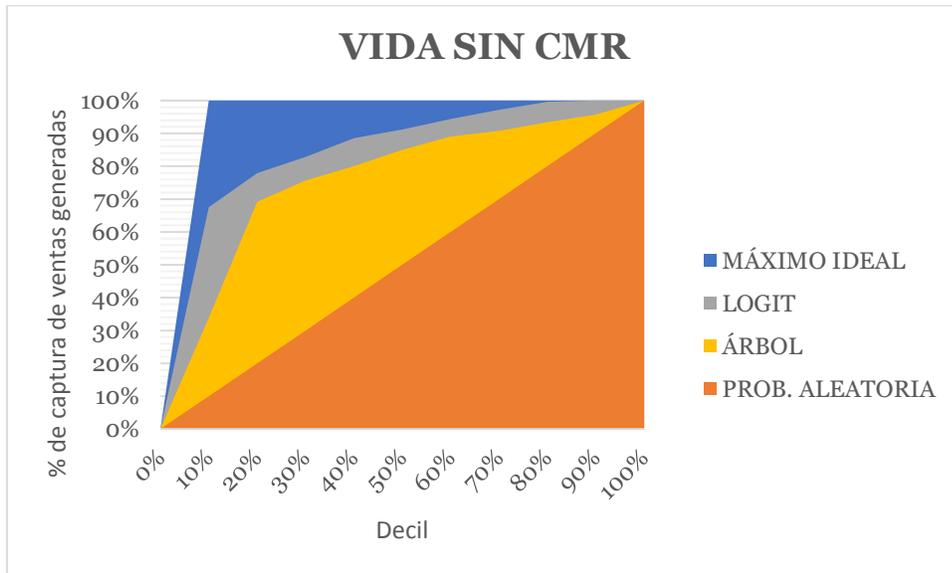


Anexo 20: Curva de Ganancia para modelo Auto sin CMR

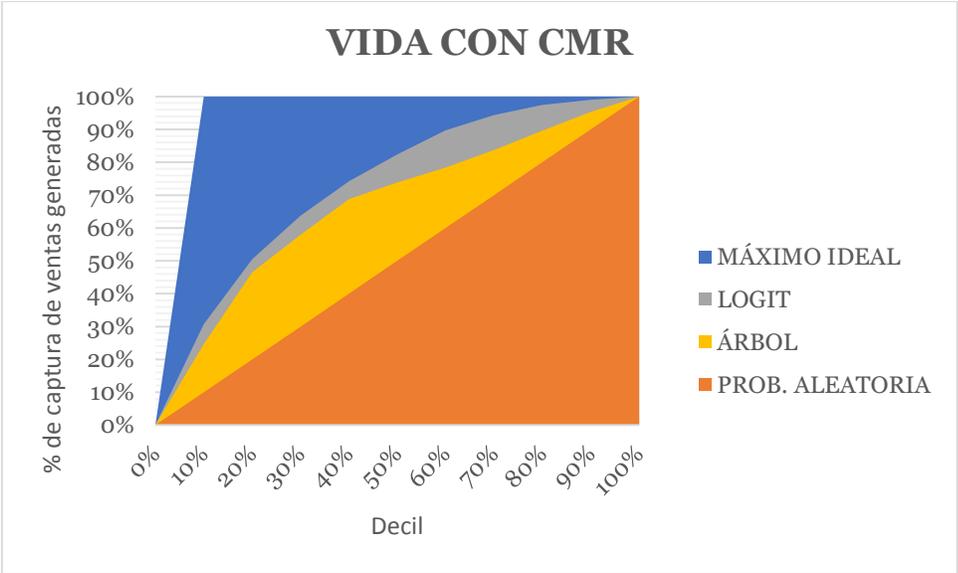
Fuente: Elaboración propia



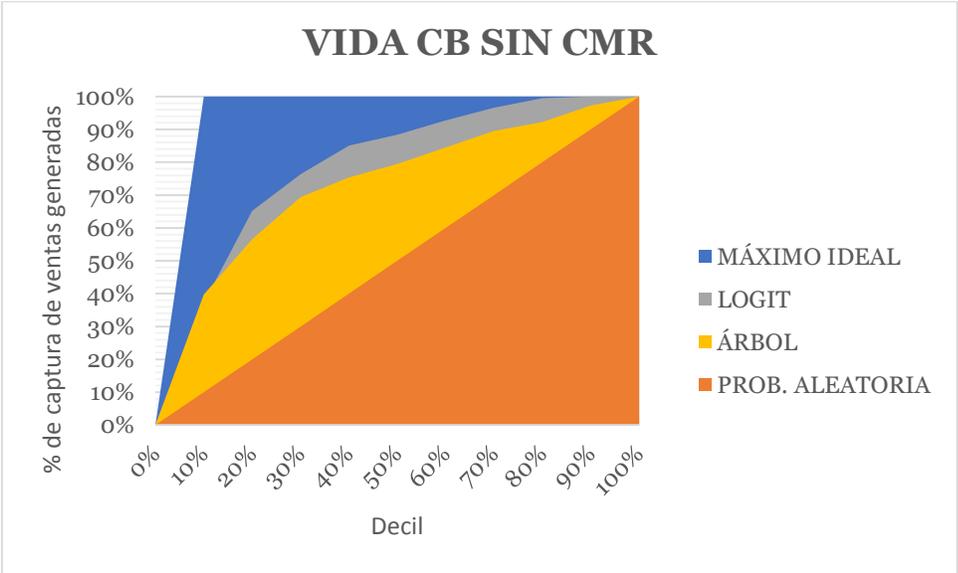
Anexo 21: Curva de Ganancia para modelo Auto con CMR
Fuente: Elaboración propia



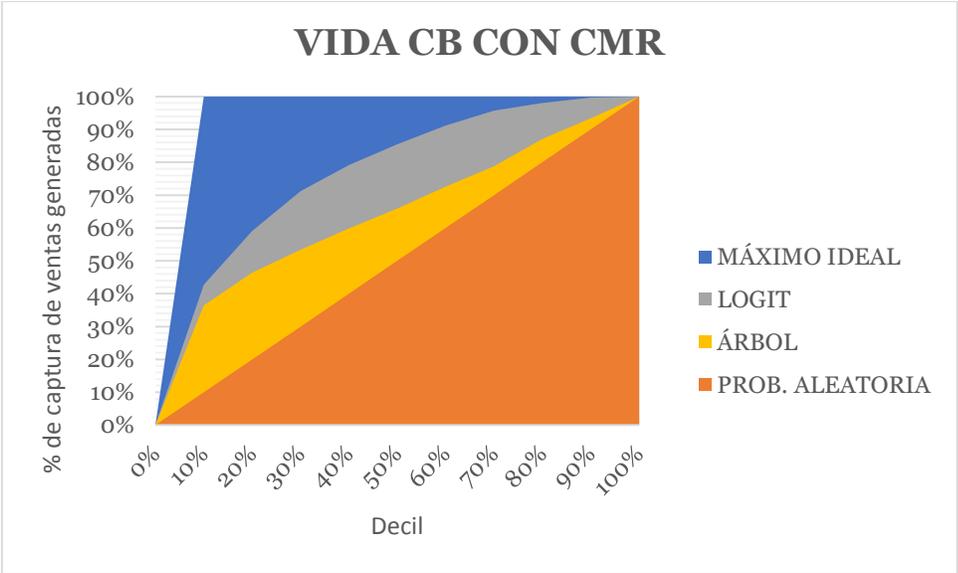
Anexo 22: Curva de Ganancia para modelo Vida sin CMR
Fuente: Elaboración propia



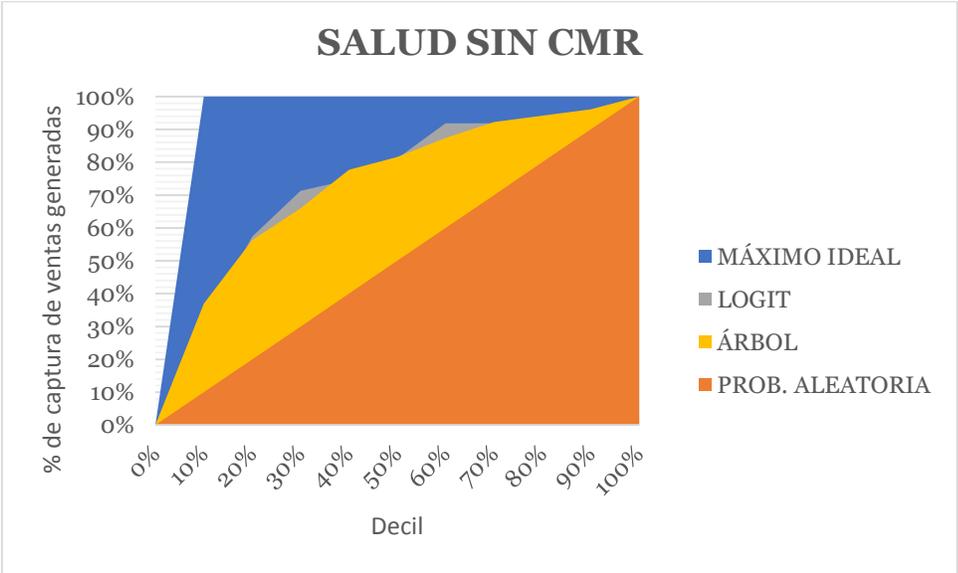
*Anexo 23: Curva de Ganancia para modelo Vida con CMR
Fuente: Elaboración propia*



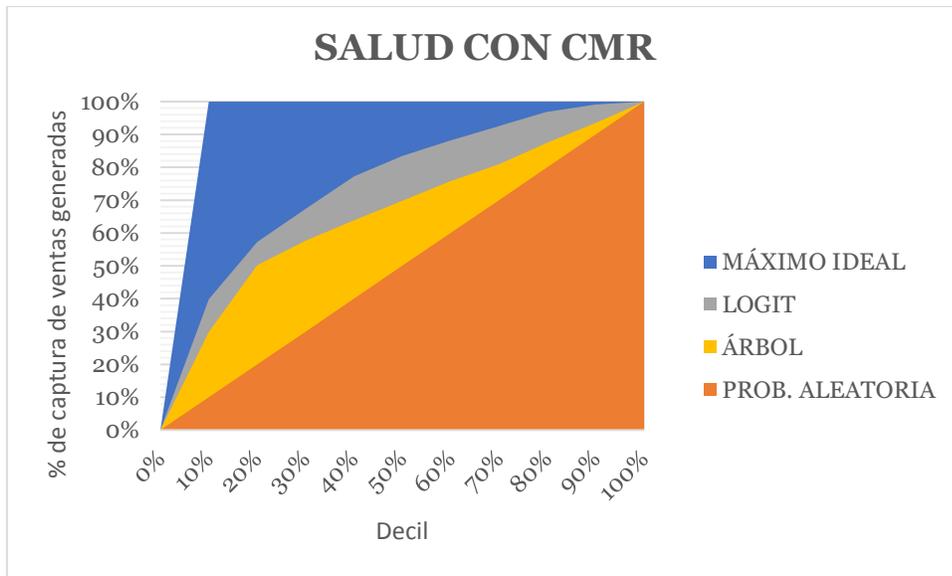
*Anexo 24: Curva de Ganancia para modelo Vida con bonificación sin CMR
Fuente: Elaboración propia*



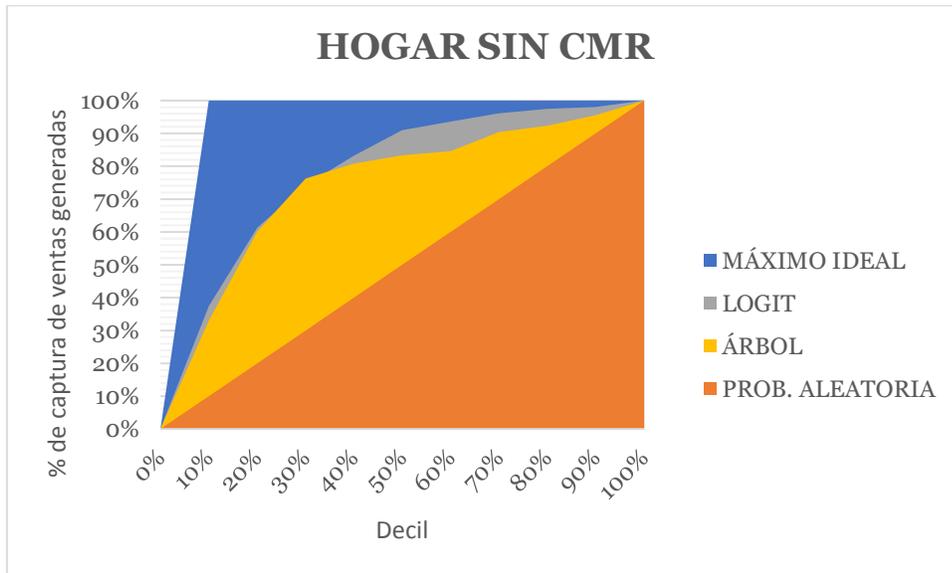
Anexo 25: Curva de Ganancia para modelo Vida con bonificación con CMR
Fuente: Elaboración propia



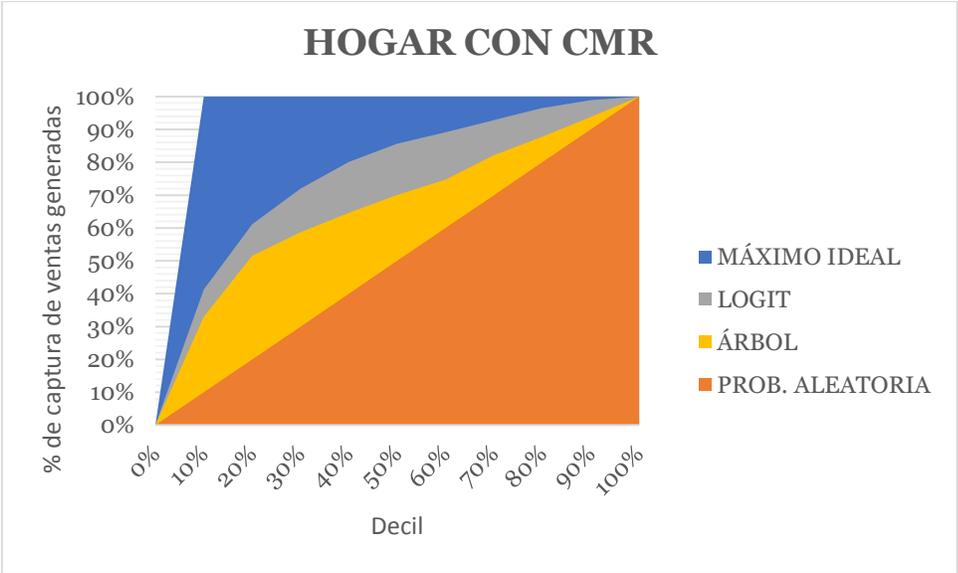
Anexo 26: Curva de Ganancia para modelo Salud sin CMR
Fuente: Elaboración propia



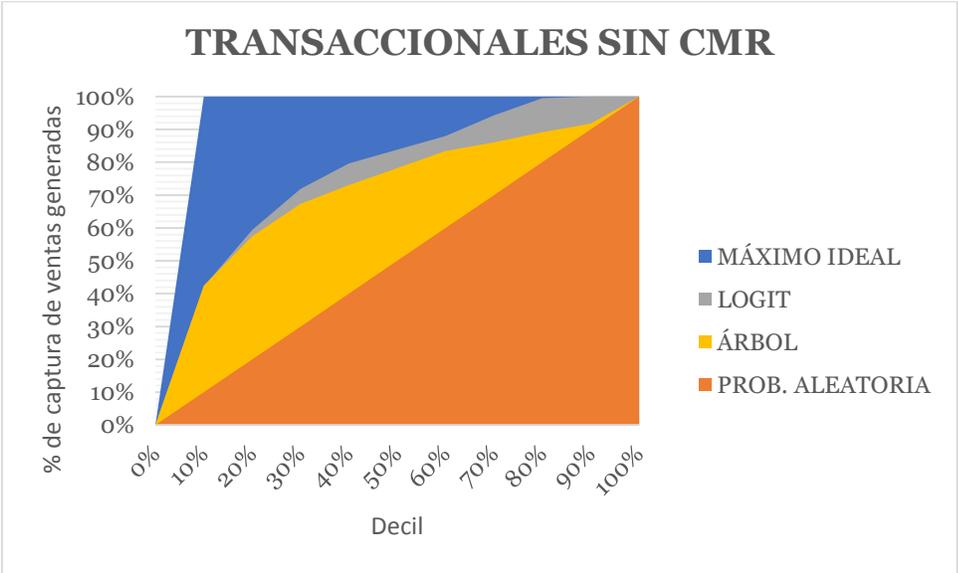
*Anexo 27: Curva de Ganancia para modelo Salud con CMR
Fuente: Elaboración propia*



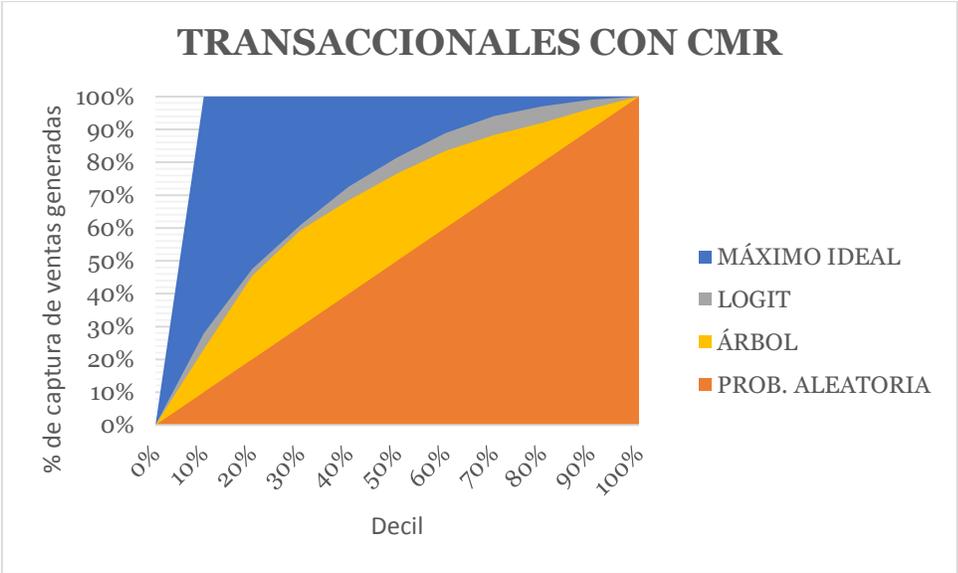
*Anexo 28: Curva de Ganancia para modelo Hogar sin CMR
Fuente: Elaboración propia*



*Anexo 29: Curva de Ganancia para modelo Hogar con CMR
Fuente: Elaboración propia*



*Anexo 30: Curva de Ganancia para modelo Transaccionales sin CMR
Fuente: Elaboración propia*



Anexo 31: Curva de Ganancia para modelo Transaccionales con CMR
Fuente: Elaboración propia