UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

MODELO PREDICTIVO DE DESAFILIACIÓN DE EMPRESAS PARA UNA CAJA DE COMPENSACIÓN DE ASIGNACIÓN FAMILIAR

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

MARÍA CONSTANZA ROJAS CERDA

PROFESOR GUÍA:
LUIS ALBERTO ABURTO LAFOURCADE

MIEMBROS DE LA COMISIÓN:
ALEJANDRA PUENTE CHANDÍA
ALEJANDRO PATRICIO MUÑOZ ROJAS

SANTIAGO DE CHILE
2015

RESUMEN DE LA MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL
Alumna: María Constanza Rojas Cerda
Fecha: 7/10/2015
Profesor Guía: Luis Aburto Lafourcade

MODELO PREDICTIVO DE DESAFILIACIÓN DE EMPRESAS PARA UNA CAJA DE COMPENSACIÓN DE ASIGNACIÓN FAMILIAR

Una permanente preocupación de las Cajas de Compensaciones es la fuga de sus clientes al igual que en la mayoría de las empresas. En Chile existen cinco Cajas que compiten entre ellas por lo que cada vez es más necesario tener estrategias que ayuden a captar y retener a sus clientes de forma más efectiva y eficiente.

El presente trabajo de título consiste en generar un modelo predictivo de desafiliación de empresas para una Caja de Compensación, el cual tiene como objetivo identificar de forma temprana aquellas empresas que tiendan a fugarse para poder gestionar de mejor forma las acciones que se realizan sobre cada una de ellas, logrando disminuir la actual tasa de fuga de la Caja en estudio.

Para generar el modelo primero se realizó una investigación exploratoria, donde se levantaron diferentes causas que llevan a las empresas a desafiliarse. Principalmente ligados a descontento con los beneficios, baja colocación de crédito y problemas operativos con la Caja. Con estos motivos se generaron diferentes variables que pudieran reflejar el comportamiento de un año atrás de las empresas seleccionando las más significativas para utilizarlas en modelos de árboles de decisión y regresión logística. Los modelos desarrollados presentaron una alta precisión en general sobre el 80% y se seleccionó el árbol de decisión Chaid exhaustivo para proponer acciones de retención dada su fácil interpretación. De él se obtuvo que las principales variables que describen el comportamiento de fuga tienen que ver con la variación de beneficios y deuda crediticia de la empresa durante el último año. Por otra parte, para focalizar de mejor manera los recursos se seleccionaron a 139 empresas que cuentan con una alta probabilidad de fuga y presentan un alto valor para la Caja, rediseñando el actual foco de empresas a las cuales se les realiza acciones de fidelización.

Las acciones de retención propuesta están ligadas a las campañas de beneficios que se realizan y la colocación de crédito, con tal de disminuir el descontento de sus clientes en estos ámbitos. Además, se propone realizar un diseño experimental para medir el impacto que tiene reasignar recursos de las empresas más leales en aquellas que tienen una mayor probabilidad de fugarse.

Finalmente, se espera que la Caja pueda aprovechar la información que se genera de cada empresa para brindarles un mejor servicio a sus trabajadores y los incentive a usar los beneficios que tiene para ellos, fortaleciendo la relación que tienen con la Caja y convirtiéndolos en clientes más leales.

Tabla De Contenido

Índice de Tablas

Índice de Ilustraciones

1 Antecedentes Generales

1.1 Descripción del Mercado de las Cajas de Compensación

Las Cajas de Compensación de Asignación Familiar son corporaciones chilenas de derecho privado, con patrimonio propio y sin fines de lucro. Su objetivo es la administración de prestaciones de seguridad social que tiendan al desarrollo y bienestar de sus trabajadores y su grupo familiar, protegiéndolo de contingencias sociales y económicas. Fueron creadas en julio de 1953 y son supervisadas por la Superintendencia de Seguridad Social (SUSESO), fiscalizadas por la Controlaría General de la República, ya que administran fondos públicos, y por la Superintendencia de Valores y Seguros (SVS). Originalmente se generaron por la iniciativa de las asociaciones de empleadores y prestaban únicamente servicios a los obreros de sus propias empresas, posteriormente se extendió a todos los trabajadores sin distinción.

Las CCAF poseen tres tipos de clientes: Pensionados, Trabajadores y Empresas. Los pensionados son independientes y son ellos quienes deciden cuándo afiliarse o desafiliarse, mientras que los trabajadores se afilian a una caja de compensación cuando el mayor porcentaje de los empleados de la empresa están de acuerdo en afiliarse a cierta caja, o bien, desafiliarse de la caja a la cual pertenecen. Estos clientes pueden afiliarse a una CCAF de dos formas: *Colectivamente* e *Individualmente.* Las afiliaciones *Colectivas*, corresponden a entidades empleadoras con sus trabajadores, mientras que *Individualmente* corresponde a los pensionados.

Las Cajas administran dos tipos de prestaciones de seguridad social: las Prestaciones Legales y las Prestaciones de Bienestar Familiar Social. Las *Prestaciones Legales*, corresponden a las administraciones de recursos traspasados por el Estado, para que la caja pague a los trabajadores beneficios como Asignación Familiar, Subsidio por Incapacidad Laboral, Licencias Médicas, entre otros. Las *Prestaciones de Bienestar Social y Adicionales* corresponden a los beneficios adicionales que entregan las cajas a los trabajadores afiliados, tales como Crédito Social y Bonos por fallecimiento, matrimonio, nacimiento o por escolaridad [1].

Actualmente, existen 5 cajas de compensación en Chile: Los Andes, La Araucana, Los Héroes, Gabriela Mistral y 18 de Septiembre, las cuales conviven en un mercado altamente competitivo debido a que su oferta de productos es poco diferenciada lo que ha generado un creciente gasto en campañas de marketing para la captación de nuevos clientes [2].

Dentro de los efectos provocados por la competencia en el sector está la alta rotación de los pensionados entre las distintas CCAF, ya que los niveles de lealtad y fidelización con una caja son bajos, lo que aumenta las tendencias a la fuga. En el caso de las empresas, la rotación es más estable, suelen cambiarse de una caja a otra debido a problemas específicos con la caja a la cual pertenecen o mejores ofertas de la competencia.

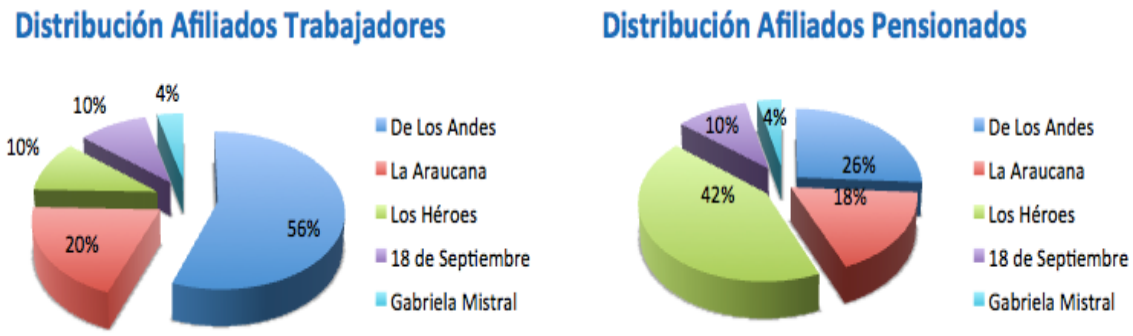
En la Tabla 1 se muestra el número promedio mensual de afiliados por caja durante el año 2011 y en la Ilustración 1 la participación de mercado de cada una de las cajas existentes.

Tabla 1 Número promedio mensual de Afiliados por CCAF (Año 2011)

C.C.A.F.	Trabajadores	Pensionados	Total
De Los Andes	2.285.455	356.935	2.642.390
La Araucana	811.150	241.315	1.052.465
Los Héroes	438.514	582.770	1.021.284
18 de Septiembre	412.476	142.498	554.973
Gabriela Mistral	155.990	48.552	204.542
Total	**4.103.585**	**1.372.070**	**5.475.655**

Fuente: SUSESO

Ilustración 1 Participación de Mercado de las CCAF de Trabajadores y Pensionados



Fuente: SUSESO

Las CCAF tienen el deber de ser autosustentables para gestionar las prestaciones sociales que les delega el Estado y para entregar el resto de los beneficios. Las fuentes de financiamiento de las cajas son, principalmente, su patrimonio conformado por el Fondo Social, los bancos y los inversionistas institucionales, a través de colocaciones de instrumentos de oferta pública en el mercado financiero.

Por otra parte, una de sus principales fuentes de ingreso corresponde al crédito social, es decir, el sistema de préstamos que otorgan las CCAF a sus afiliados, que fueron creadas para hacer frente a contingencias o satisfacer necesidades de sus clientes. Se caracterizan por entregar créditos a una tasa menor que los Bancos y con menor riesgo debido a que son descontados de las cotizaciones de los trabajadores.

1.2 Descripción de la Caja de Compensación en Estudio

La Caja de Compensación en donde se desarrolla el modelo predictivo nace por iniciativa de la Asociación de Industrias Metalúrgicas y Metalmecánicas (Asimet) para administrar el beneficio de la asignación familiar de los trabajadores de dicho sector [2]. Parte importante de los excedentes generados a través de los productos, servicios y transacciones de la CCAF permiten entregar beneficios de salud, educación, recreación y bonos en dinero a los trabajadores y pensionados afiliados a ella.

Actualmente cuenta con alrededor de 900.000 afiliados y con más de 5.500 empresas, posicionándose como una institución relevante en el país. Además, posee el liderazgo en el segmento de pensionados a diferencia del segmento trabajador en donde compite por el tercer lugar en el mercado.

La Caja actualmente coloca sus recursos y esfuerzo en las empresas con mayor cantidad de trabajadores y monto de crédito colocado, acompañado de variables como a que rubro pertenecen o la rentabilidad que les generan, con ello determina cuáles son sus empresas foco.

2 Descripción Del Proyecto Y Justificación

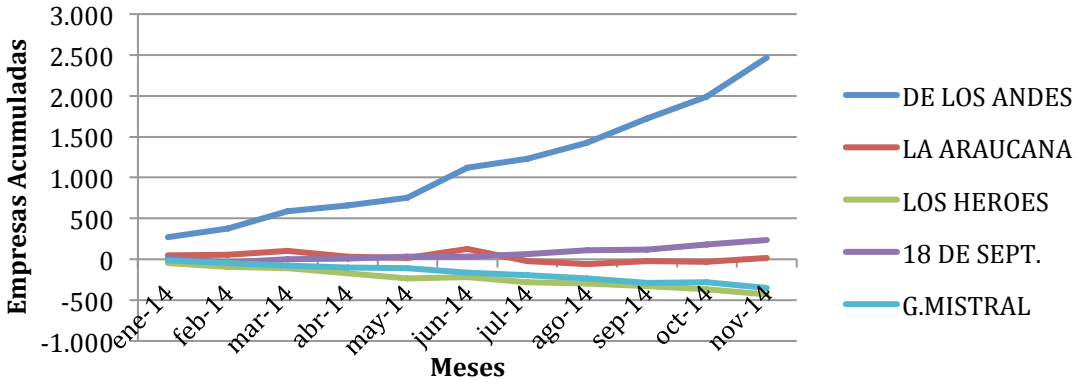
Una fuga en una Caja puede definirse como la acción en que una empresa termina su relación con la CCAF a la cual está afiliada, perdiendo los beneficios que ésta le entrega. O bien, en el caso de pensionados, cuando este anula el contrato que lo liga con la caja dejando de entregar el 1% de su pensión a cambio de la administración de sus prestaciones y perdiendo los beneficios que recibe de ella.

Las CCAF se ven afectadas por la fuga de sus clientes al igual que las mayoría de las empresas con la diferencia que en vez de perder a una persona la Caja pierde empresas, por lo que presenta relaciones B2B (Business-to-business) entre la Caja y las Empresas y además, B2C (Business-to-Consumer) entre la Caja y los consumidores o trabajadores en este caso. Siendo estos últimos uno de los activos más importantes que posee, ya que están estrechamente relacionados con las utilidades del negocio. Esto genera incentivos a mantener o aumentar la cartera de clientes con tal que las utilidades crezcan, ya que una cartera con un mayor número de trabajadores permitirá el aumento de los ingresos generados por colocación de crédito social, los que posteriormente son utilizados para generar mayores beneficios para sus afiliados.

Para aumentar la cartera de cliente es necesario realizar dos actividades comerciales: la *captación* de clientes nuevos y la *retención* de los actuales clientes. La captación de clientes se enfoca en incorporar nuevos clientes a la cartera mediante elaboradas estrategia de publicidad, alta inversión en fuerza de ventas y generación de ofertas focalizadas. Por otra parte, la retención de clientes consiste en la identificación de los clientes con mayores tendencias a la fuga y aplicar estrategias o procedimientos que aumenten el grado de fidelización y bajen los índices de fuga en la cartera [3]. Para las CCAF, es mucho más económico retener a sus clientes que captar uno nuevo debido a que es necesario que el cliente deje de estar afiliado a su caja anterior y decida cambiarse.

Cabe destacar, que la industria de las CCAF en Chile está compuesta solamente por 5 cajas, y todos los trabajadores y pensionados a lo largo del país se distribuyen en cada una de ellas, por lo que cada trabajadores que se desafilia necesariamente se afilia en otra caja de compensación, reduciendo la participación de mercado de la caja a la que pertenecía. Como se muestra en la Ilustración 2, la cantidad de empresas acumuladas que capturo cada caja durante el año 2014.

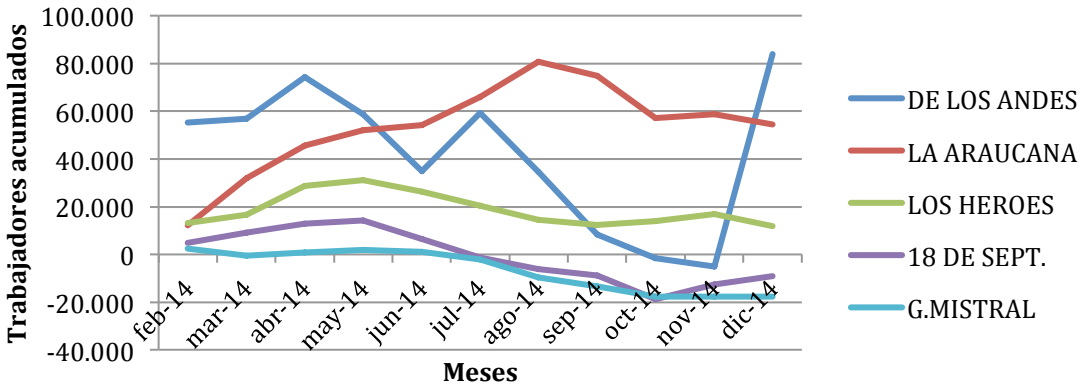
Ilustración 2 Cantidad de Empresas Acumuladas Durante el 2014.



Fuente: Elaboración Propia. Datos SUSESO.

Se puede apreciar, que la caja en estudio tendió a perder empresas afiliadas, en mayor proporción que la competencia, a pesar de esto la Caja actualmente compite con otra CCAF por el tercer lugar dentro del mercado en el segmento trabajador, por lo que dejar que siga aumentando está tendencia podría tener consecuencias significativas para la posición de la Caja en la industria. En la Ilustración 3 se presenta la cantidad de trabajadores que acumuló cada CCAF durante el 2014 con cada una de las empresas que afilio a su cartera.

Ilustración 3 Cantidad de trabajadores afiliados durante el 2014 por cada CCAF

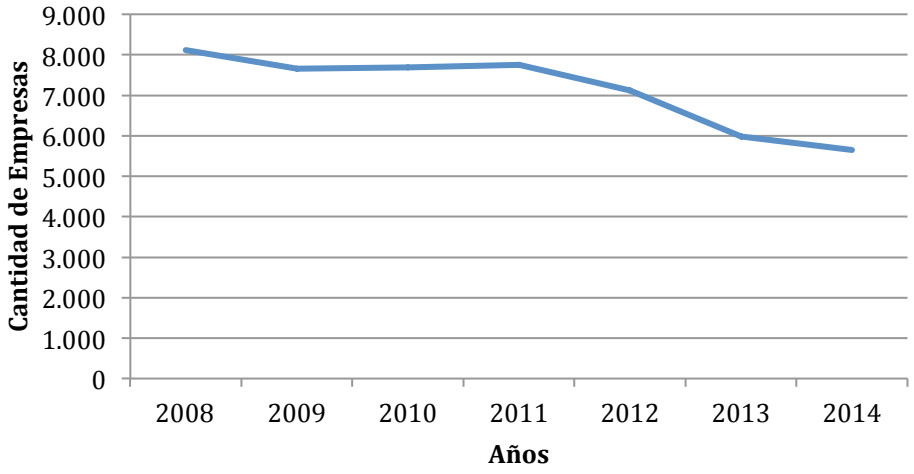


Fuente: Elaboración Propia. Datos SUSESO

Dado el comportamiento anterior, los intereses en retener afiliados y mantenerlos por el mayor tiempo posible aumentan, principalmente debido a que los créditos colocados a sus trabajadores suelen ser a varios meses o años, y la desafiliación de estos puede interrumpir con los pagos de las cuotas del crédito, quedando en mora con la Caja.

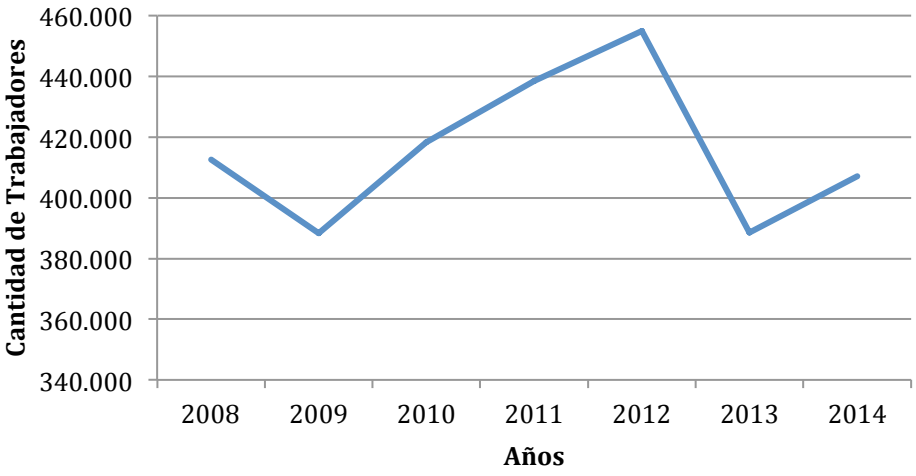
En la Ilustración 4 se presenta como ha ido decreciendo el stock de empresas con que cuenta La Caja, acompañado del gráfico en Ilustración 5, que presenta la cantidad de trabajadores afiliados durante el mismo periodo.

Ilustración 4 Evolución del Stock de Empresas Afiliadas a la Caja en estudio



Fuente 1 Elaboración Propia, Datos SUSESO

Ilustración 5 Evolución de Trabajadores Afiliados a la Caja



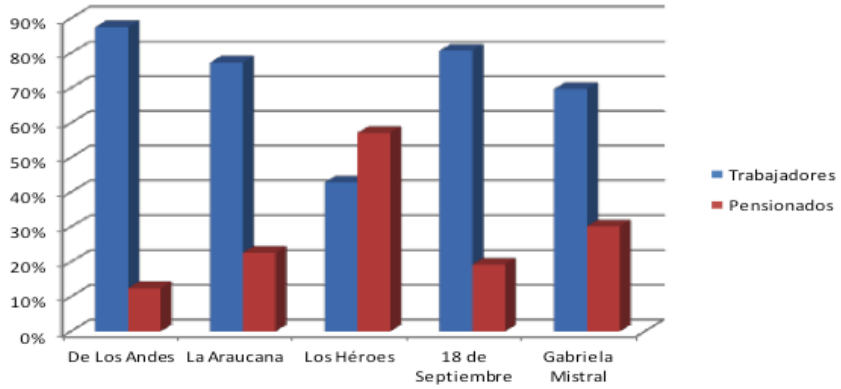
Fuente 2 Elaboración Propia, Datos SUSESO

Actualmente, la tasa de desafiliación de la Caja en estudio corresponde a un 6% aproximadamente y como se puede apreciar en la Ilustración 4 e Ilustración 5, durante el periodo 2009 y 2012, la cantidad de trabajadores aumentó, pero no así la cantidad de empresas afiliadas, esto puede deberse a que las pocas empresas que se afiliaron contaban con una gran cantidad de trabajadores. Mientras que en el periodo 2012-2013 la cantidad de empresas siguió disminuyendo principalmente con la fuga de empresas que contaban con una gran cantidad de

trabajadores. Lo más conveniente para las Cajas es que se afilien empresas con una gran cantidad de trabajadores, de esta manera se tendrá mayores posibilidades de colocar créditos.

Por otra parte, las cajas han orientado su oferta de crédito mayoritariamente al segmento de trabajadores, a excepción de la caja en estudio que ha seguido una estrategia de negocios distinta, enfocada en el segmento pensionado, como se muestra en la Ilustración 6.

Ilustración 6 Distribución de cartera por segmento pensionado o trabajador. Industria de Cajas de Compensación en porcentaje colocación de créditos por entidad.



Identificar a los posibles clientes fugados antes de que el suceso de fuga ocurra permite brindar el tiempo necesario para reaccionar y aplicar las políticas de retención necesarias de forma anticipada [3]. Además, permite entender el patrón de comportamiento de los clientes fugados mediante la relación entre las variables e identificar la importancia de cada una de ellas.

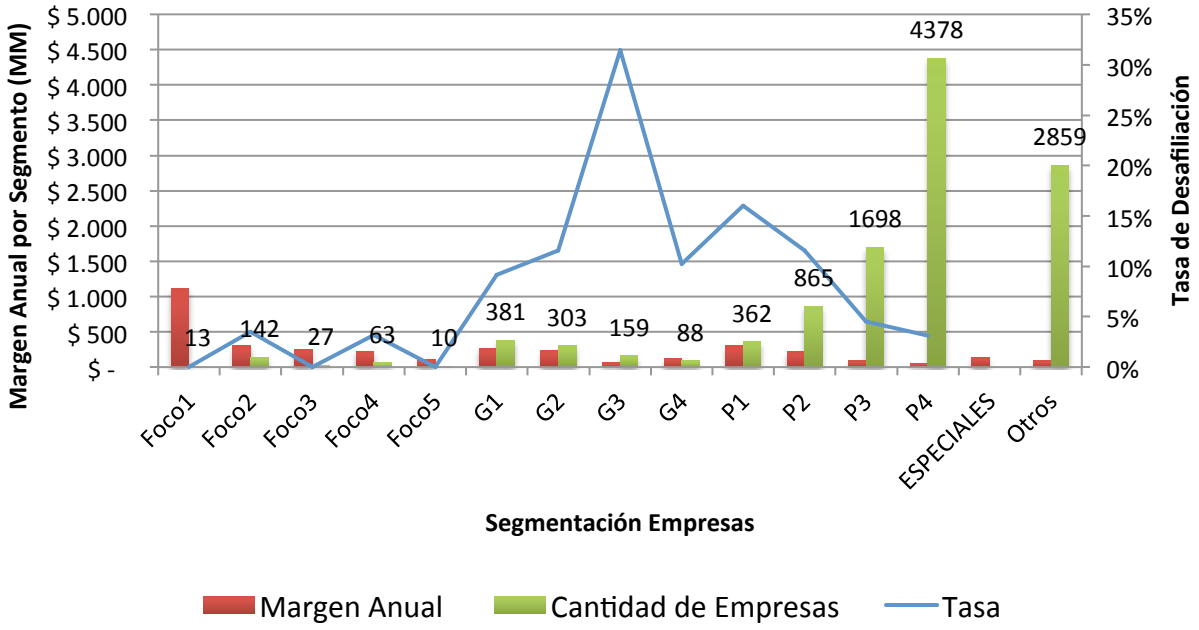
Entender los factores que explican el comportamiento de fuga de los clientes es un problema complejo y no trivial por lo que hace necesario utilizar herramientas de data mining que ayuden a determinar los atributos que están más relacionados con la fuga.

La presente memoria se enfocará solamente en la retención de trabajadores pertenecientes al segmento empresa de la Caja en estudio, ya que en el caso particular de las CCAF y las entidades financieras, incorporar a un cliente nuevo es más riesgoso ya que no se posee información de su comportamiento financiero anterior. Por otra parte, se pondrá mayor énfasis en aquellas empresas más rentables y con mayor probabilidad de fuga fortaleciendo las actuales estrategias de retención que se les realizan. Las tasas de desafiliación de acuerdo a la segmentación interna que presenta la Caja según la cantidad de trabajadores y condición crediticia de las empresas, se muestran en la Ilustración 7.

Por otra parte, en la Ilustración 7 se presenta las utilidades que deja cada Segmento Comercial en un año. La segmentación comercial además de considerar los mismos segmentos de

la condición crediticia, incluye a las empresas Foco, las cuales son empresas que la Caja considera importante ya sea porque en ellas se concentra el 40% de sus colocaciones aproximadamente o porque mantienen acuerdos especiales. Los segmentos comerciales principales corresponde a Gendarmería, Minería, Educación, Salud Privada, Servicios Públicos que son las empresas foco y G1 , G2, PYM1, PYM2 y otros que pertenecen al resto de las empresas.

Ilustración 7 Margen Total por Segmento Comercial y tasa de desafiliación total en un Año.



Como se puede apreciar, los segmentos que presentaban mayor tasa de desafiliación, P1 y P2, también presentan un gran valor para la Caja ya que dejan un margen relativamente alto, por lo que predecir que empresas de estos segmentos tienen la intención de fugarse permitiría tomar acciones en ellas y reducir las pérdidas para la CCAF en estudio. A diferencia, del Segmento G3, que presenta una alta tasa de fuga pero que sus márgenes anuales están por debajo del resto de los segmentos, por lo que generar acciones de retención sobre este segmento probablemente será más costoso que los ingresos que recibirán por mantenerlas afiliadas.

Una vez construido el modelo y se tengan las probabilidades de fuga de cada empresa en conjunto con las acciones de retención propuestas para cada grupo de empresas, se espera mejorar por lo menos en un 50% la tasa de fuga de global, es decir pasar de un 6% , como se muestra en la Tabla 2, a un 3% de fuga anual. En la tabla a continuación se muestra las ganancias anuales generadas por mejorar la tasa de desafiliación en dicho valor.

Tabla 2 Ganancia Anual por mejorar en un 50% la Tasa de Desafiliación en los Segmentos Crediticios más importantes

Estado Crediticio	Afiliadas	Desafiliadas	Cantidad de Empresas Segmento	Tasa Desafiliación Segmento	Margen Promedio Anual por Empresa	Margen Mensual Segmento	Margen Mensual Empresa Promedior6	Margen Segmento Anual	Ganancia por Mejorar 50% Fuga (tasa 3%)
G1	346	35	381	9%	$ 708.661	$ 22.500.000	$ 65.029	$ 270.000.000	$ 8.100.000
G2	268	35	303	12%	$ 762.376	$ 19.250.000	$ 63.531	$ 231.000.000	$ 6.930.000
G3	109	50	159	31%	$ 377.358	$ 5.000.000	$ 31.447	$ 60.000.000	$ 1.800.000
G4	79	9	88	10%	$ 1.397.727	$ 10.250.000	$ 116.477	$ 123.000.000	$ 3.690.000
P1	304	58	362	16%	$ 828.729	$ 25.000.000	$ 69.061	$ 300.000.000	$ 9.000.000
P2	765	100	865	12%	$ 249.711	$ 18.000.000	$ 20.809	$ 216.000.000	$ 6.480.000
P3	1621	77	1698	5%	$ 54.770	$ 7.750.000	$ 4.564	$ 93.000.000	$ 2.790.000
P4	4242	136	4378	3%	$ 12.334	$ 4.500.000	$ 1.028	$ 54.000.000	$ 1.620.000
P5	345	29	374	8%	$ -	$ -	$ -		$ -
P6	68		68	0%	$ -	$ -	$ -		$ -
GC	56	5	61	8%	$ -	$ -	$ -		$ -
R0	667	61	728	8%	$ -	$ -	$ -		$ -
Otras	2779	80	2859	3%	$ 10.143	$ 2.416.667	$ 845	$ 29.000.000	$ 870.000
Total	**11649**	**675**	**12324**	**6%**	**$ 111.652**	**$114.666.667**		**$ 1.376.000.000**	**$ 41.280.000**

Fuente: Elaboración Propia.

Se puede apreciar, que actualmente la tasa de desafiliación corresponde a un 6%, y si se mejora un 50%, es decir, que disminuya a un 3% la tasa de desafiliación las ganancias anuales aumentan 41M anuales.

Se concluye, que generar un modelo que prediga que empresas se van a fugar permitirá ser más eficiente con la asignación de recursos que se destinan a la retención de empresas, además de mejorar las relaciones con ellos evitando que prefieran otra caja. Por otra parte, permitirá entender el patrón de comportamiento que motiva la fuga de los clientes [3]. La identificación de estos clientes hará posible la aplicación de *estrategias de fidelización* y *retención* en ellos, de tal manera que La Caja pueda lograr fortalecer el segmento de trabajadores.

3 Objetivos

3.1 Objetivo General

"Generar un modelo predictivo de desafiliación de empresas, para una detección temprana de los potenciales clientes en fuga de una Caja de Compensación de Asignación Familiar".

3.2 Objetivos Específicos

- Generar indicadores que explican el comportamiento de los trabajadores a nivel de la empresa.
- Validar o rechazar las hipótesis más comunes de propensión de fuga de empresas.
- Modelar la fuga de empresas mediante modelos de árboles de decisión y regresión logística.
- Interpretar resultados e identificar a los grupos de empresas más propensas a fuga con el modelo que entregue mayor precisión.
- Proponer acciones de retención a los grupos de empresas más propensas y que dejan un mayor margen.

4 Alcances

La memoria contempla la entrega de un modelo para la detección temprana de empresas en fuga de tal manera que el área de Fidelización de la Caja pueda enfocar de mejor manera sus acciones de retención de sus clientes.

Tanto el modelo de propensión de fuga como los distintos análisis se considerarán entregables. El modelo se enfocará en identificar a las potenciales empresas fugitivas, dejando fuera del análisis al segmento pensionado de la Caja. Se trabajó con las hipótesis más importantes que se levantaron de las entrevistas en profundidad con los expertos del negocio y las cuales fueron testeadas con los diferentes modelos.

Para el modelo de propensión de fuga se utilizó una base de datos con la cartera de empresas de la Caja de los últimos 2 años debido a que las características del mercado de las CCAF cambian constantemente por lo que se consideró el comportamiento más reciente. Se estimó que con esta cantidad de datos sería suficiente para llegar a resultados representativos y aplicables a la actualidad. Además, se analizaron datos de todas las áreas o procesos que influyen en los trabajadores y las empresas (Créditos, Licencias Médicas, Asignaciones Familiares, Visitas a Sucursal y Beneficios).

Por otra parte, se utilizaron distintos modelos predictivos para encontrar los segmentos más propensos a la fuga de acuerdo a la base de datos analizada y al modelo que arrojó mejor precisión. Además, fue desarrollado en Modeler SPSS y se utilizó esta misma herramienta para probar las hipótesis planteadas.

Se propusieron estrategias de retención para los principales clientes propensos a fuga obtenido del modelo con ayuda de estudios realizados y los recursos de la empresa. Los alcances de este trabajo consisten solamente en la propuesta de retención, excluyendo la implementación de las acciones que se propongan.

5 Marco Conceptual

5.1 Concepto de Fuga de Clientes

El concepto de fuga de clientes ha sido un tema de estudio científico intensivo en los últimos años y en diferentes rubros, más aún en el de las CCAF donde la retención de clientes es vital, dado el costo de adquisición de un nuevo cliente y las campañas de marketing que se realizan. Este concepto hace referencia a la proporción o tasa de clientes que se van o dejan al proveedor de un servicio durante un período de tiempo determinado.

Existen dos tipos de fuga: las fugas *voluntarias* y las fugas *no voluntarias*. Las fugas voluntarias se asocian a la desafiliación del cliente por iniciativa propia, sin injerencia directa por parte de la institución. En el caso de las CCAF corresponde cuando más del 50% de los trabajadores están de acuerdo en desafiliarse, por lo tanto solamente existe fuga voluntaria del trabajador o decisión voluntaria de la empresa. Las compañías tienden a focalizar sus acciones en las fugas voluntarias, dado que pueden influir directamente en ellas y suelen producirse debido a problemas en la relación entre el cliente y la empresa [3].

En estudios relacionados con la retención de clientes en entidades bancarias han mostrado los beneficios de tener una cartera con un número mayor de clientes, dentro de ellas destacan que al realizar en promedio más transacciones aumentan las utilidades de la institución. Además, si se logra una mayor permanencia de un cliente se obtienen beneficios asociados a la disminución de los costos operacionales y al incremento en las transacciones [3].

La predicción de fuga es un elemento importante para la retención de clientes. Tanto la identificación de los clientes con tendencias a fuga como la determinación de su rentabilidad futura permiten focalizar los esfuerzos de retención en los clientes más apropiados. La Ilustración 8 muestra el ciclo de retención basado en una adecuada predicción de fuga [3].

Ilustración 8 Ciclo de Retención de Clientes

5.2 Modelos de Propensión y Minería de Datos

Los modelos de propensión utilizan el comportamiento histórico de los usuarios (en conjunto con información demográfica y de otras fuentes) con el fin de pronosticar su comportamiento futuro. Además, buscan identificar segmentos de clientes con mayor disposición a realizar una acción y estimar cuál es la probabilidad con que ese cliente ejecutará dicha acción.

Las técnicas de minería de datos pueden clasificarse en tres tipos de modelización: Predicción, Agrupación y Asociación. Los modelos de Predicción o modelos de *Aprendizajes Supervisados,* son utilizados para predecir comportamientos debido a que usan una variable dependiente de los atributos seleccionados. Existen distintos tipos de modelos de Aprendizaje Supervisado, entre ellos están los Árboles de Decisión, Redes Neuronales, Support Vector Machines (SVM) y las Redes Bayesianas. Estos modelos entregan una ponderación o peso de las variables independientes que explican el comportamiento de la variable dependiente. El objetivo es seleccionar el modelo que entregue la mejor precisión (accuracy) para predecir.

Los modelos de Agrupación o de *Aprendizaje No supervisados*, a diferencia de los Supervisados, no utilizan una variable dependiente, sino que generan información solamente a partir de las variables independientes. Por lo general son utilizados para agrupar datos o Cluster. Algunos de los métodos No Supervisados son Fuzzy c-means y K-means.

Las técnicas de Asociación se pueden concebir como modelos de predicción generalizados. En ellas los campos pueden actuar simultáneamente como inputs y outputs. Las reglas de asociación tratan de asociar una conclusión particular con un conjunto de condiciones.

5.2.1 Test Anova

Para seleccionar las variables que realmente influyen en la fuga se utilizará el Test Anova, el cual es una análisis de varianza que se utiliza para determinar si existen diferencias significativas entre las medias de tres o más variables independientes. El test Anova considera tres hipótesis que existe independencia, normalidad y homocedasticidad entre sus variables. El test pone a prueba la siguiente hipótesis nula

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \ \ldots = \mu_k$$
$$H_1: \mu_1 \neq \mu_2$$

Donde μ es la media del grupo y k es el numero de variables. En donde, si se obtiene un valor significativo se aprueba la hipótesis nula y si la significancia es menor a 0,05 se rechaza, lo que significa que aceptamos que hay diferencias entre las medias.

5.2.2 Árboles de Decisión

Los árboles de decisiones son uno de los modelos de minería de datos más usados, tanto por su valor en la predicción, como por su simpleza en la interpretación. Corresponden a modelos estadísticos en los que interesa explicar una variable dependiente cualitativa, en este caso una variable de binaria en donde 1 es fuga y 0 no fuga, en función de varias variables explicativas [4]. El método consiste en dividir el tablón de datos en varios subconjuntos descendientes según la proporción de casos positivos de la variable objetivo que se concentre en cada uno de ellos.

Todas las observaciones se encuentran inicialmente en un mismo grupo, en el cual hay tantos casos de fuga como no fuga. Para la variable objetivo la proporción de casos positivos posee una probabilidad p, llamada prior o probabilidad inicial del evento. El método consiste en ir dividiendo en varios subconjuntos y cada uno de ellos con distinta proporción de casos positivos. Esta división se realiza a través de la variable independiente que sea más discriminante en cada caso y por el punto de corte más óptimo. Este proceso se realiza de forma recursiva hasta que se cumplan los criterios de parada.

El árbol se representa como un nodo inicial que se divide en distintos nodos conectados. Cada nodo representará un segmento de la población que caracterizará un patrón de comportamiento respecto la variable objetivo.

Existen cuatro métodos de división para los árboles [5]:

• CHAID (Chi-square automatic interaction detector): Consiste en un rápido algoritmo de árbol estadístico y multidireccional que explora datos de forma rápida y eficaz, y crea segmentos y perfiles con respecto al resultado deseado. Permite la detección automática de interacciones mediante Chi-cuadrado. En cada paso, CHAID elige la variable independiente (predictora) que presenta la interacción más fuerte con la variable dependiente. Las categorías de cada predictor se funden si no son significativamente distintas respecto a la variable dependiente [5].

• CHAID exhaustivo: Supone una modificación de CHAID que examina todas las divisiones posibles para cada predictor y trata todas las variables por igual, independientemente del tipo y el número de categorías [5].

• Árboles de clasificación y regresión (CRT-Classification and regression trees): Consiste en un algoritmo de árbol binario completo que hace particiones de los datos y genera subconjuntos precisos y homogéneos. CRT divide los datos en segmentos para que sean lo más homogéneos posible respecto a la variable dependiente [5]. El criterio que guía el crecimiento del árbol CRT cuando el campo resultado es simbólico se conoce como impureza, el cual captura el grado en el que las respuestas dentro de un nodo se concentran dentro de una única categoría. Un nodo puro sería aquel en el que todos los casos caen dentro de la misma categoría, mientras que un nodo con un máximo de impureza tendrá el mismo número de casos para cada categoría del campo resultado. La impureza se puede definir en mediante el *índice Gini de Dispersión,* en donde si $P(t)_i$ es la proporción de casos en el nodo t que pertenecen a la categoría de i del campo resultado, entonces el índice Gini es:

$$Gini = 1 - \sum_i P(t)_i^{\,2}$$

• QUEST (Quick, unbiased, efficient, statistical tree): Consiste en un algoritmo estadístico que selecciona variables sin sesgo y crea árboles binarios precisos de forma rápida y eficaz. Con cuatro algoritmos tenemos la posibilidad de probar métodos diferentes de crecimiento de los árboles y encontrar el que mejor se adapte a nuestros datos. Es un método rápido y que evita el sesgo que presentan otros métodos al favorecer los predictores con muchas categorías. Sólo puede especificarse QUEST si la variable dependiente es nominal [5].

5.2.3 Boosting Algorithm

El principal objetivo de boosting es mejorar el desempeño de la clasificación a través de la combinación de decisiones provenientes de varios modelos de clasificación, los cuales son llamados clasificadores débiles (o algoritmos de aprendizaje débil). Boosting ha sido exitosamente aplicado para predecir la fuga de clientes en retail y compañías de comunicaciones. Los clasificadores débiles son usados como subprocesos combinados de manera de construir un clasificador altamente preciso en la data de entrenamiento.

5.2.4 Regresión Logística

Las modelos de regresión logísticas son modelos estadísticos en los que se desea conocer la relación entre una variable dependiente binaria y una o más variables explicativas. La variable respuesta es la probabilidad de éxito, es decir cuando la variable binaria es igual a 1 en el caso de fuga, luego permitirá clasificarlos en una de las categorías de esta variable [4].

La ecuación de partida en los modelos de regresión logística es:

$$P(y = 1|x) = \frac{exp(b_0 + \sum_{i=1}^{n} b_i\, x_i)}{1 + exp(b_0 + \sum_{i=1}^{n} b_i\, x_i)}$$

Donde

$$P(y = 1|x) = p\ es\ la\ probabilidad\ de\ que\ cumple\ el\ evento$$

$$x_i\ desde\ i = 1\ hasta\ n\ variables\ explicativas$$

5.2.5 Medidas de Evaluación de Desempeño del Modelo

La sensibilidad, la especificidad y la precisión son ampliamente utilizado estadísticas para cuantificar lo bueno y fiable que es una prueba. La sensibilidad evalúa que tan bien es en la detección de casos positivos y la especificidad estima que casos negativos pueden ser descartados correctamente. La curva ROC es una presentación gráfica de la relación entre sensibilidad y especificidad y que ayuda a decidir el modelo optimo a través de la determinación del mejor umbral para la prueba.

Las medidas de exactitud permiten determinar que tan correcto una prueba identifica y excluye una condición determinada. El Accuracy puede ser determinado a partir de la sensibilidad y la especificidad con la presencia de la prevalencia.

En la Ilustración 9 se presenta cada uno de los casos de acuerdo a lo obtenido por la estimación y el resultado real, en donde se explica la cantidad de casos que coinciden entre el caso real y el estimado por el modelo.

Ilustración 9 Matriz de Confusión

		Clase real	
		Clase referencia	***Clase no referencia***
Clase estimada	Clase referencia	TP	FP
	Clase no referencia	FN	TN

Por ejemplo, *True Positives* (TP) corresponden a casos donde la empresa no se fugó en la realidad y el modelo predice que esa empresa no se fugara, al igual que *los True Negatives* (TN) que serán casos en que en la realidad la empresa se fugó y el modelo arroja correctamente que será una empresa fugada. Por otra parte los *False Positives* (FP) y los *False Negatives* (FN) corresponden a los errores tipo I y tipo II, en donde son empresas que el modelo arroja que se fugarán pero en la realidad no lo hicieron (Error tipo I) o que el modelo arroja que no se fugarán pero que en realidad si se fugaron (Error tipo II) [6].

5.2.5.1 Accuracy

El Accuracy de un modelo corresponde a la proporción de los resultados verdaderos sobre la población. Los resultados positivos corresponden a los *true possitives* (TP) y *true negatives (TN),* mientras que la población corresponde al resultado de la suma de *true possitives* (TP), *false possitives* (FP), *true negatives* (TN) y *false negatives* (FN).

$$Acurracy = \frac{Número\ de\ True\ Possitives + Número\ de\ True\ Negatives}{Número\ de\ True\ Positives + False\ Positives + True\ Negatives + False\ Positives}$$

Un Accuracy del 100% significa que los valores estimados son exactamente igual a los valores reales. Por otra parte la sensibilidad y la especificidad se calculan de la siguiente forma:

$$Sensibilidad = \frac{Número\ de\ True\ Possitives}{Número\ de\ True\ Positives + False\ Negatives}$$

$$Especificidad = \frac{Número\ de\ True\ Negatives}{Número\ de\ True\ Negatives + False\ Positives}$$

Como se sugiere por las ecuaciones anteriores, la sensibilidad es la proporción de verdaderos positivos que se identifican correctamente por un modelo y demuestra lo bueno que la prueba es en la detección de casos positivo dado que son positivo. La especificidad es la proporción de los verdaderos negativos identificados correctamente por un modelo y sugiere lo bueno que la prueba en identificar los casos que realmente son negativos. Se considera que un modelo es bueno cuando su accuracy es superior al 75%. [6]

Además, de las ecuaciones anteriores, el *Accuracy* puede determinarse a partir de sensibilidad y especificidad, donde prevalencia es desconocida. La prevalencia es la probabilidad de exista cierta condición en la población en un momento dado:

$$Accuracy = Sensibilidad * Prevalencia + Especificidad * (1 - Prevalencia)$$

5.2.5.2 Precisión

La precisión o predictor de los valores positivos, esta definido como la proporción de *true positives* sobre el total de casos positivos (*True positives* y *False positives*). Por lo general, se expresa como un porcentaje [6].

$$Precisión = \frac{Número\ de\ True\ Positives}{Número\ de\ True\ Positives + False\ Positives}$$

5.2.5.3 Sensibilidad

Recall, o también llamada Sensibilidad, como se comentó en la parte anterior corresponden al porcentajes los casos positivos que un modelo predice sobre el total de caso real que ocurrieron realmente. Suele representarse como un porcentaje.

$$Sensibilidad = \frac{Número\ de\ True\ Possitives}{Número\ de\ True\ Positives + False\ Negatives}$$

Un predictor perfecto sería descrito como el 100% de sensibilidad, es decir todos los valores estimados positivos coinciden con el total de valor reales. La sensibilidad determina que tan capaz es un modelo de identificar una condición correctamente [6].

5.2.5.4 AUC

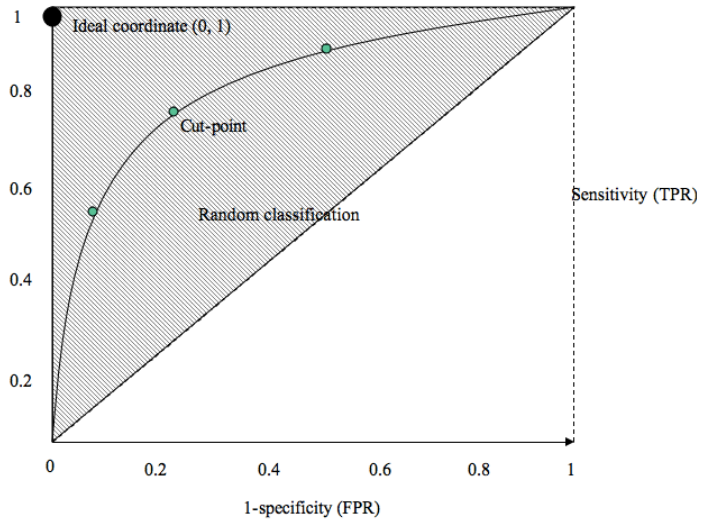
AUC o *Area Under Curve*, corresponde al área que se encuentra bajo la curva ROC o Receiver Operating Characteristic, la cual es utilizada como medida de desempeño en un modelo utilizando el True Positive Rate (TPR) o Sensibilidad y el False Positive Rate (FPR). Donde FPR es,

$$FPR = \frac{False\ Positives}{False\ Positives + True\ Negatives} = (1 - Especificidad)$$

Todas las combinaciones posibles de TPR y FPR componen un espacio ROC. Un TPR y un FPR juntos determinan un simple punto en el espacio ROC y la posición de un punto en el

espacio ROC muestra el trade off entre la sensibilidad y la especificidad, es decir, un aumento en la sensibilidad es acompañado por una disminución de la especificidad. Así, la ubicación de un punto en el espacio ROC representa si la clasificación estimada es buena o no. En una situación ideal, un punto determinado por un TPR y FPR con coordenada (0,1) representa una sensibilidad de 100% y una especificidad de 100%, llamada también clasificación perfecta. Cuando se tiene un 50% de sensibilidad y 50% de especificidad puede ser visualizado en la diagonal determinada por coordenadas (0, 0) y las coordenadas (1, 0) como se muestra en la imagen a continuación. En teoría, una conjetura al azar daría un punto a lo largo de esta diagonal. Un punto predicho que cae en el área por encima de la diagonal representa un buen modelo de clasificación, de lo contrario corresponde a una mala predicción [6].

Ilustración 10 Espacio de ROC. La área sombreada corresponde a mejor pronostico de clasificación.



Los diferentes puntos de cortes determinan la curva ROC, los puntos de la curva más cercanos a la diagonal poseen menor precisión mientras que los más lejanos y cercanos a (0,1) corresponden al caso ideal y con mejor precisión. Por lo tanto, el área bajo la curva o AUC provee una tipo de medida de precisión de un modelo y se calcula de la siguiente forma [6].

$$AUC = \int_0^1 ROC(t)dt$$

$$Donde\ t = (1 - Especificidad)\ y\ ROC(t)\ esla\ sensibilidad$$

De acuerdo a los resultados obtenidos se puede clasificar la precisión del modelo de acuerdo a la Tabla 3.

Tabla 3 Clasificación de un Modelo según el Rango AUC

Rango AUC	Clasificación
0.9<AUC<1.0	Excelente
0.8<AUC<0.9	Bueno
0.7<AUC<0.8	Bajo
0.6<AUC<0.7	Malo

La curva ROC es una buena herramienta para seleccionar un posible punto de corte óptimo para determinar la precisión de un modelo.

5.2.5.5 LIFT

Lift permite comparar entre diferentes resultados producidos por distintos algoritmos, está especialmente orientado a la evaluación de problemas de clasificación. Por ejemplo, para el caso de fuga se pueden aplicar diferentes modelos y se escoge el que arroja mejor probabilidad para clasificar a los fugados y no fugados, la diferencia de eficacia entre estos métodos es medida a través del Lift.

El Lift se calcula a través del cociente entre el porcentaje de concentración de elementos o hechos es una determinada clase, frente a la concentración que presenta la población en su conjunto.

$$Lift = \frac{Procentaje\ de\ objetivo\ en\ la\ clase}{Porcentaje\ de\ objetivo\ en\ la\ población}$$

No es un porcentaje sino un indicador de cuantas veces es mejor el modelo en la captación del hecho objetivo, que la aleatoriedad. Por lo general, es utilizado como medida para evaluar el desempeño de una regla sobre un conjunto de datos. Un alto lift indica que el modelo describe bien el comportamiento y por lo tanto, es un buen candidato para ser usado [6].

5.2.5.6 F-Measure

F-Measura corresponde a una medida que combina los conceptos de Precisión y Sensibilidad utilizado para análisis estadísticos de clasificación binaria como medida para determinar la precisión de un modelo. También denominado F-scores y se interpreta como el promedio ponderado de la Precisión y Sensibilidad, en donde su mejor valor es en 1 y el peor en 0 [6].

$$F = 2 * \frac{Precisión * Sensibilidad}{Precisión + Sensibilidad}$$

6 Marco Metodológico

En los siguientes pasos se describirá la metodología que se utilizará en la presente memoria para responder al problema de negocio planteado.

6.1 Definición del Problema y Revisión Bibliográfica

Se comenzó con una fase exploratoria, que abordó el conocimiento del negocio, debido a que las Cajas de compensación funcionan de forma distinta a las demás empresas, quizás de forma más similar a los bancos debido a que sus ingresos se obtienen a partir de los créditos otorgados a sus clientes y por lo general no se tiene un conocimiento común de la función que realizan. Además, se estudió la situación actual en la empresa y trabajos anteriores realizados en el ámbito de la fuga.

La revisión bibliográfica corresponde a estudios de tesis y artículos relacionados con fugas y retención de clientes. Además, se analizó información histórica acerca de la compañía y se investigó acerca de los diferentes modelos que se pueden aplicar al problema y como implementarlos en los Softwares estadísticos (Modeler SPSS). Con la finalidad de contextualizar a la industria de las Cajas de Compensación y plantear el proyecto.

6.2 Selección y Procesamiento de datos

En esta etapa se determinan las fuentes de los datos que se usarán. Cuáles son las bases de datos y que atributos se considerarán para el análisis. Se extrae los datos relevantes para el modelo y las variables de entrada al modelo predictivo.

6.3 Pre-procesamiento de la Información

Esta etapa está enfocada en la limpieza de los datos y eliminación de las inconsistencias que presente la base. Por una parte, dar solución a los valores faltantes y por otro, a los valores que estén fuera de rango, de tal manera de obtener una base de datos adecuada para su posterior transformación.

Los valores faltantes corresponden a la inexistencia del valor de un registro en cierta variable y los valores fuera de rango o "outlier" corresponden a observaciones numéricas distintas al resto de los datos.

6.4 Transformación de la Información

El objetivo principal de esta etapa es realizar el tratamiento preliminar de los datos, transformaciones y generación de nuevas variables a partir de las variables que sean entregadas en la base de datos original con tal de tener una estructura de datos más apropiada. El resultado debe ser una base datos totalmente numérica tanto para los análisis estadísticos como para utilizarla en los modelos.

Las transformaciones que se realizarán son: Cambios de texto o String a variables categóricas numéricas; Escalamiento y estandarización de las variables; y Creación de nuevas variables a partir de las variables originales (Generación de indicadores de las empresas).

6.5 Selección de la Técnica de Minería de Datos

En esta etapa se desarrollan los modelo predictivo, para ello se probaron distintas técnicas de minería de datos que pueden modelar el problema.

Dada las condiciones del problema primero se procedió a modelar árboles de decisiones, aprovechando su fácil interpretación encontrando segmentos de clientes con mayor propensión a fuga y a partir de su comportamiento se extrajeron resultados. Por otra parte, esté método permite reducir el número de variables independientes, esto junto al test de Anova permite seleccionar las variables más significativas para aplicar la Regresión Logística.

6.6 Evaluación e Interpretación de los Resultados

Se identificarán los patrones de comportamiento obtenidos y como estos concuerdan con la lógica del negocio. Se evaluaran los resultados de los modelos y se compararán los poderes de predicción de cada modelo sobre la fuga.

7 Desarrollo Metodológico

7.1.1 Recolección de Hipótesis de Propensión de Fuga en la Empresa

Mediante entrevista en profundidad con los expertos del negocio, se recopilarán las hipótesis más comunes de propensión de fuga de las empresas.

Como primer paso se hará una investigación dentro del área empresas y trabajadores, hablando con personas claves del negocio como Gerente de Productos y Servicios, Gerente Empresas y Trabajadores, Jefe de Calidad y Servicio, Ejecutivo Empresa, Jefe de Fidelización y Analistas de modo de poder levantar información acerca de cuáles son los motivos que ellos creen que incentivan a que una empresa decida desafiliarse y cuáles de ellas tienen mayor tendencia a fuga.

El objetivo de esta fase, es recolectar la mayor cantidad de información por parte de los expertos del negocio y en base a ella, plantear las hipótesis relevantes acerca de la fuga de las empresas.

7.1.2 Preparación de la Base

Se seleccionará una muestra de las empresas vigentes durante el año 2014 y el total de empresas fugadas durante el mismo periodo. Se analizará un año de historia de cada empresa, por lo que la base de datos cuenta con información del 2013 y 2014. Cada variable generada se calculará para 3, 6, 9 y 12 meses, con tal de escoger en qué periodo esa variable es más significativa.

7.1.2.1 Generación de Indicadores

Se incorporarán distintas variables relevantes de cada área del negocio, teniendo en cuenta las variables encontradas de la investigación del negocio, por ejemplo se incorporarán indicadores de la empresa acerca de los beneficios que han usados sus trabajadores, licencias médicas, créditos, visitas que han hecho sus trabajadores a sucursal o las visitas que les ha hecho Los Héroes a las empresas, antigüedad de la empresa afiliada, entre otras.

Además, se considerará un periodo de dos meses antes de la fecha de desafiliación de una empresa para la generación de indicadores del modelo, con tal de que se pueda predecir que empresas en dos meses a futuro se van a desafiliar y poder tomar acciones hoy para que no se desafilie en la fecha pronosticada.

7.1.3 Encontrar Segmentos Más Propensos en base al Modelo Predictivo de fuga

Se probará con distintos modelos, para determinar la probabilidad de fuga de cada empresa, se escogerá el modelo que entregue mejor precisión ya que se quiere la mejor predicción para los casos fugados. Con el modelo de árboles de decisión se puede obtener

directamente los segmentos más propensos, mientras que con la regresión logística a partir de la probabilidad que entregue el modelo se segmentará según la significancia de la propensión.

7.1.4 Prueba Hipótesis Planteadas

De acuerdo a los resultados obtenidos de los modelos y la significancia de cada variable se intentará concluir si las hipótesis que se plantean se cumplen o no. Por una parte, se realizará un Test ANOVA para comparar las medias de las variables que influyen en las hipótesis con la variable dependiente para tener una visión global si la variable es significativa sobre la fuga y por otra parte, a partir del comportamiento de las empresas y las variables que escoge en el árbol se intentará desprender conclusiones que aprueben o desvaliden las hipótesis.

7.1.5 Analizar en Conjunto la Propensión de fuga y Rentabilidad de los segmentos de Empresas.

Para determinar en qué empresas conviene enfocar las acciones de retención se analizaran en conjunto la probabilidad de fuga que arroje el modelo y la rentabilidad de la empresa en cada segmento o el valor del segmento para la Caja, dado que existen empresas que no presentan un alto margen pero si tienen un Valor especial para la Caja, por ejemplo que el rubro al que pertenecen sea Servicios Públicos.

7.1.6 Proponer Estrategia de Retención

Se propusieron diferentes estrategias de retención de empresas para los segmentos con mayor probabilidad de fuga que se obtuvieron del modelo, considerando los análisis de comportamiento de cada grupo enfocando los tipos de acciones realizar. Se priorizaron los segmentos de empresas que son más rentables para la empresa y tienen mayor probabilidad de fuga.

8 Investigación Exploratoria de la Empresa y Planteamiento de Hipótesis de Fuga de Clientes

Existen diversos motivos que pueden llevar a que una empresa se desafilie de una caja, en este caso particular, los reclamos hechos a la Caja en estudio están concentrados en la disconformidad con los productos.

Los créditos, al igual que las licencias, afectan directamente a los trabajadores, los reclamos están ligados al pago de cuotas o montos que deben pagar, en los cuales el trabajador no está de acuerdo con lo que le están cobrando o las cuotas que debe pagar.

Otros factores que pueden influir en la decisión de cambiarse de caja son la asignación de beneficios que ofrece la Caja en estudio a cada empresa y es donde compite directamente con las demás cajas. Por ejemplo, a las empresas foco en donde la Caja quiere crecer comercialmente les asigna un monto para beneficios llamado PDB, que pueden ser utilizados en fiestas de fin de año, paseos, entradas al cine, seguros para sus trabajadores, salud, entre otros. Por ejemplo, puede suceder que por motivos de presupuesto o decisiones comerciales de la Caja este beneficio no se les sea asignado nuevamente a ciertas empresas, lo cual genera una molestia en los trabajadores y pueden verse tentados por la oferta de otra caja con mayores beneficios.

El objetivo principal del área de Fidelización de Los Héroes es reducir y controlar la Fuga de clientes, para esto se realizan diversas acciones ya sea la entrega de beneficios como merchandising para posicionar a la Caja o convenios con el área de Salud. Durante años se han realizado diversas acciones para retener clientes, especialmente en el sector Pensionado en donde la caja en estudio posee el liderazgo, no así para el segmento trabajador. Es por esto que surge la necesidad de definir los segmentos de empresas más propensas a fuga con tal de implementar en ellas estrategias más fuertes de retención de clientes, optimizando recursos y esfuerzos de la Caja.

De acuerdo al tipo de empresa que tenga afiliada, la CCAF en estudio ofrece distintos beneficios, por lo general aquellas que poseen mayor colocación y tienen mejor condición crediticia se les realiza campañas para que usen beneficios más fuerte. Esto le permite enfocar más recursos en aquellos clientes que son más rentables y que desea retener por sobre aquellos que no generan ingresos para la caja y es preferible que dejen de ser afiliados ya que generan más costos administrativos que utilidades [3] provocando pérdidas para La Caja.

A modo introductorio se definirán algunos de los procesos que realizan las CCAF y que influyen en sus trabajadores:

Licencias Médicas: Los trabajadores presentan su licencia médica al Empleador, y quien debe entregarla a la CCAF que están afiliadas. La caja envía la licencia al COMPIN (Comisión de Medicina Preventiva e Invalidez) y esta entidad aprueba o rechaza la cantidad de días de la licencia. La resolución de la licencia es reenvía a la CCAF quien paga el monto por los días de licencia.

Asignación Familiar: corresponde al subsidio estatal para las cargas legales de los trabajadores. De acuerdo a la cantidad de cargas que tenga un trabajador y al tramo al que pertenezca el trabajador se le paga la Asignación familiar. El estado entrega los recursos para pagar las cargas y la Caja es quien le paga al Empleador las asignaciones familiares de cada uno de sus trabajadores.

Tabla 4 Tramos de Asignación Familiar

Tramo	Desde	Hasta	Monto por carga
A	$ 1	$ 236.094	$ 9.242
B	$ 236.095	$ 344.840	$ 5.672
C	$ 344.841	$ 537.834	$ 1.793
D	$ 537.835	-	$ 0

Crédito Social: corresponde a los créditos que ofrece la CCAF a sus afiliados. A diferencia de los bancos, ofrece una tasa menor y las cuotas a sus trabajadores son descontadas por planillas.

Beneficios: corresponden a las prestaciones adicionales que ofrece la Caja, ya sea convenio con distintas áreas para ofrecer descuentos en Salud, entretención, vida sana, educación, etc. Como también bonos que entrega a sus trabajadores.

En cuento a algunas de las acciones que se realizan con las empresas focos (Con más de 250 trabajadores) de La Caja:

Ejecutivo Empresa: Fidelizar a las empresas afiliadas a La Caja, a través de relaciones y vínculos de confianza con los actores claves de las empresas, promoviendo productos y servicios de La Caja y administrando su entrega a partir de los procedimientos establecidos por el área, con la finalidad de mantener la permanencia de las empresas y alcanzar los objetivos fijados.

- Dealer: encargado de ofrecer créditos a los trabajadores.
- Visitas de Mantención: realizadas por el ejecutivo empresa para levantar información acerca del estado de la empresa, si tiene riesgos de fuga se levantan alertas.
- PDB: corresponde a un presupuesto de dinero para aquellas empresas en las cuales se quiere crecer comercialmente, el cual lo pueden destinar a diferentes eventos durante el año, por ejemplo fiesta de fin de año.

8.1 Investigación Exploratoria

Para un mejor entendimiento del negocio se realizaron diferentes entrevistas en profundidad con expertos del negocio, de donde se levantaron algunos de los motivos que puede llevar a que un empresa se desafilie, como se muestran a continuación:

1. Subgerencia de Mantención

Está área se relaciona día a día con las empresas mediante las visitas de mantención, de donde se levanta información acerca de la situación actual de la empresa o problemas que haya tenido con la Caja. Aquí se generan alertas "Altas" como que el sindicato llamo asamblea para votar por cambio de caja debido a que la competencia tiene una mejor oferta o alertas "Medias" donde existe algún malestar debido algún servicio como por ejemplo, atrasos en el pago de una licencia médica. Por otra parte, está área resalta que las empresas a las que se le realizan visitas son aquellas que La Caja realiza acciones de retencion, dado que no importa la cantidad de trabajadores que posea la empresa, sino la colocación. Por ende, las que se visitan son aquellas con una gran cantidad de créditos colocados y trabajadores.

2. Gerencia de Empresas y Negocios

El área de empresas destaca que uno de los factores que influye en que una empresa se vaya es debido a las mejores ofertas de las otras cajas, como por ejemplo que tengan mejores beneficios o centro recreacionales, que entreguen más bonos de educación, entre otros. Por otra parte, señala que existen problemas en particular que pueden hacer que se levante una votación para desafiliarse como por ejemplo problemas con el pago de las licencias médicas, o que no se estén pagando correctamente las asignaciones familiares.

3. Gerencia de Productos –Servicios Previsionales

Todas CCAF gestionan el pago de las Asignaciones Familiares y Licencias Médicas, por lo que este servicio es un *commodity* entre las cajas. Lo que marca la diferencia en Licencias Médicas entre las cajas es cuánto tiempo se demoren en pagar la licencia y que el monto sea correcto. Dentro de los principales problemas que reclaman los trabajadores es debido a que no le pagaron el monto que solicitaron, es decir, el COMPIN aprobó menos días de los que la licencia contenía. Esto genera una molestia para el trabajador y pasa a ser La Caja el responsable del problema. Por otra parte, puede que el proceso de pago tome más días de lo normal, lo que nuevamente puede generar malestar en los trabajadores y un descontento con la caja.

Las licencias médicas están ligadas directamente a los trabajadores, la Caja en estudio le paga solamente a los afiliados que pertenecen a Fonasa. Esto genera mayores transacciones para la caja ya que lo realiza a nombre del Estado, que es quien entrega los recursos.

Pese a que la decisión de afiliarse a una caja es acogida por el voto de la mayoría de los trabajadores, puede llegar a ocurrir que se demoren en pagar una licencia médica a un trabajador de la empresa más del tiempo estimado y esta información llega a oídos del presidente del sindicato, él puede influir y generar un motivo para levantar sesión para votar por desafiliarse de la caja. Es por esto que un aliado fundamental para los agentes que visitan las empresas focos es el presidente del sindicato ya que con ellos se puede acceder a información acerca de la situación

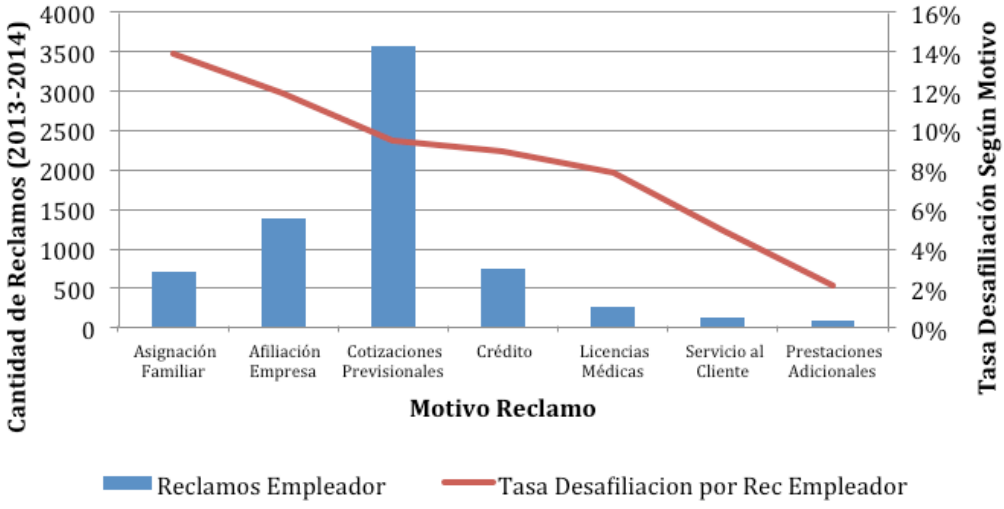
actual de la empresa, si está se encuentra conforme con el servicio o presenta intenciones de cambiarse de caja.

El pago de Asignaciones Familiares, a diferencia de los créditos y las licencias médicas, están directamente relacionadas con la empresa, ya que es ésta quien debe pagar las asignaciones familiares a sus trabajadores. La empresa de acuerdo a la renta y a las cargas familiares de sus trabajadores paga las asignaciones familiares y luego presenta la documentación correspondiente a la caja, la cual le devuelve el monto que pago a sus trabajadores con recursos que el Estado destina para ello. Suele suceder que la empresa paga la asignación familiar de acuerdo a la información que posee de sus trabajadores pero la caja posee información diferente de sus trabajadores, ya sea porque los hijos ya son mayores de edad, o el trabajador pertenece a otro estrato de renta que provocan que la asignación familiar cambie, por lo que la caja paga un monto menor de lo que ya la empresa pagó a sus trabajadores. A esto se le llama Diferencia de compensación y genera un saldo en contra para el empleador. Además, se generan problemas por la entrega de documentación, ya que provocan un atraso en los pagos y afecta la fecha de devolución del dinero a la empresa. Por lo tanto, la empresa al verse afectada por problemas de este tipo puede tomar la decisión de cambiarse de caja.

4. Gerencia de Calidad y Servicio

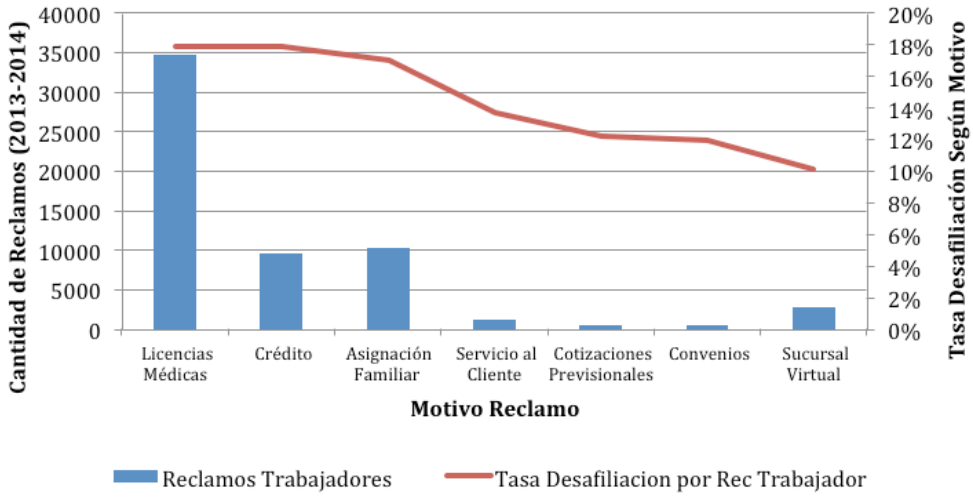
En la gerencia de calidad y servicio se encuentra el área de reclamos de donde se levantó información acerca de los motivos más comunes por los que una empresa realiza un reclamo. Los reclamos se pueden separar de acuerdo a quienes los realizan, por una parte pueden ser los trabajadores y por otra el empleador. En la Ilustración 11 e Ilustración 12 se muestran los productos con más reclamos hechos por los empleadores y trabajadores, además la línea roja corresponde a la tasa de desafiliación dentro del mismo grupo. Por lo que se puede apreciar que existen 4 productos que destacan por su cantidad de reclamos: licencias médicas, crédito, asignaciones familiares y cotizaciones previsionales.

Ilustración 11 Cantidad de Reclamos hechor por el Empleador por Producto durante el 2013-2014 y su tasa de desafiliación.



En el gráfico se aprecia que el motivo por el cual el empleador mayormente reclama y se desafilia es por asignación familiar, esto debido a que es un producto que le afecta directamente en las finanzas a la empresa.

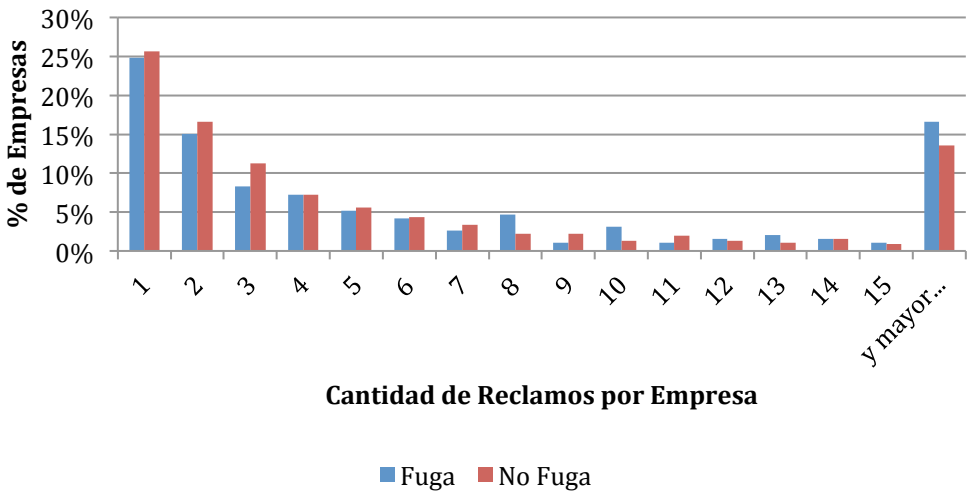
Ilustración 12 Cantidad de Reclamos hechos por los Trabajadores Durante el 2013-2014 y la tasa de desafiliación de los trabajadores por Producto.



En el caso de los trabajadores principalmente sus reclamos están concentrados en el proceso de pago de licencias además la tasa de desafiliación muestra que el 18% de los trabajadores que reclamaron por licencias médicas se desafiliaron de la Caja.

En la Ilustración 13, se muestra el porcentaje de trabajadores por cantidad de reclamos que realizaron el último año. Se puede apreciar que la cantidad de empresas fugadas que realiza reclamos menores a 3 tienden a ser menor que las afiliadas, mientras que las empresas fugadas que realizan más e 3 reclamos por trabajador tiende a superar a las empresas afiliadas.

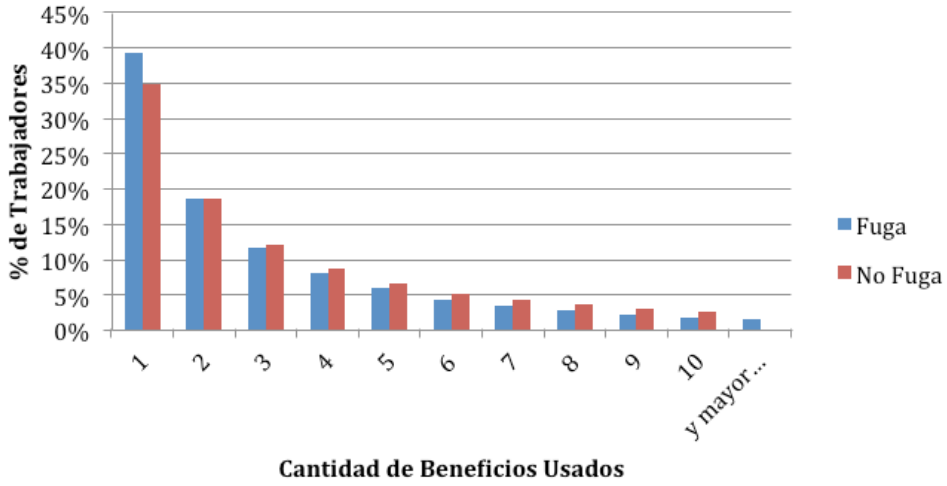
Ilustración 13 Histograma de la Cantidad de Reclamos por Empresa



5. Área de Beneficios

De los diferentes beneficios que ofrece La Caja, se destacan dos que correspondientes a convenios con entidades de Salud (farmacia Cruz Verde y Red Salud). Estos beneficios capturan el mayor porcentaje de los beneficios. En la Ilustración 14, se muestra el porcentaje de trabajadores versus la cantidad de beneficios que uso durante el último año, donde se puede apreciar que los trabajadores fugados tienden a usar 2 o menos beneficios, mientras que los trabajadores afiliados usan 2 o más beneficios cada uno.

Ilustración 14 Histograma con el porcentaje de trabajadores que usó 1 o más beneficios el último año.



8.2 Definición de Hipótesis

De acuerdo a la investigación exploratoria, se levantaron distintas motivos que pueden influir en la fuga, los cuales se agruparon en tres grupos:

1. Problemas Operativos

- Reclamos vs Uso de Beneficios: mientras mayor sea la cantidad de reclamos que realiza la empresa y menor el uso de beneficios, mayor la fuga.
- Tiempo de Pago LM: a mayor demora en el pago de LM la fuga.
- Diferencia de Compensación: a mayores errores por la diferencia de compensación mayor la fuga.
- Colocación de Créditos: A mayor colocación de créditos o penetración, menor es la fuga.

2. Mejores Ofertas de la Competencia

- CCAF de destino: mientras más se repita una caja de destino será debido a que presenta una mejor oferta.

3. Empleados Descontentos con Beneficios

- Porcentaje de Uso de Beneficios: a mayor porcentaje de trabajadores usando beneficios dentro de la empresa, menor será su descontento, por lo que presenta menor fuga.
- Beneficios promedio por trabajador: a menores beneficios promedios por trabajador, mayor será el descontento de la empresa, por lo que presenta mayor fuga.
- PDB: si existe disminución o pérdida de PDB aumenta la fuga.

Por lo tanto, a modo de resumen las hipótesis que se plantean y se demostrará su veracidad o se rechazarán son las siguientes:

H1. A mayor cantidad de reclamos mayor probabilidad de fuga.
H2. A menor uso de beneficios mayor es la fuga.
H3. A mayor tiempo de demora en el pago de licencias médicas mayor es la fuga.
H4. A mayor diferencias de compensación con la empresa mayor la fuga.
H5. A mayor colocación de créditos menor es la fuga.
H6. Perdida del PDB aumenta la fuga

9 Desarrollo de Modelo de Propensión de Fuga

Para comenzar la construcción de la base de datos del modelo se consideraron las siguientes variables más relevantes de los t*rabajadores* según la investigación de la empresa que se realizó:

Ilustración 15: Variables de las Bases de Datos a Utilizar (1)

Área	Nombre Variable	Definición
Perfil Empresa	CUENTA EMPRESA	ID de cada Empresa dentro de Los Héroes
	Fecha_Dec	
	RUT_EMP	Rut de la empresa
	Seg	Micro,PYME o GGEE (Grandes Empresas)
	Est_Cred	Condición créditicia que se le asigna a la empresa
	CCAF Destino	CCAF a la que se afilia un empresa fugada de Los Héroes
	Tipo Cuenta	Privada o Pública
	Segmentación Comercial	Gendarmería, Minería,Educación, Servicios Públicos, Salud Privada,Asimet y Otros.
	Agente	Con o sin agente empresa
	Sub_Mant	Con o sin Subgerente de Mantención
	Antigüedad	Años de afiliación a Los Héroes
	Dealer	Si cuenta con Dealer o no (Vendedor de créditos)
	Cant_trab	Cantidad de trabajadores afiliados
	Fecha_Dec	Fecha en que la empresa se desafilió
Beneficios	Rut	Rut del Trabajador
	FECHA_VTA	Fecha en que el trabajador usó el beneficio
	BENEFICIO	Tipo de Beneficio que se utilizó
	PDB	Presupuesto para eventos
Reclamos	Rut	Rut de quien realizó el reclamo
	Fecha Creación	Fecha en que el cliente realizó el reclamo
	Producto	Sobre que servicio está reclamando
	Motivo	Subconjunto del producto por el cual está reclamando
	Tipo	Si corresponde a un Reclamo o Sugerencia o Solicitud
	Tipo cliente	Si es Trabajador, Carga, Pensionado o Empleador

Ilustración 16: Variables de las Bases de Datos a Utilizar (2)

Área	Nombre Variable	Definición
Créditos	fecha_reporte	Fecha en que se ingresa el crédito al sistema
	Rut_Clie	Rut de quien solicito el crédito
	tipo_cli	Tipo de Cliente: Pensionado o Trabajador
	estado	Estado del Credito Vigente, Moroso o Castigado
	monto_bruto	Monto Bruto del Crédito Solicitado
	monto_liq	Monto liquido del Crédito Solicitado
	fecha_coloc	Fecha en que se colocó el crédito
	fecha_aprob	Fecha en que se aprobó el crédito
	fecha_pri_cuo	Fecha en que comienza a pagar el crédito
	fecha_ult_cuo	Fecha en que deberia terminar el pago del crédito
	fecha_ult_pago	Fecha de la última cuota pagada
	total_cuo	Total de cuotas del crédito
	total_cuo_mor	Total de cuotas morosas del crédito
	total_cuo_vig	Total de cuotas vigentes del crédito
	total_cuo_pa	Total de cuotas pagadas del crédito a la fecha.
	valor_pri_cuo	Valor de la primera cuota
	valor_ult_cuo	Valor de la ultima cuota
	rut_clie	Rut del cliente que solicita el crédito
	end_max	Endeudamiento maximo del cliente
	end_dispon	endeudamiento disponible del cliente
	rut_emp	Rut de la empresa a la que pertenece
	cta_emp_bt	Cuenta Empresa a la que pertenece
Licencias Médicas	RUT	Rut Trabajador
	PAGO_TOTAL	Monto que se le paga por la licencia
	FECHA PAGO	Fecha en el dinero del pago esta disponible para ser retirado por el trabajador.
	FECHA PAGO EF	Fecha efectiva en que se le realiza el pago de la LM
	C1_RUT_EMP	Rut de la empresa a la que pertenece
	A1_DIA_REP	Cantidad de días de reposo según la LM
	C1_FEC_REC	Fecha en que el Empleador recibe la LM
	C2_REC_CCA	Fecha en que la CCAF recibe la LM
	FEC_PRO_PAG	Fecha probable de pago de la LM
	PRE_DEV	Número de veces con devolución de la LM
	LIC_DEV	Licencia devuelta por el COMPIN
	TIP_RECHAZO	Tipo de Rechazo por COMPIN
	DIA_CARENCIA	Días de Carencia
	PAG_FPLA	Pago fuera de plazo
	B_DIA_AUT	Días autorizados por el COMPIN
	B_TIP_AUT	Tipo de continuidad de la LM
	B_TIP_RES	Tipo de resolucion de la LM
	B_CAU_REC	Causa de rechazo de la LM
	R.PARTNER_EMP	Cuenta Empresa a la que pertenece el trabajador

Ilustración 17: Variables de las Bases de Datos a Utilizar (3)

Área	Nombre Variable	Definición
Reportes de alertas (Visitas de Mantención)	CUENTA EMPRESA	Cuenta Empresa que fue visitada
	FECHA DE VISITA	Fecha en que se realizó la visita a la empresa
	DECLARAR RIESGO DE DESAFILIACIÓN DE LA EMPRESA	Declaración MUY ALTA,ALTA,MEDIA, BAJA o MUY BAJA
Visita Sucursal	RUT CLIENTE	Rut del cliente que visito sucursal
	TIPO CLIENTE	Tipo de Cliente que visito sucursal
	CUENTA EMPRESA	Cuenta Empresa a la que pertenece el cliente
	TOTALPACK	Es 1 si ingreso Rut en el Totalpack, 0 si no.
	TP_FECHA	Fecha en que asistio a al sucursal

9.1 Preparación de los Datos

Se recolectó una base de datos con un historial de 2 años, desde Diciembre del 2012 a Diciembre del 2014. Las bases de datos de cada área cuentan con diferentes cantidades de registros por lo que se trabajaron de forma separada.

Del total de empresas afiliadas a la Caja correspondientes a 11500 aproximadamente se tomó una muestra de 1000 empresas afiliadas vigentes durante el 2014 y se utilizaron el total de empresas desafiliadas durante el mismo periodo correspondientes a 300. Para analizar la fuga de estas empresas, se tomó como periodo de estudio un año hacían atrás, descontando dos meses desde que ocurre la fuga, como se muestra en la **¡Error! No se encuentra el origen de la referencia.**. Los dos meses ocultos son para determinar el momento en que se toma la decisión de desafiliarse y utilizar estos dos meses para tomar acciones de retención con estas empresas. Para el caso, de las empresas desafiliadas, se utilizo un año móvil a partir del mes en que se desafilio, dejando los dos meses ocultos al igual que el caso de las empresas afiliada, con tal de extraer el comportamiento de los últimos 12 meses antes de tomar la decisión de desafiliarse.

Figura 1 Periodo histórico analizado para cada empresa afiliada.

Se realizó una preparación de la base eliminando los registros nulos en rut o cuenta empresa debido a que imposibilita cruzar los diferentes registros perdiendo información. Además, se transformaron los datos de texto a variables numéricas, como por ejemplo para el caso de la variable Segmento (Seg) donde por ejemplo "1" es Micro, "2" es PYME y "3" es grandes empresas.

Con las variables que se seleccionaron para trabajar de cada área (Ilustración 15, Ilustración 16 y Ilustración 17) se procedió a generar indicadores, con tal de obtener a partir de los registros de cada trabajador uno de forma agregada de la empresa.

Se asignó la siguiente variable como variable objetivo.

$$Fuga\ (Y) = \begin{cases} 1\ Empresa\ Desafiliada \\ 0\ Empresa\ Afiliada \end{cases}$$

9.2 Indicadores Generados

A continuación se listan las variables e indicadores generados para cada empresa clasificados de acuerdo al área de donde se obtuvo la información. Cabe destacar además, que todos los registros se encontraban a nivel de trabajador, por lo que tuvieron que ser agregados a nivel de cuenta empresa ya que la decisión de desafiliación es tomada en conjunto de todos los trabajadores de una empresa.

1. Variables Demográficas

Las variables demográficas tienen por fin incluir en el entrenamiento del modelo variables que describan características generales de las empresas afiliada. Las variables demográficas incluidas en el tablón son las siguientes:

- **Antigüedad:** Años en que la empresa ha estado Afiliada a Los Héroes.
- **Quintil Antigüedad:** Quintil de la Antigüedad de la Empresa Afiliada.
- **Renta:** Promedio de la renta de los trabajadores de la empresa afiliada.
- **Segmento Comercial:** Segmento comercial al que pertenece la empresa (Gendarmería, Minería, Educación, Salud Privada, Servicios Públicos, G1 , G2, PYM1, PYM2 y otros)
- **Seg:** Clasificación por Segmento Micro, PYME o GGEE (Grandes Empresas).
- **Cant_Trab:** Cantidad de trabajadores de la empresa.
- **Estado_Trab:** Si la empresa tiene trabajadores o no.
- **Est_Cred:** Clasificación Crediticia (G1,G2,G3,G4,PYM1,PYM2,PYM3….)
- **Tipo de Cuenta:** Privada o Pública
- **Dealer:** Si la empresa tiene asignado Dealer.
- **Agente:** Si la Empresa cuenta con Agente o no.
- **Sub_Mant:** Si la empresa cuenta con Subgerente de Mantención.
- **ECE:** Si la empresa posee ejecutivo (ECE).

2. Variables de Beneficios

Las variables de Beneficios aportan al modelo datos del comportamiento de uso de Beneficios y cercanía que tienen los trabajadores con la caja. Dentro de las variables construidas a partir de los orígenes de datos se incluyeron.

- **Ben_promXX:** Variable construida con una agrupación de los últimos 3, 6, 9 y 12 meses con el promedio de beneficios usados mensualmente.
- **Porc_Uso_Ben_XX:** Porcentaje de trabajadores que usaron beneficios durante los últimos X meses, agrupada para 3, 6, 9 y 12 meses.
- **Ben_X_Y:** corresponde al porcentaje de aumento o disminución de la cantidad de beneficios que usa la empresa. Se expresa como el porcentaje de beneficios que vario entre el trimestre anterior y el siguiente. Por ejemplo, Ben_9_12 corresponde

al porcentaje de variación entre los beneficios acumulados de los meses 10,11 y 12 (trimestre 12) antes de desafiliarse y acumulado de los meses 7, 8 y 9 (trimestre 9). Se calculó para Ben_3_6, Ben_6_9 y Ben_9_12. Mientras que Ben_6_12 corresponde a la variación de beneficios acumulados entre el primer y segundo semestre.

- **Rut_Ben_X_Y:** corresponde al porcentaje de variación de cantidad de trabajadores (Rut) que usaron beneficios. Se calcula de la misma manera que Ben_X_Y.

3. Variables de Reclamos

Las variables de Reclamos aportan al modelo datos acerca de los descontentos que ha tenido la empresa con la Caja. Fueron incluidas las siguientes variables:

- **Rec_promXX:** Variable construida con una agrupación de los últimos 3, 6, 9 y 12 meses con el promedio de reclamos hechos por los trabajadores mensualmente.
- **Porc_Rec_XX:** Porcentaje de trabajadores que hicieron reclamos durante los últimos X meses, agrupada para 3, 6, 9 y 12 meses.
- **Rec_Trab_X_Y:** corresponde al porcentaje de variación de reclamos de trabajadores entre un trimestre y otro.
- **Rec_Emp_X_Y:** corresponde al porcentaje de variación de reclamos hechos por el empleador entre un trimestre y otro.
- **Rec_CredXX:** Porcentaje de reclamos crediticios en los últimos XX meses.
- **Rec_AFXX:** Cantidad de Reclamos de Asignación Familiar de los últimos XX meses.
- **Rec_LMXX:** Porcentaje de reclamos de Licencias Médicas de los últimos XX meses.

4. Variables de Contexto Crediticio

Las variables de contexto crediticio aportan el comportamiento de pago y condiciones comerciales de los productos crediticios que poseen los trabajadores afiliados de la empresa. Dentro de las variables construidas a partir de los orígenes de datos contamos con:

- **Porc_Coloc_XX:** Variable construida con una agrupación de los últimos 3, 6, 9 y 12 meses con el porcentaje de trabajadores que colocaron crédito.
- **Porc_Cred_VigXX:** Variable construida con una agrupación de los últimos 3, 6, 9 y 12 meses con el porcentaje de créditos vigentes que posee la empresa.
- **Porc_Cred_MorososXX:** Variable construida con una agrupación de los últimos 3, 6, 9 y 12 meses con el porcentaje de créditos morosos que posee la empresa.
- **Porc_Cred_CastXX:** Variable construida con una agrupación de los últimos 3, 6, 9 y 12 meses con el porcentaje de créditos castigados que posee la empresa.
- **Avance_Cred_tramoX:** Porcentaje de trabajadores que se encuentran en el Tramo 1,2,3,4,5 de su crédito.

- **Cuo_Pag:** porcentaje de cuotas pagadas del total de crédito promedio de sus trabajadores.
- **Cred_Sucursal:** Porcentaje de trabajadores que sacaron crédito en sucursal el último año.
- **Cred_Dealer:** Porcentaje de trabajadores que sacaron crédito mediante Dealer el último año.
- **Saldo_Capital_X_Y:** Porcentaje de variación entre un trimestre y el siguiente acerca del saldo Insoluto que tiene la Empresa con la Caja.
- **Saldo_Capital_6_12:** Porcentaje de variación entre el primer semestre y el segundo en cuanto a colocación de créditos.

5. Variables de Visitas

Las variables de contexto de visitas aportan al modelo datos acerca de la proximidad que tienen los trabajadores de una empresa con las sucursales de la Caja. Se incluyeron las siguientes variables.

- **Vis_SucXX:** Variable construida con una agrupación de los últimos 3, 6, 9 y 12 meses con las visitas promedio mensual que visito la sucursal el trabajador.
- **Porc_Vis_SucXX:** Porcentaje de trabajadores del total de trabajadores de la empresa que visita sucursal agrupada para 3, 6, 9 y 12 meses.

6. Variables de Licencias Médicas

Las variables de contexto de Licencias Médicas entregan información acerca de los problemas que pudieron haber tenido los trabajadores con sus Licencias Médicas y les haya dejado algún malestar con la caja. Se generaron las siguientes variables:

- **Proc_LMXX:** porcentaje de trabajadores que presentan LM acumulado para los últimos 3, 6, 9 y 12 meses.
- **Prom_LM_TrabXX:** Promedio de licencias médicas que presentaron los trabajadores de la empresa acumulado para los últimos 3, 6, 9 y 12 meses.
- **PorcLM_AtrasosXX:** Porcentaje de LM del total de LM presentadas de una empresa que tuvieron atrasos acumulados para los últimos 3, 6, 9 y 12 meses.
- **LM_Días_CorrectosXX:** Si se le asignaron correctamente los días a la LM acumulado para los últimos 3, 6, 9 y 12 meses.

7. Variables de Contexto de Asignaciones Familiares

Las variables de contexto de Asignaciones familiares nos entregan información acerca del proceso de pago de asignaciones familiares a los empleadores, si hubo una diferencia en el monto de compensación podría generar un malestar para el empleador. Se incluyo la siguiente variable:

- **Dif_AcumX_marca:** Si tuvo o no diferencias en el monto de Compensación en los últimos 3,6,9,12 meses.

9.3 Aplicación de Modelos y Resultados

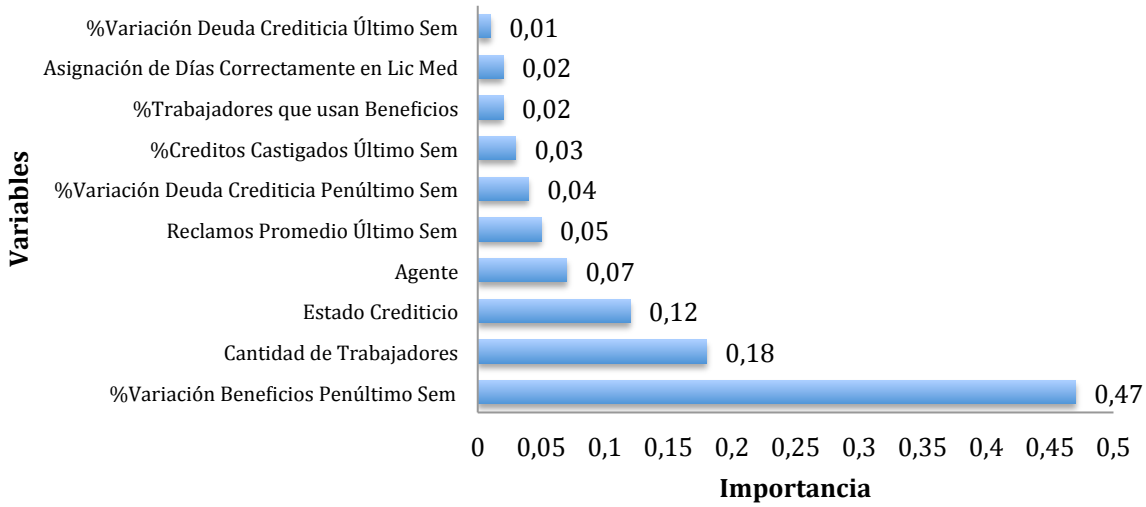
En los siguientes puntos se detallan los modelos desarrollados y sus principales resultados y conclusiones.

9.3.1 Análisis de Árbol de Decisión CHAID Exhaustivo

Se realizó un árbol de decisiones utilizando la muestra de 1300 empresas. Se usó el método de CHAID Exhaustivo con variable objetivo Fuga y sin variables forzosas. El árbol cuenta con 28 Nodos y se podó para no tener nodos con muy pocos casos, se limitó el número mínimo de registros en la rama principal con 30 casos y en la rama secundaria con 15 casos. Además, se le impuso como requisito que tuviera un costo de equivocarse, el cual fue asignado en la proporción 4:2 (Error tipo II: Error tipo I) lo que significa que tiene mayor costo equivocarse en aquellas empresas que el modelo pronostica como no fugadas y realmente se fugaron.

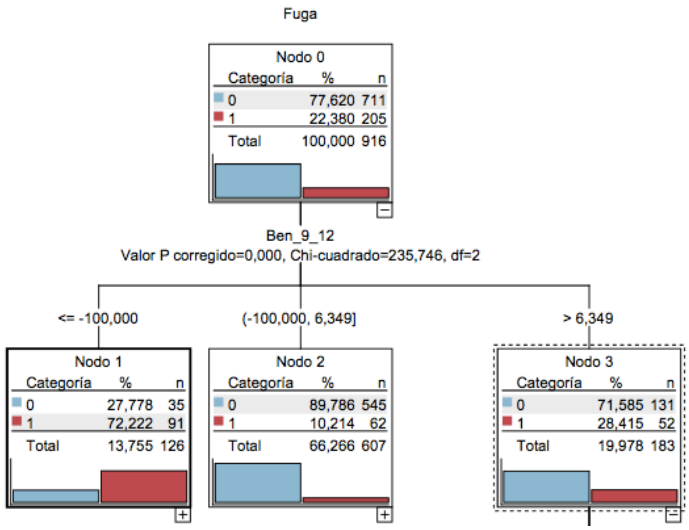
El modelo arrojó las siguientes variables ordenadas en la Ilustración 18 como las más importantes para predecir, la principal corresponde al porcentaje de variación de la cantidad beneficios usados durante el penúltimo semestre (Ben_9_12) del año móvil analizado de las empresas, siguiéndole cantidad de trabajadores, si cuenta con Agente, reclamo promedio por trabajador durante el último semestre (Rec_prom6), luego el porcentaje de variación de la deuda crediticia de la empresa afiliada con la Caja durante el penúltimo semestre (Saldo_Capital_9_12), porcentaje de créditos castigados del total durante los últimos 6 meses (%Cred_Cast6), porcentaje de trabajadores que usaron beneficios el último año (Porc_Uso_Ben12) y por último, la variación de la deuda crediticia durante el último semestre.

Ilustración 18 Importancia de las Variables del Modelo



La importancia de la variable tiene en cuenta la totalidad del árbol y se calcula sobre la partición de validación o comprobación, y la suma de la importancia de las variables es 1.0 teniendo en cuenta todos los campos. La importancia de la variable no se relaciona con la precisión del modelo, sino que es una medida de cuanto afecta una variable en la predicción del modelo, en este caso la variación del uso de beneficios, la cantidad de trabajadores y el estado crediticio son las que más afectan la predicción del modelo. En el Anexo 14.2 se muestra el árbol CHAID exhaustivo y las variables predictivas más significativas que seleccionó mediante el test Chi-cuadrado o Test de Pearson. A continuación se analizará el comportamiento de las empresas obtenido del árbol.

Ilustración 19 Primer Nivel Árbol de Clasificación

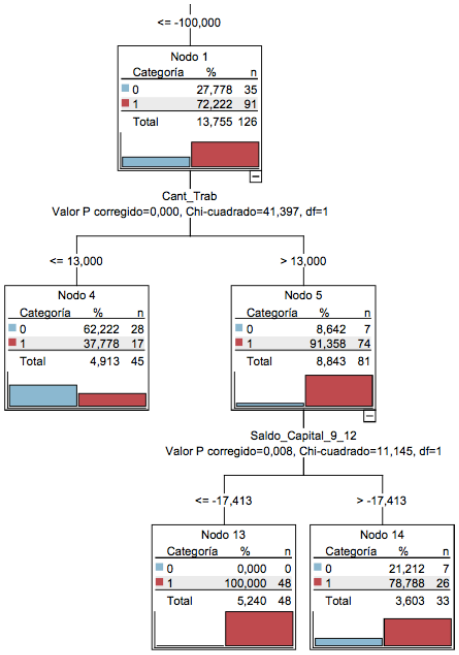


La variable que abre el primer nivel del árbol, Ilustración 19, es el porcentaje de variación de la cantidad de beneficios usados durante el penúltimo semestre (Ben_9_12). Se puede apreciar que el Nodo 1 con el porcentaje de variación de beneficios menor o igual a -100%, es decir, las empresas que dejaron de usar la totalidad de los beneficios, explica el 13,8% de las empresas del modelo, en donde un 72,2% son empresas desafiliadas y en un 27,8% las empresas afiliadas.

Luego el Nodo 2 explica un 66,3% de las empresas de la base y con un 10,2% de probabilidad estas empresas se fugaron y un 89,7% se mantuvieron afiliadas por lo que esta rama concentra a la mayoría de las empresas leales.

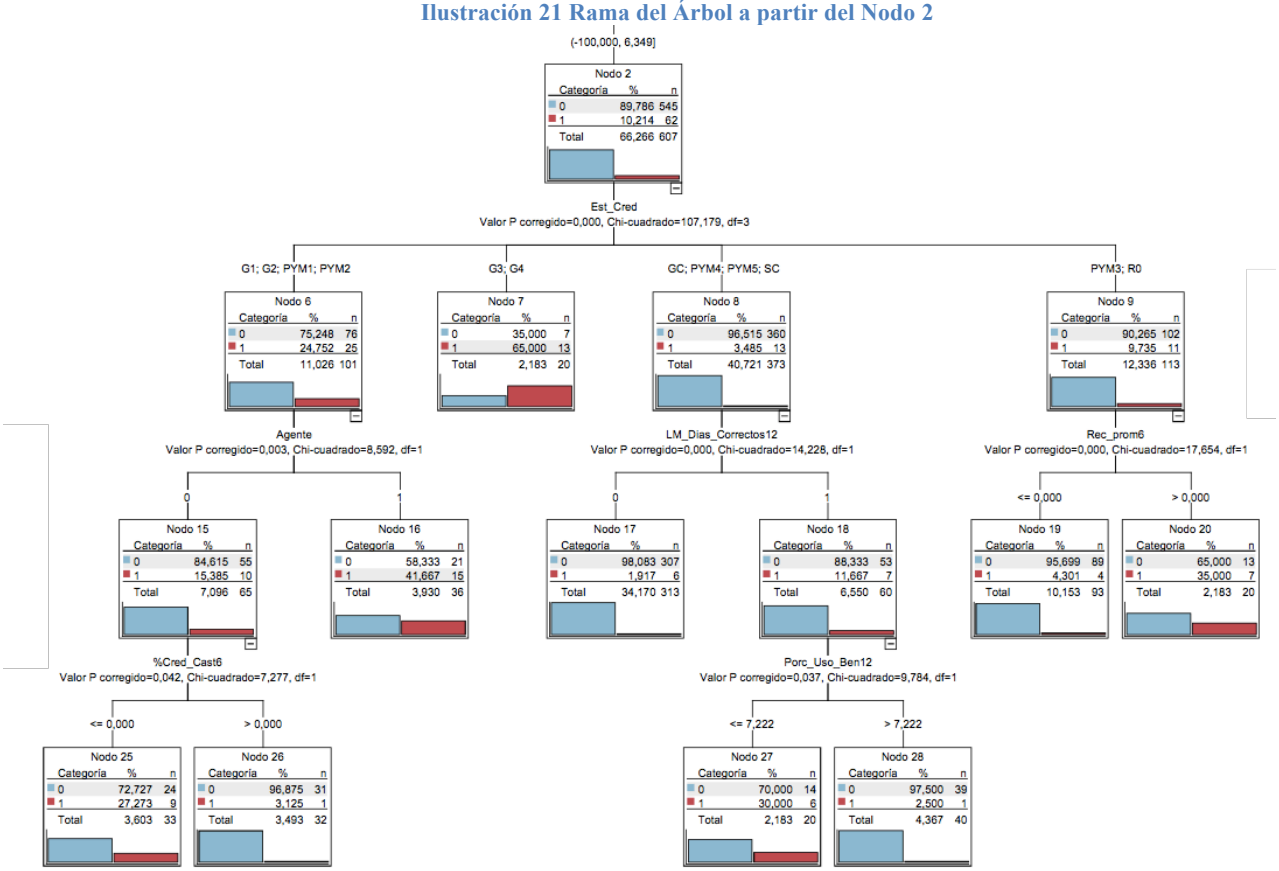
El nodo 3, concentra el 19,9% de las empresas que tuvieron un aumento en más de un 6% en la cantidad de beneficios usados durante el primer semestre del ultimo año y explica que con un 71,6% estás empresas se mantendrán afiliadas a la Caja. Dado que la primera variable que escoge como más significativa el modelo esta relacionada a la variación del uso de beneficios parece lógico que las empresas que han dejado de usarlos tiendan a desafiliarse y las que han aumentado mantengan su contrato de afiliación.

Ilustración 20 Rama del Árbol a partir del Nodo 1



El Nodo 1, Ilustración 20, escoge como variable más significativa (P<0,05) a la cantidad de trabajadores, separando a las empresas que tienen muy pocos trabajadores de las que tienen un cantidad más considerable (mayor a 13 trabajadores). Estas últimas, vuelven a ser clasificadas de acuerdo a el porcentaje de variación de la deuda crediticia con la Caja, en donde las empresas que su deuda a disminuido en más de un 17% van a tener probabilidad de un 100% (Nodo 13) de desafiliarse y las que están sobre este valor tendrán un 78,8% (Nodo 14) de probabilidad de fugarse. Estos nodos presentan las reglas con mayor probabilidad de fuga y están ligados fuertemente a las dos hipótesis planteada acerca del Bajo uso de beneficios y la baja colocación de crédito o pago de la deuda con la Caja.

Ilustración 21 Rama del Árbol a partir del Nodo 2



En la Ilustración 21, el Nodo 2 en donde las empresas han mantenido relativamente constante la variación de cantidad de beneficios usados, escoge como variable significativa el Estado Crediticio de las empresas, se aprecia que la fuga aumenta para los segmento G1,G2,PYM1 y PYM2 (Nodo 6) y G3 y G4 (Nodo 7).

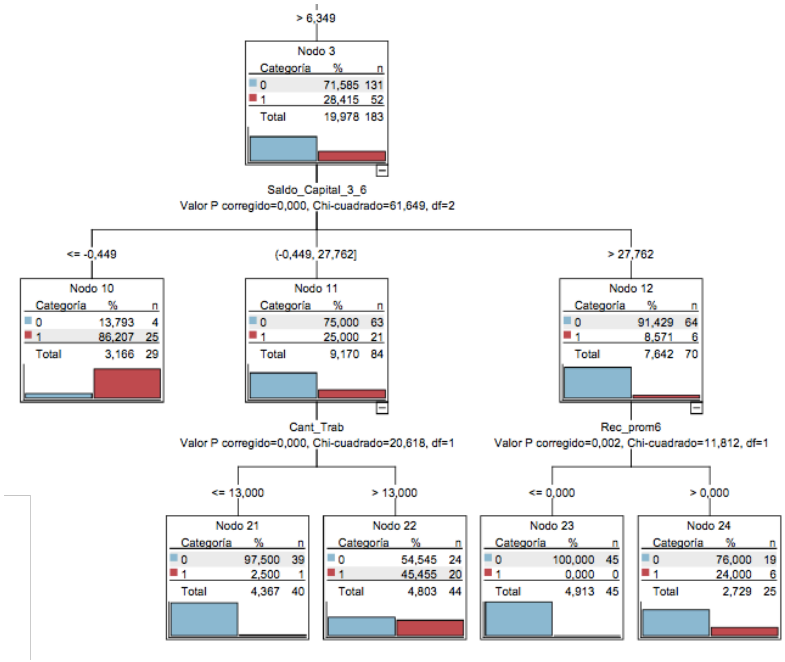
El Nodo 6 que contiene las empresas foco y selecciona como variable más significativa la variable Agente, donde las empresas que cuentan con agente y las que no tienen probabilidades similares de fugarse (Nodo 16), lo que puede deberse a que los agentes no estén marcando la diferencia en un servicio preferencial para las empresas foco. Por otra parte, las empresas que no cuentan con Agente (Nodo 15) es nuevamente dividido por la variable de porcentaje de créditos castigados (%Cred_Cast6), en donde para ambos casos, la probabilidad de mantenerse afiliadas es mayor a la de fugarse.

El Nodo 8, agrupa a todas las empresas que tienen menor valor o mayor riesgo crediticio para la Caja, y tienden a quedarse con una probabilidad del 96,5%. Este nodo escoge como variable significativa para continuar clasificando si tuvo o no problemas con la asignación de los días en el pago de sus licencias médicas durante el último año (LM_Dias_Correctos12), es decir, problemas operativos con la Caja y la probabilidad de fuga aumenta levemente cuando se presenta este comportamiento. Por ultimo, este nodo es nuevamente segmentado por la variable porcentaje de trabajadores que usó beneficios durante el último año (Porc_Uso_Ben12), en donde

las empresas en que menos del 7% de sus trabajadores usa beneficio tienen mayor probabilidad de desafiliarse, lo que se relaciona directamente con una de las hipótesis de benéficos.

Por último, el Nodo 9 concentra a las empresas PYM3 que tiene un riesgo medio y valor relativamente bajo para la Caja junta a las empresas R0 que corresponden a las bloqueadas crediticiamente. El nodo es clasificado por la variable de reclamos promedio por trabajador durante el último semestre (Rec_prom6) aumentando la probabilidad de fuga en aquellas empresas que tienen un promedio de reclamos por trabajador mayor a 0, es decir, han realizado reclamos alguna vez. Este ultimo punto también se relaciona con la hipótesis de reclamos, demostrando que a mayor cantidad de reclamos por trabajador la fuga tiende aumentar, bajo estas reglas o condiciones.

Ilustración 22 Rama del Árbol a partir del Nodo 3



El Nodo 3, en la Ilustración 22, corresponde a las empresas que aumentaron el uso de sus beneficios durante el penúltimo semestre en más de un 6% y representa a un 19,9% del total de las empresas de la base, teniendo un 91,4% de probabilidad de quedarse. Este nodo escoge como variable significativa para seguir clasificando a el porcentaje de variación de la deuda crediticia de la empresa con la Caja (Saldo_Capital_3_6) del ultimo semestre, en donde el Nodo 10 agrupa a todas las empresas que su deuda con la caja disminuyo en 1%. En este nodo predomina la fuga con una probabilidad del 86,2% de desafiliarse y el nodo concentra a un 3,1% de las empresas de la base.

El Nodo 11 representa a aquellas empresas que mantuvieron relativamente constante su deuda crediticia con la caja y tienen un 75% de probabilidad de quedarse. Este nodo es nuevamente dividido por cantidad de trabajadores, separando las empresas de menor tamaño, que

tienen mayor probabilidad de quedarse de las empresas con mayor a 13 trabajadores en donde la probabilidad de fugarse o no es relativamente de una 50% para cada opción.

Finalmente el Nodo 12, representa a las empresas que han aumentado su deuda crediticia con la Caja en más de un 27% y tendrán un 91,1% de probabilidad de quedarse. El nodo es nuevamente dividido por la cantidad de reclamos promedio por trabajador del ultimo semestre (Rec_prom6), en donde aumenta la probabilidad de fuga en el Nodo 24 que son las empresas que presentan reclamos.

En la sección Anexos 14.3 se presenta un tabla que resume las reglas de los nodos terminales para las empresas No fugadas y el comportamiento que presentan, mientras que en el Anexo 14.4 muestra las reglas para las empresas fugadas.

Se concluye del Árbol que las dos variables que más influyen en determinas la fuga de las empresas son las que reflejan el comportamiento crediticio y uso de beneficios del ultimo año y en casos particulares se incorporan variables que reflejan problemas operativos de las empresas con la Caja.

9.3.1.1 Medidas de Evaluación de desempeño del Árbol CHAID exhaustivo

Para analizar la precisión del modelo se analizó la matriz de confusión, que se encuentra en el Anexo 14.5, en donde la partición de entrenamiento arroja un 85,5% de precisión y un 77,6% y 85,9% para la partición de comprobación y validación respectivamente. En la Ilustración 23, se muestra la matriz de confusión para la partición de validación del modelo, del cual se desprenderán las medidas de evaluación de desempeño del modelo.

Ilustración 23 Matriz de Confusión Árbol Chaid Exhaustivo para base de validación

Clasificación

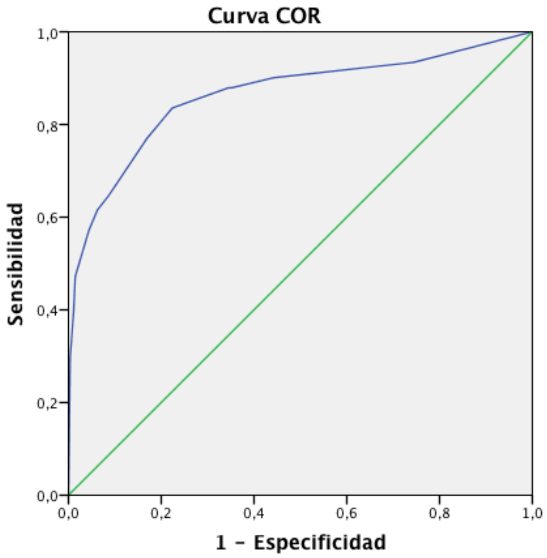
		Pronosticado		
	Observado	0	1	Porcentaje correcto
Entrenamiento	0	620	91	87,2%
	1	41	164	80,0%
	Porcentaje global	72,1%	27,8%	85,5%
Prueba	0	121	19	86,4%
	1	6	32	84,2%
	Porcentaje global	71,3%	28,6%	85,9%

Método de crecimiento: EXHAUSTIVE CHAID
Variable dependiente: Fuga

En la Ilustración 24 se presenta la Curva ROC del árbol CHAID exhaustivo, en donde la línea azul representa el resultado del modelo y la línea verde el resultado de un modelo al azar. En el eje horizontal se encuentran ordenados los casos o empresas de acuerdo a su probabilidad de fuga agrupados en percentiles, mientras que el eje vertical representa la sensibilidad de cada

percentil acumulado, es decir, que tan bien predice el modelo las fugas sobre el total de fugas reales. En el caso contrario, que tan bien predice el modelo las empresas que se quedan sobre el total de empresas que se mantuvieron afiliadas. Se puede apreciar que la línea azul se encuentra muy inclinada al principio dado que su poder predictivo es bueno para las empresas con alta probabilidad y a medida que se acerca al caso (1,1) su sensibilidad tiene a mantenerse. Además, junto al gráfico se presenta el valor AUC del área bajo la curva igual a 0,896 , lo que de acuerdo a la Tabla 3 clasifica al modelo como bueno.

Ilustración 24 Curva ROC y Valor AUC Árbol CHAID exhaustivo

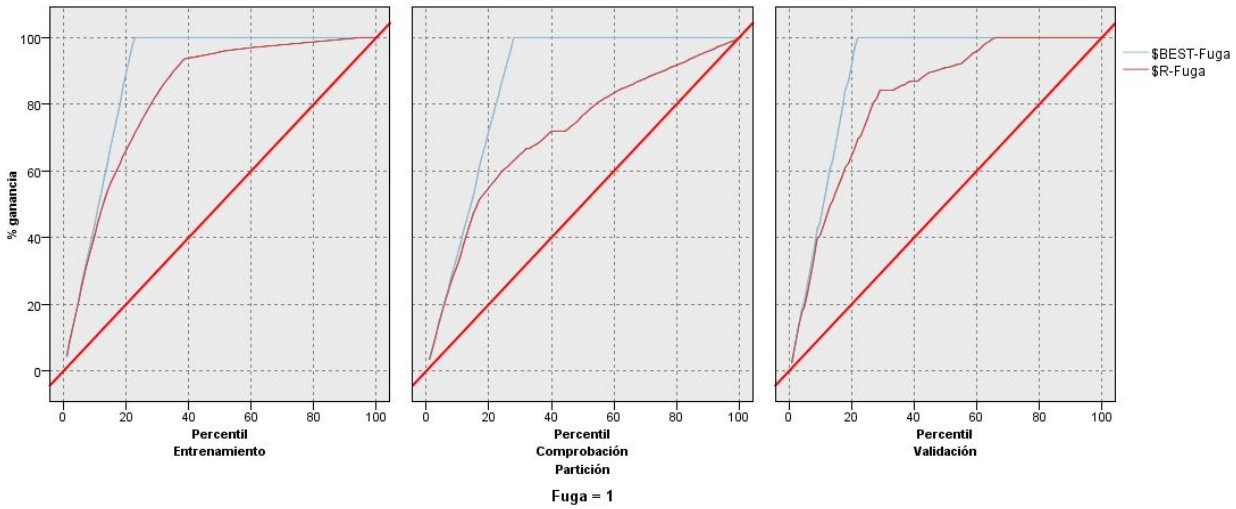


Los segmentos de diagonal se generan mediante empates.

AUC o Área bajo la curva
,863

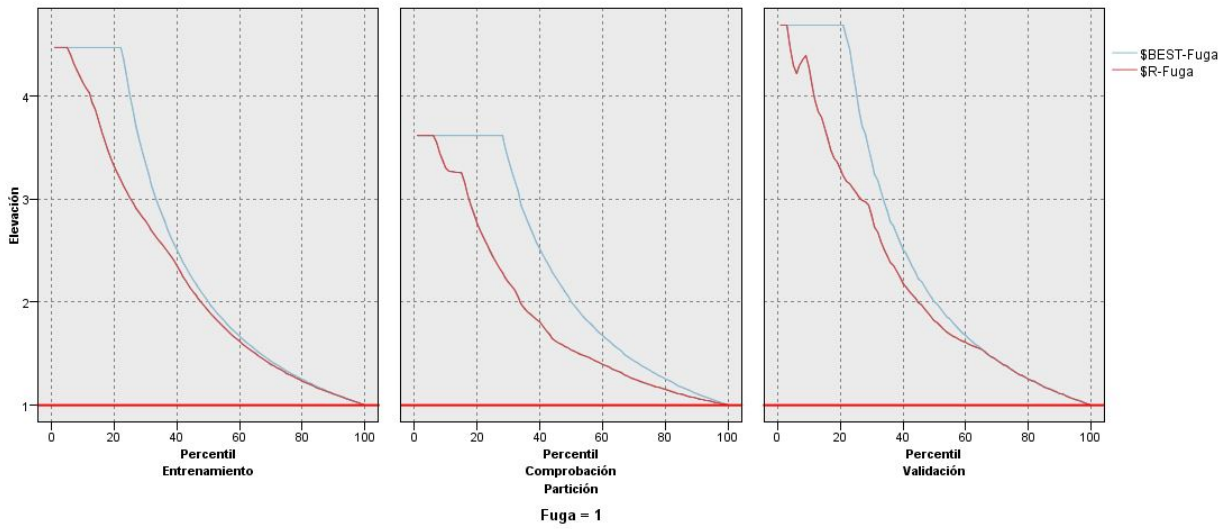
En Ilustración 25 se presentan el porcentaje de ganancias en el eje vertical, es decir el acumulado de aciertos, mientras que en el eje horizontal se presenta la probabilidad de fuga ordenado por percentiles (por la confianza en orden descendiente). La línea diagonal representa la línea base, es decir, lo que esperaríamos si fuese el pronóstico de la variable fuga en un modelo que predijese al azar, mientras que la línea azul representa el mejor de los modelos posibles y por último la línea de al medio muestra los resultados del árbol Chaid Exhaustivo antes explicado. Se aprecia que el modelo predice mejor que el azar, pero en la partición de comprobación tiende a disminuir su porcentaje de aciertos. Por otra parte, el de validación se acerca a la línea azul que es el modelo ideal, y mantiene su inclinación pronunciada al principio, es decir, predice bien para las empresas que se van a mantener afiliadas.

Ilustración 25 Gráfico de Ganancias para la categoría Fuga



En la Ilustración 26, se muestra el gráfico de Lift o de elevación que explica cuanto más probable es tomar una acción sobre un segmento de empresas específico que sobre un segmento al azar de clientes o empresas de la Caja. Por ejemplo, si tomamos al 20% de las empresas con mayor probabilidad de fuga de la Caja tendrá 4 veces más efecto la acción que al no usar el modelo (línea roja).

Ilustración 26 Gráfico de Lift o Elevación del Modelo



A modo de resumen y a partir de los resultados de la Ilustración 23, se generó la Tabla 5 en donde se muestran los valores para cada indicador de precisión del modelo CHAID exhaustivo.

Tabla 5 Medidas de Desempeño para Árbol CHAID exhaustivo

	CHAID Exhaustivo
Accuracy	85,9%
Sensibilidad	86,4%
Especificidad	84,2%
Precisión	95,3%
FPR	0,158
F-measure	0,906
Lift	1,211

9.3.2 Análisis de Árbol de Decisión CHAID Exhaustivo con Boosting

Para mejorar el rendimiento del Árbol CHAID exhaustivo se le aplicó Boosting, el cual es un método utilizado para mejorar el rendimiento de cualquier algoritmo de aprendizaje, en teoría se puede utilizar para reducir significativamente el error de un algoritmo débil, en este caso el árbol de decisión explicado en la sección anterior. Este método combina clasificadores débiles para producir un clasificador fuerte, pero sacrificando rapidez en la ejecución del modelo y facilidad de interpretar los resultados. Se seleccionó una cantidad de 10 árboles para mejorar el modelo anterior, por lo que se ejecuta repetidamente hasta 10 el modelo sobre distintas distribuciones de los datos de la base de entrenamiento y luego combina los clasificadores seleccionados para obtener un único modelo, árbol en este caso, más fuerte o con mayor precisión [7].

En la Tabla 6 Matriz de Confusión para árbol CHAID exhaustivo con boosting **¡Error! No se encuentra el origen de la referencia.** se muestra los resultados obtenidos en la matriz de confusión para este modelo, donde la base de entrenamiento tiene un 98,6% de precisión y la partición de validación un 90,4% . En la sección Anexos 14.8, se muestra con detalle la matriz de confusión del árbol con boosting.

Tabla 6 Matriz de Confusión para árbol CHAID exhaustivo con boosting

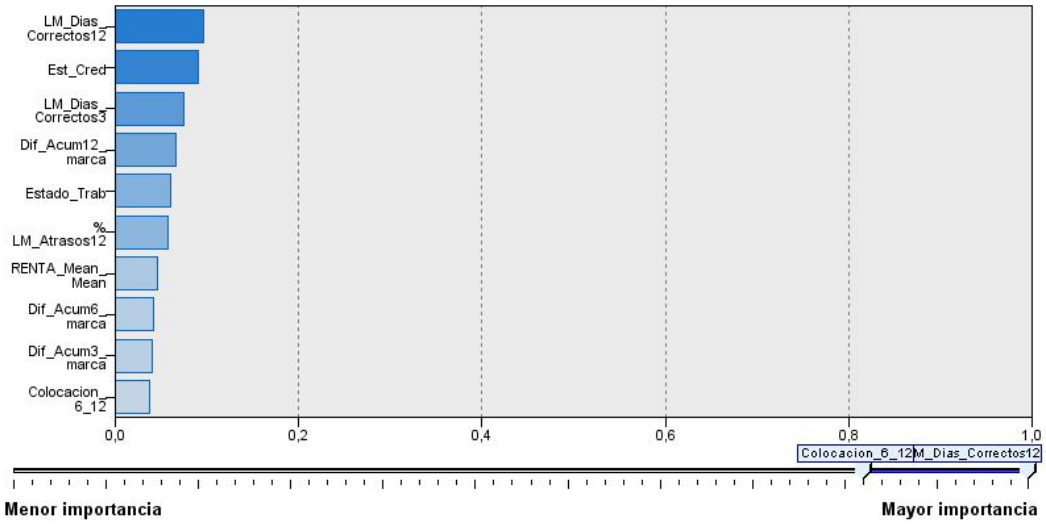
Clasificación

		Pronosticado		
Observado		0	1	Porcentaje correcto
Entrenamiento	0	705	6	99,1%
	1	6	199	97,0%
	Porcentaje global	77,6%	22,3%	98,6%
Prueba	0	133	7	86,4%
	1	10	28	84,2%
	Porcentaje global	80,3%	19,6%	90,45%

Método de crecimiento: EXHAUSTIVE CHAID
Variable dependiente: Fuga

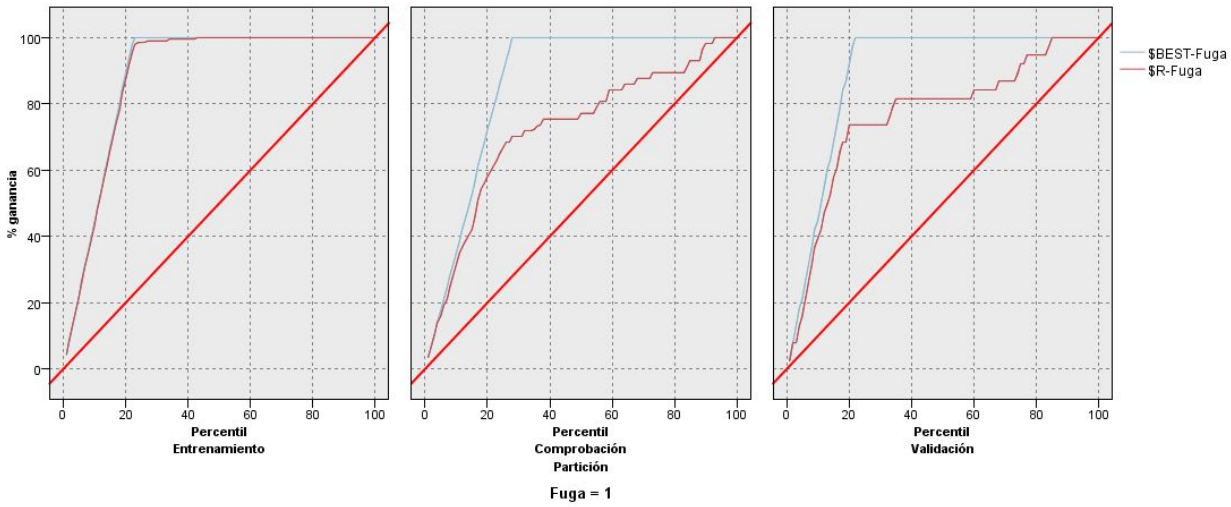
En la Ilustración 27 se muestran las variables más importantes a la hora de predecir para el modelo. La importancia de cada una de estas variables toma un valor entre 0 y 1 y la suma de estos valores es 1, asignándole un peso a cada una de ellas. Del gráfico se puede apreciar que las variables que más toman importancia del árbol con Boosting, son los problemas con la asignación de días correctamente en las licencias médicas (LM_días_correctos), los problemas con diferencias en el pago de las compensaciones a los empleadores (Dif_acum) y el porcentaje de licencias médicas que tuvieron atraso en la fecha de pago.

Ilustración 27 Importancia de los predictores para el árbol CHAID exhaustivo con boosting



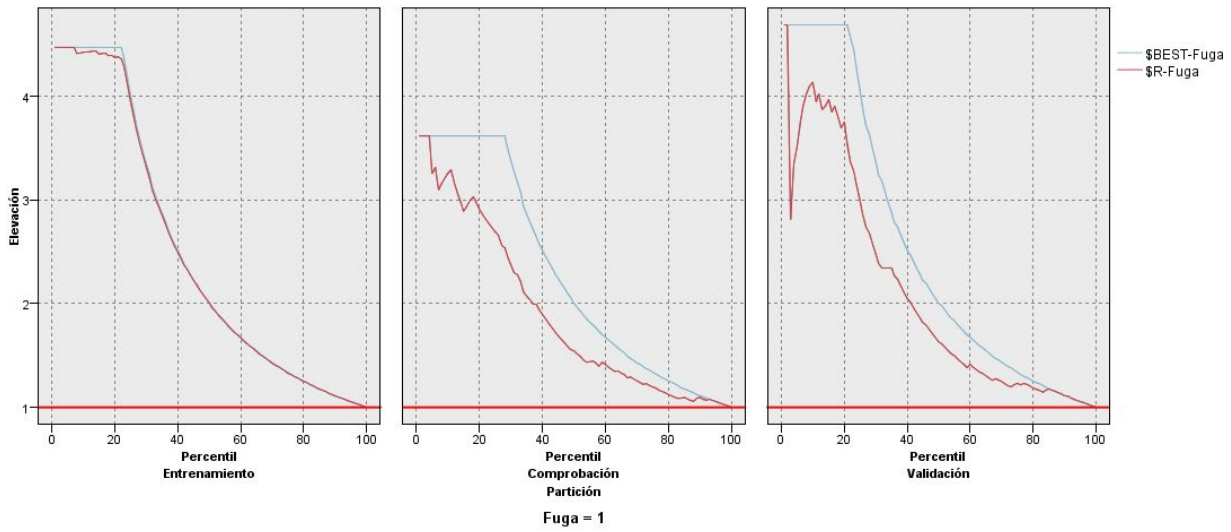
En la Ilustración 28 se muestra el gráfico de ganancia para cada partición de la base, en donde se aprecia que la base de entrenamiento presenta un precisión casi perfecta siguiendo el comportamiento de la línea azul, a diferencia de la matriz de comprobación y validación en donde disminuye la precisión y se puede apreciar la sensibilidad que presenta el modelo con Boosting a los datos.

Ilustración 28 Gráfico de ganancias para la variable fuga para el árbol CHAID Exhaustivo con boosting



En la Ilustración 29 se muestra el gráfico de elevación o Lift del modelo, si se analiza la partición de validación muestra que seleccionando al 10% más propenso a fuga se tendrá 4 veces una mejor llegada de las acciones de retención que si no sé tuviera el modelo.

Ilustración 29 Gráficos de lift o elevación para las particiones de entrenamiento, comprobación y validación del árbol CHAID exhaustivo con boosting.



En la Tabla 7 se muestra el resultado de los principales indicadores obtenidos por el modelo, en donde se rescata su alta precisión y sensibilidad, pero un bajo poder de especificidad, es decir, de detectar las empresas fugadas que realmente se fugaron.

Tabla 7 Medidas de Desempeño para Árbol CHAID Exhaustivo con Boosting

	CHAID Boosting
Accuracy	90,4%
Sensibilidad	95,0%
Especificidad	73,7%
Precisión	93,0%
FPR	0,263
F-measure	0,940
Lift	1,183

9.3.3 Análisis de Regresión Logística

Con la finalidad de obtener una estimación no sesgada o ajustada de la relación entre la variable dependiente y las múltiples variables independientes, se aplicó la regresión logística a la base de datos generada originalmente, pero antes de ingresar todas las variables al modelo se realizó un análisis bivariante entre las variables independientes y la variable fuga para seleccionar que variables realmente influyen sobre variable objetivo.

En la sección Anexos 14.6 se muestra la prueba de homogeneidad de varianza entre las variables y en Anexos 14.7 el Test ANOVA, ambas tablas fueron utilizadas para determinar que variables eran significativas en el comportamiento de la variable fuga, seleccionando aquellas que su significancia fuera menor a 0,05. Las variables seleccionadas de carácter categóricas o cuantitativas, fueron procesadas con tal de facilitar el análisis de la regresión logística, por ejemplo, la variable segmentación comercial fue transformada a diferentes variables dicotómicas para cada categoría.

Una vez obtenida la base final se modelo la regresión logística utilizando la técnica hacia delante para introducir las variables, en donde en la iteración 18 se obtuvo el modelo final. En la **¡Error! No se encuentra el origen de la referencia.**, se muestra cuáles son las variables más significativas y su peso relativo (Exp(B)) sobre la probabilidad de fuga. Cabe destacar, que la variable Subgerente de Mantención (Sub_Mant) toma una particular relevancia dentro del modelo, dado su alto valor en el Exp(B) , por lo que tiene una alta importancia relativa en el riesgo de fuga. Esto puede significar que se estén realizando buenas o malas acción en las mantención de empresas afectando significativamente en su motivo de desafiliación. Luego, le siguen variables como el porcentaje de trabajadores que reclamaron el último semestre (Rec_trab6) y el porcentaje de variación de la deuda crediticia de la empresa con la caja del último semestre (Saldo_Capital_3_6).

Tabla 8 Parámetros obtenidos de la regresión logística para las variables más significativas

Variables en la ecuación

		B	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
								Inferior	Superior
Paso 18[a]	Colocacion_6_12_transformed	,014	,006	5,961	1	,015	1,014	1,003	1,025
	Saldo_Capital_3_6_transformed	,073	,018	16,355	1	,000	1,076	1,039	1,115
	Saldo_Capital_6_9_transformed	,048	,012	15,504	1	,000	1,049	1,024	1,074
	Saldo_Capital_9_12_transformed	,028	,009	10,365	1	,001	1,029	1,011	1,046
	Saldo_Capital_6_12_transformed	-,059	,016	13,493	1	,000	,943	,914	,973
	Ben_6_12_transformed	,030	,006	28,709	1	,000	1,031	1,019	1,042
	RENTA_Mean_Mean_transformed	,039	,006	49,161	1	,000	1,040	1,029	1,052
	Ben_prom12_transformed	-,056	,008	45,013	1	,000	,946	,930	,961
	Vis_Prom6_transformed	-,016	,006	6,673	1	,010	,984	,973	,996
	Est_Crediticio_2(1)	-2,240	,636	12,396	1	,000	,106	,031	,370
	Est_Crediticio_3(1)	-3,735	,527	50,276	1	,000	,024	,009	,067
	Est_Crediticio_4(1)	-2,626	1,145	5,262	1	,022	,072	,008	,682
	Est_Crediticio_5(1)	-2,691	1,309	4,224	1	,040	,068	,005	,883
	Dif_Acum12_marca(1)	-1,785	,257	48,208	1	,000	,168	,101	,278
	Rec_Trab6	,301	,149	4,106	1	,043	1,352	1,010	1,809
	Agente(1)	-3,328	,616	29,184	1	,000	,036	,011	,120
	Sub_Ment(1)	1,870	,556	11,320	1	,001	6,491	2,183	19,299
	Cred_Cast6_transformed	-,023	,006	13,535	1	,000	,977	,965	,989
	Constante	10,454	2,138	23,914	1	,000	34685,504		

a. Variables especificadas en el paso 18: Rec_Trab6.

En la Tabla 9 se muestran los resultados de la matriz de confusión del modelo el cual presenta buena precisión global con un 88,2% para la matriz de validación, pero, un especificidad o detección de las fugas del real de fugas observadas de un 67,9% por lo que no es un fuerte predictor para detectar las fugas, más bien, es fuerte en detectar las empresas que se mantienen afiliadas.

Tabla 9 Matriz de Confusión de la Regresión Logística

Clasificación

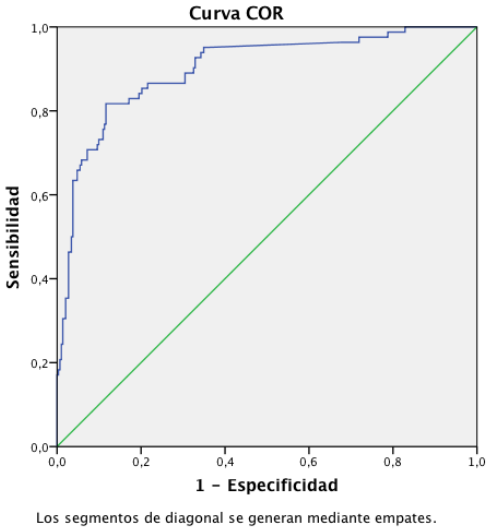
		Pronosticado		
	Observado	0	1	Porcentaje correcto
Entrenamiento	0	680	28	96,1%
	1	82	136	62,4%
	Porcentaje global	82,3%	17,7%	88,1%
Prueba	0	274	18	93,8%
	1	26	55	67,9%
	Porcentaje global	80,4%	196%	88,2%

Método de crecimiento: EXHAUSTIVE CHAID
Variable dependiente: Fuga

En la Ilustración 30 se muestra el gráfico de la Curva ROC para la partición de validación de la regresión logística, en donde se puede aprecia que la detección de verdaderos positivos es mucho mayor que la detección de falsos positivos, es decir, existen más empresas declaradas como no fugadas cuando realmente se fugaron. Por otra parte, cuenta con un AUC de 0,9 lo que

significa que ante dos empresas una fugada y otra que no, el modelo tiene una alta probabilidad de clasificarlos correctamente.

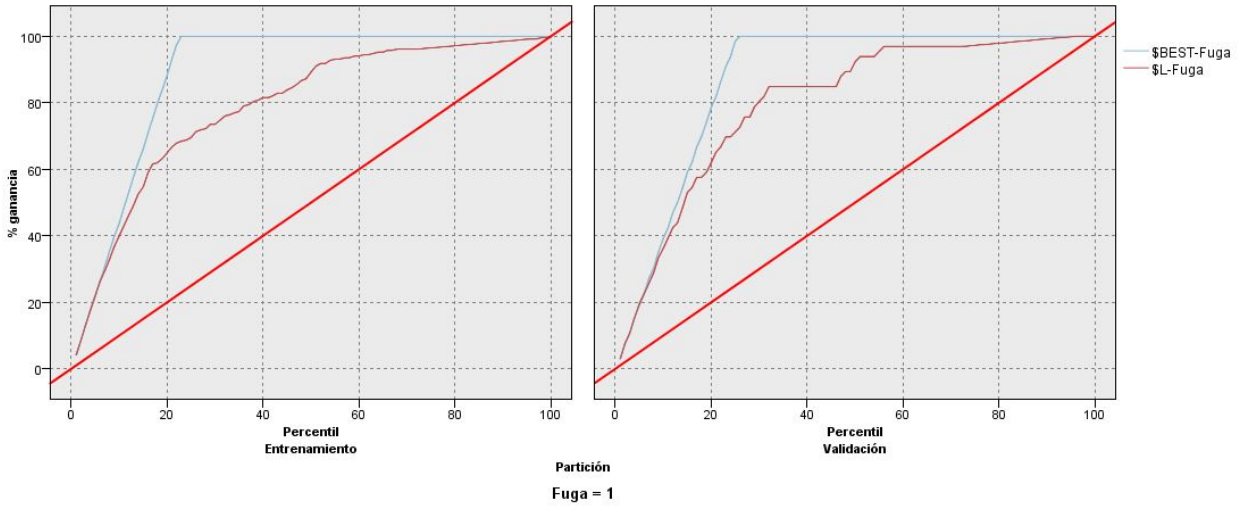
Ilustración 30 Gráfico de la Curva ROC de la Regresión Logística y Área bajo la Curva o AUC



AUC o Área Bajo la Curva
0,900

En la Ilustración 31 se muestra el gráfico de ganancia para la regresión logística, en donde se aprecia que el ajuste del modelo es bueno y logra detectar el 60% de las fugas con el 20% de empresas más probables a la fuga (Gráfico a la derecha, partición de validación).

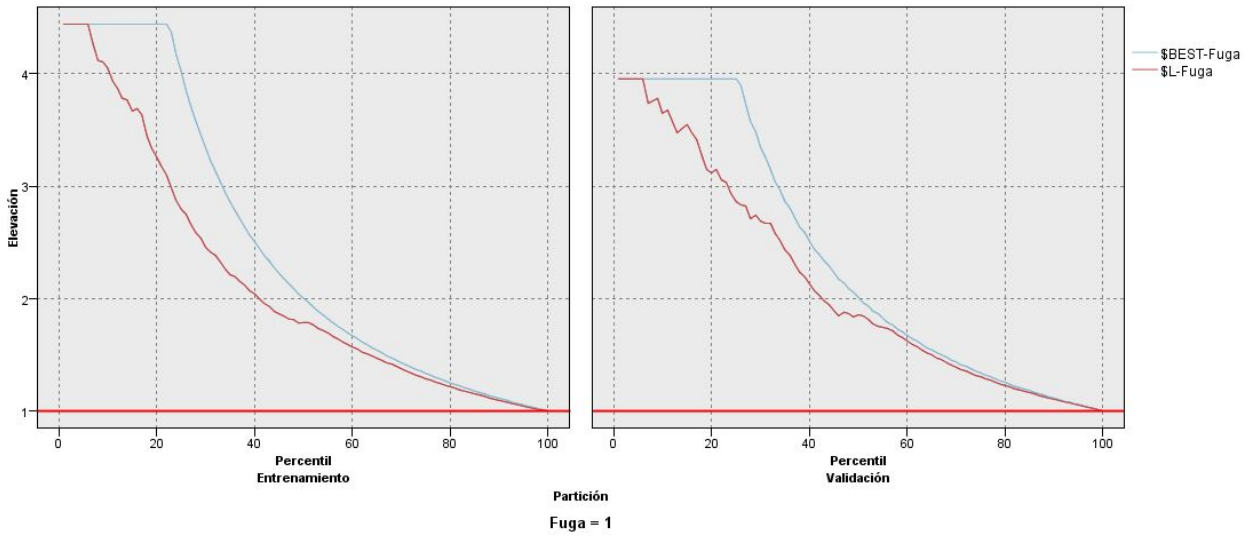
Ilustración 31 Gráfico de ganancia para la partición de entrenamiento y validación de la Regresión Logística



En la Ilustración 32 se presenta el gráfico de elevación o Lift para la regresión logística, que muestra que tan efectivo es un modelo en cuanto a la respuesta que se tendrá de los clientes ante una acción de retención cuando se focalizan sobren los clientes más propensos. Al igual que en los modelos anteriores, entre utilizar la regresión logística (línea roja) y no usarlo (línea roja constante en 1, en donde las empresas se seleccionan al azar o sin un modelo predictivo) se tendrá mayor efecto o respuesta a las acciones de retención con la regresión logística, ya que si se

toma al 20% de las empresas con mayor probabilidad de fuga (eje X) y se aplican acciones de retención se tendrá un efecto 3 veces mejor (Eje Y) al de no utilizar el modelo.

Ilustración 32 Gráfico de Lift o Elevación para las particiones de Entrenamiento y Validación de la Regresión Logística



En la Tabla 10 se muestra un resumen de los principales indicadores para medir el desempeño de la regresión logística. Cabe destacar que el modelo presenta una alta precisión, es decir, su capacidad de detectar las empresas que realmente se quedan, pero no así para clasificar como fugadas a las empresas que realmente se van a fugar.

Tabla 10 Medidas de desempeño para modelo Regresión Logística

	Regresión Logística
Accuracy	88,2%
Sensibilidad	91,3%
Especificidad	75,3%
Precisión	93,8%
FPR	0,247
F-measure	0,926
Lift	1,167

9.3.4 Análisis y Elección del Modelo Final

Para escoger el modelo final, se compararon las medidas de desempeño de los modelos obtenidas por cada uno de ellos. En la Tabla 11 se presentan los resultados de los tres modelos aplicados, más los indicadores para el caso del Árbol CHAID exhaustivo pero sin costos, con tal de poder comparar el caso en que no se considera un costo asociado al equivocarse al predecir.

De la Tabla 11 se desprende que todos los modelos tienen buen accuracy y sensibilidad, es decir, la capacidad de predecir correctamente aquellos casos en que realmente las empresas no se fueron. Este valor es superior al 90% en el árbol con Boosting, la regresión logística y el árbol sin costos, pero no así para el árbol con costo, ya que al incorporar un costo al predecir incorrectamente se pierde un poco en acurracy , pero se gana en especificidad, que es que tan bien detecta a las empresas fugadas del total, siendo este punto realmente el desafío de los modelos.

Tabla 11 Medida de Desempeño para los Modelos Aplicados

	CHAID exhaustivo con costos	CHAID Boosting	Regresión Log	CHAID exhaustivo sin costos
Accuracy	86,0%	90,4%	88,2%	85,7%
Sensibilidad	86,4%	95,0%	91,3%	93,8%
Especificidad	84,2%	73,7%	75,3%	61,5%
Precisión	95,3%	93,0%	93,8%	88,0%
FPR	0,158	0,263	0,247	0,385
F-measure	0,906	0,940	0,926	0,908
Lift	1,211	1,183	1,167	1,173

Al momento de comparar varios modelos conviene fijarse en el valor del indicador F-measure, que relaciona los conceptos de sensibilidad y precisión, mientras más cercano a 1 mejor es el modelo. En este caso el mejor modelo sería el Árbol CHAID exhaustivo con Boosting, ya que utilizarlo mejora el desempeño del árbol original. Pese a esto, se va a escoger el Árbol CHAID exhaustivo con costos dada su fácil interpretación, ya que va a ser de utilidad para analizar los segmentos más propensos a la fuga y su comportamiento. Por otra parte, las medidas de desempeño no varían demasiado, por lo tanto se sacrificará accuracy por un mayor poder de interpretabilidad de los datos, con tal de analizar el comportamiento de las empresas y poder desprender acciones de retención a partir de su comportamiento.

9.4 Prueba Hipótesis Planteadas

Para validar las hipótesis planteadas en un comienzo, se utilizaran dos herramientas: Por una parte, se validará la significancia de las variables influyentes en cada hipótesis sobre la variable fuga mediante un test ANOVA y en segundo lugar, se recurrirá a los resultados obtenidos por el árbol CHAID exhaustivo en la sección 9.3.1. A partir del comportamiento de las empresas y las variables más significativas que escoge el árbol se determinó si las hipótesis pueden validarse o rechazarse. A continuación se analizan cada una de las hipótesis planteadas.

H1.*A mayor cantidad de reclamos mayor probabilidad de fuga.*

En la Tabla 12, se muestra el test de Homogeneidad de varianza para las variables relacionadas con los reclamos que realizan los trabajadores y el empleador de las empresas, además de variables que reflejan la variación de reclamos entre los trimestre. Se aprecia que la

mayoría de las variables presentan un significancia menor a 0,05 y además en la sección Anexos 14.7 se muestra el Test ANOVA para da una de ellas. Al analizar amabas tablas en conjunto se puede rechazar la hipótesis nula del Test de Anova, confirmando que el comportamiento de estas variables entre las empresas fugadas y no fugadas en temas de reclamos es distinto e influyente en la desafiliación.

Las variables de reclamo que no son significativas son Reclamo promedio por trabajador (Rec_promX) que captura los reclamos como un estado de los trabajadores, la cual no es influyente en la fuga y la cantidad de Reclamos crediticios y de Licencias Médicas durante los últimos 9 y 12 meses, dado que son casos muy particulares y corresponden al acumulado del ultimo año lo cual no es significativo.

Por otra parte, del Árbol CHAID exhaustivo , Anexo 14.2, se puede rescatar que una de las variables que el modelo considera significativa para clasificar un nodo es la variable Reclamo promedio del último semestre (Rec_prom6), se clasifica al nodo entre empresas que tuvieron reclamo promedio 0 y las que tuvieron reclamo mayor a 0, por lo que considera significativo separar el comportamiento entre las empresas que han manifestado algún malestar y las que no, por lo tanto se concluye que a mayor cantidad de reclamos mayor es la fuga.

Tabla 12 Prueba de Homogeneidad de varianzas para las variables de Reclamos

Prueba de homogeneidad de varianzas				
	Estadístico de Revene	df1	df2	Sig.
Rec_Emp_3_6	65,017	1	1298	,000
Rec_Emp_6_9	6,137	1	1298	,013
Rec_Emp_9_12	5,986	1	1298	,015
Rec_Emp_6_12	66,787	1	1298	,000
Rec_trab_3_6	163,656	1	1298	,000
Rec_trab_6_9	64,231	1	1298	,000
Rec_trab_9_12	3,351	1	1298	,067
Rec_trab_6_12	151,557	1	1298	,000
Rec_Empleador_Acum3	76,404	1	1298	,000
Rec_Empleador_Acum6	68,168	1	1298	,000
Rec_Empleador_Acum9	11,524	1	1298	,001
Rec_Empleador_Acum12	7,045	1	1298	,008
Rec_prom3	,854	1	1298	,356
Rec_prom6	,969	1	1298	,325
Rec_prom9	1,061	1	1298	,303
Rec_prom12	1,080	1	1298	,299
Porc_Rec3	15,021	1	1298	,000
Porc_Rec6	10,903	1	1298	,001
Porc_Rec9	1,222	1	1298	,269
Porc_Rec12	,957	1	1298	,328
Rec_Trab3	146,680	1	1298	,000
Rec_Trab6	60,712	1	1298	,000
Rec_Trab9	9,215	1	1298	,002
Rec_Trab12	4,855	1	1298	,028
Rec_Cred3	27,244	1	1298	,000
Rec_Cred6	21,705	1	1298	,000
Rec_Cred9	1,055	1	1298	,305
Rec_Cred12	1,801	1	1298	,180
Rec_AF3	194,126	1	1298	,000
Rec_AF6	62,249	1	1298	,000
Rec_AF9	36,248	1	1298	,000
Rec_AF12	10,537	1	1298	,001
Rec_LM3	19,151	1	1298	,000
Rec_LM6	1,895	1	1298	,169
Rec_LM9	,044	1	1298	,834
Rec_LM12	,138	1	1298	,710

H2.*A menor uso de beneficios mayor es la fuga.*

Al igual que en el caso de los reclamos, se analizaron las variables relacionadas con los beneficios que usan los trabajadores, para ello se analizó la Tabla 13 y el test de Anova de las variables que se encuentra en Anexos 14.7. Se puede apreciar que los beneficios promedios de cada periodo (Ben_promX) y el Porcentaje de trabajadores que usa beneficios (Porc_Uso_BenX) no son significativos, pero si lo es el porcentaje de variación de las cantidad de beneficios que usan los trabajadores entre cada trimestre y semestre.

Tabla 13 Prueba de homogeneidad de varianzas para las variables de Beneficios

Prueba de homogeneidad de varianzas				
	Estadístico de Levene	df1	df2	Sig.
Ben_3_6	96,838	1	1298	,000
Ben_6_9	143,302	1	1298	,000
Ben_9_12	,217	1	1298	,041
Ben_6_12	66,091	1	1298	,000
Porc_Uso_Ben3	1,114	1	1298	,291
Porc_Uso_Ben6	1,146	1	1298	,285
Porc_Uso_Ben9	1,157	1	1298	,282
Porc_Uso_Ben12	1,178	1	1298	,278
Ben_prom3	1,194	1	1298	,275
Ben_prom6	1,180	1	1298	,277
Ben_prom9	1,192	1	1298	,275
Ben_prom12	1,212	1	1298	,271

Además, incluyendo al análisis anterior los resultados del Árbol CHAID exhaustivo, Anexo 14.2, en donde la primera variable que clasifica al árbol es la variación de beneficios del penúltimo semestre (Ben_9_12), se puede decir que el cambio en la cantidad de beneficios que usa una empresa si influye en la fuga. Se concluye que un disminución en la cantidad de beneficios que suelen usar las empresas aumenta la fuga, validándose la hipótesis.

H3.*A mayor tiempo de demora en el Pago Licencias Médicas mayor es la fuga.*

Se analizaron las variables relacionadas con licencias médicas, en donde la prueba de homogeneidad de varianza de la Tabla 14 y el test de ANOVA en el Anexo 14.7, arrojan que las variables son significativas con la variable fuga, a excepción de si tuvieron algún problema con los días de las licencias durante el último año (LM_Dias_Correctos12) que puede deberse a un error en la variable dado que todas las anteriores son significantes, o bien, a que el periodo anual de está variable no es significativo.

Tabla 14 Prueba de Homogeneidad de Varianzas para las variables de Licencias Medicas

Prueba de homogeneidad de varianzas				
	Estadístico de Levene	df1	df2	Sig.
LM_Atrasos3	169,016	1	1298	,000
LM_Atrasos6	136,113	1	1298	,000
LM_Atrasos9	108,587	1	1298	,000
LM_Atrasos12	107,635	1	1298	,000
LM_Dias_Correctos3	140,646	1	1298	,000
LM_Dias_Correctos6	46,100	1	1298	,000
LM_Dias_Correctos9	18,088	1	1298	,000
LM_Dias_Correctos12	,094	1	1298	,059

Dado que la variable LM_Día_Correctos12 si fue seleccionada por el árbol para clasificar un nodo de la rama que incluye al comportamiento general de la mayoría de las empresas para clasificar si tuvieron un problema operativo, en este caso licencias médicas, se considerará que la hipótesis es válida.

H4.*A mayor diferencias de compensación con la empresa aumenta la fuga.*

Nuevamente se analizó la significancia de las variables relacionadas con el proceso de pago de compensaciones a las empresas, se obtuvo que todas las variables son significativas con respecto a la fuga (En la Tabla 15 se encuentra la prueba de homogeneidad de varianzas y en Anexos 14.7 el test ANOVA), es decir, independiente del periodo si la empresa tuvo alguna diferencia en su monto compensado dentro del último año puede influir en la fuga, lo cual es coherente dado que si existe alguna diferencia en el monto compensado esto influye directamente en el presupuesto del empleador quien puede incentivar a que sus trabajadores se cambien de caja.

Tabla 15 Prueba de Homogeneidad de varianzas para las variables de Diferencia de Compensación

Prueba de homogeneidad de varianzas

	Estadístico de Levene	df1	df2	Sig.
Dif_Acum3_marca	142,994	1	1298	,000
Dif_Acum6_marca	123,144	1	1298	,000
Dif_Acum9_marca	112,151	1	1298	,000
Dif_Acum12_marca	119,443	1	1298	,000

Pese a que estás variables no fueron incluidas en el árbol, cabe mencionar que la variable tendía aparecer en el árbol cuando la poda era menor, es decir, cuando se analizaban casos más particulares de empresas, por lo que se asumirá que la hipótesis es valida.

H5.*A mayor Colocación de Créditos menor es la probabilidad de fuga.*

El análisis de las variables de créditos reflejó que las variables que muestran un cambio en el comportamiento de las empresas, es decir, porcentaje de variación de colocación (Colocación_6_12) y el porcentaje de variación de la deuda crediticia de la empresa con la caja (Saldo_Capital_X_X) son significativas con la variable fuga según la prueba de homogeneidad de varianzas en la Tabla 16 Prueba de Homogeneidad de varianzas para las variables de Créditos que muestran un estado como el porcentaje de colocación de créditos (Porc_ColocX) y colocación promedio (ColocX) no son significativas.

Tabla 16 Prueba de Homogeneidad de varianzas para las variables de Créditos

Prueba de homogeneidad de varianzas

	Estadístico de Levene	df1	df2	Sig.
Colocacion_6_12	149,338	1	1298	,000
Porc_Coloc3	9,668	1	1298	,002
Porc_Coloc6	6,745	1	1298	,010
Porc_Coloc9	4,072	1	1298	,044
Porc_Coloc12	1,686	1	1298	,194
Saldo_Capital_3_6	120,546	1	1298	,000
Saldo_Capital_6_9	86,934	1	1298	,000
Saldo_Capital_9_12	13,696	1	1298	,000
Saldo_Capital_6_12	27,099	1	1298	,000
Coloc3	2,388	1	1298	,123
Coloc6	,814	1	1298	,367
Coloc9	,961	1	1298	,327
Coloc12	1,046	1	1298	,307

El porcentaje de variación de la deuda de la empresa con la caja (Saldo_CapitalX_X) es escogida por el árbol para clasificar varios nodo, por lo que a mayor colocación de créditos o un

aumento de la deuda de la empresa con la caja la fuga es menor y en el caso contrario, cuando un empresa comienza a pagar su deuda y disminuye la solicitud de créditos tiende a aumentar la fuga, lo que valida la hipótesis.

H6.*Perdida del PDB aumenta la fuga*

El PDB es el monto que se le asigna a ciertas empresas en las cuales la Caja quiere crecer comercialmente, de acuerdo a la Tabla 17 se puede validar que es un variable significativa. Según la prueba de homogeneidad de las varianzas y el test ANOVA (Anexos 14.7) se puede asegurar que al 95% de confianza que el comportamiento entre ambos grupos fugadas y no fugadas es distinto.

Tabla 17 Prueba de homogeneidad de varianzas para las variables de PDB

Prueba de homogeneidad de varianzas				
	Estadístico de Levene	df1	df2	Sig.
PDB_2013	94,418	1	1298	,000
PDB_2014	34,083	1	1298	,000
PDB_Aumento	45,960	1	1298	,000

Por otra parte, estas variables no fueron incluidas en el árbol dado su pequeño impacto sobre el total de empresas, ya que aproximadamente 200 empresas del total presentan este beneficio. Se concluye que está variable es valida solamente para las empresas focos, pero no se puede extrapolar este resultado a todas las empresas.

10 Análisis de Acciones sobre Segmentos Más Propensos a la Fuga

Las acciones de retención de empresas que realiza la Caja en estudio varían de acuerdo a las distintas características que tiene la empresa, siendo la principal variable para escoger sobre cuales se realizan acciones el Stock de crédito que tiene la empresa. De acuerdo a esto y la cantidad de trabajadores que tenga la empresa, la caja la va considerar como empresa foco o no, teniendo un trato y acciones diferentes entre foco y no foco.

Las empresas foco por lo general cuentan con más de 250 trabajadores y poseen un Agente, Ejecutivo Cuenta Empresa (ECE) o Subgerente mantención, quien será el principal responsable de los acontecimiento que ocurran sobre estas empresas. Si la empresa presenta problemas operativos, es decir, atrasos en los pagos de sus licencias medicas o problemas con su pago de Compensación, el Agente, ECE o Subgerente será el principal responsable en levantar una alerta o solucionar el caso particular que tenga la empresa.

Si las empresas están disminuyendo sus créditos, la Caja con las empresas foco aumenta la acción de los Dealers, quien son agentes externos subcontratados que van a vender créditos a las empresas, de tal manera de aumentar la colocación de la empresa que se quiere retener. Por otra parte, si son empresas no foco, se realizan campañas de call center en donde se realizan llamados para ofrecer créditos con menores tasas a aquellos trabajadores con mayor probabilidad a aceptarlos. Además, se incluyen a estos trabajadores en campañas de mailing en donde se les envía una carta con ofertas de créditos a los que tienen posibilidad de aceptar el crédito.

Además, en cuanto a temas de beneficios y difusión, en las empresas foco los Subgerentes de Mantención, encargados de ir a visitarlas cada cierto tiempo llevan material de *merchandising* para incentivar el uso de los beneficios y productos que esta ofreciendo la caja, como por ejemplo descuentos especiales en los Centros Vacacionales, o bien, descuento en ciertas instituciones de salud. En el caso de las empresas no foco, se hacen campañas mucho más masivas a nivel de sucursal donde se dan a conocer mediante revistas, flayer o trípticos que muestran los beneficios que tiene la Caja para sus afiliados, además entregan merchandising o regalos a aquellos pensionados que tienen probabilidad de desafiliarse de forma independiente. Otro canal fuerte que están utilizando para difundir beneficios es mediante Facebook y correo electrónico, en donde dan a conocer información relevante de beneficios, entretención y convenios y mediante Facebook realizan concursos sorteando premios a sus afiliados como entradas al cine o conciertos, respectivamente.

El Subgerente de mantención realiza visitas periódicamente a las empresas foco y levanta una alerta de acuerdo al riesgo de fuga que pueda tener la empresa, en el caso de ser media o alta, el Agente o Subgerente pueden incentivar a estas empresas a gastar de su presupuesto para beneficios o eventos (PDB) con tal de disminuir su interés de desafiliarse. Por otra parte, al relacionarse con el área de recursos humanos de la empresa el ejecutivo o subgerente pueden

estar más reactivo en caso de cualquier problema que tenga la empresa, por lo que su labor y las horas que le dedica a cada empresa de su cartera es crucial.

Es importante aclarar que todas estas acciones se hacen en distintos periodos de tiempo y no todas a la misma vez, dado que los recursos son limitados se tiende a priorizar solamente a aquellas empresas que tienen mayor stock, sin tomar en cuenta que existen empresas que son realmente leales a la caja y no se cambiarán, como también existen empresas que no son tan leales y necesitan mayores recursos para ser retenidas.

La mayoría de estas acciones son reactivas pues atacan la fuga una vez que la empresa ha mostrado indicios de querer desafiliarse, en el caso de las empresas foco. Además, varias de estas acciones no tienen un segmento foco definido de modo de rentabilizar las acciones masivas en los segmentos más propensos o trabajadores menos leales que si generar ingresos para la caja.

11 Propuestas de Estrategia de Retención

Como se pudo ver anteriormente, Ilustración 7, los segmentos más rentables para la caja son las empresas foco, las grandes empresas G1 y G2 y las pequeñas y medianas empresas PYM1 y PYM2 por lo que estos segmentos comerciales cruzadas con las reglas más propensas serán los Segmentos a los cuales se les propondrá acciones de retención.

11.1 Segmentos más Propensos a la Fuga

En la Tabla 18 se muestras las 3 reglas del Árbol CHAID con mayor probabilidad de fuga y cuales son las principales característica de las empresas que fueron asignadas a esas rama. A continuación se describirán las principales propuestas para cada segmento.

Tabla 18 Reglas para las Empresas con mayor propensión a la fuga

Variable Fuga_Nodo	Regla	Característica 1	Característica 2	Característica 3	Pred_Fuga	Probabilidad Fuga	Porcentaje de Empresas del total que presentan está regla.
12	Ben_9_12 <= -100 and Cant_Trab > 13 and Saldo_Capital_9_12 <= -17,413	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre disminuyo en un 100%.	Cantidad de Trabajadores mayor a 13	La deuda Crediticia de la Empresa con la caja en el penúltimo semestre disminuyo en más de un 17%	1	100%	5,2%
16	Ben_9_12 > 6,349 and Saldo_Capital_3_6 <= -0,449	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre aumento en más de un 6%.	La deuda crediticia de la Empresa con la caja disminuyo en más de un 1% en el ultimo semestre.		1	86%	3,2%
13	Ben_9_12 <= -100 and Cant_Trab > 13 and Saldo_Capital_9_12 > -17,413	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre disminuyo en un 100%.	Cantidad de Trabajadores mayor a 13	La deuda Crediticia de la Empresa con la caja se mantuvo relativamente constante	1	79%	3,6%

11.1.1 Empresas con Bajos Beneficios y Colocación

El nodo 12 en la Tabla 18 son aquellas empresas que tienen un 100% de probabilidad de fugarse y tienen mas de 13 trabajadores las cuales han disminuido sus beneficios en el ultimo año y su deuda crediticia, por lo que están dando indicios de que están dejando de usar los dos principales servicios de la caja. Además, corresponden a un 5,2% de toda la base, por lo que representan a las mayores causas de desafiliación.

De acuerdo al valor de cada empresa que sea asignada en este nodo se debe decidir si se realizarán campañas para incentivar estos dos últimos ámbitos. Por una parte incluirlos en las campañas de call center con tal de aumentar la colocación de créditos y ofrecerles mejores tasas, de tal manera que tengan intereses de mantener una relación con la caja y por otra parte, incentivar el uso de beneficios que ofrece la caja, como dar a conocer mejor los beneficios y donde pueden usarlos, ya sea mediante Facebook o sucursal o mediante una campaña de correos electrónicos más agresiva a estos trabajadores pertenecientes a estas empresas con mayor riesgo de desafiliarse, que sea diseñada y enfocada para ellos que están descontentos o no conocen los beneficios. O bien, regalar algún producto a estos trabajadores que los fidelice.

11.1.2 Empresas con Baja Colocación

El nodo 16 obtenido en el árbol que se muestra en la Tabla 18, corresponde aquellas empresas que están usando sus beneficios, y se ve reflejado en el aumento de ellos en el ultimo semestre, pero están disminuyendo su colocación de créditos, es decir, pagando su deuda y no solicitando más créditos. Representan a un 3,2% del total de las empresas.

Nuevamente primero que nada se debe cruzar el valor de cada empresas para realizar las acciones en las empresas con rentabilidad altas para la caja. Al igual que el punto anterior, la acción a realizar sobre estas empresas será una campaña agresiva en créditos, que fomente la colocación de crédito con ofertas más atractivas para los trabajadores.

11.1.3 Empresas con Bajos Beneficios

El nodo 13 de la Tabla 18, muestra aquellas empresas que tienen un 79% de probabilidad de fugarse. Estas empresas tienen más de 13 trabajadores y han disminuido el uso de beneficios en un 100% al igual que el nodo 12, con la diferencia que estas empresas han mantenido su deuda con la caja relativamente constante, es decir, o no han pedido créditos, o bien, han pedido nuevos créditos y pagado los antiguos.

Se propone, que una vez cruzadas esta probabilidad de fuga con el valor de las empresas, aplicar una campaña agresiva de beneficios, al igual que en la sección anterior, diseñar un campaña enfocada en los trabajadores pertenecientes a las empresas con mayor probabilidad de fuga.

11.2 Segmento más Propensos a No Fuga

Por otra parte, además de determinar el comportamiento de las empresas que van a fugarse se tiene el comportamiento de las empresas leales a la Caja y que dan indicios de que no pretenden desafiliarse.

En la Tabla 19 se encuentran las reglas para las empresas más leales, y que representan el mayor porcentaje de las empresas. Dado que los recursos son limitados, se propone que aquellas

empresas que son más leales quitarles recursos para asignárselos a las con mayor propensión a fuga y que tengan igual rentabilidad.

Variable Fuga_Nodo	Reglas	Característica 1	Característica 2	Característica 3	Característica 4	Pred_ Fuga	Probabilidad Fuga	Porcentaje de Empresas del total que presentan está regla.
3	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and LM_Dias_Correctos12 = 0 and (Est_Cred ="SC" or Est_Cred = "PYM4" or Est_Cred = "PYM5"or Est_Cred = "GC")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	PYM4,PYM5, GC,SC	Días Correctos asignados en las Licencias Médicas en el ultimo año		0	98,0%	34,1%
6	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and Rec_prom6 <= 0 and (Est_Cred= "R0" or Est_Cred="PYM3")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	PYM3,R0	Trabajadores sin Reclamos en los últimos 6 meses		0	95,7%	10,2%
9	Ben_9_12 > 6,349 and Saldo_Capital_3_6 > 27,762 and Rec_prom6 <= 0	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre aumento en más de un 6%.	La deuda Crediticia de la Empresa con la Caja Aumento en más de un 27% en el ultimo semestre.	Trabajadores sin reclamos en los últimos 6 meses		0	100,0%	4,9%
8	Ben_9_12 > 6,349 and Saldo_Capital_3_6 > -0,449 and Saldo_Capital_3_6 <= 27,762 and Cant_Trab <= 13	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre aumento en más de un 6%.	La deuda Crediticia de la Empresa con la caja se mantuvo relativamente constante	Cantidad de Trabajadores menor a 13.		0	97,5%	4,4%
5	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and LM_Dias_Correctos12 = 1 and Porc_Uso_Ben12 > 7,222 and (Est_Cred = "SC" or Est_Cred = "PYM4" or Est_Cred = "PYM5" or Est_Cred = "GC")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	PYM4,PYM5, GC,SC	Días Asignados Incorrectamente en las Licencias Médicas en el ultimo año	Porcentaje de Trabajadores que usaron Beneficios mayor a un 7%	0	97,5%	4,3%
1	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and Agente = 0 and %Cred_Cast6 <= 0 and (Est_Cred ="G1" or Est_Cred = "G2" or Est_Cred = "PYM1"or Est_Cred = "PYM2")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	G1,G2,PYM1, PYM2	Sin Agente	Sin Créditos Castigados en los últimos 6 meses	0	72,7%	3,6%
2	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and Agente = 0 and %Cred_Cast6 > 0 and (Est_Cred ="G1" or Est_Cred = "G2" or Est_Cred = "PYM1"or Est_Cred = "PYM2")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	G1,G2,PYM1, PYM2	Con Agente	Con Créditos Castigados en los últimos 6 meses	0	96,8%	3,5%

Tabla 19 Reglas de comportamiento para las Empresas con Mayor probabilidad de No Fuga

Como se puede apreciar, el nodo 2 de la Tabla 19, son empresas que tienen un 96,8% de probabilidad de quedarse y cuentan con Agente. Además, son empresas que pertenecen al segmento G1, G2, PYM1 y PYM2 los cuales son segmentos que tienen una alta probabilidad de fuga y alto margen como se muestra en la Ilustración 7. Por lo tanto, se pretende reasignar los recursos con que cuentan estas empresas, por ejemplo, las horas hombres que dedica el agente a estás empresas y reasignarlo a aquellas empresas que tienen mayor probabilidad de fuga y el mismo margen. Otra propuesta, es recarterizar la cartera de empresas con las que trabaja cada Agente, Ejecutivo o Subgerente de mantención, con tal de que le den prioridad aquellas empresas que están en riesgo.

Dado que no se sabe que costo puede tener aplicar esta acción en aquellas empresas más leales, es decir, como influye quitarle el Agente a una empresa o disminuir la atención en ellas y que impacto puede generar, se propone realizar un diseño experimental.

El diseño experimental consiste en poder validar la hipótesis, en este caso la de que reasignar a los agentes no afecta la fuga de las empresas leales. En donde la única variable independiente que se manipulará será la asignación o no de agente a las empresas leales. Se tomarán dos grupos al azar de estás empresas más leales, uno será el grupo de control al cual no se le aplicará tratamiento, mientras que al segundo grupo se le quitarán los agentes, con tal de validar si existe un aumento en la probabilidad de fuga en estas empresas leales una ves que deja de tener agente.

El objetivo del diseño experimental es obtener cual es el costo de reasignar los ejecutivos a las empresas con mayor probabilidad de fuga, con tal de identificar que tan fuerte es la lealtad de las empresas y que tanto aumenta la fuga en aquellas empresas que dejan de recibir este servicio.

11.3 Rentabilidad de las Acciones de Retención

Para medir la rentabilidad de las acción se maximizará la utilidad de la retención de empresas, para ello se utilizará como beneficio el margen que genera cada empresa y el costo de cada acción que se le realizará, como se muestra a continuación.

$$máx \sum_{1}^{N} (Beneficio_i - Costo\ Acción_i) * X_i + Beneficio_i * (1 - X_i)$$

$$s.a.$$

$$P_1 \geq P_2 \geq P_3 \ldots\ldots P_{N-1} \geq P_N$$

En donde,

$$N = Total\ de\ Empresas$$

$$Beneficio_i = El\ margen\ que\ genera\ la\ empresa\ i\ por\ retenerla$$

$$Costo\ Acción_i = Costo\ por\ generar\ la\ acción\ i$$

$$X_j = \begin{cases} 1\ si\ se\ le\ realiza\ acción\ a\ la\ empresa \\ 0\ si\ no\ se\ le\ realiza\ acción\ a\ la\ empresa \end{cases}$$

$$P_j = Probabilidad\ de\ fuga\ de\ la\ empresa$$

Por otra parte, la matriz de costos se muestra en Tabla 20, en donde se tendrá un $Beneficio_j$ cuando se retienen las empresas que el modelo arroja la predicción correctamente y se realiza un acción para retenerlas con un $Costo_f$ y un $Beneficio_i$ de las empresas que seguirán afiliadas. Mientras que se tendrá una perdida de $-Beneficio_k$ por aquellas empresas que el modelo no detecta como fugadas y predice que no se fugan, o bien, tendrá un gasto innecesario en aquellas empresas que el modelo arroja como fuga, cuando realmente no se fugarán aplicando un acción de retención con un $-Costo_s$.

Tabla 20 Matriz de Beneficios y Costos al Aplicar un Acción a cada segmento

		Real	
		No Fuga	Fuga
Pronóstico	No Fuga	$Beneficio_i$	$-Beneficio_k$
		0	0
	Fuga	$Beneficio_j$	$Beneficio_j$
		$-Costo_s$	$-Costo_f$

Con la información anterior se optimizará la utilidad de donde se obtendrá el porcentaje de empresas con mayor probabilidad de fuga a las cuales se les debe realizar acciones de retención.

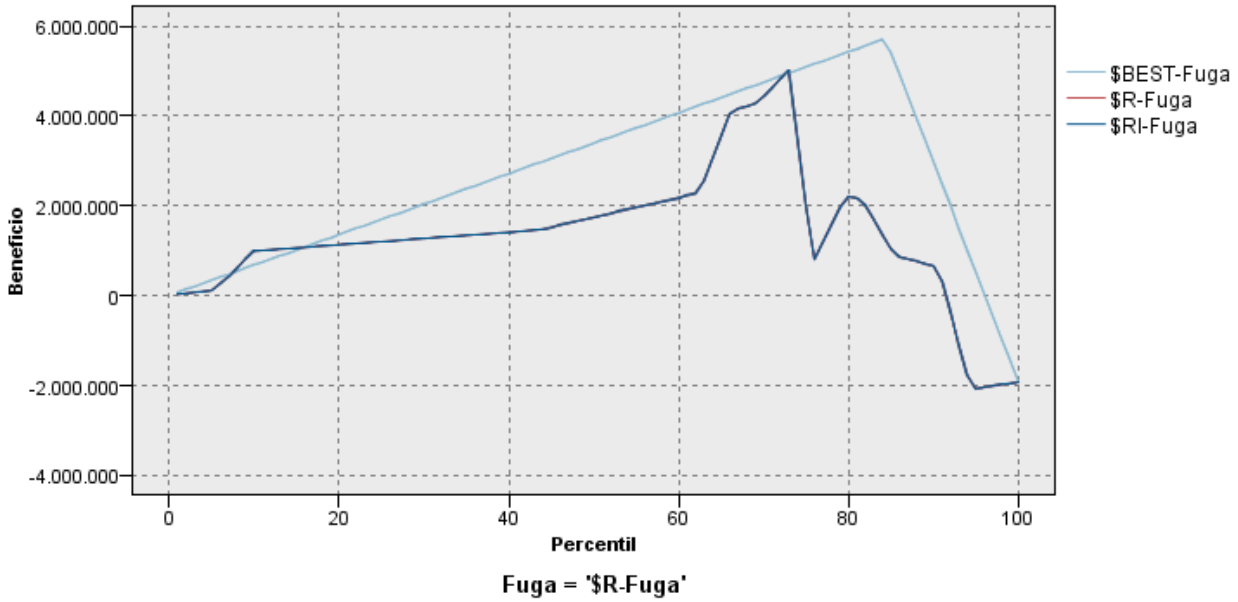
Los costos asignados serán variables por acción y por empresa, ya que dependerá de la cantidad de trabajadores que tenga. Además, se tiene el costo promedio de una campaña de Beneficios o de Crédito por trabajador la cual será ponderada por la cantidad de trabajadores de la empresa. El beneficio utilizado será el margen mensual de cada empresa que se muestra en la Tabla 2 y se utilizarán los costos aproximados para cada campaña como se muestra en la Tabla 21, correspondiente a un costo referencial. El costo de una campaña de beneficios incluye la campaña de marketing y el costo por el beneficio en si. En promedio el costo de un beneficio es de $500 pesos y la tasa de efectividad de una campaña es del 5%, lo que corresponde a un costo de $25 pesos por beneficio y se asumió un costo de $35 en el costo de envió y generación de la campaña.

Tabla 21 Costos por Acción de Retención promedio por Trabajador en un mes

	Costo Promedio por Trabajador (Pesos)
Campaña de Beneficios	$60
Campaña de Créditos	$120
Campaña de Beneficio y Créditos simultaneas	$180
Reasignación de Agente	$180

En el Ilustración 33, se muestra el gráfico con el beneficio generado al aplicar las acciones de retención a las empresas. La línea azul oscuro corresponde al modelo y alcanza su máximo cuando se aplican acciones de retención sobre el 72% de empresas con mayor probabilidad de fuga y se obtiene un beneficio de 4.8M mensual al realizar la acción.

Ilustración 33 Gráfico de Beneficios Obtenidos al Aplicar Acciones de Retención



Se aprecia además en la Ilustración 33, que al considerar a más de un 90% de las empresas más probables a fugarse los beneficios pasan a ser negativos, por lo que hubiese convenido no realizar acciones sobre las empresa.

Los beneficios generados por las acciones de retención en comparación al esperado al mejorar en un 50% la tasa actual de fuga que se muestra en la Tabla 2, donde la ganancia anual sería de 41M lo que corresponde a 3.4M mensuales, que es un poco menor al beneficio que arroja al optimizar la utilidad 4,8M, siendo montos relativamente similares. Se concluye que aplicar acciones sobre el 73% de empresas más propensas a la fuga o un porcentaje menor traerá beneficios positivos a la Caja.

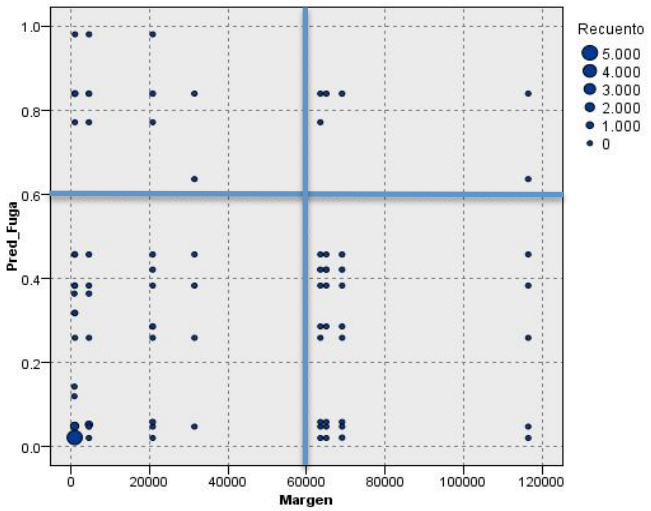
11.4 Grupo de Empresas más Propensas y Prioritarias para realizar acciones de retención

Dado que los recursos son limitados, y tomar acciones sobre el 73% más propenso de todas las empresas, aproximadamente 8 mil empresas, es demasiado costoso, se realizarán filtros para escoger exactamente a que empresas se desea retener.

En la Ilustración 34 se muestra el gráfico de dispersión de las empresas de acuerdo al valor que tienen para los héroes (margen mensual) en el Eje X versus la probabilidad de fuga que posee en el Eje Y. Como criterios para realizar el filtro será tomar las empresas con probabilidad de fuga superior a 0,6 ya que menor a esto la probabilidad es similar a utilizar un modelo de azar y las empresas que dejan un margen mensual sobre 60 mil pesos, es decir, las empresas que están sobre el 50% con mayor valor para la caja. Estas empresas quedan el cuadrante superior derecho

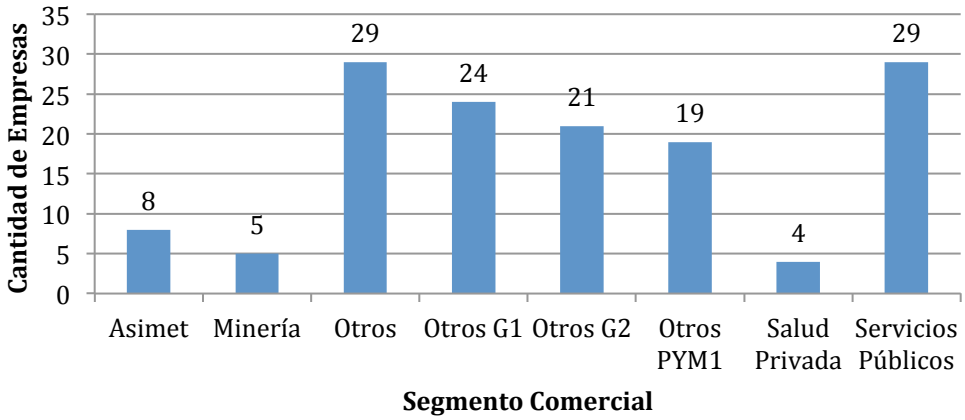
de la Ilustración 34 y corresponden a las empresas de mayor valor y con mayor probabilidad de fuga.

Ilustración 34 Gráfico de Dispersión de empresas de acuerdo valor y probabilidad de fuga.



A las empresas seleccionadas del cuadrante superior derecho se les aplico nuevamente un filtro considerando solamente aquellas empresas que tuviesen sobre 100 trabajadores. El resultado a esta selección corresponde a 139 empresas, las cuales se proponen como el Segmento foco para realizar las acciones descritas en los capítulos anteriores. En la Ilustración 35, se muestra la distribución de estas empresas de acuerdo a su segmento comercial.

Ilustración 35 Cantidad de Empresas más propensas a fuga por Segmento Comercial



La tasa de fuga real de estas 139 empresas corresponde a un 2,2% y en la Tabla 22 se muestra la tasa de desafiliación real para cada segmento.

Tabla 22 Tasa de Desafiliación Real Segmento comercial en un mes

Etiquetas de fila	No Fuga	Fuga	Total general	Tasa Desafiliación Real

Asimet	8		8	0,0%
Minería	5		5	0,0%
Otros	28	1	29	3,4%
Otros G1	23	1	24	4,2%
Otros G2	21		21	0,0%
Otros PYM1	19		19	0,0%
Salud Privada	4		4	0,0%
Servicios Públicos	28	1	29	3,4%
Total general	**136**	**3**	**139**	**2,2%**

Solamente 2 de estas empresas no cuentan con agente, por lo que son dos posibles candidatas a reasignarles recursos o ser utilizadas en el diseño experimental antes mencionado.

Por lo tanto, se pretende que al aplicar las acciones antes propuestas se logre que en este grupo de empresas de mayor valor y probabilidad de fuga reducir la tasa de fuga a la mitad a un 1,1%. En promedio el valor de estas empresas es de $75.578 pesos, por lo que reducir la tasa a la mitad, quiere decir que en vez de perder 3 empresas mensualmente de este grupo ahora serán 1,5, lo que en terminos anuales significa que retenemos 18 empresas con un alto valor. En terminos monetarios se están ganando $1.360.412 por retenerlas en promedio.

12 Conclusiones y Recomendaciones

Se concluye que a partir de las hipótesis de negocio planteadas que el comportamiento en cuanto al uso de beneficios y la deuda crediticia que mantiene la empresa con la Caja, son indicadores claves del estado de la empresa y fueron aceptados con gran certeza. Además, son las principales variables que clasifican el comportamiento de cada empresa con respecto a la caja. En cambio, indicadores a nivel de problemas operativos fueron validados en base a casos muy particulares del árbol y casi al final de la rama, diferenciando la decisión de desafiliarse condicionado a reglas anteriores. Por otra parte, que existan mejores ofertas en la competencia no pudo ser validado como un factor relevante en la fuga debido a que no se cuenta con información histórica de estos eventos, que pudieran dar indicios que tienen relación con la fuga de empresas.

Por otra parte, a partir de los periodos de tiempos de cada variable, se concluyo que para temas de reclamos y problemas operativos son más significativas las que han ocurrido en un plazo cercano a la fecha de desafiliación. Además, si la empresa tuvo un problema con la caja hace un año no es tan relevante como el del ultimo semestre. En cambio, las variables de estado que acumulan el comportamiento anual no tienen un poder predictivo tan fuerte como las variables dinámicas que reflejan un cambio entre los trimestres. Otro *insight* que se extrae de los datos, es que aquellas variables que reflejan si tuvo o no algún problema operativo, son claves para descartar a las empresas leales, ya que estás se caracterizan por tener un comportamiento positivo en la mayoría de los ámbitos: usan beneficios, colocan créditos, no han reclamado o no han tenido problemas operativos. Es por esto, que la caja debe estar atenta a los cambios negativos que se produzcan en los distintos ámbitos.

Actualmente, la Caja pone la mayoría de sus recursos en sus empresas foco, dejando fuera a segmentos que por cantidad de empresas generan un margen anual mayor. En particular, se pudo observar que las mayores tasas de fuga están concentradas en estos segmentos a los que no se les realizan las acciones adecuadas, por lo que se propone enfocar más recursos en estas empresas que son de menor tamaño que las foco pero en conjunto tienen un impacto mayor en los ingresos de la Caja.

La mayoría de los modelos desarrollados presentaron buena precisión y accurancy y finalmente se escogió al árbol CHAID exhaustivo con costos, debido a su fácil interpretación y buena precisión en general. A partir de las reglas obtenidas en el árbol se construyeron distintos perfiles de empresas y a los nodos con las más propensas a fuga se les propuso acciones de retención de acuerdo a su comportamiento o problemas que haya tenido. El modelo entrega un grupo de 139 empresas prioritarias para realizar acciones de retención las cuales además de presentar una alta probabilidad de fuga también presentan un alto margen para la Caja. Definir a estas empresas permite no malgastar recursos en aquellas empresas que no son tan importantes sino en aquellas que realmente le generan un valor a la Caja, rediseñando su foco o la asignación de recursos en sus empresas que actualmente realiza.

Además de las estrategias de retención planteadas, se propone probar que pasa si se reasignan recursos de las empresas leales a empresas similares pero con mayor riesgo de fuga. Determinar el efecto de esta acción puede ayudar optimizar mucho mejor los recursos de la empresa y reasignar personas, horas trabajadas y acciones a las empresas en que es necesario. Por lo tanto, el modelo además de detectar que empresas posiblemente se fuguen, funciona como una herramienta para determinar que empresas están satisfechas con el servicio y se mantendrán leales.

Por lo tanto, se propone como trabajo futuro implementar un diseño experimental como seguimiento de los efectos de reasignar Agentes de las empresas foco a no foco, pero con rentabilidad similar, para posteriormente medir su efectividad y utilidades. Para las empresas no foco, se propone utilizar una estrategia masiva pero focalizada, ya que actualmente las campañas de beneficio no son focalizadas en este segmento, con tal de fomentar el uso de beneficio en aquellas empresas que su comportamiento refleja una tendencia a la disminución. Además, las campañas de créditos que son realizadas mediante un call center tienden a ser más costosas, pero generan un mayor impacto o beneficio cuando se colocan créditos, por lo que genera un valor agregado el saber en que empresas es necesario aumentar su colocación, cruzado con la probabilidad de que el trabajador acepte el crédito se podrá aumentar el impacto en la acción.

Finalmente se concluye, que el tomar acciones de retención con las empresas más rentables y propensas a fuga, le permitirá a la Caja rentabilizar su cartera priorizando aquellas empresas que le generan más valor. Con la investigación realizada en este trabajo queda demostrado que enfocando las acciones de retención en las empresas más probables a fuga y más rentables aumentarán los beneficios esperado, por lo que es importante poner énfasis en las estrategias que se realizarán para retenerlas.

13 Bibliografía

Gobierno de Chile. (2014, Octubre) Superintendencia de Seguridad Social. [Online]. 1] www.suseso.cl

Caja de Compensación Los Héroes. (2012) Primer Reporte de Sustentabilidad. 2] Documento.

Jaime Miranda, Pablo Rey, and Richard Weber, "Prediccion de Fugas de Clientes 3] para una Institucion Financiera mediante Support Vector Machines," Revista Ingeniería de Sistemas, vol. XIX, Octubre 2005.

Raquel García Fernández, Métodos de Predicción de Fuga con Grandes Volúmenes 4] de Datos, TRabajo de Fin de Grado, Grado de Estadistica ed., Facultad de Ciencias, Ed. Valladolid, España: Universidad De Valladolid, 2007.

Vanesa Berlanga Silvente, María José Rubio Hurtado. , and Ruth Vilà Baños, "Cómo 5] aplicar Árboles de Decisión en SPSS," REIRE, Revista d'Innovació i Recerca en Educació, vol. 6, no. 1, pp. 65-79, 2013.

Wen Zhu, Nancy Zeng , and Ning Wang, Sensitivity, Specificity, Accuracy, 6] Associated Confidence Interval and ROC Analysis with Practical SAS® Implementations.: NESUG , 2010.

Yoav Freund and Robert E. Schapire. (1996, Enero) Experiments with a New 7] Boosting Algorithm. Documento.

IBM SPSS, Modelización Avanzada de Datos con Clementine. 8]

Raquel García Fernández, Métodos de predicción de fuga con grandes volúmenes de 9] datos, Trabajo Fin de Grado - Grado en Estadística ed.

Richard Bagozzi, "The CHAID Approach to Segmentation Modeling: CHI-squared 10] Automatic Interaction Detection," in Advanced Methods of Marketing Research.: Blackwell, 1994.

MORINEAU A. ALUJA T., Aprender de los datos: el análisis de componentes 11] principales, una aproximación desde el data mining. Barcelona: EUB, 1999.

Juan Carlos Alcaide, Fidelización de Clientes. Madrid: ESIC EDITORIAL, 2010. 12]

David Hand, Heikki Mannila, and Padhraic Smyth, Principles of Data Mining. 13] Cambridge, Massachusetts, EEUU: The MIT press, 2001.

T. Vafeiadis a, "A comparison of machine learning techniques for customer," 14] ELSEVIER, 2015.

K.I. Diamantaras b, G. Sarigiannidis a, K.Ch. Chatzisavvas a T. Vafeiadis a, "Credit 15] card churn forecasting by logistic regression and decision tree," ELSEVIER, 2015.

14 Anexos

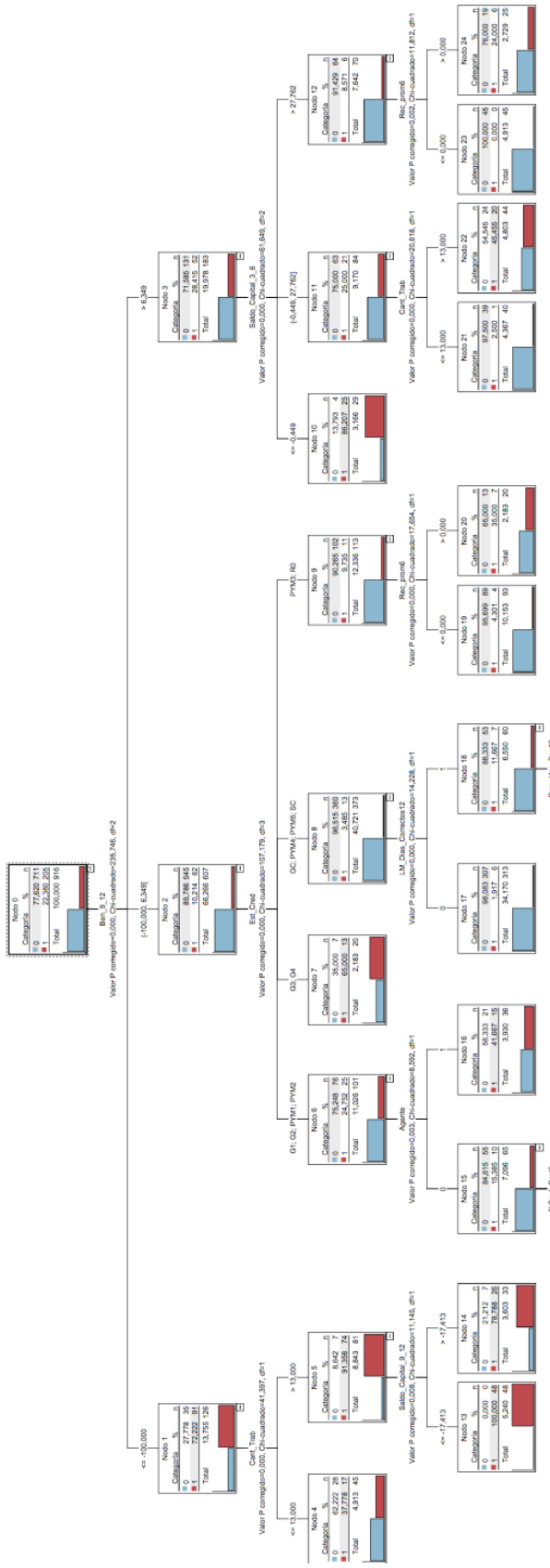
14.1 Indicadores de la Empresa

Área	Nombre Variable	Temporalidad	Definición	Categorización	
Beneficios	Porc_Uso_Ben	3-6-9-12	Corresponde al porcentaje de trabajadores de la empresa que usaron beneficios durante ese periodo.		
	Ben_prom	3-6-9-13	Corresponde a los beneficios acumulados de la empresa durante esos meses divido por la cantidad de trabajadores, generando el promedio de beneficios por trabajador dentro de la empresa.		
	Porc_Uso_ben_rango	3-6-9-14	Corresponde la variable Porc_Uso_Ben categorizada	1	Porc_Uso_Ben3 = 0.0
				2	Porc_Uso_Ben3 > 0.0 and Porc_Uso_Ben3 <= 10
				3	Porc_Uso_Ben3 > 10 and Porc_Uso_Ben3 <= 20
				4	Porc_Uso_Ben3 > 20 and Porc_Uso_Ben3 <= 30
				5	Porc_Uso_Ben3 > 30 and Porc_Uso_Ben3 <= 40
				6	Porc_Uso_Ben3 > 40 and Porc_Uso_Ben3 <= 50
				7	Porc_Uso_Ben3 > 50
	Ben_prom_rango	3-6-9-15	Corresponde a la variable Ben_prom categorizada	1	Ben_prom3 = 0.0
				2	Ben_prom3> 0.0 and Ben_prom3 <= 2.5
				3	Ben_prom3> 2.5 and Ben_prom3<=5
				4	Ben_prom3> 5 and Ben_prom3<=7.5
				5	Ben_prom3> 7.5 and Ben_prom3<=10
				6	Ben_prom3> 10 and Ben_prom3<=15
				7	Ben_prom3<15
	PDB_2013	-	Empresa presenta PDB durante el año 2013	1	Con PDB
				0	Sin PDB
	PDB_2014	-	Empresa presenta PDB durante el año 2014	1	Con PDB
				0	Sin PDB
	PDB_Aumento	-	Registra los cambios en el monto del PDB durante los años 2013 y 2014	2	'PDB 2014'>'PDB 2013'
				1	'PDB 2014' = 'PDB 2013'
				0	'PDB 2014'<'PDB 2013'
Reclamos	Porc_Rec	3-6-9-12	Corresponde al porcentaje de trabajadores que hicieron reclamos respecto el total de trabajdores de la empresa.		
	Rec_prom	3-6-9-12	Corresponde al promedio de reclamos por trabajador dentro de la emrpesa		
	%Rec_Trab	3-6-9-12	Corresponde a la variable Porc_Rec categorizada	1	Porc_Rec3 = 0.0
				2	Porc_Rec3 > 0.0 and Porc_Rec3 < 10
				3	Porc_Rec3 >= 10 and Porc_Rec3 < 20
				4	Porc_Rec3 >= 20 and Porc_Rec3 < 30
				5	Porc_Rec3 >= 30 and Porc_Rec3 < 40
				6	Porc_Rec3 >= 40 and Porc_Rec3 < 50
				7	Porc_Rec3 >= 50
	Rec_Trab_prom	3-6-9-12	Corresponde a la variable Rec_prom categorizada	1	Rec_prom3 = 0.000
				2	Rec_prom3 > 0.0 and Rec_prom3 < 0.25
				3	Rec_prom3 >= 0.25 and Rec_prom3 < 0.5
				4	Rec_prom3 >= 0.5 and Rec_prom3 < 0.75
				5	Rec_prom3 >= 0.75 and Rec_prom3 < 1
				6	Rec_prom3 >= 1
	Rec_Cred				
	Rec_AF				
	Rec_LM				

Área	Nombre Variable	Temporalidad	Definición	Categorización	
Reporte de Alertas	Visita_Acum	3-6-9-12			
	Visita_prom	3-6-9-12	Corresponde a la variable Visita_Acum categorizada	1	Vista_Acum3=0.0
				2	Vista_Acum3 > 0.0 and Vista_Acum3 <= 5
				3	Vista_Acum3 > 5 and Vista_Acum3 <= 10
				4	Vista_Acum3 > 10 and Vista_Acum3 <= 20
				5	Vista_Acum3 > 20 and Vista_Acum3 <= 30
				6	Vista_Acum3 > 30 and Vista_Acum3 <= 40
				7	Vista_Acum3 > 40 and Vista_Acum3 <= 50
				8	Vista_Acum3 > 50
	MA	3-6-9-12	Declaración de riesgo en visita de MUY ALTA	1	Con Declaración MA
				0	Sin Declaración MA
	A	3-6-9-12	Declaración de riesgo en visita de ALTA	1	Con Declaración A
				0	Sin Declaración A
	M	3-6-9-12	Declaración de riesgo en visita de MEDIA	1	Con Declaración M
				0	Sin Declaración M
Créditos	Colocación	3-6-9-12	Cantidad de créditos colocados promedio mensual		
	%Coloc	3-6-9-12	Porcentaje de colocación de créditos dentro de la empresa, es decir, cantidad de trabajadores que colocaron crédito en el periodo.		
	%Coloc_rango	3-6-9-12	Corresponde a la variable %Coloc categorizada	1	'%Coloc3' = 0
				2	'%Coloc3' >= 0 and '%Coloc3' < 5
				3	'%Coloc3' >= 5 and '%Coloc3' < 10
				4	'%Coloc3' >= 10 and '%Coloc3' < 20
				5	'%Coloc3' >=20
	Coloc_prom	3-6-9-12	Corresponde a la variable "Colocación" categorizada	1	Colocacion3 = 0.0
				2	Colocacion3 >= 0 and Colocacion3 < 2.5
				3	Colocacion3 >= 2.5 and Colocacion3 < 5
				4	Colocacion3 >= 5 and Colocacion3 < 10
				5	Colocacion3 >= 10 and Colocacion3 < 20
				6	Colocacion3 >= 20
	%Cred_Vig	3-6-9-12	Porcentaje de trabajadores con creditos vigentes del total de trabajadores de la empresa		
	%Cred_Vig_rango	3-6-9-12	Corresponde a la variable %Cred_Vig categorizada	1	'%Cred_Vig3' = 0
				2	'%Cred_Vig3' >= 0 and '%Cred_Vig3' < 5
				3	'%Cred_Vig3' >= 5 and '%Cred_Vig3' < 10
				4	'%Cred_Vig3' >= 10 and '%Cred_Vig3' < 20
				5	'%Cred_Vig3' >=20
	%Cred_Mor	3-6-9-12	Porcentaje de trabajadores con creditos morosos del total de trabajadores de la empresa		
	%Cred_Mor_rango	3-6-9-12	Corresponde a la variable %Cred_Mor categorizada	1	'%Cred_Mor6' = 0
				2	'%Cred_Mor6' >= 0 and '%Cred_Mor6' < 5
				3	'%Cred_Mor6' >= 5 and '%Cred_Mor6' < 10
				4	'%Cred_Mor6' >= 10 and '%Cred_Mor6' < 20
				5	'%Cred_Mor6' >=20
	%Cred_Cast	3-6-9-12	Porcentaje de trabajadores con creditos Castigados del total de trabajadores de la empresa		
	%Cred_Cast_rango	3-6-9-12	Corresponde a la variable %Cred_Cast categorizada	1	'%Cred_Cast6' = 0
				2	'%Cred_Cast6' >= 0 and '%Cred_Cast6' < 5
				3	'%Cred_Cast6' >= 5 and '%Cred_Cast6' < 10
				4	'%Cred_Cast6' >= 10 and '%Cred_Cast6' < 20
				5	'%Cred_Cast6' >=20
	Cuo_Pag	3-6-9-12			
	%Cuo_Pag	3-6-9-12	Categorización variable Cuo_Pag	1	Cuo_Pag3=0.000
				2	Cuo_Pag3 > 0.000 and Cuo_Pag3 < 25
				3	Cuo_Pag3 >= 25 and Cuo_Pag3 < 50
				4	Cuo_Pag3 >= 50 and Cuo_Pag3 < 75
				5	Cuo_Pag3 >= 75
	Avance_cred_tramo1				
	Avance_cred_tramo2				
	Avance_cred_tramo3				
	Avance_cred_tramo4				
	Avance_cred_tramo5				
	%Avance_cred_tramo1/2/3/4/5	1-2-3-4-5	Categorización de la variable Avance_cred_TramoX	1	Avance_cred_tramo3=0.000
				2	Avance_cred_tramo3 > 0.000 and Avance_cred_tramo3<= 25
				3	Avance_cred_tramo3 > 25 and Avance_cred_tramo3<= 50
				4	Avance_cred_tramo3> 50 and Avance_cred_tramo3<= 75
				5	Avance_cred_tramo3 > 75 and Avance_cred_tramo3<= 100
				6	Avance_cred_tramo3 > 100

Área	Nombre Variable	Temporalidad	Definición	Categorización	
Visita Sucursal	Porc_Vis_Suc	3-6-9-12	Porcentaje de trabajadores que visitaron la sucursal durante el periodo del total de trabajadores de la empresa		
	%Vis_Suc_rango	3-6-9-12	Categorización de la variable Porc_Vis_Suc	1	Porc_Vis_Suc3=0.000
				2	Porc_Vis_Suc3>0.000 and Porc_Vis_Suc3<=5
				3	Porc_Vis_Suc3>5 and Porc_Vis_Suc3<=10
				4	Porc_Vis_Suc3>10 and Porc_Vis_Suc3<=20
				5	Porc_Vis_Suc3>20 and Porc_Vis_Suc3<=30
				6	Porc_Vis_Suc3>30
RFM	Puntuación RFM				
Licencias Médicas	%LM	3-6-9-12	Porcentaje de trabajadores que presenta LM a la CCAF		
	%LM_rango	3-6-9-12	Categorización de %LM	1	'%LM3'=0.000
				2	'%LM3' > 0.000 and '%LM3' <= 5
				3	'%LM3' > 5 and '%LM3' <= 10
				4	'%LM3' > 10 and '%LM3' <= 15
				5	'%LM3' > 15 and '%LM3' <= 20
				6	'%LM3' >20
	Prom_LM_Trab	3-6-9-12	Promedio de LM que presenta un trabajador dentro de la empresa		
	Prom_LM_Trab_rango	3-6-9-12	Categorización de la variable Prom_LM_Trab	1	Prom_LM_Trab3=0.000
				2	Prom_LM_Trab3>0.000 and Prom_LM_Trab3<=0.5
				3	Prom_LM_Trab3>0.5 and Prom_LM_Trab3 <= 1
				4	Prom_LM_Trab3>1
	%LM_Atrasos	3-6-9-12	Porcentaje de LM que se pagaron con atrasos respecto del total de LM presentadas en el mismo periodo		
	%LM_Atrasos_rango	3-6-9-12	Categorización %LM_Atrasos	1	'%LM_Atrasos3'=0.000
				2	'%LM_Atrasos3'>0.000 and '%LM_Atrasos3'<=25
				3	'%LM_Atrasos3'>25 and '%LM_Atrasos3'<=50
				4	'%LM_Atrasos3'>50 and '%LM_Atrasos3'<=75
				5	'%LM_Atrasos3'>75 and '%LM_Atrasos3'<=100

14.2 Árbol CHAID Exhaustivo



14.3 Reglas de los Nodos terminales del Árbol de Decisión para empresas No Fugadas

Variable Fuga_Nodo	Reglas	Característica 1	Característica 2	Característica 3	Característica 4	Pred_Fuga	Probabilidad Fuga	Porcentaje de Empresas del total que presentan está regla.
3	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and LM_Dias_Correctos12 = 0 and (Est_Cred ="SC" or Est_Cred = "PYM4" or Est_Cred = "PYM5"or Est_Cred = "GC")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	PYM4,PYM5,GC,SC	Días Correctos asignados en las Licencias Médicas en el ultimo año		0	0,98	34,1%
6	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and Rec_prom6 <= 0 and (Est_Cred= "R0" or Est_Cred="PYM3")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	PYM3,R0	Trabajadores sin Reclamos en los últimos 6 meses		0	0,957	10,2%
9	Ben_9_12 > 6,349 and Saldo_Capital_3_6 > 27,762 and Rec_prom6 <= 0	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre aumento en más de un 6%.	La deuda Crediticia de la Empresa con la Caja Aumento en más de un 27% en el ultimo semestre.	Trabajadores sin reclamos en los últimos 6 meses		0	1	4,9%
8	Ben_9_12 > 6,349 and Saldo_Capital_3_6 > -0,449 and Saldo_Capital_3_6 <= 27,762 and Cant_Trab <= 13	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre aumento en más de un 6%.	La deuda Crediticia de la Empresa con la caja se mantuvo relativamente constante	Cantidad de Trabajadores menor a 13.		0	0,975	4,4%
5	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and LM_Dias_Correctos12 = 1 and Porc_Uso_Ben12 > 7,222 and (Est_Cred = "SC" or Est_Cred = "PYM4" or Est_Cred = "PYM5" or Est_Cred = "GC")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	PYM4,PYM5,GC,SC	Días Asignados Incorrectamente en las Licencias Médicas en el ultimo año	Porcentaje de Trabajadores que usaron Beneficios mayor a un 7%	0	0,975	4,3%
1	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and Agente = 0 and %Cred_Cast6 <= 0 and (Est_Cred ="G1" or Est_Cred = "G2" or Est_Cred = "PYM1"or Est_Cred = "PYM2")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	G1,G2,PYM1,PYM2	Sin Agente	Sin Créditos Castigados en los últimos 6 meses	0	0,727	3,6%
2	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and Agente = 0 and %Cred_Cast6 > 0 and (Est_Cred ="G1" or Est_Cred = "G2" or Est_Cred = "PYM1"or Est_Cred = "PYM2")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	G1,G2,PYM1,PYM2	Con Agente	Con Créditos Castigados en los últimos 6 meses	0	0,968	3,5%
10	Ben_9_12 > 6,349 and Saldo_Capital_3_6 > 27,762 and Rec_prom6 > 0	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre aumento en más de un 6%.	La deuda Crediticia de la Empresa con la Caja Aumento	Trabajadores con reclamos los últimos 6 meses.		0	0,76	2,7%
4	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and LM_Dias_Correctos12 = 1 and Porc_Uso_Ben12 <= 7,222 and (Est_Cred="SC" or Est_Cred= "PYM4" or Est_Cred= "PYM5" or Est_Cred= "GC")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	PYM4,PYM5,GC,SC	Días Asignados Incorrectamente en las Licencias Médicas en el ultimo año	Porcentaje de Trabajadores que usaron Beneficios menor a un 7%	0	0,7	2,2%
7	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and Rec_prom6 > 0 and (Est_Cred= "R0"or Est_Cred= "PYM3")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	PYM3,R0	Trabajadores con reclamos en los últimos 6 meses		0	0,65	2,2%

14.4 Reglas de los Nodos terminales del Árbol de Decisión para empresas Fugadas

Variable Fuga_Nodo	Regla	Característica 1	Característica 2	Característica 3	Pred_Fuga	Probabilidad Fuga	Porcentaje de Empresas del total que presentan está regla.
12	Ben_9_12 <= -100 and Cant_Trab > 13 and Saldo_Capital_9_12 <= -17,413	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre disminuyo en un 100%.	Cantidad de Trabajadores mayor a 13	La deuda Crediticia de la Empresa con la caja en el penúltimo semestre disminuyo en más de un 17%	1	100%	5,2%
11	Ben_9_12 <= -100 and Cant_Trab <= 13	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre disminuyo en un 100%.	Cantidad de trabajadores menor a 13.		1	38%	4,9%
17	Ben_9_12 > 6,349 and Saldo_Capital_3_6 > -0,449 and Saldo_Capital_3_6 <= 27,762 and Cant_Trab > 13	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre aumento en más de un 6%.	La deuda Crediticia de la Empresa con la Caja se mantuvo relativamente constante	Cantidad de Trabajadores mayor a 13.	1	45%	4,8%
14	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and Agente = 1 and (Est_Cred= "G1"or Est_Cred= "G2"or Est_Cred= "PYM1"or Est_Cred= "PYM2")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	G1,G2,PYM1,PYM2	Con Agente	1	42%	3,9%
13	Ben_9_12 <= -100 and Cant_Trab > 13 and Saldo_Capital_9_12 > -17,413	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre disminuyo en un 100%.	Cantidad de Trabajadores mayor a 13	La deuda Crediticia de la Empresa con la caja se mantuvo relativamente constante	1	79%	3,6%
16	Ben_9_12 > 6,349 and Saldo_Capital_3_6 <= -0,449	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre aumento en más de un 6%.	La deuda crediticia de la Empresa con la caja disminuyo en más de un 1% en el ultimo semestre.		1	86%	3,2%
15	Ben_9_12 > -100 and Ben_9_12 <= 6,349 and(Est_Cred= "G3" or Est_Cred="G4")	Porcentaje de variación de la Cantidad de Beneficios Usados el penúltimo semestre es Constante	G3,G4		1	65%	2,2%

14.5 Matriz de Confusión Árbol CHAID Exhaustivo

Comparando $R-Fuga con Fuga

'Partición'	1_Entrenamiento		2_Comprobación		3_Validación	
Correctos	784	85,59%	160	77,67%	153	85,96%
Erróneos	132	14,41%	46	22,33%	25	14,04%
Total	916		206		178	

Matriz de coincidencias para $R-Fuga (las filas muestran las reales)

'Partición' = 1_Entrenamiento	0	1
0	620	91
1	41	164

'Partición' = 2_Comprobación	0	1
0	122	27
1	19	38

'Partición' = 3_Validación	0	1
0	121	19
1	6	32

14.6 Prueba de Homogeneidad de Varianza para las Variables influyentes en las Hipótesis

Prueba de homogeneidad de varianzas

	Estadístico de Levene	df1	df2	Sig.
Colocacion_6_12	149,338	1	1298	,000
Porc_Coloc3	9,668	1	1298	,002
Porc_Coloc6	6,745	1	1298	,010
Porc_Coloc9	4,072	1	1298	,044
Porc_Coloc12	1,686	1	1298	,194
Saldo_Capital_3_6	120,546	1	1298	,000
Saldo_Capital_6_9	86,934	1	1298	,000
Saldo_Capital_9_12	13,696	1	1298	,000
Saldo_Capital_6_12	27,099	1	1298	,000
Ben_3_6	96,838	1	1298	,000
Ben_6_9	143,302	1	1298	,000
Ben_9_12	,217	1	1298	,041
Ben_6_12	66,091	1	1298	,000
Rec_Emp_3_6	65,017	1	1298	,000
Rec_Emp_6_9	6,137	1	1298	,013
Rec_Emp_9_12	5,986	1	1298	,015
Rec_Emp_6_12	66,787	1	1298	,000
Rec_trab_3_6	163,656	1	1298	,000
Rec_trab_6_9	64,231	1	1298	,000
Rec_trab_9_12	3,351	1	1298	,067
Rec_trab_6_12	151,557	1	1298	,000

Porc_Uso_Ben3	1,114	1	1298	,291
Porc_Uso_Ben6	1,146	1	1298	,285
Porc_Uso_Ben9	1,157	1	1298	,282
Porc_Uso_Ben12	1,178	1	1298	,278
PDB_2013	94,418	1	1298	,000
PDB_2014	34,083	1	1298	,000
PDB_Aumento	45,960	1	1298	,000
Ben_prom3	1,194	1	1298	,275
Ben_prom6	1,180	1	1298	,277
Ben_prom9	1,192	1	1298	,275
Ben_prom12	1,212	1	1298	,271
Rec_Empleador_Acum3	76,404	1	1298	,000
Rec_Empleador_Acum6	68,168	1	1298	,000
Rec_Empleador_Acum9	11,524	1	1298	,001
Rec_Empleador_Acum12	7,045	1	1298	,008
Rec_prom3	,854	1	1298	,356
Rec_prom6	,969	1	1298	,325
Rec_prom9	1,061	1	1298	,303
Rec_prom12	1,080	1	1298	,299
Porc_Rec3	15,021	1	1298	,000
Porc_Rec6	10,903	1	1298	,001
Porc_Rec9	1,222	1	1298	,269
Porc_Rec12	,957	1	1298	,328
Rec_Trab3	146,680	1	1298	,000
Rec_Trab6	60,712	1	1298	,000
Rec_Trab9	9,215	1	1298	,002
Rec_Trab12	4,855	1	1298	,028
Rec_Cred3	27,244	1	1298	,000
Rec_Cred6	21,705	1	1298	,000
Rec_Cred9	1,055	1	1298	,305
Rec_Cred12	1,801	1	1298	,180
Rec_AF3	194,126	1	1298	,000
Rec_AF6	62,249	1	1298	,000
Rec_AF9	36,248	1	1298	,000
Rec_AF12	10,537	1	1298	,001
Rec_LM3	19,151	1	1298	,000
Rec_LM6	1,895	1	1298	,169
Rec_LM9	,044	1	1298	,834
Rec_LM12	,138	1	1298	,710
Coloc3	2,388	1	1298	,123
Coloc6	,814	1	1298	,367

Coloc9	,961	1	1298	,327
Coloc12	1,046	1	1298	,307
LM_Atrasos3	169,016	1	1298	,000
LM_Atrasos6	136,113	1	1298	,000
LM_Atrasos9	108,587	1	1298	,000
LM_Atrasos12	107,635	1	1298	,000
LM_Dias_Correctos3	140,646	1	1298	,000
LM_Dias_Correctos6	46,100	1	1298	,000
LM_Dias_Correctos9	18,088	1	1298	,000
LM_Dias_Correctos12	,094	1	1298	,759
Dif_Acum3_marca	142,994	1	1298	,000
Dif_Acum6_marca	123,144	1	1298	,000
Dif_Acum9_marca	112,151	1	1298	,000
Dif_Acum12_marca	119,443	1	1298	,000

14.7 Test ANOVA para las variables influyentes en las Hipótesis

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Colocacion_6_12	Entre grupos	1804845,099	1	1804845,099	54,213	,000
	Dentro de grupos	43212826,472	1298	33291,854		
	Total	45017671,571	1299			
Porc_Coloc3	Entre grupos	10383,984	1	10383,984	3,706	,054
	Dentro de grupos	3636605,686	1298	2801,699		
	Total	3646989,670	1299			
Porc_Coloc6	Entre grupos	10806,150	1	10806,150	3,025	,082
	Dentro de grupos	4636686,831	1298	3572,178		
	Total	4647492,981	1299			
Porc_Coloc9	Entre grupos	9529,264	1	9529,264	2,062	,151
	Dentro de grupos	5998537,524	1298	4621,369		
	Total	6008066,788	1299			
Porc_Coloc12	Entre grupos	7631,577	1	7631,577	1,122	,290
	Dentro de grupos	8829260,961	1298	6802,204		
	Total	8836892,538	1299			
Saldo_Capital_3_6	Entre grupos	8017582031055 689700,000	1	8017582031055 689700,000	33,220	,000

	Dentro de grupos	3132741108616 18800000,000	1298	2413513951168 09536,000		
	Total	3212916928926 74460000,000	1299			
Saldo_Capital_6_9	Entre grupos	7322474072985 3360000,000	1	7322474072985 3360000,000	25,434	,000
	Dentro de grupos	3736901647449 938000000,000	1298	2878968911748 796400,000		
	Total	3810126388179 791500000,000	1299			
Saldo_Capital_9_12	Entre grupos	7530363721522 58180,000	1	7530363721522 58180,000	3,517	,061
	Dentro de grupos	2779275534896 33670000,000	1298	2141198409011 04512,000		
	Total	2786805898617 85930000,000	1299			
Saldo_Capital_6_12	Entre grupos	1095586156755 33104,000	1	1095586156755 33104,000	6,955	,008
	Dentro de grupos	2044599620968 5406000,000	1298	1575192311994 2532,000		
	Total	2055555482536 0937000,000	1299			
Ben_3_6	Entre grupos	151870694,053	1	151870694,053	34,238	,000
	Dentro de grupos	5757648722,62 2	1298	4435784,840		
	Total	5909519416,67 5	1299			
Ben_6_9	Entre grupos	211816015,300	1	211816015,300	47,445	,000
	Dentro de grupos	5794805718,75 2	1298	4464411,185		
	Total	6006621734,05 2	1299			
Ben_9_12	Entre grupos	2510,457	1	2510,457	,001	,079
	Dentro de grupos	4741159934,20 6	1298	3652665,589		
	Total	4741162444,66 3	1299			
Ben_6_12	Entre grupos	187668276,978	1	187668276,978	23,710	,000
	Dentro de grupos	10273930171,9 63	1298	7915200,441		
	Total	10461598448,9 41	1299			
Rec_Emp_3_6	Entre grupos	20753,912	1	20753,912	14,629	,000
	Dentro de grupos	1841410,879	1298	1418,652		
	Total	1862164,790	1299			
Rec_Emp_6_9	Entre grupos	6875,400	1	6875,400	5,610	,018
	Dentro de grupos	1590642,654	1298	1225,457		
	Total	1597518,054	1299			
Rec_Emp_9_12	Entre grupos	15203,139	1	15203,139	5,488	,019
	Dentro de grupos	3595841,971	1298	2770,294		
	Total	3611045,109	1299			
Rec_Emp_6_12	Entre grupos	17866,585	1	17866,585	6,438	,011
	Dentro de grupos	3602146,548	1298	2775,151		
	Total	3620013,132	1299			
Rec_trab_3_6	Entre grupos	584697,710	1	584697,710	51,930	,000
	Dentro de grupos	14614658,083	1298	11259,367		
	Total	15199355,793	1299			
Rec_trab_6_9	Entre grupos	297474,135	1	297474,135	23,959	,000
	Dentro de grupos	16116125,286	1298	12416,121		
	Total	16413599,422	1299			

Rec_trab_9_12	Entre grupos	11602,698	1	11602,698	1,848	,174
	Dentro de grupos	8148104,072	1298	6277,430		
	Total	8159706,770	1299			
Rec_trab_6_12	Entre grupos	1329021,617	1	1329021,617	55,053	,000
	Dentro de grupos	31334787,913	1298	24140,823		
	Total	32663809,531	1299			
Porc_Uso_Ben3	Entre grupos	168833,125	1	168833,125	,340	,560
	Dentro de grupos	643853539,122	1298	496035,084		
	Total	644022372,247	1299			
Porc_Uso_Ben6	Entre grupos	489610,884	1	489610,884	,347	,556
	Dentro de grupos	1833122005,104	1298	1412266,568		
	Total	1833611615,989	1299			
Porc_Uso_Ben9	Entre grupos	542432,675	1	542432,675	,354	,552
	Dentro de grupos	1987888867,080	1298	1531501,438		
	Total	1988431299,755	1299			
Porc_Uso_Ben12	Entre grupos	586743,338	1	586743,338	,358	,550
	Dentro de grupos	2129090636,230	1298	1640285,544		
	Total	2129677379,568	1299			
PDB_2013	Entre grupos	2,432	1	2,432	25,776	,000
	Dentro de grupos	122,491	1298	,094		
	Total	124,923	1299			
PDB_2014	Entre grupos	,690	1	,690	8,951	,003
	Dentro de grupos	100,003	1298	,077		
	Total	100,692	1299			
PDB_Aumento	Entre grupos	,322	1	,322	3,136	,077
	Dentro de grupos	133,121	1298	,103		
	Total	133,442	1299			
Ben_prom3	Entre grupos	464,818	1	464,818	,371	,542
	Dentro de grupos	1624858,084	1298	1251,817		
	Total	1625322,903	1299			
Ben_prom6	Entre grupos	4955,895	1	4955,895	,368	,544
	Dentro de grupos	17462069,376	1298	13453,058		
	Total	17467025,271	1299			
Ben_prom9	Entre grupos	11619,267	1	11619,267	,376	,540
	Dentro de grupos	40159721,752	1298	30939,693		
	Total	40171341,018	1299			
Ben_prom12	Entre grupos	21555,704	1	21555,704	,383	,536
	Dentro de grupos	73023088,209	1298	56258,157		
	Total	73044643,913	1299			
Rec_Empleador_Acum3	Entre grupos	2,560	1	2,560	21,217	,000
	Dentro de grupos	156,643	1298	,121		
	Total	159,203	1299			
Rec_Empleador_Acum6	Entre grupos	4,675	1	4,675	20,202	,000
	Dentro de grupos	300,386	1298	,231		
	Total	305,061	1299			
Rec_Empleador_Acum9	Entre grupos	2,843	1	2,843	4,623	,032
	Dentro de grupos	798,279	1298	,615		
	Total	801,122	1299			
Rec_Empleador_Acum12	Entre grupos	5,616	1	5,616	6,292	,012
	Dentro de grupos	1158,454	1298	,892		
	Total	1164,070	1299			
Rec_prom3	Entre grupos	,265	1	,265	,195	,659
	Dentro de grupos	1765,712	1298	1,360		

	Total	1765,977	1299			
Rec_prom6	Entre grupos	1,282	1	1,282	,236	,627
	Dentro de grupos	7062,114	1298	5,441		
	Total	7063,396	1299			
Rec_prom9	Entre grupos	3,068	1	3,068	,281	,596
	Dentro de grupos	14180,273	1298	10,925		
	Total	14183,341	1299			
Rec_prom12	Entre grupos	6,001	1	6,001	,286	,593
	Dentro de grupos	27240,612	1298	20,987		
	Total	27246,613	1299			
Porc_Rec3	Entre grupos	190,600	1	190,600	8,375	,004
	Dentro de grupos	29540,888	1298	22,759		
	Total	29731,488	1299			
Porc_Rec6	Entre grupos	358,993	1	358,993	7,351	,007
	Dentro de grupos	63387,079	1298	48,834		
	Total	63746,071	1299			
Porc_Rec9	Entre grupos	174,958	1	174,958	1,560	,212
	Dentro de grupos	145537,564	1298	112,124		
	Total	145712,522	1299			
Porc_Rec12	Entre grupos	267,392	1	267,392	2,099	,148
	Dentro de grupos	165384,538	1298	127,415		
	Total	165651,930	1299			
Rec_Trab3	Entre grupos	14,927	1	14,927	68,101	,000
	Dentro de grupos	284,516	1298	,219		
	Total	299,443	1299			
Rec_Trab6	Entre grupos	18,135	1	18,135	45,086	,000
	Dentro de grupos	522,108	1298	,402		
	Total	540,243	1299			
Rec_Trab9	Entre grupos	9,761	1	9,761	15,204	,000
	Dentro de grupos	833,316	1298	,642		
	Total	843,077	1299			
Rec_Trab12	Entre grupos	13,739	1	13,739	17,926	,000
	Dentro de grupos	994,854	1298	,766		
	Total	1008,593	1299			
Rec_Cred3	Entre grupos	12,457	1	12,457	7,340	,007
	Dentro de grupos	2202,906	1298	1,697		
	Total	2215,362	1299			
Rec_Cred6	Entre grupos	10,900	1	10,900	6,050	,014
	Dentro de grupos	2338,411	1298	1,802		
	Total	2349,311	1299			
Rec_Cred9	Entre grupos	2,512	1	2,512	,386	,535
	Dentro de grupos	8447,546	1298	6,508		
	Total	8450,058	1299			
Rec_Cred12	Entre grupos	8,012	1	8,012	,673	,412
	Dentro de grupos	15443,258	1298	11,898		
	Total	15451,270	1299			
Rec_AF3	Entre grupos	6,487	1	6,487	47,966	,000
	Dentro de grupos	175,556	1298	,135		
	Total	182,043	1299			
Rec_AF6	Entre grupos	,115	1	,115	15,190	,000
	Dentro de grupos	9,836	1298	,008		
	Total	9,951	1299			
Rec_AF9	Entre grupos	,095	1	,095	8,956	,003
	Dentro de grupos	13,828	1298	,011		
	Total	13,923	1299			
Rec_AF12	Entre grupos	,020	1	,020	2,628	,105
	Dentro de grupos	9,931	1298	,008		
	Total	9,951	1299			

Rec_LM3	Entre grupos	14,694	1	14,694	6,147	,013
	Dentro de grupos	3102,556	1298	2,390		
	Total	3117,249	1299			
Rec_LM6	Entre grupos	10,209	1	10,209	1,122	,290
	Dentro de grupos	11805,738	1298	9,095		
	Total	11815,947	1299			
Rec_LM9	Entre grupos	3,510	1	3,510	,176	,675
	Dentro de grupos	25917,447	1298	19,967		
	Total	25920,957	1299			
Rec_LM12	Entre grupos	1,088	1	1,088	,036	,850
	Dentro de grupos	39375,573	1298	30,336		
	Total	39376,661	1299			
Coloc3	Entre grupos	1289,333	1	1289,333	1,379	,240
	Dentro de grupos	1213654,168	1298	935,019		
	Total	1214943,500	1299			
Coloc6	Entre grupos	41617,405	1	41617,405	,213	,645
	Dentro de grupos	254012114,679	1298	195695,004		
	Total	254053732,084	1299			
Coloc9	Entre grupos	144699,804	1	144699,804	,261	,610
	Dentro de grupos	720098260,022	1298	554775,239		
	Total	720242959,826	1299			
Coloc12	Entre grupos	365192,175	1	365192,175	,290	,590
	Dentro de grupos	1635432573,702	1298	1259963,462		
	Total	1635797765,877	1299			
LM_Atrasos3	Entre grupos	30358,817	1	30358,817	72,249	,000
	Dentro de grupos	545419,110	1298	420,200		
	Total	575777,927	1299			
LM_Atrasos6	Entre grupos	32540,830	1	32540,830	72,310	,000
	Dentro de grupos	584120,571	1298	450,016		
	Total	616661,401	1299			
LM_Atrasos9	Entre grupos	32687,614	1	32687,614	66,398	,000
	Dentro de grupos	639002,334	1298	492,298		
	Total	671689,949	1299			
LM_Atrasos12	Entre grupos	37832,223	1	37832,223	77,296	,000
	Dentro de grupos	635302,045	1298	489,447		
	Total	673134,268	1299			
LM_Dias_Correctos3	Entre grupos	25,334	1	25,334	136,169	,000
	Dentro de grupos	241,493	1298	,186		
	Total	266,827	1299			
LM_Dias_Correctos6	Entre grupos	28,001	1	28,001	136,675	,000
	Dentro de grupos	265,922	1298	,205		
	Total	293,922	1299			
LM_Dias_Correctos9	Entre grupos	25,590	1	25,590	119,171	,000
	Dentro de grupos	278,721	1298	,215		
	Total	304,311	1299			
LM_Dias_Correctos12	Entre grupos	31,365	1	31,365	144,664	,000
	Dentro de grupos	281,423	1298	,217		
	Total	312,788	1299			
Dif_Acum3_marca	Entre grupos	7,339	1	7,339	48,199	,000
	Dentro de grupos	197,642	1298	,152		
	Total	204,981	1299			
Dif_Acum6_marca	Entre grupos	9,169	1	9,169	50,736	,000
	Dentro de grupos	234,581	1298	,181		
	Total	243,750	1299			
Dif_Acum9_marca	Entre grupos	9,447	1	9,447	49,971	,000

	Dentro de grupos	245,396	1298	,189		
	Total	254,843	1299			
Dif_Acum12_marca	Entre grupos	16,865	1	16,865	86,282	,000
	Dentro de grupos	253,708	1298	,195		
	Total	270,572	1299			

14.8 Matriz de confusión del Árbol de Decisión CHAID exhaustivo con Boosting

Comparando $R-Fuga con Fuga

'Partición'	1_Entrenamiento		2_Comprobación		3_Validación	
Correctos	904	98,69%	174	84,47%	161	90,45%
Erróneos	12	1,31%	32	15,53%	17	9,55%
Total	916		206		178	

Matriz de coincidencias para $R-Fuga (las filas muestran las reales)

'Partición' = 1_Entrenamiento	0	1
0	705	6
1	6	199

'Partición' = 2_Comprobación	0	1
0	138	11
1	21	36

'Partición' = 3_Validación	0	1
0	133	7
1	10	28