



# Técnicas de Minería de Datos para la Retención de Clientes en el Sector Asegurador

Trabajo presentado para el XV Premio de Investigación  
sobre Seguros y Fianzas 2008,

Mtro. Miguel Macías Rodríguez  
"ANIVDELAREV"



COMISIÓN NACIONAL DE  
SEGUROS Y FIANZAS

CNSF

# XV

Premio de Investigación sobre  
Seguros y Fianzas 2008

Tercer Lugar  
Categoría de Seguros

SHCP

## ÍNDICE

RESEÑA .....	1
INTRODUCCIÓN.....	2
CAPÍTULO I. EL SEGURO Y LA INDUSTRIA ASEGURADORA EN MÉXICO.....	5
1.1 Antecedentes y principios del seguro.....	5
1.1.1 Naturaleza del seguro.....	5
1.1.2 Antecedentes históricos.....	5
1.1.3 Principios del seguro.....	6
1.2 Marco jurídico del seguro.....	7
1.2.1 Primeras reglamentaciones.....	7
1.2.2 Códigos reguladores en México (1854 – 1892).....	7
1.2.3 Primera Ley del Seguro en México.....	8
1.2.4 Segunda Ley del Seguro en México.....	8
1.2.5 Ley General de Sociedades de Seguros (1926).....	9
1.2.6 Ley General de Instituciones y Sociedades Mutualistas de Seguros.....	9
1.2.7 Ley sobre el Contrato de Seguros.....	10
1.3 La industria aseguradora en México.....	11
1.3.1 Los inicios de la industria.....	11
1.3.2 La época moderna.....	12
1.3.3 La industria aseguradora en México (1995 – 2001).....	14
1.3.4 La industria aseguradora en México (2002 – 2003).....	17
1.3.5 La industria aseguradora en México (2004 – 2006).....	19
1.3.6 Actualidad de la industria aseguradora en México.....	20
CAPÍTULO II. LA IMPORTANCIA DE LA RETENCIÓN DE CLIENTES.....	22

2.1	El comportamiento de los clientes.....	22
2.1.1	Características básicas.....	22
2.1.2	El concepto de lealtad.....	23
2.1.3	Clasificación de los clientes.....	23
2.1.4	El comportamiento de los clientes en el sector seguros.....	24
2.2	Indicadores básicos para la retención de clientes.....	26
2.2.1	Tasa de retención de clientes.....	26
2.2.2	Tasa de deserción de clientes.....	26
2.2.3	Vida media de la relación de un cliente con la empresa.	27
CAPÍTULO III.	EL RECONOCIMIENTO DE PATRONES A TRAVÉS DE LA MINERÍA DE DATOS.....	29
3.1	Minería de datos.....	29
3.1.1	Antecedentes y necesidades.....	29
3.1.2	Extracción de conocimiento en bases de datos.....	29
3.1.3	Técnicas de Minería de Datos.....	31
	3.1.3.1 Reglas de asociación.....	31
	3.1.3.2 Clasificación y predicción.....	32
	3.1.3.3 Clustering.....	32
3.2	Árboles de decisión.....	33
3.2.1	Características básicas.....	33
3.2.2	El algoritmo ID3.....	34
3.2.3	El algoritmo C4.5.....	36
3.3	Aplicaciones del minado de datos.....	38
3.3.1	Campo de acción.....	38
3.3.2	El caso de la industria aseguradora.....	39
3.4	WEKA.....	40
3.4.1	Software de minería de datos en Java.....	40
3.4.2	WEKA Knowledge Explorer.....	41
3.4.3	El algoritmo J4.8.....	42
CAPÍTULO IV.	LA EMPRESA EN ESTUDIO Y SUS PRODUCTOS AFECTADOS.....	44

4.1	Introducción.....	44
4.1.1	Empresa en estudio.....	44
4.2	Descripción de los productos.....	44
4.2.1	Accidentes Personales en moneda nacional.....	44
4.2.2	Vida Individual en moneda nacional.....	44
4.2.3	Canales de venta.....	45
4.3	Niveles de cancelación.....	45
4.3.1	Histórico de cancelación.....	45
4.3.2	Tasa de retención y vida media.....	45
CAPÍTULO V.    PREPARACIÓN DE LOS DATOS.....		47
5.1	Proceso de extracción de conocimiento.....	47
5.2	Selección del objetivo.....	47
5.2.1	Problema observado.....	47
5.2.2	Meta del proyecto.....	47
5.2.3	Fuente de datos.....	47
5.2.4	Muestra inicial.....	48
5.3	Preproceso de datos.....	48
5.3.1	Introducción.....	48
5.3.2	Atributos preprocesados.....	49
5.3.2.1	Producto.....	49
5.3.2.2	Edad de emisión.....	49
5.3.2.3	Sexo del asegurado.....	50
5.3.2.4	Estado civil.....	50
5.3.2.5	Entidad federativa.....	51
5.3.2.6	Código postal.....	51
5.3.2.7	Suma asegurada.....	52
5.3.2.8	Prima de recibo.....	52
5.3.2.9	Forma de pago.....	53
5.3.2.10	Día de facturación.....	53
5.3.2.11	Punto de venta.....	53
5.3.2.12	Tipo de cuenta.....	53

5.3.2.13	Meses pagados.....	54
5.4	Transformación de datos.....	55
5.4.1	Atributos finales.....	55
5.4.2	Formato ARFF.....	55
5.4.3	Seguros.arff.....	56
CAPÍTULO VI.	MODELO DE MINERÍA DE DATOS.....	57
6.1	Minado de datos.....	57
6.1.1	Introducción.....	57
6.1.2	Aplicación del algoritmo J4.8.....	57
6.1.2.1	Parámetro «binarySplits».....	59
6.1.2.2	Parámetro «confidenceFactor».....	59
6.1.2.3	Parámetro «debug».....	59
6.1.2.4	Parámetro «minNumObj».....	59
6.1.2.5	Parámetro «numFolds».....	59
6.1.2.6	Parámetro «reducedErrorPruning».....	59
6.1.2.7	Parámetro «saveInstanceData».....	59
6.1.2.8	Parámetro «seed».....	59
6.1.2.9	Parámetro «subtreeRaising».....	59
6.1.2.10	Parámetro «unpruned».....	59
6.1.2.11	Parámetro «useLaplace».....	60
6.2	Interpretación de resultados.....	60
6.2.1	Árbol de decisión obtenido.....	60
6.2.2	Coefficiente de Kappa.....	61
6.2.3	Detalle de precisión por clase.....	61
CONCLUSIONES.....		64
BIBLIOGRAFÍA.....		66
ANEXO A. REGLAS DE DECISIÓN.....		69

## LISTA DE GRÁFICAS

Gráfica 1.1	Participación de capital extranjero en la Prima Directa 1995-2001.....	15
Gráfica 1.2	Prima Directa y PIB (Crecimiento Real) 1995-2001.....	17
Gráfica 2.1	Modelo simple de comportamiento de compra.....	23
Gráfica 2.2	Clasificación de clientes.....	24
Gráfica 2.3	Tasa de retención de clientes.....	26
Gráfica 2.4	Tasa de deserción de clientes.....	27
Gráfica 2.5	Vida media de la relación de un cliente con la empresa.....	27
Gráfica 3.1	Pasos que componen al proceso de extracción de conocimiento en bases de datos (KDD).....	30
Gráfica 3.2	Clustering de datos bancarios.....	32
Gráfica 3.3	Árbol de decisión para la compra de computadoras.....	33
Gráfica 3.4	Una estructura de árbol para los objetos en C.....	35
Gráfica 3.5	Elevación de subárboles.....	38
Gráfica 3.6	WEKA (Waikato Environment for Knowledge Analysis).....	40
Gráfica 3.7	WEKA Knowledge Explorer.....	41
Gráfica 4.1	Histórico de cancelación en Accidentes Personales en M.N.....	45
Gráfica 4.2	Histórico de cancelación en Vida Individual en M.N.....	46
Gráfica 5.1	Atributo "Producto".....	49
Gráfica 5.2	Atributo "Edad de emisión".....	49
Gráfica 5.3	Atributo "Sexo del asegurado".....	50
Gráfica 5.4	Atributo "Estado civil".....	50
Gráfica 5.5	Atributo "Zona geográfica de venta" (Antes "Entidad federativa").....	52
Gráfica 5.6	Atributo "Prima mensualizada" (Antes "Prima de recibo").....	52
Gráfica 5.7	Atributo "Día de facturación".....	53
Gráfica 5.8	Atributo "Tipo de cuenta".....	54
Gráfica 5.9	Atributo "Trimestres de cancelación" (Antes "Meses pagados").....	54

Gráfica 5.10	Archivo Seguros.arff.....	56
Gráfica 6.1	Carga de datos en WEKA.....	57
Gráfica 6.2	Clasificador J4.8.....	58
Gráfica 6.3	Parámetros del algoritmo J4.8.....	58
Gráfica 6.4	Árbol de decisión obtenido.....	60

## LISTA DE CUADROS

Cuadro 1.1	Número de compañías autorizadas para operar en el Sector Asegurador 1995-2001.....	15
Cuadro 1.2	Instituciones de Seguros integradas a Grupos Financieros 1995-2001.....	16
Cuadro 1.3	Número de empresas autorizadas para operar en el Sector Asegurador 2002 – 2003.....	18
Cuadro 1.4	Resultados de Operación 2002 – 2003.....	19
Cuadro 1.5	Número de empresas autorizadas para operar en el Sector Asegurador 2004 – 2006.....	19
Cuadro 1.6	Tamaño potencial del mercado.....	21
Cuadro 2.1	Fases de ciclo de vida y la contratación de seguros.....	25
Cuadro 5.1	Atributos seleccionados en la primera muestra de datos.....	48
Cuadro 5.2	Zonas geográficas de venta.....	51
Cuadro 5.3	Atributos finales.....	55
Cuadro 6.1	Evaluación en los datos de prueba.....	61
Cuadro 6.2	Detalle de precisión por clase.....	61
Cuadro 6.3	Matriz de confusión.....	63

## RESEÑA

La Minería de Datos (*Data Mining*) es un grupo de técnicas provenientes de la Estadística, la Inteligencia Artificial y la Biología, entre otras disciplinas, utilizadas en la exploración y análisis de grandes volúmenes de datos, cuya finalidad es la de encontrar patrones de interés que aporten conocimiento. Las aplicaciones de estas técnicas tienen un vasto campo de acción en áreas como la Astronomía (*clustering* y clasificación de cuerpos celestes), la Medicina (asociación de síntomas), la Biología Molecular (predicción de sustancias cancerígenas, genoma humano, etc.), la Mercadotecnia (clientes potenciales y selección de productos) y la Manufactura (diagnóstico de fallas), por mencionar solo algunas.

En el caso particular de la industria aseguradora, países como Estados Unidos y Canadá se han dado cuenta del enorme potencial de las técnicas de Minería de Datos para lograr beneficios como el crecimiento en el mercado o el incremento en la rentabilidad. En este sentido, existe un gran número de investigaciones que apoyan el uso del *Data Mining* dentro de la industria del seguro. En contraste, en el sector asegurador latinoamericano y en particular en el mexicano, existen muy pocos estudios formales que demuestran la factibilidad del uso de las técnicas de *Data Mining* dentro del ámbito de seguros.

El **objetivo** del presente trabajo de investigación es mostrar la factibilidad del uso de la Minería de Datos para construir modelos predictivos que apoyen los programas de retención de clientes en empresas aseguradoras.

La **importancia** radica en la aportación que puede dar el *Data Mining* a los programas de retención de clientes del sector asegurador, los cuales a su vez, pueden impactar directamente la rentabilidad de las compañías de seguros.

# INTRODUCCIÓN

## Antecedentes

A mediados de la década de los 90's el sector asegurador mexicano experimentó importantes cambios en su estructura; el número de participantes en la industria se incrementó de manera significativa con respecto a años anteriores, de tal forma que hacia finales de 1996 había alrededor de 60 instituciones aseguradoras en el país. La entrada de nuevos participantes se explica principalmente por las mejores condiciones para la inversión extranjera derivadas de la modificación de la Ley General de Instituciones y Sociedades Mutualistas de Seguros (LGISMS) y de la firma del Tratado de Libre Comercio con América del Norte (TLCAN).

Desde entonces, el crecimiento que ha tenido este sector ha sido muy importante, llegando a ser en los últimos años superior al de inflación e inclusive, superior al crecimiento del PIB. No obstante este dinamismo, una constante que se presenta en diversas instituciones de aseguradoras del país es el alto porcentaje en la cancelación de pólizas, alcanzando en algunos productos niveles de cancelación del 80% con respecto al total emitido. Al mismo tiempo, cerca del 70% de las cancelaciones se realizan durante el primer año de vigencia del seguro, lo cual significa un alto costo para las empresas.

En función a lo expuesto anteriormente, surge la necesidad en las instituciones de Seguros del país, de contar con modelos predictivos que apoyen y fortalezcan sus programas de retención de clientes.

## Objetivo general

El objetivo general del presente trabajo es mostrar la factibilidad del uso de la Minería de Datos para construir modelos predictivos que apoyen los programas de retención de clientes en instituciones de seguros.

## Objetivos específicos

- Construir un modelo de Minería de Datos para una de las principales empresas de Seguros del país el cual clasifique a sus clientes en función al tiempo de permanencia con su contrato de seguros.
- Identificar que tipo de clientes tienden a permanecer por más tiempo con su contrato de seguros así como los clientes que tienden a cancelar de manera anticipada sus contratos.

## Alcances y limitaciones

El presente trabajo de investigación ha sido enfocado hacia el sector asegurador mexicano; en particular, se ha elegido a una importante empresa aseguradora del país para construir el modelo de Minería de Datos.

Por otra parte, el modelo predictivo que se pretende construir se basa en la técnica de Minería de Datos conocida como Inducción de Árboles de Decisión, la cuál, es la más popular dentro de las técnicas de clasificación.

## **Organización del trabajo**

El presente trabajo esta organizado en seis capítulos, los cuales son detallados a continuación:

En el Capítulo 1 se da un amplio panorama del seguro y la industria aseguradora en México, desde los conceptos básicos del seguro y sus antecedentes históricos, hasta la actualidad del sector asegurador mexicano.

El Capítulo 2 contiene conceptos importantes de lealtad y retención de clientes, los cuales son utilizados en el Capítulo 4 para describir el entorno de la empresa en estudio. Además, en este capítulo se incluye una pequeña sección sobre el comportamiento de clientes en el mercado de seguros, la cual se considera importante para entender cómo es la dinámica de venta dentro de la industria aseguradora.

En el Capítulo 3 se ofrece un panorama básico de la Minería de Datos y las técnicas que la componen. En este capítulo también se muestran las ideas básicas de los algoritmos ID3 y C4.5, los cuales se utilizan en la mayoría de los paquetes de software que inducen árboles de decisión. Finalmente, se incluye un apartado en donde se mencionan las aplicaciones más comunes de la Minería de Datos, así como los casos en los que estas técnicas han sido empleadas en beneficio de la industria aseguradora de otros países.

En el Capítulo 4 se presenta una breve descripción de la empresa en estudio, la cual sirvió como base para construir el modelo predictivo de Minería de Datos. En esta sección también se mencionan las características principales de los productos de seguros que fueron utilizados para construir el modelo predictivo. Dentro de estas características principales destacan el tipo de producto, la forma de pago, el canal de distribución, así como los niveles de cancelación asociados al producto.

El Capítulo 5 contiene el proceso de preparación de datos necesario para construir el modelo de *data mining*. En él se muestra la selección del objetivo, el preproceso de datos y la transformación de los mismos. Finalmente, se presenta el archivo final de datos, en formato ARFF, el cual sirvió para la construcción del modelo.

El Capítulo 6 contiene el proceso de minado de datos propiamente dicho. En el se encuentran los parámetros utilizados en el algoritmo J4.8 de WEKA, así como los resultados de la inducción. Al final del capítulo se ofrece la interpretación de dichos resultados.

Por último, se encuentran las conclusiones derivadas del presente trabajo, así como la información bibliográfica utilizada en el mismo. La parte final de la investigación esta compuesta por un anexo que contiene sesenta y siete reglas de decisión, las cuales se proponen como modelo predictivo para el apoyo de programas de retención de clientes en la empresa de seguros en estudio.

*“Si buscas resultados distintos,  
no hagas siempre lo mismo”.*

*Albert Einstein*

## CAPÍTULO I. EL SEGURO Y LA INDUSTRIA ASEGURADORA EN MÉXICO

---

### 1.1 Antecedentes y principios del seguro

#### 1.1.1 Naturaleza del seguro

Todos los días y a todas horas el ser humano se enfrenta a situaciones riesgosas que pueden ocasionarle pérdidas económicas imprevistas. El riesgo de fallecer, sufrir un accidente o una enfermedad o de perder algún bien material, es algo latente que está fuera de los límites del control humano.

Ante tal situación, el seguro surge como una solución sustancial, que si bien no elimina el riesgo de sufrir alguna pérdida, sí garantiza al que la sufre directamente, que no tendrá que soportar la carga económica por sí solo. En este sentido, el seguro reduce la incertidumbre y el riesgo a un grado de seguridad y protección relativas.<sup>1</sup>

Diversos autores han definido al seguro de distintas formas, algunos hacen hincapié en el principio de solidaridad que lo sustenta, otros mencionan que el seguro se fundamenta en un principio de contraprestación.

De acuerdo con Riegel (1980), "el seguro puede ser definido como una combinación de individuos que aceptan pagar pequeñas contribuciones para remunerar a quienes sufran pérdidas que pueden ser previstas y calculadas".

Palacios (1996) señala por su parte, que el seguro como institución, es un sistema de protección del hombre y de su patrimonio frente a diversos hechos que amenazan su integridad, su vida, su interés y su propiedad. Como contrato, el seguro es el convenio entre dos partes, la compañía o entidad aseguradora y el asegurado o contratante, mediante la cual la primera se compromete a cubrir económicamente la pérdida o daño que el segundo puede sufrir durante la vigencia del contrato.

#### 1.1.2 Antecedentes históricos

El origen y evolución del seguro son paralelos a la evolución y desarrollo del comercio. La mayoría de las modalidades del seguro nacieron como mecanismos de protección de las actividades comerciales.

Minzoni (1994) señala que en las culturas antiguas como La India, Egipto, Grecia y Roma, aunque el seguro no fue conocido como una institución definida con caracteres propios, se

---

<sup>1</sup> El *riesgo* es la posibilidad de tener una pérdida o un daño económicamente desfavorable. Se considera además como el "antecedente primigenio y fundante de la institución del seguro". Por otro lado, la *incertidumbre* es la falta de conocimiento o de seguridad sobre la ocurrencia de algo. Para que un riesgo pueda ser considerado como asegurable, "se exige la incertidumbre del suceso, en el sentido de que su ocurrencia exacta no se conozca [...] aunque se sepa que ha de ocurrir". Quiñones, C. E. (1999). La Individualización del Riesgo en el Contrato de Seguro Marítimo. *Revista de Derecho*, 11, 75-92.

encuentran vestigios que hacen suponer que el espíritu de la prevención ya tenía algún arraigo. Como ejemplo de esto tenemos a las Leyes de Rodas, las cuales fueron las bases del Derecho mercantil de Atenas. En una parte de dichas Leyes se habla «de las obligaciones de los cargadores de contribuir a la indemnización de los graves daños causados en perjuicio común en caso de tempestad o rescate de buque apresado por enemigos o piratas».<sup>2</sup>

Sin embargo, no fue sino hasta la Edad Media cuando se comenzaron a sentar las bases del seguro tal y como se conoce hoy en día. El gran desarrollo de la navegación en la época de la Europa Medieval, contribuyó con el surgimiento del seguro marítimo.

De esta forma, “el seguro marítimo –aún sin reglamentación– fue la primera actividad en la que se desarrolló el seguro y el documento histórico que puede considerarse como la primera póliza de seguro marítimo, lleva fecha 23 de octubre de 1347”.<sup>3</sup>

### 1.1.3 Principios del seguro

Los principios del seguro son los fundamentos doctrinarios en los que se basa la práctica del seguro. Palacios (1996) reconoce seis principios básicos:

- a) *Principio de mutualidad.*<sup>4</sup> *Las pérdidas de pocos son cubiertas por la contribución de muchos.* Según este principio, que parece ser el primero o más antiguo, las primas pagadas por una colectividad de asegurados sirven para reponer, reparar o indemnizar las pérdidas de quienes sufran siniestros.
- b) *Principio de buena fe.* Llamada la máxima buena fe, según la cual tanto el asegurado como el asegurador tienen que obrar solo con la verdad; el primero describiendo o exponiendo el riesgo tal cual es, sin eludir ni ocultar nada, a fin de que el asegurador – al apreciarlo debidamente– lo cubra con equidad. Es obligación también del asegurador obrar de buena fe en la atención del siniestro.
- c) *Principio de indemnización.* *El seguro no es para ganar, el seguro es para no perder.* Según el contrato de seguro una pérdida o daño que ocurra obliga al asegurador a resarcirlo o indemnizarlo a fin de poner el bien en las mismas condiciones en que se encontraba inmediatamente antes del siniestro, pues el seguro no puede ser objeto de lucro. Lo que debe esperar el asegurado es no perder.
- d) *Principio de interés asegurable.* Según este principio, no puede asegurarse ningún evento incierto que no cause daño o pérdida si tal evento ocurre (la apuesta o lotería es un evento incierto, pero no causa daño, entonces no hay interés asegurable). Los seguros de daños son de estricta indemnización y es preciso que, para su validez, en el momento de su contratación, exista un interés asegurado, es decir, que el asegurado, si ocurre el siniestro, ha de sufrir un daño que es lo único que justifica la percepción de la indemnización.
- e) *Principio de contribución.* En el caso de que un mismo bien tenga dos o más coberturas, en aplicación de los principios de indemnización y del interés asegurable, el

<sup>2</sup> En Minzoni, A. (1994). *Crónica de Doscientos Años del Seguro en México*. México: CNSF, también se hace mención de algunos indicios del seguro en el México prehispánico, ya que algunos pueblos, principalmente los mayas y chichimecas, manejaban conceptos como la *indemnización* y el *pago de deudas*.

<sup>3</sup> Minzoni, A. *op. cit.*, p. 22.

<sup>4</sup> También se conoce como *principio de solidaridad*.

resarcimiento de la pérdida o daño será efectuado por todos los aseguradores en proporción a los capitales por cada uno contratados, de modo que la contribución total no sobrepase la máxima pérdida sufrida por el asegurado.

- f) *Principio de subrogación.* Mediante este principio, el asegurador tiene el derecho de repetir contra un tercero responsable, si lo hay, el importe de la indemnización pagada, es decir, se subroga en los derechos del asegurado y recobra el valor de los daños del causante responsable. Por la aplicación de este principio, el asegurador es también dueño de los recursos de robo, restos de un salvamento o producto de la venta de estos.

## 1.2 Marco jurídico del seguro

### 1.2.1 Primeras reglamentaciones

En el año de 1345 surgió la primera reglamentación que reguló de alguna forma la actividad aseguradora, en específico el seguro marítimo. Este instrumento jurídico fue la Ordenanza de Barcelona, “dedicada a extirpar fraudes, daños, discusiones y debates que podían tener lugar en dicha ciudad en ocasión de los seguros sobre buques, mercancías, efectos y bienes y también en beneficio de los asegurados y aseguradores”.<sup>5</sup>

Minzoni (1994) señala que una serie de estatutos y ordenanzas siguieron a esta ordenanza, entre ellos la Recopilación de las Indias, una especie de código redactado para la América española, que tuvo como objeto la reglamentación del Derecho marítimo y mercante. Así mismo, menciona que una legislación completa sobre el seguro marítimo fue otorgada por Felipe II en 1536, la cual trataba de «Aseguradores, Riesgos y Seguros de la Carrera de las Indias». Esas disposiciones fueron perfeccionadas por las Ordenanzas de Bilbao de 1725.

### 1.2.2 Códigos reguladores en México (1854 – 1892)

Las Ordenanzas de Bilbao regularon la actividad del seguro en México hasta 1854, año en que se instituyó en el México independiente el primero Código de Comercio. Dicho código fue preparado por el celebre jurista Teodosio Lares, y se le denominó Código Lares. “Su vigencia dependió de la actitud de los siguientes gobiernos, se vio derogado y posteriormente restablecido el 15 de julio de 1863.”<sup>6</sup>

El Código Civil de 1870, promulgado por el Presidente Juárez, reforma y amplía el de Lares y dentro de seis artículos define y norma el «contrato de seguro», al cual se le define como «aquel por el cual una de las partes se obliga mediante cierto precio, a responder o indemnizar a la otra del daño que podrían causarle ciertos casos fortuitos a que está expuesta». Se aclara también que el contrato de seguro es nulo si no se otorgan en escritura pública y se establece que cualquier persona o compañía capaz de obligarse puede fungir como asegurador. En lo referente al «precio», se menciona que éste puede ser fijado libremente por las partes y puede pagarse de una sola vez o en plazos.<sup>7</sup>

---

<sup>5</sup> Minzoni, A. *op. cit.*, p. 22.

<sup>6</sup> Ledesma, J. (1981). Panorama del Derecho Mexicano en el Siglo XIX. *Jurídica*, 13 (11), 625-652.

<sup>7</sup> Minzoni, A. *op. cit.*, p. 42.

Posteriormente, en los Códigos de 1884 y 1889 aparecieron algunos aspectos que especificaban el carácter mercantil del contrato de seguro, y no fue sino hasta 1892, durante la época del Porfiriato, cuando se promulgo por primera vez una Ley del Seguro en México.

### 1.2.3 Primera Ley del Seguro en México

“En la exposición de motivos del proyecto (de Ley del Seguro,) presentado a las Cámaras en 1892, para el funcionamiento de las Compañías (de Seguros), se optaba por una libertad absoluta para su organización sin necesidad de permiso especial ni concesión de autoridad”.<sup>8</sup>

No obstante, la naciente Ley preveía un incipiente control por parte de las autoridades, sobre todo en el aspecto de normar las obligaciones de los aseguradores con sus asegurados.

Minzoni (1994) señala que la mencionada «Ley sobre Compañías de Seguros» refería la necesidad de fijar las prescripciones fundamentales a que tenían que someterse las sociedades nacionales y extranjeras de Seguros de Vida, contra incendio y otros riesgos, que en todas partes habían sido objeto de una legislación más o menos rigurosa, y cuyo propósito era garantizar los intereses de las personas que con ellas contrataban; además, era necesario modificar los impuestos que habían gravado con desigualdad y sin sistema las diversas operaciones de las expresadas compañías.

Con la promulgación de esta Ley, el Estado Mexicano decide por primera vez intervenir en la actividad aseguradora del país. Su importancia histórica radica en establecer los primeros “lineamientos y reglas que permitieron un principio de sano desarrollo del seguro”.<sup>9</sup>

### 1.2.4 Segunda Ley del Seguro en México

“En la exposición de motivos de 1910 para una Ley sobre Compañías de Seguros de Vida, se habla ya de la rapidez con que esta actividad se ha desarrollado y se plantea como una necesidad inaplazable la de vigilar que se llenen los fines científicos del seguro. Es decir, que al aumentar tan considerablemente el caudal recogido por las compañías de seguros, aparece la preocupación de que estas sumas pudieran no ser aplicadas a los fines precisos, propios del seguro manejado científicamente y así no se ofrecieran amplias garantías a los asegurados”.<sup>10</sup>

Es así como el 25 de mayo de 1910 se promulga la Ley Relativa de la Organización de Compañías de Seguros sobre la Vida, “la cual regulaba por primera vez, en forma moderna, el seguro de vida contratado en México”.<sup>11</sup>

Un aspecto importante que marca esta Ley, es la prohibición a particulares para realizar la practica del seguro, dejándose dicha actividad solamente a Sociedades Anónimas o Cooperativas autorizadas por el Estado. También “se determinan ya las reservas que han de constituirse y el régimen de inversión de las mismas y para compañías extranjeras se fija la obligación de establecer depósitos en manos del Estado, que garanticen su gestión en el país. [...] (Se) obliga además a que las Compañías produzcan anualmente amplios informes de sus cuentas y hagan todas sus publicaciones convenientes, para que puedan apreciarse si los bienes que poseen responden a las obligaciones asumidas. Se trata, en suma, de la primera

<sup>8</sup> Elizondo, J. (1996). *Vigilancia y Diagnóstico en Aseguradoras de Daños*. México: IMCP.

<sup>9</sup> Minzoni, A. *op. cit.*, p. 46.

<sup>10</sup> Elizondo, J. *op. cit.*, p. 264.

<sup>11</sup> Minzoni, A. *op. cit.*, p. 55.

Ley especializada en seguros, como base para un desarrollo más sano de esta actividad comercial”.<sup>12</sup>

### 1.2.5 Ley General de Sociedades de Seguros (1926)

“(A) casi dos décadas desde que había sido promulgada la segunda Ley del Seguro, la cual se relacionaba con los seguros de vida, [...] (se) hacía evidente la necesidad de reformar las leyes de seguros, a fin de que abarcaran toda la actividad aseguradora.

Se debe al Presidente Calles [...] la Ley General de Sociedades de Seguros, la cuál constituyó el primer documento legal y completo que regulaba la actividad aseguradora en todos los ramos en los que una institución de seguros pudiera operar.”<sup>13</sup>

“En su exposición de motivos plantea como una finalidad importante, la de establecer la vigilancia del Estado sobre los actos de las sociedades de seguros y sus contratos, para garantizar principalmente a los asegurados, ya que estos, difícilmente podrán juzgar sobre aspectos técnicos, que exigen conocimiento muy especializados. [...] (Es así como en 1926 se cuenta) con una Ley que en sus nueve capítulos se ocupa de la creación, el desarrollo y la desaparición de las Compañías de Seguros, tanto de vida, accidentes, riesgos profesionales y daños, ya sean nacionales o extranjeras”.<sup>14</sup>

### 1.2.6 Ley General de Instituciones y Sociedades Mutualistas de Seguros <sup>15</sup>

La evolución de la legislación mexicana en materia seguros continuó durante los años posteriores a 1926. Diversos decretos y reglamentos fueron promulgados como resultado de la preocupación del Estado por proteger los intereses del público.<sup>16</sup>

El 31 de agosto de 1935, bajo la Presidencia del General Lázaro Cárdenas, fue promulgada la Ley General de Instituciones de Seguros. Esta Ley, con algunas modificaciones, fue la que reguló la actividad aseguradora hasta el año de 1980.

De acuerdo con Minzoni (1994) dicha Ley significó un cambio completo de la legislación sobre materia de seguros. Tomó en cuenta los artículos de la Ley de 1926 y del decreto de 1931 y de su reglamento, procurando una mejor sistematización, y ampliando y adicionando otras disposiciones nuevas, relativas a la inversión de las reservas en México, las cuales fueron el motivo por el que las aseguradoras extranjeras se retiraron del mercado. Así, una medida de política económica general fue el punto de partida del desarrollo del seguro mexicano.

En las modificaciones de 1981 a la Ley General de Instituciones de Seguros, se buscó implantar una serie de medidas “tendientes a promover un desarrollo equilibrado; lograr controles más eficaces por parte de las autoridades; prever mejores sistemas y mecanismos para una mayor sanidad en las instituciones; flexibilizar el marco operativo para que las instituciones [...] (mostraran) un mejor servicio al público, (así como) obtener adecuados rendimientos en la

---

<sup>12</sup> Elizondo, J. *op. cit.*, p. 264.

<sup>13</sup> Minzoni, A. *op. cit.*, p. 64.

<sup>14</sup> Elizondo, J. *op. cit.*, p. 265.

<sup>15</sup> Denominación reformada por el decreto del 3 de enero de 1990.

<sup>16</sup> Los aspectos que preocupaban más al Estado mexicano eran la suficiencia en primas y la constitución de reservas de previsión.

inversión de sus reservas y, en consecuencia, [...] (pudiera) reducirse el costo de los seguros,”.<sup>17</sup>

Dichas modificaciones regularon la actividad del seguro durante toda la década de los 80's, tiempo durante el cual el Estado mantuvo una fuerte intervención en todas las fases de la operación del seguro.

Sin embargo, ante los propósitos de cambio y modernización que surgieron en el país a finales de los años ochenta, “era necesario que la legislación del seguro en vigor, [...] sufriera cambios que permitieran a la industria del seguro del país una mayor competitividad en lo interno y en lo externo, [...] por ello, siguiendo los criterios de liberalización y desregulación ya aplicados a otros sectores de la economía, con fecha 3 de enero de 1990 el Diario Oficial de la Federación publicó un decreto del Ejecutivo federal promulgando la Ley General de Instituciones y Sociedades Mutualistas de Seguros, mismo que entró en vigor el día siguiente de su publicación”.<sup>18</sup>

“(Posteriormente,) las reglas ya vigentes del Tratado de Libre Comercio de América del Norte, dieron pie para adecuar el artículo 29 de la Ley General de Instituciones y Sociedades Mutualistas de Seguros en lo relativo al origen mexicano o extranjero de los accionistas, [...] (estableciendo que) pueden ser autorizadas instituciones mayoritariamente mexicanas y mayoritariamente extranjeras, en este último caso se consideran como filiales de institución financiera del exterior”.<sup>19</sup>

“Asimismo, [...] (esta) Ley contiene una serie de disposiciones que se relacionan con inversiones, con la seguridad de la operación, con la suficiencia de primas, con las diferentes reservas a constituir; con retención de riesgos y (la) cesión en reaseguro, con (el) registro de reaseguradoras extranjeras [...], (con el) traspaso de carteras, (con la) vigilancia, etcétera. Finalmente, los artículos del 80 al 107 se refieren a las (instituciones) mutualistas”.<sup>20</sup>

### 1.2.7 Ley sobre el Contrato de Seguros<sup>21</sup>

El gobierno del General Lázaro Cárdenas se ocupó de promulgar la primera Ley referente al contrato de seguros, que fue publicada en el Diario Oficial de la Federación el 31 de agosto de 1935, a cinco días de haberse publicado la Ley de Instituciones de Seguros.

Contiene 196 artículos que norman de una manera amplia y clara todo lo relacionado con el contrato de seguros, sus alcances, obligaciones de la empresa aseguradora y de los asegurados. Inspirada en leyes de otros países europeos, en donde el seguro ya tenía algún siglo de operar, y también en los estatutos de algunas compañías locales establecidas durante la época de Maximiliano, es una Ley que desde el ya lejano año de 1935 se encuentra en vigor en México, sólo con algunas muy modestas modificaciones.

La Ley, sin definir lo que es contrato de seguro, en su artículo 1º establece:

*Por el contrato de seguro, la empresa aseguradora se obliga, mediante una prima, a resarcir un daño o a pagar una suma de dinero al verificarse*

<sup>17</sup> Elizondo, J. *op. cit.*, p. 266.

<sup>18</sup> Minzoni, A. *op. cit.*, p. 136.

<sup>19</sup> Elizondo, J. *op. cit.*, p. 268-269.

<sup>20</sup> Minzoni, A. *op. cit.*, p. 138.

<sup>21</sup> *Ibid.* p. 83.

*la eventualidad prevista por el contrato.*

A continuación, entre las obligaciones del todavía no asegurado, la Ley establece:

*El proponente estará obligado a declarar por escrito a la empresa aseguradora, de acuerdo con el cuestionario relativo, todos los hechos importantes para la apreciación del riesgo que puedan influir en las condiciones convenidas, tales como las conozca o deba conocer en el momento de la celebración del contrato.*

Continúa el texto legal:

*Cualquier omisión o inexacta declaración de los hechos a que se refiere el artículo anterior, facultará a la empresa aseguradora para considerar rescindido de pleno derecho el contrato aunque no haya influido en la realización del siniestro.*

En lo que respecta a la aceptación de un riesgo, la Ley es muy estricta:

*El contrato de seguro se perfecciona desde el momento en que el proponente tuviere conocimiento de la aceptación de la oferta. [Por lo que] la empresa aseguradora no podrá eludir la responsabilidad por la realización del riesgo, por medio de cláusulas en que convenga que el seguro no estará en vigor sino después del pago de la primera prima o fracción de ella.*

En el capítulo de «prescripciones», se establece que «todas las acciones que se deriven de un contrato de seguro prescribirán en dos años, contados desde la fecha del acontecimiento que les dio origen».

Sobre los contratos de seguros contra los daños, la Ley establece que: «Todo interés económico que una persona tenga en que no se produzca un siniestro, podrá ser objeto de contrato de seguro contra los daños», aclarando lo que debe entenderse por seguro contra incendio, de provechos esperados, de transporte terrestre, seguro contra responsabilidades y una serie de disposiciones especiales relacionadas con el contrato de seguros sobre las personas.

Con relación a este tipo de seguros, la Ley ordena: «El contrato de seguros sobre personas comprende todos los riesgos que puedan afectar a la persona del asegurado en su existencia (vida), integridad personal (accidentes) salud o vigor vital (enfermedades)».

### **1.3 La industria aseguradora en México**

#### **1.3.1 Los inicios de la industria**

“El seguro en México, como en otros muchos lugares del mundo, habría de nacer en el mar, y en este caso, en el puerto de Veracruz, ciudad que a fines del siglo XVIII gozaba de gran prosperidad comercial [...]. Fue en el año de 1789, cuando se constituyó la primera compañía de seguros en ese puerto, que se denominó Compañía de Seguros Marítimos de Nueva España,

con el propósito de cubrir los riesgos de los que se denominaba en España como La Carrera de las Indias”.<sup>22</sup>

Sin embargo, se considera que el inicio formal del sector asegurador mexicano se remonta a finales del siglo XIX, cuando se regula en el Código Civil el contrato de seguro, y posteriormente, se promulga la primera Ley que regía a las compañías de seguros de esa época.<sup>23</sup>

De esta forma, la actividad aseguradora en el México de finales del siglo XIX comenzó a tener un desarrollo muy significativo, destacándose por su número, las compañías que operaban el seguro contra incendio, las cuales en su mayoría eran instituciones extranjeras.

Para los primeros años del siglo XX, las compañías que operaban el seguro de vida habían crecido considerablemente, por esta razón, se hacía cada vez más necesaria una Ley que regulara dicha actividad.

Minzoni (1994) señala que en la «Memoria de los Sesenta Años de La Latinoamericana», se hace patente la situación que prevalecía en aquella época:

*“En un ambiente de ignorancia casi absoluta, de los beneficios del seguro de vida, dentro de un sistema pobre de economía nacional, nació La Latinoamericana en el año de 1906, y sus fundadores, considerando que era oportuno el momento para organizar una empresa de seguros de vida y, sin más miras que la de buscar la protección de las clases desamparadas, establecieron esta empresa el 30 de abril del mencionado año.”*

Ante tales condiciones “se promulga en 1910 una Ley especializada que establece que el seguro no lo podrán practicar particulares, sino únicamente las sociedades que sean autorizadas por la Ley y que tendrán la forma de Sociedades Anónimas o Cooperativas”.<sup>24</sup>

A partir de entonces, el desarrollo de la industria aseguradora mexicana ha ido en constante ascenso, lo cual ha llevado a los diferentes gobiernos posrevolucionarios, a reglamentar cada vez más cuidadosamente la actividad aseguradora.

### 1.3.2 La época moderna

Con la promulgación de la Ley General de Sociedades de Seguros, durante el gobierno del Presidente Plutarco Elías Calles, se reguló por primera vez de manera completa la actividad aseguradora. Para entonces operaban ya en el país un considerable número de compañías de seguros de vida, accidentes, riesgos profesionales y daños, tanto nacionales como extranjeras.

Entre los considerandos de la Ley General de Instituciones de Seguros del 31 de agosto de 1935, que con algunas modificaciones es la que regía esta actividad hasta 1980, ya se tomaba en cuenta que las compañías extranjeras no invertían en el país ni siquiera una parte de las

<sup>22</sup> Minzoni, A. *op. cit.*, p. 26.

<sup>23</sup> Según datos de la Asociación Mexicana de Instituciones de Seguros, las aseguradoras más antiguas de México, cuya existencia se extiende hasta nuestros días, se fundaron en aquellos años.

<sup>24</sup> Elizondo, J. *op. cit.*, p. 264.

primas que cobraban y de manera visionaria, se estableció en el cuerpo de la Ley la obligación de invertir una buena parte de esa recaudación: hasta por el importe de las reservas técnicas.<sup>25</sup>

Al ser promulgada esta Ley “las compañías extranjeras amenazaron con retirarse del mercado nacional de seguros; (pero) México no se arredró y, vencido el plazo que les concedió [...] para obtener la nueva autorización en términos de la Ley, [...] quedaron eliminadas las que no quisieron acogerse a la nueva legislación.

Documentos de diferente índole, señalan que a finales de 1935 operaban en el país 66 instituciones aseguradoras, de las cuales 52 eran extranjeras, en su mayoría inglesas (27), ocho alemanas y las restantes pertenecientes a Italia, Suiza, Canadá, Francia, España, Holanda y los Estados Unidos de Norteamérica”.<sup>26</sup>

En opinión de Elizondo (1996), “la mexicanización de la actividad aseguradora [...] debe calificarse como un acto visionario de los legisladores que comprendían el desarrollo económico que llegaría a alcanzar nuestro país, como pronto se constató, quedando en el país los beneficios de esta actividad que no sólo aprovechó a sus accionistas, sino que ha constituido una fuente importante de tributación fiscal”.

Durante las décadas posteriores, el entorno de la actividad aseguradora se vio enriquecido con diversos decretos que buscaban estimular la participación de las empresas aseguradoras en el proceso de desarrollo económico del país, fomentando la canalización de sus recursos hacia las actividades productivas y de beneficio social.

En enero de 1981 la Ley General de Instituciones de Seguros fue substancialmente modificada, con ello, se buscaba promover un desarrollo equilibrado en el sector asegurador. Elizondo (1996) destaca las principales reformas en cuatro rubros:

*1º. Reformas a la estructura del sector asegurador.- Se suprimen las sucursales de compañías extranjeras de seguros. Se dan las bases para regular la actividad de los agentes de seguros, de los ajustadores de seguros y de las distintas figuras que intervienen en las operaciones de reaseguro.*

*2º. Reformas al régimen de operación de las empresas aseguradoras.- Se pretende regular la operación de las instituciones, señalando en forma limitativa las operaciones que pueden realizar, en su carácter de aseguradoras. Como entidades financieras adecuando su operación, en lo posible, al régimen de las operaciones activas de las instituciones de crédito. Se reconoce pues una doble función: como aseguradoras y como fuentes de financiamiento.*

*3º. Medidas para una mayor sanidad en el sector.- Se procura con ellas ampliar y definir las funciones de inspección y vigilancia de las autoridades del sector, redefinir y complementar los regímenes de revocación de autorizaciones y el de sanciones, para procurar una mayor seguridad en la realización de las operaciones de las instituciones y sociedades mutualistas de seguros.*

---

<sup>25</sup> Elizondo, J. *op. cit.*, p. 265.

<sup>26</sup> Minzoni, A. *op. cit.*, p. 83.

*4º. Ajustes diversos.- En la redefinición de la Ley, se reclasificaron las disposiciones que el texto anterior contenía, reubicando varias de ellas en nuevos títulos bajo el criterio de agrupar todos los artículos afines en un mismo rubro, de acuerdo a la especialización de los mismos.*

Para mediados de los años ochenta “la política mexicana de incorporación deliberada a la globalización económica comienza [...] con la política de liberalización comercial. [...] Al culminar la apertura comercial, en 1988 se inició la desregulación financiera [...]. Su importancia radica en que inaugura un proceso amplio e intenso de reformas que afectará prácticamente a todo el sistema financiero mexicano, pues combinará de inmediato diversas disposiciones que afectarán a todos los mercados, intermediarios e instrumentos financieros, por medio, en particular, de un rápido proceso de reprivatización bancaria, de la integración de agrupaciones financieras y de la apertura de la cuenta de capitales”.<sup>27</sup>

Es así como en 1990 se inicia la etapa de desregulación del sector asegurador mexicano, la cual permite a las compañías de seguros pertenecer a Grupos Financieros, además de abrirse a la inversión extranjera, prohibida en el país desde 1965. Más adelante en 1993, se da la autorización para el establecimiento de filiales de compañías extranjeras para realizar operaciones de seguros en territorio mexicano.

Minzoni (1994) señala que la composición del mercado asegurador mexicano a mediados de 1992, no había cambiado en comparación con la composición correspondiente a finales de 1988, pero a raíz de la apertura comercial, doce instituciones aseguradoras presentaban ya participación de capital extranjero, la cual oscilaba entre el 10% y el 49% del capital pagado. En este sentido, el continente europeo (Inglaterra, Italia, Francia, España y Suiza) tenía una mayor presencia en el mercado asegurador que la correspondiente a los Estados Unidos de Norteamérica.

La puesta en marcha del TLCAN en enero de 1994 entrañó la entrada al país de bancos e intermediarios financieros no bancarios estadounidenses y canadienses, es decir, la entrada parcial en el mercado mexicano dentro de un periodo de transición a partir de inicios de 1994 de bancos comerciales y de inversión, almacenes, afianzadoras y aseguradoras. Debido a la crisis de diciembre de 1994, muchos intermediarios, especialmente bancos, abrirían su propiedad a los inversionistas extranjeros con la finalidad de capitalizarse y se permitió la entrada de nuevos intermediarios financieros conforme a un programa más acelerado.<sup>28</sup>

### **1.3.3 La industria aseguradora en México (1995 – 2001)<sup>29</sup>**

En los años de 1995 a 2001 el sector asegurador mexicano experimento importantes cambios en su estructura, incrementándose el número de empresas participantes hasta alcanzar un total de 70 a finales del año 2001 (ver Cuadro 1.1).

<sup>27</sup> Sánchez R. y Sánchez A. (2000). Desregulación y Apertura del Sector Financiero Mexicano. *Comercio Exterior*, 50 (8), 686-697.

<sup>28</sup> Sánchez, R. *op. cit.*, p. 693.

<sup>29</sup> Los cuadros y gráficas que se presentan en este apartado fueron tomados de Loyola, J. y Pérez, A. (2002). *Evolución del Sector Asegurador Mexicano de 1995 a 2001* (Documento Descriptivo No.15). México: CNSF, Dirección General de Desarrollo e Investigación.

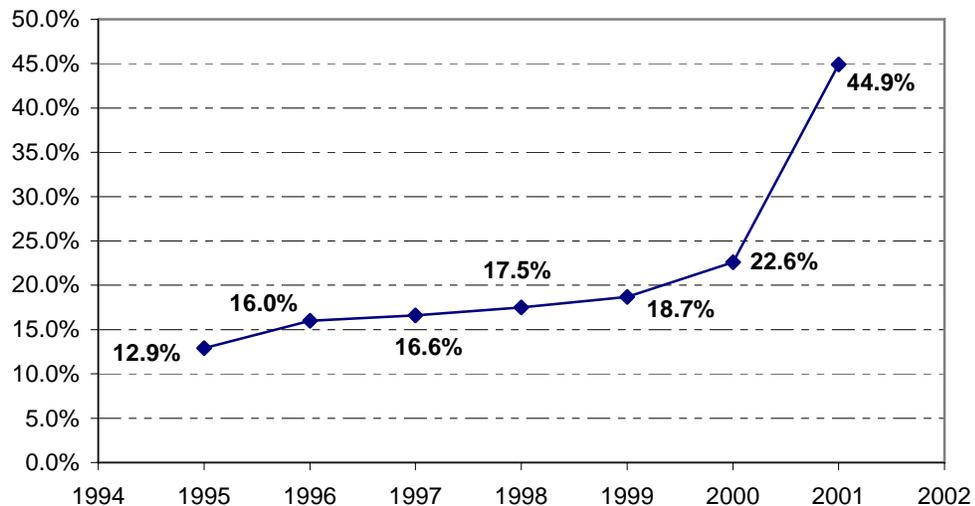
**Cuadro 1.1 Número de compañías autorizadas para operar en el Sector Asegurador 1995-2001**

	1995	1996	1997	1998	1999	2000	2001
Nacionales	2	2	2	2	2	2	2
Privadas	49	54	62	60	60	62	63
Mutualistas	3	3	3	3	3	3	2
Reaseguradoras	2	2	3	3	3	3	3
<b>Total</b>	<b>56</b>	<b>61</b>	<b>70</b>	<b>68</b>	<b>68</b>	<b>70</b>	<b>70</b>
Filiales	13	18	23	25	27	30	33
Grupos Finan.	14	15	17	18	17	15	14

Fuente: CNSF.

La entrada de nuevas entidades financieras al sector, se explica por las mejores condiciones para la inversión extranjera, derivadas de: la modificación de la Ley General de Instituciones y Sociedades Mutualistas de Seguros (LGISMS) de 1993, la firma del Tratado de Libre Comercio con América del Norte (TLCAN) en 1994, así como el Tratado de Libre Comercio firmado con la Unión Europea (UE) y la extensión de los beneficios a todos los miembros de la OCDE adquiridos en el TLCAN.<sup>30</sup> Este efecto se aprecia claramente en la Gráfica 1.1, donde se muestra el constante crecimiento de la participación del capital extranjero en el mercado asegurador mexicano, durante los años 1995 a 2001.

**Gráfica 1.1 Participación de capital extranjero en la Prima Directa 1995-2001**



Fuente: CNSF.

La apertura del sector asegurador no solamente privilegió una mejora sustancial en las condiciones de entrada a la inversión extranjera sino, también favoreció una estructura más abierta del sector asegurador, permitiendo la participación en el mercado de nuevas compañías. Aunado a esto, la estructura del sector se vio modificada por el inicio del nuevo

<sup>30</sup> Loyola, J. *op. cit.*, p. 4.

Sistema de Pensiones en julio de 1997, que propició la creación de empresas especializadas en estos seguros.<sup>31</sup>

Por otra parte, a partir de 1990, año en que se otorgó la autorización del Gobierno Federal para la constitución de Grupos Financieros, se ha constatado una clara tendencia hacia la integración de los distintos servicios prestados por los diversos intermediarios que participan en la industria de los servicios financieros (ver Cuadro 1.2). El sector asegurador se ha insertado plenamente en este proceso con el objeto de fortalecer la posición de las instituciones de seguros en el mercado, así como elevar su capacidad competitiva al aprovechar la sinergia de pertenecer a un grupo financiero.<sup>32</sup>

**Cuadro 1.2 Instituciones de Seguros integradas a Grupos Financieros 1995-2001**

Compañía	De Grupo Financiero:	1995	1996	1997	1998	1999	2000	2001
ABA seguros	Abaco	*	*	*	*	*	*	-
AIG México Seguros Interamericana	Bital	*	-	-	-	-	-	-
Aseguradora GBM Atlántico	GBM Atlántico	*	*	*	*	*	*	*
Aseguradora Interacciones	Interacciones	*	*	*	*	*	*	*
C.B.I. Seguros	CBI	*	*	*	*	*	*	*
Kemper de México	Sofimex	*	*	*	*	*	-	-
Pensiones Bancomer	Bancomer	-	*	*	*	*	*	*
Pensiones Banorte Generali	Banorte	-	-	*	*	*	*	*
Pensiones Bital	Bital	-	-	*	*	-	-	-
Royal & Sunalliance	BBV-Probursa	*	*	*	*	*	*	*
Seguros Afirme	Afirme	-	-	-	*	*	*	*
Seguros Banamex	Banamex Accival	*	*	*	*	*	*	*
Seguros Bancomer	Bancomer	-	-	*	*	*	*	*
Seguros Banorte Generali	Banorte	-	-	-	*	*	*	*
Seguros Banpaís	Asemex Banpaís	*	*	-	-	-	-	-
Seguros Bital	Bital	-	*	*	*	*	*	*
Seguros Comercial América	Asemex Banpaís	*	-	-	-	-	-	-
Seguros Inbursa	Inbursa	*	*	*	*	*	*	*
Seguros Margen	Inbursa	*	*	*	-	-	-	-
Seguros Monterrey New York Life	Margen	-	*	*	*	*	-	-
Seguros Santander Mexicano	Santander Serfin	*	*	*	*	*	*	*
Seguros Serfín	Santander Serfin	*	*	*	*	*	*	*

\* Forma parte del Grupo Financiero.

Fuente: CNSF.

Durante el periodo 1995-2001 el incremento real acumulado de la industria en su conjunto, medido a través de la emisión de primas directas, alcanzó un 39.5%, superior al 20.6% real logrado por la economía.<sup>33</sup> Estas cifras implican crecimientos anuales promedio de 5.6% para todo el sector, "cifra que se ve afectada por las caídas de la producción de la industria aseguradora para los años de 1995 y 1996, que fueron de 13.5% y 4.9% real, respectivamente, como resultado de la crisis observada en esos años".<sup>34</sup> (Ver Gráfica 1.2).

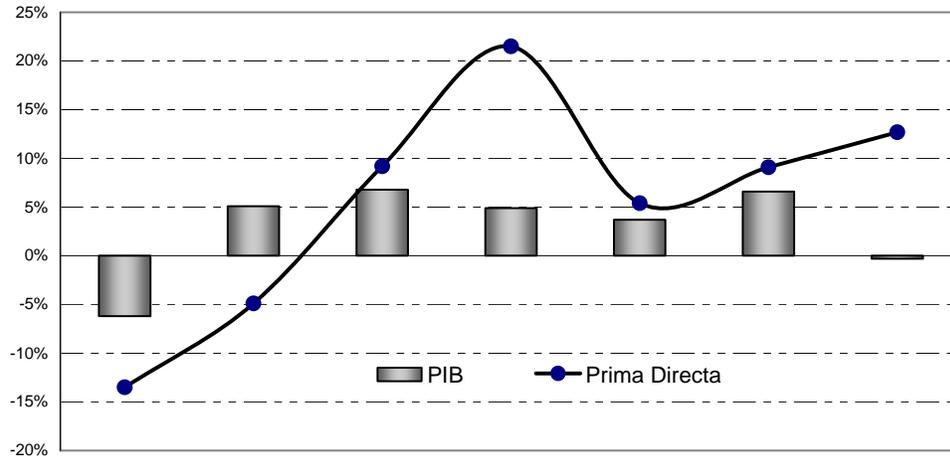
<sup>31</sup> Loyola, J. *op. cit.*, p. 5.

<sup>32</sup> *Ibid.* p. 6.

<sup>33</sup> El nivel de profundización del seguro en una economía se analiza a través de dos indicadores: La relación de primas directas a Producto Interno Bruto (Primas/PIB), que indica la importancias relativa del sector en la economía; y la relación de primas directas entre el número de habitantes (Primas Per Cápita), que mide el gasto promedio en seguros por habitante en una economía. Romero, H. (1993). El Seguro en el Mundo, 1990 - 1991. *Actualidad en Seguros y Fianzas*, 1 (9), 105-121.

<sup>34</sup> Loyola, J. *op. cit.*, p. 7.

**Gráfica 1.2 Prima Directa y PIB (Crecimiento Real)<sup>35</sup>  
1995-2001**



	1995	1996	1997	1998	1999*	2000*	2001
<b>PIB</b>	-6.2%	5.1%	6.8%	4.9%	3.7%	6.6%	-0.3%
<b>Prima Directa</b>	-13.5%	-4.9%	9.2%	21.5%	5.4%	9.1%	12.7%

\* Sin el efecto de los seguros dotales  
Fuente: CNSF e INEGI.

De este modo, se puede observar que a partir de 1996 el comportamiento del sector asegurador mexicano ha sido muy satisfactorio, ya que ha venido creciendo de manera consistente en términos reales. Dicho crecimiento no sólo ha superado al de la inflación sino que también ha sido superior al ritmo de crecimiento económico, lo cual implica que el sector de seguros haya aumentado su participación dentro de la economía nacional.<sup>36</sup>

#### 1.3.4 La industria aseguradora en México (2002 – 2003)

De las 81 empresas que integraban el sector asegurador mexicano al 31 de diciembre de 2002, 11 instituciones estaban integradas a grupos financieros, 36 eran filiales de instituciones financieras del exterior. Adicionalmente, 4 compañías presentaban capital mayoritariamente extranjero al mismo tiempo que pertenecían a algún grupo financiero.<sup>37</sup> (Ver Cuadro 1.3).

<sup>35</sup> Las cifras presentadas en esta gráfica contemplan los criterios para el registro contable de las primas de planes dotales a corto plazo y de planes de seguros flexibles que entraron en vigor a partir del primer trimestre de 2001, con el fin de eliminar el efecto contable de estos seguros sobre la emisión de prima directa al mes de septiembre de 2000.

<sup>36</sup> Loyola, J. *op. cit.*, p. 9.

<sup>37</sup> CNSF, *Memoria de Actividades 2002 y Perspectivas para el 2003*. (2003). México: Autor.

**Cuadro 1.3 Número de empresas autorizadas para operar  
en el Sector Asegurador  
2002 – 2003**

	<b>2002</b>	<b>2003</b>
Nacionales	1	1
Privadas	75	80
Mutualistas	2	1
Reaseguradoras	3	3
<b>Total</b>	<b>81</b>	<b>85</b>
Filiales	36	35
Filiales y G. Fin.	4	5
Grupos Finan.	11	9
Resto	30	36

*Fuente: CNSF.*

Al finalizar el cuarto trimestre de 2003, el sector asegurador estaba conformado por 85 compañías, de las cuales una era institución nacional de seguros, una sociedad mutualista de seguros y el resto instituciones de seguros privadas. De las 85 instituciones, 9 empresas estaban incorporadas a algún grupo financiero, mientras que 35 presentaban capital mayoritariamente extranjero, con autorización para operar como filiales de instituciones financieras del exterior. Adicionalmente, 5 compañías presentaban capital mayoritariamente extranjero al mismo tiempo que pertenecían a algún grupo financiero.<sup>38</sup>

Al cierre de 2002, las primas directas reportaron un crecimiento real de 12.1% respecto al año 2001. Sin embargo, para el cuarto trimestre de 2003 el comportamiento de la prima directa mostró un decremento del 9% respecto al registrado en 2002.

Según especialistas de la Comisión Nacional de Seguros y Fianzas, el decremento experimentado por el sector asegurador en el 2003 se explica en buena medida por la pérdida de dinamismo que experimentaron tanto los Seguros de Pensiones derivados de las leyes de Seguridad Social, que registraron un decremento real de 65.9% con respecto al año anterior, como al descenso de los seguros de Vida.

No obstante, y a pesar del decremento de primas directas registrado en 2003, el sector asegurador alcanzó una utilidad técnica de 8,664.7 millones de pesos en el ejercicio, lo cual significó un crecimiento de 27.3%, en términos reales, en relación a la presentada el año anterior (ver Cuadro 1.4).

En el mismo ejercicio se puede apreciar que los productos financieros netos del conjunto de las instituciones de seguros sumaron 12,796.8 millones de pesos, lo que representó un incremento en términos reales de 29.4%. En consecuencia, el sector asegurador mostró una utilidad neta del ejercicio de 9,429.3 millones de pesos, lo que representa un crecimiento real de 74.2% respecto al 2002.

<sup>38</sup> Comisión Nacional de Seguros y Fianzas. (2004a). *Boletín de Análisis Sectorial*, 3 (9).

**Cuadro 1.4 Resultados de Operación  
2002 – 2003**

<b>Resultados de Operación</b>	<b>Dic-02</b>	<b>Dic-03</b>	<b>Var. Real (%)</b>
Primas Retenidas	107,168.00	98,106.50	-12.0
Incr. Neto a Rvas. Riesgos en Curso	20,350.20	15,977.90	-24.5
Costo Neto de Adqs.	15,179.56	16,040.90	1.6
Costo Neto de Siniest.	65,089.90	57,423.00	-15.2
Utilidad Técnica	6,548.40	8,664.70	27.3
Incr. Neto otras Reservas.	-2,567.40	-580.50	78.3
Gastos de Operación	12,441.20	11,971.20	-7.5
Utilidad de Operación	-3,325.50	-2,726.00	21.2
Prod. Financieros Netos	9,509.80	12,796.80	29.4
Utilidad Antes de ISR y PTU	6,184.30	10,070.80	56.6
ISR y PTU	979.50	641.50	-37.0
<b>Utilidad del Ejercicio</b>	<b>5,204.80</b>	<b>9,429.30</b>	<b>74.2</b>

*Fuente: CNSF.*

### 1.3.5 La industria aseguradora en México (2004 – 2006)

Al finalizar el cuarto trimestre de 2004, el sector asegurador estaba conformado por 87 compañías, de las cuales una era institución nacional de seguros, una sociedad mutualista de seguros y el resto instituciones de seguros privadas. De las 87 instituciones, 6 empresas estaban incorporadas a algún grupo financiero, mientras que 38 presentaban capital mayoritariamente extranjero, con autorización para operar como filiales de instituciones financieras del exterior.<sup>39</sup> (Ver Cuadro 1.5).

**Cuadro 1.5 Número de empresas autorizadas para operar  
en el Sector Asegurador 2004 – 2006**

	<b>2004</b>	<b>2005</b>	<b>2006</b>
Nacionales	1	1	1
Privadas	83	82	87
Mutualistas	1	1	1
Reaseguradoras	2	2	2
<b>Total</b>	<b>87</b>	<b>86</b>	<b>91</b>
Filiales	38	39	43
Grupos Finan.	6	6	6
Resto	43	41	42

*Fuente: CNSF.*

<sup>39</sup> Comisión Nacional de Seguros y Fianzas. (2005). *Boletín de Análisis Sectorial*, 4 (13).

Para el cuarto trimestre de 2006, el sector asegurador estaba conformado por 91 compañías, de las cuales una era institución nacional de seguros, una sociedad mutualista de seguros y el resto, instituciones de seguros privadas. De las 91 instituciones, 6 empresas estaban incorporadas a algún grupo financiero, mientras que 43 presentaban capital mayoritariamente extranjero, con autorización para operar como filiales de instituciones financieras del exterior. Adicionalmente 7 compañías presentaban capital mayoritariamente extranjero al mismo tiempo que pertenecían a algún grupo financiero.<sup>40</sup>

El total de las primas emitidas por el sector asegurador, al cierre del cuarto trimestre de 2004, ascendió a 140,771.0 millones de pesos; al cierre de 2005 ascendió a 140,886.0 millones de pesos; mientras que al cierre de 2006 las primas emitidas sumaron 166,240.5 millones de pesos.

El incremento de prima directa en términos reales fue significativo al cierre de 2006, alcanzando un 13.5% de crecimiento con respecto al mismo periodo del año anterior. Según especialistas, hacia finales del 2006 el sector era 42 veces más grande de lo que lo era en el año 2000, creciendo incluso por encima del PIB.

### 1.3.6 Actualidad de la industria aseguradora en México

En los últimos años el mercado de seguros mexicano se ha caracterizado por ser uno de los más dinámicos de América Latina. Según datos de la Federación Interamericana de Empresas de Seguros (FIDES), México y Brasil recaudan cerca del 60% de las primas de la región latinoamericana. Sin embargo, estos dos grandes mercados se explican de manera diferente: si bien Brasil basa su situación más en los seguros de daños distintos del de vida, es en éste en el que se apoya buena parte de la situación relativa del mercado mexicano (a través del cual se intermedian el 42% de las primas).<sup>41</sup>

Al finalizar el cuarto trimestre de 2007, el sector asegurador estaba conformado por 95 compañías, de las cuales una era institución nacional de seguros, una sociedad mutualista de seguros y el resto, instituciones de seguros privadas. De las 95 instituciones, 6 empresas estaban incorporadas a algún grupo financiero, mientras que 46 presentaban capital mayoritariamente extranjero, con autorización para operar como filiales de instituciones financieras del exterior. Adicionalmente 8 compañías presentaban capital mayoritariamente extranjero al mismo tiempo que pertenecían a algún grupo financiero.<sup>42</sup>

El total de las primas emitidas por el sector asegurador, al cierre del cuarto trimestre de 2007, ascendió a 193,606 millones de pesos, de los cuales, el 98.3% (190,330 millones de pesos) correspondió al seguro directo, mientras que el restante 1.7% (3,276 millones de pesos) se debió a primas tomadas en reaseguro. Así, el sector asegurador registró en la prima directa un incremento en términos reales de 12.1% respecto al mismo periodo del año anterior.<sup>43</sup>

Según especialistas de la Comisión Nacional de Seguros y Fianzas, el crecimiento total del sector de seguros en 2007 es producto de los resultados de las diferentes operaciones y ramos que lo componen, de esta manera, la operación de Vida sin considerar los seguros de Pensiones derivados de las Leyes de Seguridad Social contribuyó al crecimiento del sector asegurador en

<sup>40</sup> Comisión Nacional de Seguros y Fianzas. (2007). *Boletín de Análisis Sectorial*, 6 (21).

<sup>41</sup> Federación Interamericana de Empresas de Seguros. (2003). *Memoria FIDES 2002*. Tegucigalpa: Graficentro Editores.

<sup>42</sup> Comisión Nacional de Seguros y Fianzas. (2008). *Boletín de Análisis Sectorial*, 7 (24).

<sup>43</sup> CNSF. (2008). *op. cit.*, p. 2.

su conjunto con 8.2 puntos porcentuales (pp.) Adicionalmente, los seguros de Pensiones derivados de las Leyes de Seguridad Social, la operación de Accidentes y Enfermedades, el ramo de Autos y el conjunto de los demás ramos de la operación de Daños, incluyendo Incendio y Terremoto, también presentaron contribuciones positivas al crecimiento.

Así, al finalizar el mes de diciembre de 2007, el sector asegurador presentó una utilidad técnica de 9,607.8 millones de pesos, lo que representó un incremento en términos reales de 28.6% respecto a diciembre de 2006. Una vez considerado el incremento neto a otras reservas técnicas y descontados los gastos de operación, el resultado muestra una pérdida de operación de 2,831.9 millones de pesos.<sup>44</sup>

Los productos financieros sumaron 18,789.0 millones de pesos, lo que representó un decremento en términos reales de 7.6%. Consecuentemente, el sector asegurador mostró una utilidad neta del ejercicio de 12,570.2 millones de pesos, lo cual implicó un incremento en términos reales de 19.5% con respecto a diciembre de 2006.<sup>45</sup>

Los resultados del sector asegurador durante 2007 reflejan el dinamismo y el crecimiento que le han caracterizado en los últimos años. De seguir esta tendencia, la penetración del sector asegurador mexicano seguramente alcanzará niveles cercanos al 3% para el año 2010, lo anterior, según datos dados a conocer por el Lic. Manuel Aguilera Verduzco en la XIII Convención Nacional de Aseguradores, realizada en la Ciudad de México en mayo de 2003 (ver Cuadro 1.6).

**Cuadro 1.6 Tamaño potencial del mercado<sup>46</sup>**

	Penetración (Primas/PIB)			Tamaño del Mercado (millones de pesos de 2002)		
	Escenario Bajo	Escenario Medio	Escenario Alto	Escenario Bajo	Escenario Medio	Escenario Alto
<b>2002</b>	2.0%	2.0%	2.0%	124,694	124,694	124,694
<b>2005</b>	2.3%	2.4%	2.6%	176,313	162,751	155,970
<b>2010</b>	2.8%	3.0%	3.3%	272,266	247,514	231,013
<b>2015</b>	3.4%	3.8%	4.2%	421,594	381,443	341,291
<b>2020</b>	4.0%	4.5%	5.0%	610,636	549,572	488,509

Fuente: CNSF.

Así, el sector asegurador mexicano se muestra como un sector "confiable que esta cumpliendo con sus compromisos, al mismo tiempo que registra un crecimiento sano"<sup>47</sup> en las diversas operaciones de seguros.

<sup>44</sup> CNSF. (2008). *op. cit.*, p. 9.

<sup>45</sup> CNSF. (2008). *op. cit.*, p. 9.

<sup>46</sup> Aguilera, M. (2003). *Cambio estructural y desarrollo del sector asegurador mexicano*. Trabajo presentado en la XIII Convención Nacional de Aseguradores, Mayo 19, Ciudad de México.

## CAPÍTULO II. LA IMPORTANCIA DE LA RETENCIÓN DE CLIENTES

---

### 2.1 El comportamiento de los clientes

#### 2.1.1 Características básicas

Se define como *comportamiento del consumidor* al «conjunto de actividades que lleva a cabo una persona física o jurídica desde que siente la necesidad hasta el momento en que efectúa la compra del bien y posteriormente lo utiliza o consume».<sup>47</sup>

Durante este proceso, el tiempo es un factor crítico ya que las empresas buscan estar en la mente de los consumidores antes que nadie, evitando con ello la intromisión de posibles competidores. Pero no solamente es importante anticiparse, también se debe considerar cuáles son las características básicas del consumidor estándar, de tal forma que puedan orientarse esfuerzos para obtener clientes de una manera más eficiente.

Cuesta (2003) distingue tres características básicas del ser humano en su papel de consumidor:

*Variable*, esto es, que un consumidor actúa de forma diferente en función de una serie de variables endógenas y exógenas, y así le afectará favorablemente o desfavorablemente en función de sus características personales, el tipo de producto o servicio que se le está ofreciendo, en qué momento se le está ofreciendo, quién se lo está ofreciendo, etc.

*Cambiante*, entre otras cosas, en función del ciclo de vida del producto, ya que existen diferentes porcentajes de clientes más predispuestos a aceptar los productos cuando estos son especialmente novedosos y otros, por el contrario, no quieren aceptar los productos hasta que estos están muy probados.

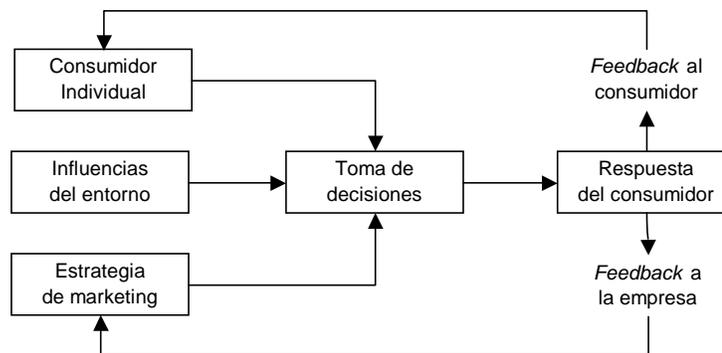
*Complejo*, esto es, que el comportamiento también puede cambiar en función de una serie de variables exógenas, es decir, en función del entorno político, social, cultural, tecnológico, etc., y la influencia de la empresa sobre el mismo; y de otra serie de variables endógenas, muy especialmente de la propia experiencia del consumidor con el producto y con la empresa que se lo ha suministrado y su comportamiento hacia él, en el grado de cumplimiento de las expectativas despertadas a través de diferentes medios o fórmulas sobre el mismo (ver Gráfica 2.1).

---

<sup>47</sup> Islas, L. (2004, Agosto 31). Repunta Vida Individual y Crece 17.6% el Sector; Seguros de Personas Representa 56.9% del Total. Durante el Primer Semestre de 2004. *El Asegurador*, pp. E1, E6.

<sup>48</sup> Cuesta, F. (2003). *Fidelización. Un paso más allá de la retención*. Madrid: McGraw-Hill.

**Gráfica 2.1 Modelo simple de comportamiento de compra**



Fuente: Cuesta, F. (2003). *op. cit.*, p. 17.

### 2.1.2 El concepto de lealtad

Se puede decir, de una manera simple, que la lealtad es el reconocimiento y fidelidad hacia algo o hacia alguien. En este sentido, un cliente brinda lealtad a un producto o servicio, cuando reconoce en él una serie de valores que lo hacen diferente.

Oliver (1997) menciona que la lealtad supone "un profundo compromiso para volver a comprar o ser cliente de un producto o servicio [...], provocando así, repetidas compras de la misma marca o conjunto de marcas, pese a que las influencias situacionales y los esfuerzos de marketing tengan el potencial de generar un cambio en el comportamiento de los clientes".

Fornell y Wernerfelt (1987) señalan que la lealtad es también una estrategia defensiva, cuyo objetivo es minimizar tanto la rotación o salida de clientes, como los cambios de marca por parte del consumidor.

En cualquier caso, e independientemente del enfoque que se utilice, la lealtad del cliente ofrece importantes beneficios para la empresa. Al respecto, diversos autores han puesto de manifiesto los beneficios que se desprenden al tener clientes leales, por ejemplo, Reichheld (1996) asegura que el atraer a nuevos clientes cuesta aproximadamente cinco veces más de lo que cuesta retener a los clientes actuales; por su parte, Cuesta (2003) señala que retener un 2% de clientes tiene el mismo efecto en las utilidades de una compañía que si se redujeran los costos en un 10%.

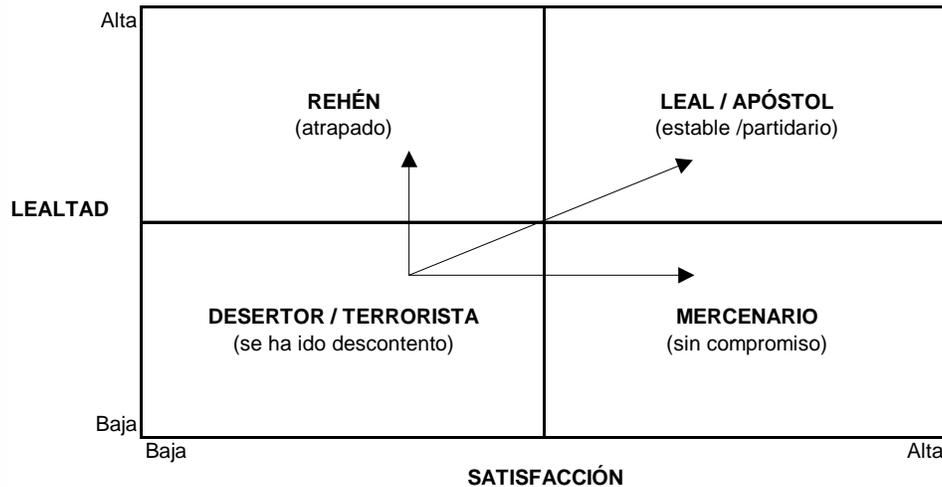
De esta forma, uno de los principales retos para cualquier empresa debería centrarse en la conversión de sus clientes, buscando tener cada vez menos clientes desertores y, por el contrario, incrementar el número de clientes leales.

### 2.1.3 Clasificación de los clientes

Cuesta (2003) hace una clasificación de clientes en función de dos dimensiones: el grado de lealtad y el grado de satisfacción. En ella distingue a cuatro tipos de clientes: el desertor o terrorista, el mercenario, el rehén y el leal o apóstol (ver Gráfica 2.2).

El desertor o terrorista se ha ido de la empresa insatisfecho. El problema no solo radica en que se ha ido, sino en que puede influir de manera negativa en los clientes que aún permanecen en la empresa.

**Gráfica 2.2 Clasificación de clientes**



Fuente: Cuesta, F. (2003). *op. cit.*, p. 21.

El rehén se caracteriza por una satisfacción baja, sin embargo, permanece fiel a la empresa porque no tiene otra alternativa. Este tipo de cliente tarde o temprano termina por abandonar la empresa.

El mercenario es un cliente poco leal y sin compromisos. Permanece con aquella empresa que lo satisfaga más.

Finalmente, el leal o apóstol, el cual se caracteriza por su alta satisfacción y su fuerte lealtad hacia la empresa. Como se ha mencionado, este tipo de cliente es el que ofrece más beneficios, de ahí la importancia de conservar su lealtad por largo tiempo.

Si bien lo ideal para una empresa sería convertir a todos sus clientes hacia la lealtad, tal vez el grupo de mayor cuidado lo sea el de los clientes desertores, ya que estos tienen la capacidad de influir de manera negativa en los clientes que aún permanecen en la empresa. Por tal motivo, "el reto de la empresa con los clientes desertores debe centrarse en conocer las razones por las que se marchan".<sup>49</sup>

#### 2.1.4 El comportamiento de los clientes en el sector seguros<sup>50</sup>

Como en cualquier otro mercado, en el de los seguros, los consumidores individuales configuran sus decisiones de comprar influenciados en mayor o menor medida por factores culturales, sociales, personales y psicológicos.

<sup>49</sup> Cuesta, F. *op. cit.*, p. 21.

<sup>50</sup> Merlo, J. y Martín, J. L. (2000). *El Marketing y el Seguro*. Madrid: Comuniland.

Los factores culturales tienen una influencia decisiva en el comportamiento del consumidor individual de seguros. En efecto, el hombre se ha ido alejando, según avanzaba la civilización, de la influencia de los instintos animales y los ha sustituido por pautas de comportamiento, aprendidas y transmitidas de generación en generación, que constituyen su entorno cultural. Estas pautas determinan sus decisiones de compra, más aún en el caso de los seguros que son productos evolucionados de esa misma cultura.

Los factores sociales también determinan las decisiones de compra de seguros. Los consumidores, además de pertenecer a entornos culturales concretos, forman grupos sociales más o menos amplios que influyen su conducta. Estos grupos van desde el núcleo familiar hasta los grupos llamados de referencia como los profesionales, los religiosos, los deportivos, etc.

Los factores personales como la edad, la personalidad, las aficiones, etc., influyen en las decisiones de compra de cada individuo. Por ejemplo, un joven de quince años no adquirirá seguros de automóviles ni de vida o jubilación. Sin embargo, influirá indirectamente en la contratación, por parte de sus padres, de seguros escolares, de dote<sup>51</sup>, de viaje, médicos, etc. Además, como miembro de pertenencia a una familia, sus decisiones de compra están influenciadas por la fase del ciclo familiar en que se encuentre (ver Cuadro 2.1).

**Cuadro 2.1 Fases de ciclo de vida y la contratación de seguros**

Fases	Comportamiento
Soltero, joven vive en la familia de orientación.	Consumidor potencial de seguros escolares, de viaje, médicos y de autos.
Soltero, joven no vive en familia.	Consumidor potencial de los seguros del grupo anterior más seguro de hogar.
Pareja recién formada, joven, sin hijos.	Consumidor potencial de seguros médicos, de viaje, de autos, de hogar y de vida riesgo para hipoteca.
Pareja joven, con hijos hasta 10 años.	Consumidores potenciales de seguros médicos, de viaje, de autos, de hogar, escolares, de accidentes y de vida riesgo.
Pareja madura, con hijos mayores de 10 años.	Consumidores potenciales de seguros médicos, de viaje, de autos, de hogar, escolares, de accidentes, de vida riesgo.
Pareja madura sin hijos dependientes y soltero mayor.	Consumidores potenciales de seguros médicos, de viaje, de autos, de hogar, de accidentes, de vida mixto.

*Fuente: Merlo, J. y Martín, J. L. (2000). op. cit., p. 49.*

La ocupación del individuo influye también sus comportamientos de compra. Por ejemplo, en una pareja madura con hijos mayores de 10 años, si el padre ha alcanzado un puesto de responsabilidad de administración de una empresa, será consumidor potencial de un seguro de responsabilidad profesional.

En las fases de ciclo de vida aplicables, un arquitecto o un médico, será consumidor potencial de un seguro de responsabilidad profesional.

<sup>51</sup> En México se conocen como *seguros dotales*.

El nivel de renta influye en las decisiones de compra en cualquiera de las fases del ciclo de vida. Si no hay capacidad de compra de un automóvil, no habrá consumos de un seguro de autos. Según el nivel de renta serán los consumos potenciales de seguros de accidentes, viajes, médicos, etc.

La personalidad y el estilo de vida influyen en las decisiones de compra de seguros. Por ejemplo, un conductor seguro de sí mismo, con fuerte personalidad, será más propenso a contratar un seguro a terceros o con franquicia.<sup>52</sup> Una persona con estilo de vida de alto nivel será más propensa a contratar un seguro médico privado.

Por último, los factores psicológicos como la motivación, la aptitud, el aprendizaje o la actitud también influyen en las decisiones de compra de los individuos.

## 2.2 Indicadores básicos para la retención de clientes

### 2.2.1 Tasa de retención de clientes

La retención de clientes se basa en la premisa básica de que un cambio a favor o en contra puede hacer la diferencia entre utilidad y pérdida.<sup>53</sup> Dado lo anterior, muchas empresas centran importantes esfuerzos en programas de retención de clientes, los cuales no siempre resultan efectivos.

Huete (1997) señala que una empresa fideliza a sus clientes en la medida en que es capaz de lograr una tasa de retención alta de los mismos. Si se define la retención como la repetición de compra en un determinado periodo de tiempo, se tendría que fidelizar a los clientes es simple y sencillamente conseguir tasas altas de repetidores, o lo que es lo mismo, tasas bajas de no repetidores (desertores). En el mismo sentido, la tasa de deserción de una empresa es el porcentaje de clientes que no repiten en un periodo de tiempo determinado. La relación entre la tasa de retención y la tasa de deserción se muestra en la Gráfica 2.3.

**Gráfica 2.3 Tasa de retención de clientes**

$$\text{Tasa de retención} = 1 - \text{Tasa de deserción}$$

*Fuente: Huete, L. M. (1997). Revitalizando los servicios. Barcelona: IESE/Folio. p. 21.*

### 2.2.2 Tasa de deserción de clientes

La tasa de deserción es un indicador de gestión de primera magnitud al que conviene dedicar más atención [...]. Conceptualmente, su significado es claro: equivale al porcentaje de clientes que dejan de serlo entre un periodo y otro. No señala el número de clientes de un periodo

<sup>52</sup> En México el *seguro a terceros* se conoce como *seguro con cobertura limitada*, mientras que el concepto de *franquicia* se maneja comúnmente como *deducible*.

<sup>53</sup> Carazo, L. M. (1997, Junio 11). Retención de clientes. *Reforma*, pp. 32.

comparado con el número de clientes del periodo siguiente. Eso sería mas bien un indicador de crecimiento. Lo que [...] interesa captar con la tasa de deserción es la magnitud de los abandonos que se producen entre los clientes. El interés en los abandonos viene suscitado por la consideración de que -¡salvo excepciones!- es deseable evitarlos.<sup>54</sup> (Ver Gráfica 2.4).

**Gráfica 2.4 Tasa de deserción de clientes**

$$\text{Tasa de deserción} = 1 - \frac{\text{Clientes (t} \rightarrow \text{t+1)}}{\text{Clientes (t)}}$$

*Fuente: Huete, L. M. (1997). op. cit., p. 23.*

La tasa deserción de clientes puede ser establecida para el conjunto total de clientes de una empresa o, lo que es mejor, desagregada por segmentos. Esta información por segmentos permite afinar en las actuaciones de fidelización. A raíz de esa información, muchas veces se decide prestar menos atención a un segmento con mayor propensión a la fidelización. También se podría decir, a raíz de esa información, el reconceptualizar la oferta para los segmentos en los que la tasa de fidelización es más baja.<sup>55</sup>

A título orientativo, la tasa de deserción anual normal de una empresa con una calidad de servicio media es del 15% al 25%. Una empresa peor posicionada tendría una tasa superior.<sup>56</sup>

### 2.2.3 Vida media de la relación de un cliente con la empresa

La tasa de deserciones permite computar la vida media de la relación de un cliente con la empresa. Tal y como se muestra en la Gráfica 2.5, la vida media de un cliente con la empresa se obtiene al dividir uno por la tasa de deserción expresada en tantos por uno<sup>57</sup>. Así, una tasa de deserción del 25% equivale a una vida media de los clientes de cuatro años, y una vida media de veinte años equivaldría a una tasa de deserción del 5%. Todo ello, claro está, si el periodo relevante es de un año. Si no lo fuera, todo sería igual excepto que, en vez de años, hablaríamos de lustros, trimestres, semanas, días, etc.<sup>58</sup>

**Gráfica 2.5 Vida media de la relación de un cliente con la empresa**

$$\text{Vida media de la relación de un cliente con la empresa} = \frac{1}{\text{Tasa de deserción}}$$

*Fuente: Huete, L. M. (1997). op. cit., p. 25.*

<sup>54</sup> Huete, L. M. *op. cit.*, p. 21.

<sup>55</sup> *Ibid.* p. 23.

<sup>56</sup> Huete, L. M. *op. cit.*, p. 23.

<sup>57</sup> Aunque esta afirmación puede parecer intuitiva, su demostración matemática exige la utilización de ecuaciones diferenciales e integrales.

<sup>58</sup> *Ibid.* p. 24.

Huete (1997) menciona que este mismo análisis puede hacerse si se pondera la importancia relativa del negocio que deserta. Por ejemplo, si un 10% del volumen total del negocio abandona durante un periodo de un año, se puede concluir que la vida media equivale a diez años. En ocasiones, la cifra que se obtiene de este análisis es más significativa que la que resulta de evaluar clientes individuales.

La vida media de un cliente permite visualizar mejor la importancia de la fidelización. Hasta el momento, el razonamiento seguido es simple: si fidelizar es conseguir repetidores, hay que disminuir el número de desertores. Esto, a su vez, determina la vida media de la relación de los clientes con la empresa. La idea que se pretende subrayar es la de que fidelizar es alargar la vida media de los clientes y que ello implica reducir el número de desertores.<sup>59</sup>

A nadie con un poco de perspicacia se le escapa la diferencia que existe entre tener una clientela media con pocos años de relación con la empresa y otra que lleve más años de relación. Equivale a tener clientes que abandonan enseguida o clientes que sistemáticamente acuden al mismo proveedor.<sup>60</sup>

---

<sup>59</sup> Huete, L. M. *op. cit.*, p. 25.

<sup>60</sup> *Ibíd.* p. 25.

## CAPÍTULO III. EL RECONOCIMIENTO DE PATRONES A TRAVÉS DE LA MINERÍA DE DATOS

---

### 3.1 Minería de datos

#### 3.1.1 Antecedentes y necesidades

El éxito de una organización en un entorno globalizado depende en gran medida del conocimiento de sus miembros y de sus habilidades para hacer negocios. Sin embargo, las organizaciones no solo necesitan de conocimiento, sino también necesitan tener la capacidad de actualizar dinámicamente el conocimiento y ponerlo en práctica.<sup>61</sup>

En esencia, deben ser capaces de resolver problemas, aprender de experiencias propias y también de experiencias de otros, transfiriendo el conocimiento y la información de forma rápida y eficiente a toda la organización.

En nuestros días, la gran mayoría de las organizaciones tienen la capacidad de generar grandes volúmenes de datos, esto de una forma rápida y creciente como resultado de sus operaciones diarias, no obstante, pocas son las organizaciones que producen información de verdadero valor.

Dentro de este contexto, surge la “urgente necesidad de contar con una nueva generación de teorías y herramientas computacionales que ayuden al ser humano a extraer información útil (conocimiento) de los crecientes volúmenes de datos digitales”.<sup>62</sup> Estas teorías y herramientas están enmarcadas en un nuevo campo de estudio denominado «Extracción de Conocimiento en Bases de Datos» (*Knowledge Discovery in Databases*, *KDD* por sus siglas en inglés).<sup>63</sup>

#### 3.1.2 Extracción de conocimiento en bases de datos

Históricamente, a la noción de encontrar patrones útiles en los datos se le ha dado una gran variedad de nombres, como «minería de datos», «extracción de conocimiento», «descubrimiento de información», «recolección de información», «arqueología de datos», y «procesamiento de patrones en datos». El término «minería de datos» se ha usado con mayor frecuencia en las comunidades de estadística, análisis de datos y sistemas de administración de la información. Dicho término también ha ganado popularidad en el campo de las bases de datos.<sup>64</sup> Sin embargo, en estricto sentido, la «minería de datos» y la «extracción de conocimiento en bases de datos» no son conceptos equivalentes.

---

<sup>61</sup> Venugopal, V. y Baets, W. (1995). Intelligent Support Systems for Organizational Learning. *The Learning Organization*, 2 (3), 22-34.

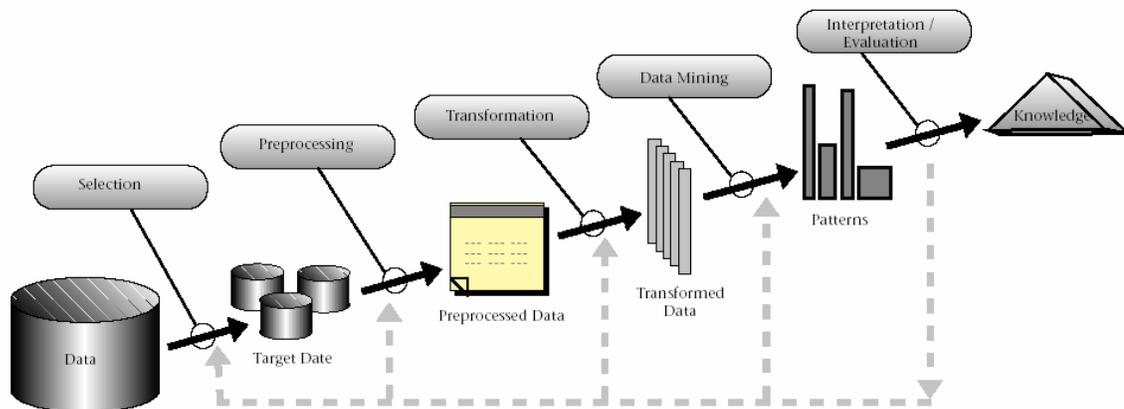
<sup>62</sup> Fayyad, U., Piatetsky-Shapiro, G. y Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *Artificial Intelligence Magazine*, 17 (3), 37-54.

<sup>63</sup> Aunque la traducción correcta de “*Knowledge Discovery in Databases*” es “*Descubrimiento de Conocimiento en Bases de Datos*”, en este trabajo se utilizará el termino “*Extracción de Conocimiento en Bases de Datos*” para referirse al proceso *KDD*.

<sup>64</sup> *Ibíd.* p. 39.

Según Fayyad, Piatetsky-Shapiro y Smyth (1996) la «extracción de conocimiento en bases de datos» (*KDD*) se refiere a todo el proceso de descubrir conocimiento útil en datos, mientras que la «minería de datos» (*data mining*) tiene que ver con la aplicación de algoritmos específicos para extraer patrones de los datos (ver Gráfica 3.1).

**Gráfica 3.1 Pasos que componen al proceso de extracción de conocimiento en bases de datos (*KDD*)**



Fuente: Fayyad, U., Piatetsky-Shapiro, G. y Smyth, P. (1996). *op. cit.*, p. 41.

Así, los pasos que componen al proceso *KDD* son cinco: selección del objetivo, preproceso de datos, transformación, minado de datos e interpretación de los resultados.

La selección del objetivo tiene como finalidad estudiar el problema y decidir cuál es la meta del proyecto. Una vez definido el problema, se identifican las fuentes de datos internas o externas y se selecciona el subconjunto de datos necesarios para la aplicación de un algoritmo de minería de datos.

El preproceso de datos consiste en estudiar los datos seleccionados para entender el significado de los atributos y para detectar errores de integración, por ejemplo, datos repetidos con distinto nombre o datos que significan lo mismo en diferente formato.

Una vez que se tienen los datos preprocesados, se procede a la transformación final de los mismos, esto con el fin de que se ajusten al formato de entrada del algoritmo seleccionado.

El siguiente paso es el minado de datos propiamente dicho. Aquí se aplican los diferentes algoritmos de análisis a los datos ya transformados. La finalidad en esta etapa es encontrar patrones útiles e interesantes en los datos.

Por último, se procede a interpretar y evaluar los resultados obtenidos en la etapa de minado de datos. Aquí, el usuario debe valorar los resultados conseguidos y, de ser necesario, aplicar una y otra vez los algoritmos de *Data Mining* hasta encontrar información útil y valiosa. Esto último hace que el proceso *KDD* sea un proceso iterativo y de búsqueda continua, en donde el conocimiento y la intuición del usuario juegan un papel fundamental en el proceso.

### 3.1.3 Técnicas de Minería de Datos

En los últimos años han existido muchos avances en las investigaciones y desarrollos relacionados con la minería de datos, como resultado, se han desarrollado diversas técnicas y sistemas relativos al *data mining*. Diferentes esquemas de clasificación pueden ser usados para categorizar métodos y sistemas de minado de datos, como el tipo de base de datos a estudiar (relacional, orientada a objetos, multimedia, etc.), el tipo de conocimiento que se quiere extraer (reglas de asociación, reglas de clasificación, *clustering*, etc.), así como las técnicas que serán aplicadas en el proceso (basadas en patrones, teoría estadística, teoría matemática, enfoques integradores, etc.).<sup>65</sup>

En la práctica, los métodos de *data mining* más utilizados caen dentro de la categoría de «tipo de conocimiento a extraer». Las técnicas de minado de datos pertenecientes a esta categoría buscan hacer predicción y/o descripción de un fenómeno determinado.<sup>66</sup>

La predicción implica utilizar algunas variables o campos de una base de datos para predecir valores desconocidos o futuros de otras variables de interés, mientras que la descripción está enfocada a encontrar patrones humano-interpretables que describen los datos. Aunque las fronteras entre la predicción y la descripción no sean muy marcadas (parte de los modelos predictivos pueden ser descriptivos, al grado de que estos sean entendibles, y viceversa), la distinción es útil para entender la meta general de la extracción de conocimiento. La importancia relativa de la predicción y la descripción para aplicaciones particulares de minado de datos puede variar considerablemente. Las metas de la predicción y la descripción se pueden lograr utilizando una variedad de métodos particulares de *data mining*.<sup>67</sup>

#### 3.1.3.1 Reglas de asociación

Mediante el minado de reglas de asociación se pueden encontrar interesantes relaciones de asociación o correlación en los datos. Dada la gran cantidad de datos que continuamente se recolectan y almacenan, muchas industrias se han interesado por encontrar reglas de asociación en sus bases de datos. El descubrimiento de interesantes relaciones de asociación en grandes cantidades de registros transaccionales, puede ayudar en diversos procesos de toma de decisiones relacionados con el negocio, tales como el diseño de catálogos, la venta cruzada, y el análisis *loss-leader*.<sup>68</sup>

Una regla de asociación es un criterio que implica ciertas relaciones de asociación entre distintos objetos de una base de datos, tales como “ocurren juntos” o “uno implica al otro”. Matemáticamente se representa como una implicación de la forma  $A \Rightarrow B$ , en donde  $A$  y  $B$  representan conjuntos de atributos con intersección vacía ( $A \cap B = \emptyset$ ), de tal forma que la regla se presenta en un conjunto de transacciones  $D$  con una confianza del  $\alpha\%$ .

Un ejemplo de regla de asociación sería: “40% de las transacciones que contienen niños también contienen pañales”. En este caso el 40% es el nivel de confianza de la regla.

<sup>65</sup> Chen, M.S., Han, J. y Yu, P.S. (1996). Data Mining: An Overview from Database Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8 (6), 866-883.

<sup>66</sup> Los algoritmos predictivos también se conocen con el nombre de *algoritmos supervisados*. Se caracterizan porque su desarrollo ocurre en dos fases: la fase de *entrenamiento*, que es donde se construye el modelo de predicción utilizando un subconjunto de datos, y la fase de *prueba*, en la cual se evalúa el modelo construido utilizando los datos restantes.

<sup>67</sup> Fayyad, U., Piatetsky-Shapiro, G. y Smyth, P. *op. cit.*, p. 44.

<sup>68</sup> Han, J. y Kamber, M. (2000). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers.

### 3.1.3.2 Clasificación y predicción

La clasificación y la predicción son dos formas de análisis de datos que pueden ser usadas para extraer modelos que describen importantes clases de datos o predicen valores futuros.<sup>69</sup>

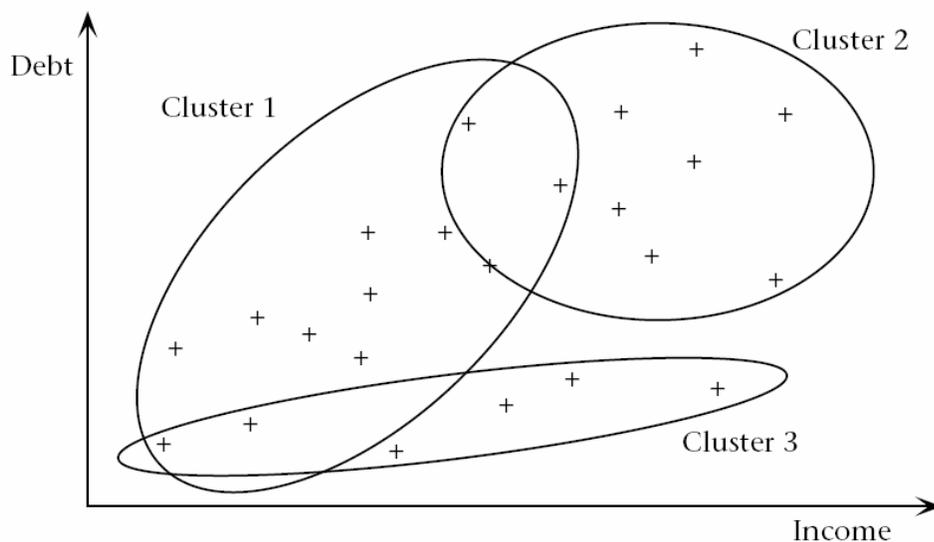
En la clasificación de datos se desarrolla una descripción o modelo para cada una de las clases presentes en la base de datos. Existen muchos métodos de clasificación tales como los árboles de decisión, los métodos estadísticos, las redes neuronales, y los conjuntos difusos, entre otros.

La predicción puede ser vista como la construcción y uso de modelos para evaluar las clases de una muestra sin clasificaciones, o para evaluar el valor, o rango de valores, que un atributo debería de tener para una muestra determinada.<sup>70</sup> En el primer caso se dice que se está haciendo una clasificación de los datos, mientras que en el segundo, se dice que se está haciendo una predicción de valores, en otras palabras, la clasificación se refiere a poder predecir valores discretos o nominales, mientras que la predicción, propiamente dicha, se refiere al pronóstico de valores continuos.

### 3.1.3.3 Clustering

El *Clustering* identifica grupos de datos que son "similares". La similitud puede medirse mediante funciones de distancia especificadas por los usuarios o por expertos (ver Gráfica 3.2).

Gráfica 3.2 *Clustering* de datos bancarios



Fuente: Fayyad, U., Piatetsky-Shapiro, G. y Smyth, P. (1996). *op. cit.*, p. 45.

<sup>69</sup> *Ibid.* p. 279.

<sup>70</sup> Han, J. y Kamber, M. *op. cit.*, p. 281.

Cuando se utiliza la técnica de *clustering*, se obtiene un diagrama en el cual se muestra como los datos caen dentro de distintos grupos (*clusters*). En el caso más simple, se asocia a cada dato un *cluster*, dibujando los datos o instancias en un diagrama de dos dimensiones.

Algunos algoritmos de *clustering* permiten a una instancia pertenecer a uno a más *clusters*, como resultado, el diagrama en dos dimensiones muestra como se traslapan los subconjuntos de datos (como un Diagrama de Venn). Otros algoritmos asocian instancias a *clusters* de manera probabilística, así para cada instancia, existe una probabilidad asociada o un "grado de pertenencia" con el cual se asigna a un determinado *cluster*.<sup>71</sup>

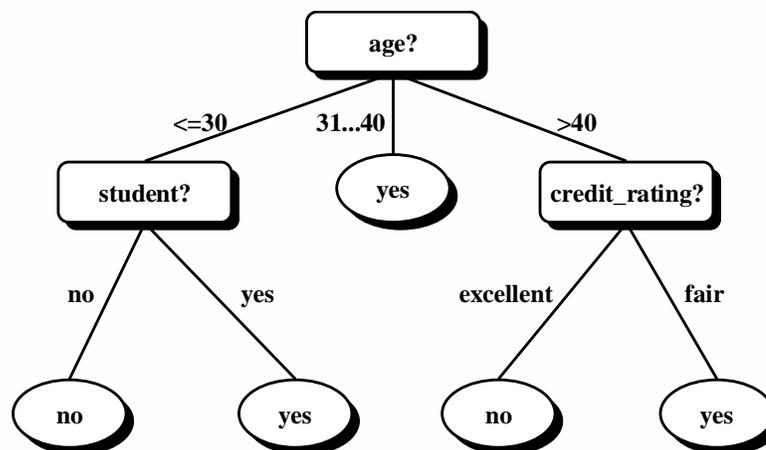
Algunos ejemplos de la aplicación de las técnicas de *clustering*, en el contexto de la extracción del conocimiento, incluyen la identificación de subgrupos homogéneos de clientes en bases de datos de *marketing*, así como la identificación de subcategorías de medidas del espectro infrarrojo del cielo.<sup>72</sup>

## 3.2 Árboles de decisión

### 3.2.1 Características básicas

Un «árbol de decisión» es un diagrama de flujo, con estructura de árbol, en donde los nodos internos representan validaciones sobre los atributos, las ramas representan las salidas de las validaciones, y los "nodos hoja" representan las clases. El nodo en la parte superior del árbol se le conoce como nodo raíz. Un típico ejemplo de un árbol de decisión se muestra en la Gráfica 3.3. Ahí se representa el concepto de comprar computadoras, esto es, el modelo predice si un cliente comprará o no una computadora en *AllElectronics*. En la gráfica los nodos internos están denotados por rectángulos, mientras que los "nodos hoja" están denotados por óvalos.<sup>73</sup>

Gráfica 3.3 Árbol de decisión para la compra de computadoras



Fuente: Han, J. y Kamber, M. (2000). *op. cit.*, p. 284.

<sup>71</sup> Witten, I. H. y Frank, E. (2000). *Data Mining: Practical machine learning tools and techniques with java implementations*. San Francisco: Morgan Kaufmann Publishers.

<sup>72</sup> Fayyad, U., Piatetsky-Shapiro, G. y Smyth, P. *op. cit.*, p. 45.

<sup>73</sup> Han, J. y Kamber, M. *op. cit.*, p. 284.

Para clasificar una instancia “desconocida”, se sigue el flujo del árbol hacia abajo, de acuerdo a los valores que tengan los atributos para cada nodo, y cuando se llega a un “nodo hoja”, la instancia se clasifica de acuerdo a la clase asignada por dicho nodo.<sup>74</sup>

Existen diversos métodos para la inducción de árboles de decisión (ID3, C4, C4.5, Bayesiano, CART, etc.), cada uno de ellos ofrece diferentes capacidades, pero en general, dichos algoritmos son apropiados para solucionar «problemas de clasificación». Estos problemas presentan las siguientes características:

- El objeto de estudio del problema está representado por un conjunto fijo de atributos.
- La función objetivo del problema tiene valores discretos.
- Los atributos son descritos mediante conceptos disyuntivos.
- Puede haber errores en los datos de entrenamiento, tanto en la clasificación como en los valores de los atributos.
- Los datos de entrenamiento, con los cuales se construye el modelo, pueden tener valores desconocidos para algunos atributos.

Es importante señalar que los árboles de decisión presentan una gran ventaja respecto a otras técnicas de clasificación. Esta ventaja consiste en poder representar al conocimiento obtenido mediante el uso de reglas de decisión.<sup>75</sup>

### 3.2.2 El algoritmo ID3<sup>76</sup>

La idea básica del algoritmo ID3 tiene su fundamento en la iteración. Un subconjunto del conjunto total de datos de entrenamiento, al cual se le conoce como *window*, es elegido de manera aleatoria para formar un árbol de decisión; este árbol clasifica de manera correcta todos los objetos que pertenecen a *window*. El resto de los objetos, dentro del conjunto de datos de entrenamiento, es clasificado utilizando dicho árbol. Si el árbol da una respuesta correcta para todos estos objetos, entonces también es correcto para el conjunto total de datos de entrenamiento, terminando así el proceso. Si no, una selección de objetos clasificados incorrectamente es adicionada al subconjunto *window*, y el proceso continúa.

El problema radica en saber cómo construir un árbol de decisión para una colección arbitraria de objetos,  $C$ . Si  $C$  es un conjunto vacío o contiene objetos de una sola clase, el árbol de decisión es el más simple, con una sola hoja etiquetada con la clase. Por otro lado, sea  $T$  alguna validación sobre un objeto con posibles salidas  $O1, O2, \dots, Ow$ . Cada objeto en  $C$  dará una de estas salidas para  $T$ , así  $T$  produce una partición  $\{C1, C2, \dots, Cw\}$  de  $C$ , en donde a cada  $Ci$  le corresponde una salida  $Oi$ . Esta idea se representa en la Gráfica 3.4. Si en esta gráfica a cada subconjunto  $Ci$  se le pudiera asociar su respectivo árbol de decisión, el resultado sería un árbol de decisión para todo el conjunto  $C$ . Más aun, si más de dos  $Ci$  son subconjuntos no vacíos, entonces cada subconjunto  $Ci$  es más pequeño que  $C$ . En el peor de los casos, esta estrategia de “divide y vencerás” generará subconjuntos con un solo objeto que satisface los requerimientos de una clase en particular. Así, dada una validación, siempre se tendrá que

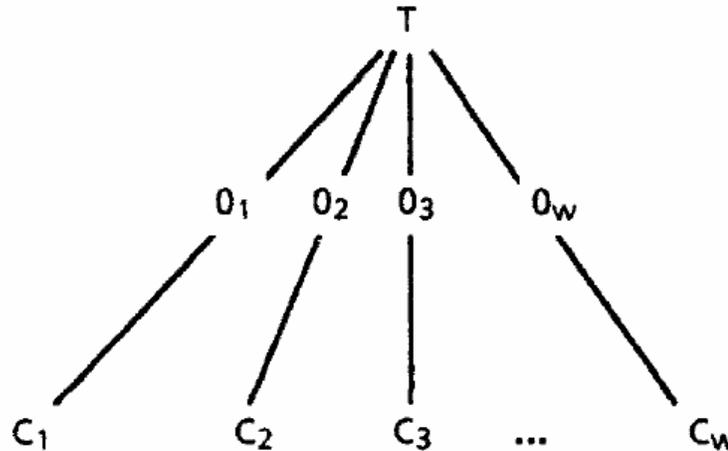
<sup>74</sup> Witten, I. H. y Frank, E. *op. cit.*, p. 58.

<sup>75</sup> Las *reglas de decisión* suelen ser cláusulas de lógica proposicional donde las premisas son condiciones sobre los atributos y el consecuente es la clase asignada en caso de que la regla se cumpla.

<sup>76</sup> Quinlan, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1 (1), 81-106.

existe una partición no trivial de cualquier conjunto de objetos, este procedimiento producirá un árbol de decisión que clasificará correctamente a cada objeto en  $C$ .

**Gráfica 3.4 Una estructura de árbol para los objetos en  $C$**



*Fuente: Quinlan, J. R. (1986). op. cit., p. 89.*

El elegir una validación es crucial si se requiere un árbol de decisión sencillo. Por el momento, una validación será limitada a ramificar los valores de un atributo, así que al escoger una validación se escoge un atributo que sirve como la raíz del árbol. Los primeros programas de inducción, en la serie de algoritmos ID, utilizaron una función de evaluación muy rudimentaria que trabajaba razonablemente bien. Debido a las sugerencias de Peter Gacs,<sup>77</sup> el algoritmo ID3 adoptó un método basado en información el cual depende de dos supuestos. Sea un conjunto  $C$  que contiene  $p$  objetos de la clase  $P$  y  $n$  de la clase  $N$ . Los supuestos son:

- [1] Cualquier árbol de decisión que clasifique correctamente a  $C$ , clasificará los objetos en la misma proporción que su representación en  $C$ . Un objeto arbitrario se clasificará dentro de la clase  $P$  con probabilidad  $p/(p+n)$  o dentro de la clase  $N$  con probabilidad  $n/(p+n)$ .
- [2] Cuando un árbol de decisión se usa para clasificar un objeto, se asocia una clase. Así, un árbol de decisión puede verse como un generador de mensajes "P" o "N", con la información esperada, necesaria para generar este mensaje, dada por

$$I(p, n) = - \frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

Si el atributo  $A$  con valores  $\{A1, A2, \dots, Av\}$  se elige como la raíz del árbol de decisión, habrá una partición  $\{C1, C2, \dots, Cv\}$  en  $C$ , donde cada  $Ci$  contendrá objetos en  $C$  que tienen valores  $Ai$  del atributo  $A$ . Sea  $Ci$  el cual contiene  $pi$  objetos de la clase  $P$  y  $ni$  de la clase  $N$ . La información esperada requerida para el subárbol de  $Ci$  es  $I(pi, ni)$ . La información esperada

<sup>77</sup> Computer Science Department, Boston University.

requerida para el árbol con el atributo  $A$  como nodo raíz, se obtiene mediante promedios ponderados

$$E(A) = \sum_{i=1}^v \frac{p_i + n_i}{p + n} I(p_i, n_i)$$

donde el peso de la  $i$ -ésima rama, es la proporción de objetos en  $C$  que pertenecen a  $C_i$ . Por lo tanto, la información ganada en la ramificación de  $A$  es

$$\text{gain}(A) = I(p, n) - E(A)$$

Una buena regla empírica parecería ser escoger aquel atributo que al ramificar gane la mayor cantidad de información.<sup>78</sup> ID3 examina todos atributos candidatos y escoge  $A$  para maximizar  $\text{gain}(A)$ , construye el árbol como se describió anteriormente, y entonces utiliza el mismo proceso de manera recursiva para formar árboles de decisión para los subconjuntos residuales  $C_1, C_2, \dots, C_v$ .

### 3.2.3 El algoritmo C4.5

El algoritmo C4.5 forma parte de la familia de los «Árboles Inducidos de Arriba hacia Abajo» (*Top Down Induction of Decision Trees, TDIDT* por sus siglas en inglés). Pertenecen a los métodos inductivos del *Machine Learning*, los cuales aprenden a partir de ejemplos preclasificados. Tanto el algoritmo ID3 como el C4.5 fueron propuestos por Ross Quinlan<sup>79</sup>, el primero en la década de los ochenta y el segundo en 1993. El algoritmo C4.5 es una extensión del algoritmo ID3, el cual trabaja únicamente con valores discretos en los atributos. En cambio, el algoritmo C4.5 permite trabajar con valores continuos, separando los posibles resultados en dos ramas: una para aquellos  $A_i \leq N$  y otra para todos los  $A_i > N$ .

De esta forma, C4.5 genera un árbol de decisión a partir de datos mediante particiones realizadas de manera recursiva. El algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta con mayor ganancia de información. Para cada atributo discreto, se considera una prueba con  $n$  resultados, siendo  $n$  el número de valores posibles que puede tomar el atributo. Para cada atributo continuo, se realiza una prueba binaria sobre cada uno de los valores que toma el atributo.

El criterio de descomposición que C4.5 utiliza por omisión, es el "porcentaje de ganancia", el cual representa una medida basada en información que toma en cuenta diferentes números (y diferentes probabilidades) en los resultados de validación. Sea  $C$  el número de clases y  $p(D, j)$  la proporción de casos en  $D$  que pertenecen a la  $j$ -ésima clase. La incertidumbre residual de la clase en la cual un determinado caso está contenido en  $D$ , puede ser expresada de la siguiente forma

<sup>78</sup> Dado que  $I(p, n)$  es una constante para todos los atributos, maximizar la ganancia es equivalente a maximizar  $E(A)$ , la cual es la información mutua del atributo  $A$  y la clase.

<sup>79</sup> University of Sydney.

$$Info(D) = - \sum_{j=1}^C p(D, j) \times \log_2(p(D, j))$$

y la correspondiente ganancia de información para una validación  $T$  con  $k$  salidas, se expresa como

$$Gain(D, T) = Info(D) - \sum_{i=1}^k \frac{|D_i|}{|D|} \times Info(D_i) .$$

La información ganada por una validación esta fuertemente afectada por el número de salidas y se maximiza cuando hay un caso en cada subconjunto  $D_i$ . De otra forma, la información potencial, obtenida al particionar un conjunto de casos, esta basada en conocer el subconjunto  $D_i$  en el cual cae un determinado caso;

$$Split(D, T) = - \sum_{i=1}^k \frac{|D_i|}{|D|} \times \log_2 \left( \frac{|D_i|}{|D|} \right)$$

Esta descomposición de información tiende a incrementarse con el número de salidas de una validación. El criterio de porcentaje de ganancia evalúa la conveniencia de tomar una validación como la proporción de información ganada contra el total de información en la partición. El porcentaje de ganancia para cada posible validación es calculado y se selecciona la partición que haya obtenido el mayor porcentaje de ganancia.<sup>80</sup>

La estrategia de partición recursiva descrita anteriormente, da como resultado un árbol de decisión que es consistente con los datos de entrenamiento. En la práctica, los datos suelen ser de poca calidad – esto debido a que los atributos registran valores erróneos y los casos son mal clasificados. Este “ruido” en los datos lleva a construir árboles muy complejos, por lo que hay que tomar en cuenta dichas anomalías. La mayoría de los sistemas podan el árbol inicial, identificando subárboles que contribuyen en muy poco con la precisión de las predicciones, para finalmente reemplazarlos por un nodo hoja.<sup>81</sup>

Existen dos enfoques para podar árboles: la pre-poda (*preprunning*), en donde se detiene el crecimiento del árbol cuando la ganancia de información obtenida no supera una cota determinada y, la post-poda (*postprunning*), la cual se aplica sobre algunas ramas, una vez que se ha terminado de construir el árbol.

La pre-poda, no pierde tiempo en construir una estructura que luego será simplificada en el árbol final, busca la mejor manera de descomponer un subconjunto y evaluar la partición desde el punto de vista de la ganancia de información, reducción de errores, etc. Si esta evaluación es menor que un límite predeterminado, la división se descarta y el árbol para el subconjunto es simplemente la hoja más apropiada. Tiene la desventaja de que no es fácil detener un particionamiento en el momento adecuado, así, un límite muy alto puede terminar con la partición antes de que los beneficios de particiones siguientes parezcan evidentes, mientras que un límite demasiado bajo resulta en una simplificación demasiado corta.

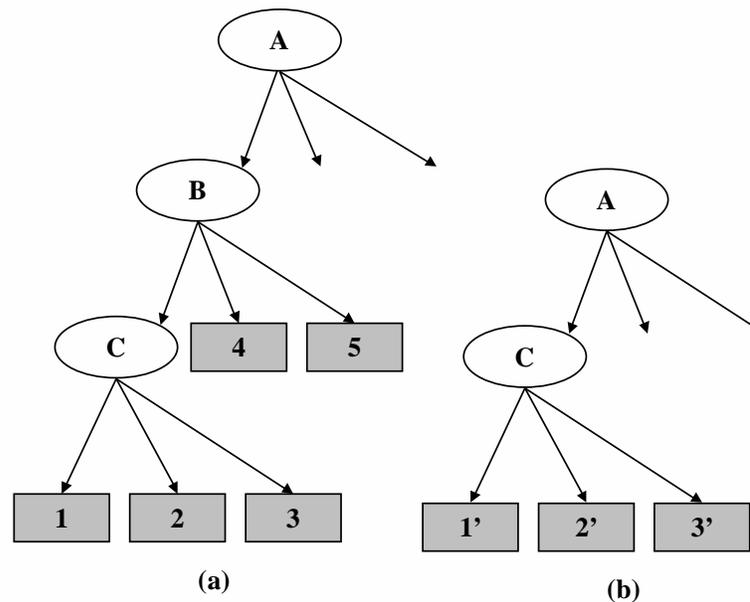
<sup>80</sup> Quinlan, J. R. (1996). Improved Use of Continuous Attributes in C4.5. *Journal of Artificial Intelligence Research*, 4 (1), 77-90.

<sup>81</sup> *Ibid.* p. 79.

La post-poda es utilizada en los algoritmos ID3 y C4.5. Una vez construido el árbol se procede a su simplificación según los criterios propios de cada uno de los algoritmos.

Witten y Frank (2000) señalan que dentro de la estrategia de post-poda existen dos tipos de operación: el reemplazo de subárboles (*subtree replacement*) y la elevación de subárboles (*subtree raising*). En la primera, la idea básica consiste en seleccionar un conjunto de subárboles y sustituirlos por nodos hoja; mientras que en la segunda, un subárbol eleva su nivel para sustituir a otro (ver Gráfica 3.5).

**Gráfica 3.5 Elevación de subárboles**



Fuente: Witten, I. H. y Frank, E. (2000). *op. cit.*, p. 163.

### 3.3 Aplicaciones del minado de datos

#### 3.3.1 Campo de acción

Las aplicaciones de las diversas técnicas del *Data Mining* tienen un vasto campo de acción en áreas como la Astronomía (*clustering* y clasificación de cuerpos celestes), la Medicina (asociación de síntomas), la Biología Molecular<sup>82</sup> (predicción de sustancias cancerígenas, genoma humano, etc.), los aspectos climatológicos (predicción de tormentas), la manufactura (diagnóstico de fallas), la Mercadotecnia (clientes potenciales, selección de productos, etc.), las inversiones en casas de bolsa y banca (análisis de clientes, aprobación de préstamos, etc.) y la detección de fraudes (telefónicos, seguros, electricidad, etc.) por mencionar solo algunas.

En Economía, por ejemplo, son muchas las posibles aplicaciones de los métodos de clasificación que intentan comprender la pertenencia a un cierto grupo, tal es el caso de la separación de clientes en fieles y desertores, la clasificación de las empresas en rentables y no rentables<sup>83</sup>, la

<sup>82</sup> Véase, Febles, J. P. y González, A. (2002). Aplicación de la minería de datos en la bioinformática. *ACIMED*, 10 (2).

<sup>83</sup> Véase, Díaz, B. y Morillas, A. (2003). *Minería de Datos y Lógica Difusa: Una aplicación al estudio de la rentabilidad económica de las empresas agroalimentarias en Andalucía*. Trabajo presentado en el VI Encuentro de Economía Aplicada, Junio 5, Granada.

determinación de las claves que conducen a un nuevo producto al éxito o al fracaso, así como la categorización del riesgo asociado a una solicitud de crédito, tanto en el caso de los créditos personales como en los empresariales.

En Internet, el *Data Mining* aporta básicamente dos aplicaciones para el apoyo en la toma de decisiones: en primer lugar, el *Data Mining* se puede utilizar para analizar los datos generados por un *website* en la red; y en segundo lugar, el *Data Mining* se puede utilizar para monitorear y detectar la aparición de anomalías y problemas potenciales en la red antes de que estos ocurran.<sup>84</sup>

Como se puede apreciar, el campo de acción del *Data Mining* es muy extenso y variado, se podría decir que es aplicable a cualquier campo del saber, sin embargo, no siempre puede ser aplicado con éxito, y esto se debe a que muy a menudo no se cuenta con los suficientes datos que permitan obtener información de interés.

### 3.3.2 El caso de la industria aseguradora

El *data mining* se está utilizando en la industria aseguradora para analizar y predecir con precisión cada fase del ciclo de vida del cliente, los prospectos que podrían convertirse en clientes, así como los productos que estos mismos comprarían, sin olvidar a aquellos que tienden a cancelar sus pólizas.<sup>85</sup>

Las primeras aplicaciones se hicieron en compañías aseguradoras de los Estados Unidos en la década de los 90's. Por ejemplo, en 1996, IBM desarrolló un paquete de software para predecir la rentabilidad de los clientes de una institución de seguros de Los Ángeles, California (*Farmers Insurance Group*). El software, llamado *Underwriting Profitability Analysis Business Tool*, utilizaba avanzados algoritmos de estadística y matemáticas para dividir a los clientes en categorías basadas en riesgo. Gracias a este tipo de desarrollos, y al rápido crecimiento del *hardware*, el sector asegurador de los Estados Unidos ha tomado ventaja de las tecnologías de la minería de datos para poder analizar enormes cantidades de información de todos sus clientes.<sup>86</sup>

En palabras de John Kuntz, Vicepresidente de Marketing en *Prudential Insurance Company*, "el *Data Mining* es todo aquello que nos ayude a entender mejor a nuestro cliente, ya sea una simple distribución de frecuencias o un complejo modelo de redes neuronales".

El conocer y entender más al cliente implica tener información más precisa de su comportamiento y sus preferencias. Por ejemplo, la mayoría de los aseguradores saben que si un cliente tiene dos o más pólizas con una compañía es más probable que renueve, que aquel que solo tiene una sola póliza. ¿Pero cómo saber que tipo de clientes están dispuestos a comprar más de dos pólizas? Muchas compañías aseguradoras norteamericanas y canadienses utilizan la técnica de *data mining* llamada "análisis de asociación", con ello pueden saber con más certeza que tipo de productos y servicios deben ofrecerse a tal o cual cliente.<sup>87</sup>

De este modo, los aseguradores de países como Estados Unidos y Canadá, se han dado cuenta del enorme potencial de las técnicas del *data mining* para lograr beneficios como el crecimiento

<sup>84</sup> Véase, Vega, A. (2000). Web-mining en el marco de los procesos de negocio y marketing de la empresa. *Comunicaciones de Telefónica I+D*, 19 (1), 125-130.

<sup>85</sup> Wincuin, J. (1999). Tapping the mountain of consumer information. *LIMRA'S MarketFacts*, 18 (6), 26-29.

<sup>86</sup> Rosen, M. (1998). There's gold in that there data. *Insurance & Technology*, 23 (12), 28-32.

<sup>87</sup> Turney, M. (2003). Intelligence behind data. *Canadian Underwriter*, 70 (5), 44-45.

en el mercado o el incremento en la rentabilidad. En este sentido, existe un gran número de investigaciones que apoyan el uso del *data mining* dentro de la industria del seguro.<sup>88</sup>

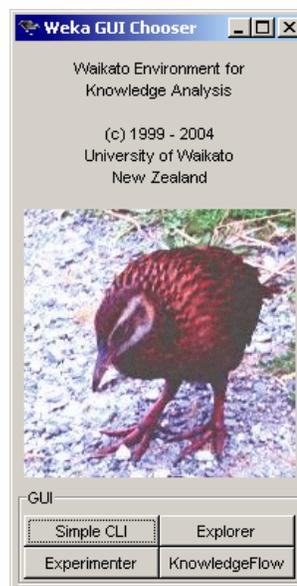
En contraste, en el sector asegurador latinoamericano y en particular en el mexicano, existen muy pocos estudios formales que demuestran la factibilidad del uso de las técnicas de *data mining* dentro del ámbito de seguros, esto puede ser preocupante, ya que tal vez se esté perdiendo una magnífica oportunidad de impulsar el crecimiento y la rentabilidad de las empresas aseguradoras de la región, a través del uso de las técnicas de minería de datos.

### 3.4 WEKA

#### 3.4.1 Software de minería de datos en Java

WEKA<sup>89</sup> fue desarrollado en la Universidad de Waikato en Nueva Zelanda. El sistema está desarrollado en Java, un lenguaje de programación orientado a objetos disponible para la mayoría de las plataformas de cómputo. WEKA ha sido probado bajo sistemas operativos como Linux, Windows y Macintosh. El lenguaje Java permite ofrecer una interfase uniforme a los diferentes algoritmos de aprendizaje, todos ellos con métodos de preproceso, postproceso y evaluación de resultados de los diferentes esquemas de aprendizaje.<sup>90</sup> (Ver Gráfica 3.6).

**Gráfica 3.6 WEKA (Waikato Environment for Knowledge Analysis)**



Hay diferentes niveles en los que WEKA puede ser utilizado. El primero, ofrece implementaciones de los algoritmos de aprendizaje pertenecientes al estado del arte, los cuales

<sup>88</sup> Véase, Smith, K. A., Willis, R. J. y Brooks, M. (2000). An analysis of customer retention and insurance claim patterns using data mining: a case study. *Journal of the Operational Research Society*, 51 (5), 532-541.

<sup>89</sup> Waikato Environment for Knowledge Analysis.

<sup>90</sup> Witten, I. H. y Frank, E. *op. cit.*, p. 265.

pueden ser aplicados a conjuntos de datos mediante una línea de comandos. También se incluyen una gran variedad de herramientas para transformar datos, como los algoritmos de discretización. Asimismo se puede preprocesar un conjunto de datos, alimentándolo dentro de un esquema de aprendizaje, analizando el clasificador resultante y su desempeño, todo ello sin escribir una sola línea de código.<sup>91</sup>

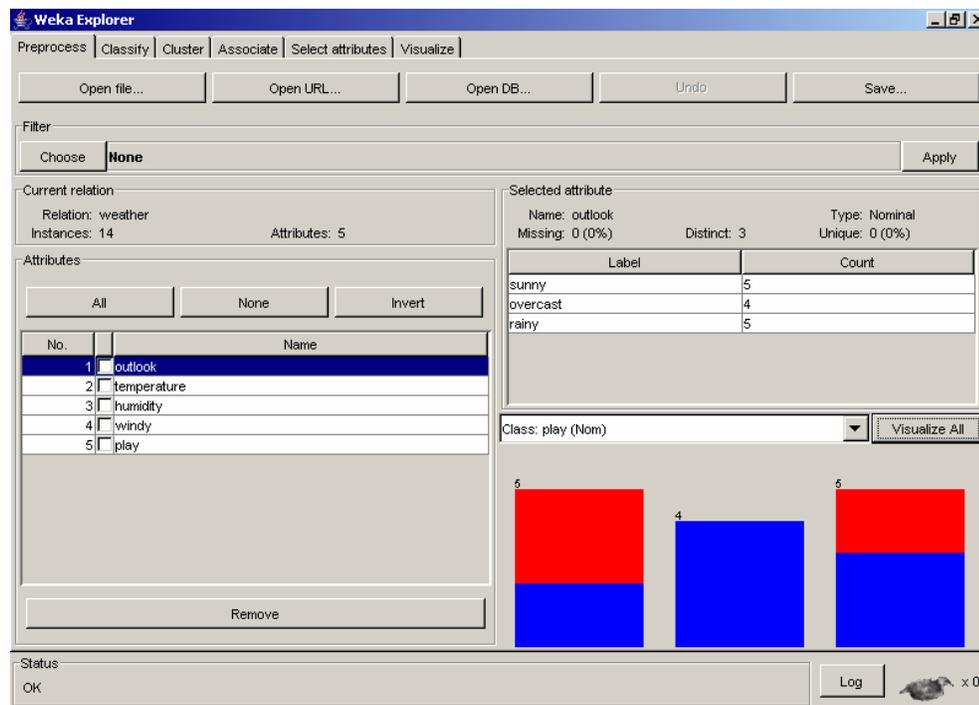
Cabe señalar que el enfoque principal de WEKA esta en los algoritmos de clasificación y filtrado. Sin embargo, también incluye implementaciones de algoritmos de reglas de asociación y *clustering*.

### 3.4.2 WEKA Knowledge Explorer

Dentro del ambiente gráfico de WEKA se encuentra el módulo exploración de datos (*WEKA Knowledge Explorer*) el cual permite preprocesar, clasificar, asociar y visualizar datos de una manera fácil e intuitiva (ver Gráfica 3.7).

WEKA cuenta con una gran variedad de algoritmos de clasificación (*classifiers*) entre los que destacan los métodos bayesianos (AODE, Bayes Net, Naive Bayes, Naive Bayes Simple, Naive Bayes Multinomial, etc.), las reglas de clasificación (Tablas de Decisión, Nnge, OneR, PART, Ridor, ZeroR, etc.), los métodos de regresión (Regresión Lineal, Regresión Logística, etc.) y los árboles de decisión (ADTree, Decisión Stump, ID3, J48, LMT, Random Forest, Random Tree, REP Tree, etc.).

Gráfica 3.7 WEKA Knowledge Explorer



<sup>91</sup> Witten, I. H. y Frank, E. *op. cit.*, p. 265-266.

### 3.4.3 El algoritmo J4.8

El algoritmo J4.8 induce árboles de decisión. Es la implementación en WEKA del algoritmo C4.5 revisión 8, la cual fue la última versión pública de esta familia de algoritmos, posteriormente apareció la primera implementación comercial, es decir, el algoritmo C5.0.

Dentro de las opciones que J4.8 soporta están:

- La poda de árboles
- La especificación de factores de confianza para la poda
- La especificación de un mínimo de instancias en las hojas
- La poda de árboles con error reducido
- La especificación del número de datos en podas con error reducido
- El uso de particiones binarias en atributos nominales

A continuación se presenta la salida de un árbol de decisión producido por el algoritmo J4.8:

```
==== Run information ====

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    weather
Instances:   14
Attributes:  5
              outlook
              temperature
              humidity
              windy
              play
Test mode:   10-fold cross-validation

==== Classifier model (full training set) ====

J48 pruned tree
-----

outlook = sunny
| humidity <= 75: yes (2.0)
| humidity > 75: no (3.0)
outlook = overcast: yes (4.0)
outlook = rainy
| windy = TRUE: no (2.0)
| windy = FALSE: yes (3.0)

Number of Leaves :      5

Size of the tree :      8

Time taken to build model: 0.07 seconds

==== Stratified cross-validation ====
```

=== Summary ===

Correctly Classified Instances	9	64.2857 %
Incorrectly Classified Instances	5	35.7143 %
Kappa statistic	0.186	
Mean absolute error	0.2857	
Root mean squared error	0.4818	
Relative absolute error	60 %	
Root relative squared error	97.6586 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.778	0.6	0.7	0.778	0.737	yes
0.4	0.222	0.5	0.4	0.444	no

=== Confusion Matrix ===

```
a b <-- classified as
7 2 | a = yes
3 2 | b = no
```

## **CAPÍTULO IV. LA EMPRESA EN ESTUDIO Y SUS PRODUCTOS AFECTADOS**

---

### **4.1 Introducción**

En el presente capítulo se muestra el entorno de la empresa en estudio, así como la problemática a la que se enfrenta hoy en día: los altos niveles de cancelación en dos de sus principales productos.

#### **4.1.1 Empresa en estudio**

La empresa en estudio nace en la década de los 90's para funcionar como institución de seguros en los ramos de vida, accidentes y enfermedades. La estrategia que ha seguido desde su inicio es la de ofrecer productos simples, integrales e innovadores en forma masiva; con lo cual, ha podido llegar a segmentos de la población que anteriormente no habían sido atendidos por las compañías tradicionales de seguros.

### **4.2 Descripción de los productos**

#### **4.2.1 Accidentes Personales en moneda nacional**

Fue el primer producto que lanzó la empresa, inicialmente se comercializó a través de correo directo y posteriormente se integró al portafolio de productos que se ofrecen en sus sucursales.

Este producto es un seguro temporal a un año, renovable automáticamente hasta los 70 años de edad, con una edad mínima de contratación de 18 años; la cobertura es por muerte accidental<sup>92</sup>; no aplican deducibles, coaseguros ni periodos de espera; y no se requiere de ningún examen médico para su contratación.

#### **4.2.2 Vida Individual en moneda nacional**

El producto se lanzó inicialmente en sus sucursales. Su introducción fue todo un éxito ya que omitió el proceso de suscripción que se manejaba hasta ese momento.

Consiste en un seguro temporal a un año, renovable automáticamente hasta los 66 años de edad, siempre que se contrate entre los 18 y 60 años de edad; cubre el fallecimiento del asegurado por enfermedad o accidente cubiertos, pudiendo otorgar anticipos de suma

---

<sup>92</sup> Se entiende por accidente a aquel acontecimiento proveniente de una causa externa, súbita, violenta y fortuita, que ocurra mientras se encuentre en vigor la cobertura de la póliza, y que produzca lesiones corporales o la muerte en la persona del Asegurado. No se considerarán accidentes las lesiones corporales o la muerte provocadas intencionalmente por el Asegurado.

asegurada en caso de enfermedad terminal.<sup>93</sup> Se puede contratar en Plan Individual o Familiar, en este último, se otorga un porcentaje de descuento sobre la prima de cada persona.

### 4.2.3 Canales de venta

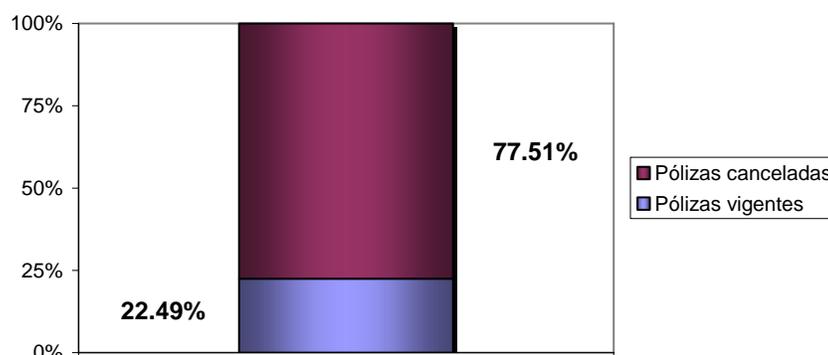
Ambos productos se ofrecen a través de los distintos canales de venta con los que cuenta la empresa, sin embargo, el mayor volumen de colocación lo sigue representando la distribución por sucursales.

## 4.3 Niveles de cancelación

### 4.3.1 Histórico de cancelación

Desde su año de lanzamiento, el producto de Accidentes Personales se ha mantenido como un producto de gran colocación en las sucursales, sin embargo, los niveles históricos de cancelación han sido muy altos, llegando incluso al 77.51% del total emitido (ver Gráfica 4.1).

**Gráfica 4.1 Histórico de cancelación en Accidentes Personales en M.N.**



Fuente: Elaboración propia.

El mismo efecto se aprecia en el producto de Vida Individual, y aunque el lanzamiento de este último fue posterior, los niveles históricos de cancelación son muy parecidos al producto de Accidentes Personales, manteniéndose por arriba del 75% (ver Gráfica 4.2).

### 4.3.2 Tasa de retención y vida media

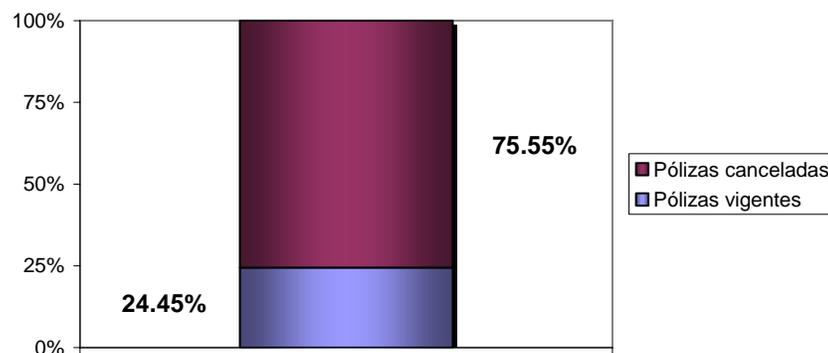
Como se mencionó en el Capítulo 2, la tasa de retención de clientes es el indicador básico para saber en que medida se esta fidelizando a los clientes de una empresa. Conceptualmente,

<sup>93</sup> Se entiende por Enfermedad Terminal a aquel padecimiento o enfermedad que ocasionará necesaria o muy probablemente la muerte del Asegurado dentro de un lapso no mayor a seis meses contados a partir de la fecha en que haya sido diagnosticada dicha enfermedad.

equivale al porcentaje de consumidores o clientes que se mantienen en una empresa entre un periodo y otro. Aunque normalmente este periodo se toma de forma anual, cualquier otra partición del tiempo es válida (mensual, trimestral, semanal, etc.), todo depende de las necesidades y del enfoque del análisis.

En el caso de la empresa en estudio, se obtuvieron las tasas de retención para los cuatro trimestres del 2004. Con ello, se calculó la tasa de retención trimestral promedio para Accidentes Personales, la cual fue del 24.32%; de igual forma, en Vida Individual esta misma tasa llegó al 26.73%.

**Gráfica 4.2 Histórico de cancelación en Vida Individual en M.N.**



Fuente: Elaboración propia.

Todos estos datos nos muestran que la empresa en estudio está teniendo serios problemas para fidelizar a sus clientes, y es que en ambos productos la vida media de las pólizas está apenas por arriba de los tres meses, lo cual resulta, hasta cierto punto, costoso para la institución.

Ante tal panorama, la empresa debe empezar a preocuparse mucho más por fidelizar a sus clientes, de lo contrario perderá una magnífica oportunidad en materia de retención de clientes, lo cual podría impactar directamente su rentabilidad.

## CAPÍTULO V. PREPARACIÓN DE LOS DATOS

---

### 5.1 Proceso de extracción de conocimiento

Como se vio en el Capítulo 3, los pasos que componen al «proceso de extracción de conocimiento en bases de datos» (*KDD*) son cinco:

1. Selección del objetivo
2. Preproceso de datos
3. Transformación de datos
4. Minado de datos
5. Interpretación de resultados

Los primeros tres pasos están orientados básicamente a la preparación de los datos, mientras que los últimos dos, corresponden tanto al minado de datos, propiamente dicho, como a la interpretación de los resultados arrojados por los algoritmos de minado.

### 5.2 Selección del objetivo

#### 5.2.1 Problema observado

La selección del objetivo tiene como finalidad estudiar el problema y decidir cuál es la meta del proyecto. En nuestro caso, el problema observado es el de los altos niveles de cancelación de pólizas que se presentan en una de las principales empresas de seguros de nuestro país.

#### 5.2.2 Meta del proyecto

Con base en lo anterior, se ha fijado una meta para el proyecto, y esta consiste en encontrar un modelo de minería de datos que permita predecir el comportamiento de los clientes y con ello, evitar al máximo la cancelación de contratos de seguros, esto tanto en la parte de Vida Individual como en la de Accidentes Personales.

#### 5.2.3 Fuente de datos

Una vez definido el problema, se deben identificar fuentes de datos y posteriormente, seleccionar un subconjunto que sea útil para la aplicación de un algoritmo de minería de datos.

Para efectos del presente trabajo, la empresa en estudio proporcionó datos relativos a pólizas canceladas de Vida Individual y Accidentes Personales. Por lo que toca al algoritmo de minado de datos, se eligió el algoritmo J4.8, el cual induce árboles de decisión. Se ha optado por la inducción de árboles de decisión porque además de ser la técnica más común dentro las técnicas de clasificación de datos, representa una gran ventaja con respecto a las demás

técnicas de clasificación: poder representar el conocimiento extraído en un conjunto de reglas de decisión de fácil entendimiento.

#### 5.2.4 Muestra inicial

Como primer paso, se extrajo una muestra de los datos de pólizas canceladas. La información se generó en un archivo de texto para tener mayor flexibilidad al momento de exportar los datos. Esta muestra inicial de datos constó de 204,205 registros, los cuales correspondían a pólizas de Vida Individual y Accidentes Personales canceladas entre el 1 de Julio de 2004 y el 31 de Diciembre del mismo año (ver Cuadro 5.1).

**Cuadro 5.1 Atributos seleccionados en la primera muestra de datos**

Muestra Inicial
Producto
Edad de emisión
Sexo del asegurado
Estado civil
Entidad federativa
Código postal
Suma asegurada
Prima de recibo
Forma de pago
Día de facturación
Punto de venta
Tipo de cuenta
Meses pagados

*Fuente: Elaboración propia.*

Cabe señalar que esta primera muestra contenía información no restringida de los contratos de seguros, por lo que en todo momento se respetó la confidencialidad de los datos del cliente. Posteriormente, dicha información se cargó a una Base de Datos Relacional para su mejor manejo. De esta forma, se dejó todo preparado para entrar a la siguiente fase del «proceso de extracción de conocimiento»: el preproceso de datos.

### 5.3 Preproceso de datos

#### 5.3.1 Introducción

El preproceso de datos consiste en estudiar los datos seleccionados para entender el significado de los atributos, detectar errores de integración, estandarizar datos, hacer agrupaciones, etc.

Para la muestra inicial, el preproceso consistió básicamente en la estandarización y el agrupamiento de datos, ya que, como se mencionó anteriormente, dicha muestra fue proporcionada por la empresa en estudio.

### 5.3.2 Atributos preprocesados

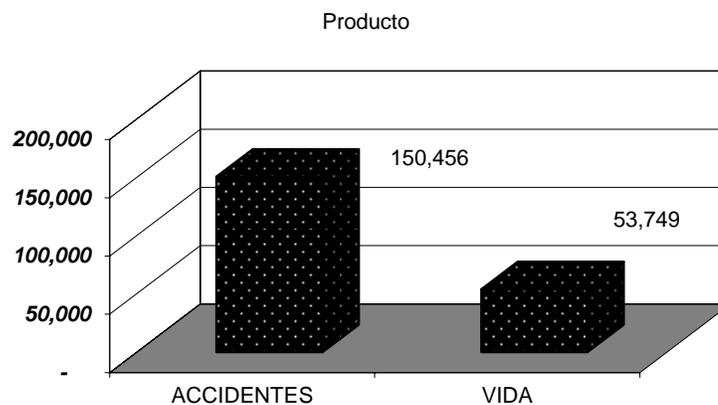
#### 5.3.2.1 Producto

El atributo fue estandarizado, cambiando los valores internos del sistema por los valores "VIDA" y "ACCIDENTES", los cuales corresponden a los dos productos en estudio, es decir, Vida Individual y Accidentes Personales respectivamente (ver Gráfica 5.1).

#### 5.3.2.2 Edad de emisión

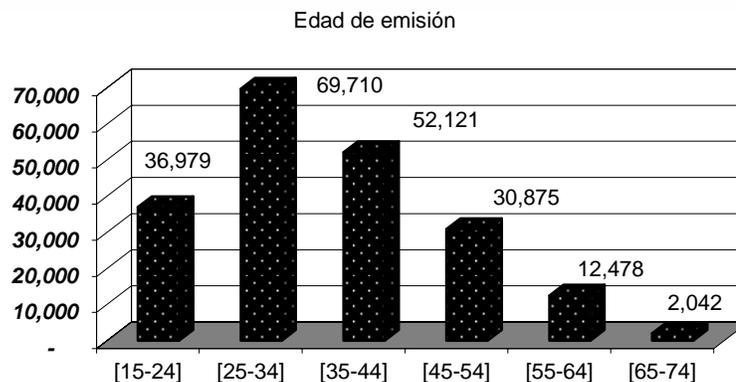
El atributo fue agrupado en seis categorías, situándose el mayor número de registros en la categoría de 25 a 34 años (ver Gráfica 5.2).

**Gráfica 5.1 Atributo "Producto"**



Fuente: Elaboración propia.

**Gráfica 5.2 Atributo "Edad de emisión"**

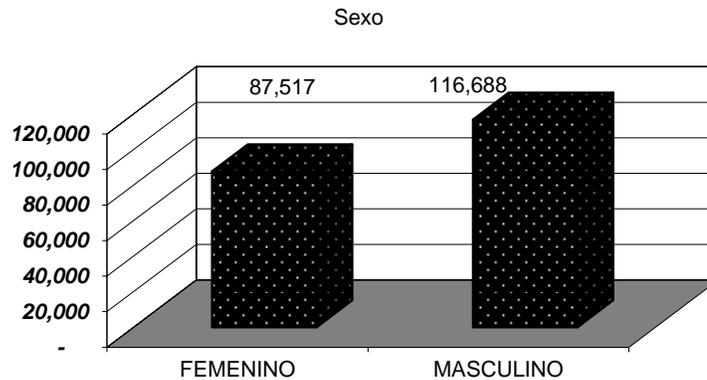


Fuente: Elaboración propia.

### 5.3.2.3 Sexo del asegurado

El atributo fue estandarizado, cambiando los valores internos del sistema por los valores "FEMENINO" y "MASCULINO" (ver Gráfica 5.3).

**Gráfica 5.3 Atributo "Sexo del asegurado"**

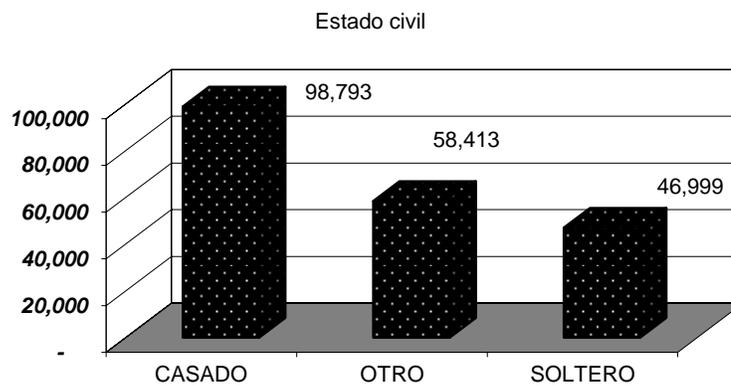


Fuente: Elaboración propia.

### 5.3.2.4 Estado civil

El atributo fue estandarizado y agrupado. Originalmente los valores extraídos del sistemas eran "CASADO", "DESCONOCIDO", "DIVORCIADO", "SOLTERO" y "VIUDO", sin embargo, los casos de personas viudas o divorciadas eran considerablemente menores al resto de las categorías, por esta razón, se agruparon las categorías minoritarias para dejar únicamente los valores "CASADO", "OTRO" y "SOLTERO" (ver Gráfica 5.4).

**Gráfica 5.4 Atributo "Estado civil"**



Fuente: Elaboración propia.

### 5.3.2.5 Entidad federativa

El atributo fue estandarizado y reagrupado. Como primer paso, se cambiaron los valores internos del sistema por valores de tipo cadena (ver Cuadro 5.2). Posteriormente, se agruparon los estados en zonas geográficas de venta, esto con la finalidad orientar el estudio hacia la búsqueda de comportamientos regionales (ver Gráfica 5.5).

### 5.3.2.6 Código postal

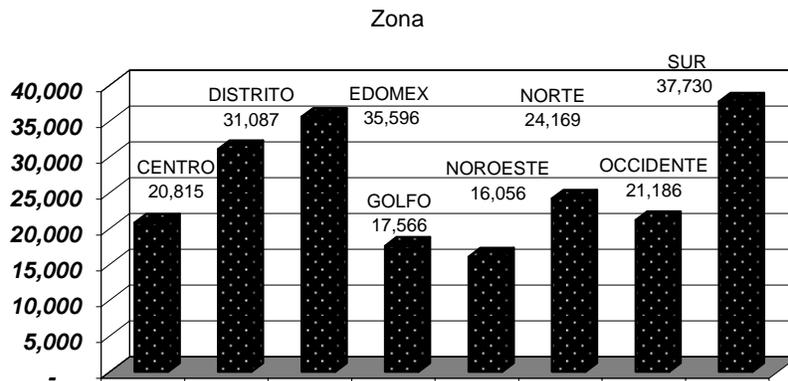
Este atributo fue omitido ya que representa una redundancia en la información (esta directamente relacionado con la zona geográfica de venta).

**Cuadro 5.2 Zonas geográficas de venta**

CENTRO	AGUASCALIENTES	NORTE	CHIHUAHUA	
	GUANAJUATO		COAHUILA	
	HIDALGO		DURANGO	
	QUERETARO		NUEVO LEON	
	SAN LUIS POTOSI		TAMPICO	
	TLAXCALA		OCCIDENTE	COLIMA
	ZACATECAS			JALISCO
GOLFO	CAMPECHE	MICHOACÁN		
	QUINTANA ROO	NAYARIT		
	TABASCO	SUR	CHIAPAS	
	VERACRUZ		GUERRERO	
	YUCATAN		MORELOS	
DISTRITO	DISTRITO FEDERAL		OAXACA	
EDOMEX	ESTADO DE MÉXICO		PUEBLA	
NOROESTE	BAJA CALIFORNIA	NOROESTE	BAJA CALIFORNIA SUR	
	SINALOA		SONORA	

Fuente: Elaboración propia.

**Gráfica 5.5 Atributo "Zona geográfica de venta"  
(Antes "Entidad federativa")**



Fuente: Elaboración propia.

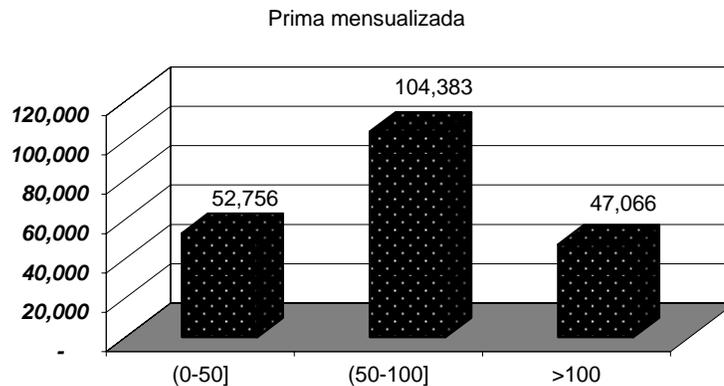
### 5.3.2.7 Suma asegurada

Este atributo fue omitido ya que representa una redundancia en la información (esta directamente relacionado con la prima de recibo).

### 5.3.2.8 Prima de recibo

El atributo fue recalculado y agrupado. En primera instancia, se recalculó la prima en los casos en que el pago no fuera de forma mensual, esto con la finalidad de estandarizar la forma de pago. Posteriormente, se agruparon los registros en tres categorías de prima mensualizada: menor o igual a \$50.00, mayor a \$50.00 y menor o igual a \$100.00, y finalmente, mayor a \$100.00 (ver Gráfica 5.6).

**Gráfica 5.6 Atributo "Prima mensualizada"  
(Antes "Prima de recibo")**



Fuente: Elaboración propia.

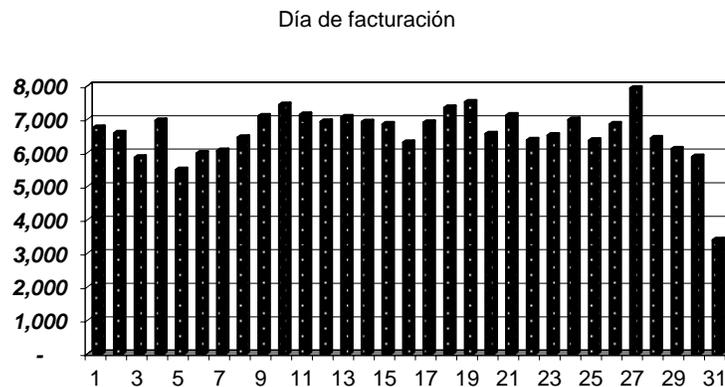
### 5.3.2.9 Forma de pago

Este atributo fue omitido ya que se mensualizó la prima de recibo.

### 5.3.2.10 Día de facturación

El atributo no sufrió ningún cambio. Indica el día del mes en que el sistema efectúa los cargos a la cuenta del cliente (ver Gráfica 5.7).

**Gráfica 5.7 Atributo "Día de facturación"**



Fuente: Elaboración propia.

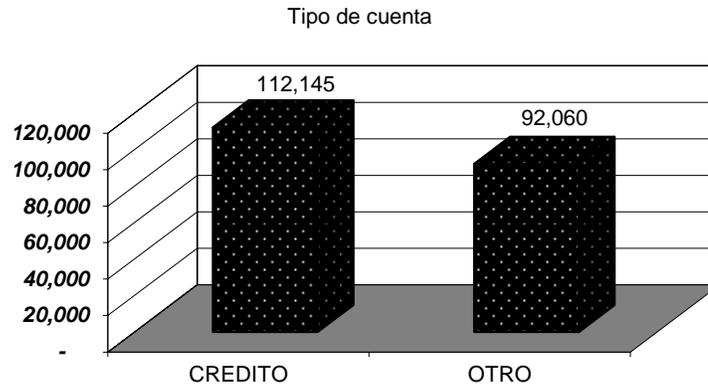
### 5.3.2.11 Punto de venta

Este atributo fue omitido ya que el 98% de los registros pertenecían a pólizas comercializadas a través de la red de sucursales.

### 5.3.2.12 Tipo de cuenta

El atributo fue estandarizado y agrupado. Originalmente los valores extraídos del sistema eran "CHEQUES", "CREDITO", "DEBITO" y "OTRO", sin embargo, los casos de personas con cuentas de cheques y tarjetas de débito eran considerablemente menores al de los clientes con tarjetas de crédito, por esta razón se reagruparon y se dejaron los valores "CREDITO" y "OTRO" (ver Gráfica 5.8).

**Gráfica 5.8 Atributo "Tipo de cuenta"**

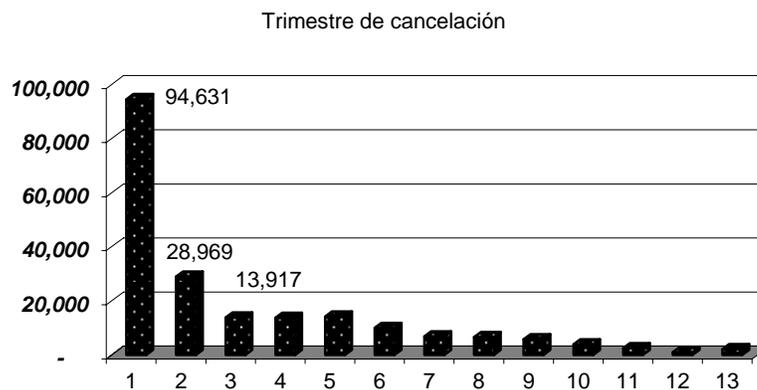


Fuente: Elaboración propia.

### 5.3.2.13 Meses pagados

El atributo fue agrupado. Originalmente la muestra contenía en este campo el número de meses que pagó el asegurado. Debido a que existían demasiadas categorías con esta definición, se optó por agrupar la información en trimestres. De esta forma, el atributo ya agrupado nos indica en qué trimestre se canceló el contrato de seguro (ver Gráfica 5.9).

**Gráfica 5.9 Atributo "Trimestre de cancelación" (Antes "Meses pagados")**



Fuente: Elaboración propia.

La problemática a la que se enfrenta la empresa en estudio se ve claramente en la gráfica anterior, debido a que la gran mayoría de los contratos emitidos en sucursales se cancelan durante el primer trimestre de vigencia. En consecuencia, el reto para la empresa aseguradora

es claro: conservar vigentes el mayor número de pólizas de seguros que tienden a cancelarse durante el primer trimestre de vigencia.

## 5.4 Transformación de datos

### 5.4.1 Atributos finales

Una vez que la muestra inicial de datos fue preprocesada se obtuvieron los atributos finales para la construcción del modelo (ver Cuadro 5.3).

**Cuadro 5.3 Atributos finales**

Muestra Final
Producto
Edad de emisión
Sexo del asegurado
Estado civil
Zona geográfica
Prima mensualizada
Día de facturación
Tipo de cuenta
Trimestre de cancelación

*Fuente: Elaboración propia.*

### 5.4.2 Formato ARFF

El siguiente paso fue la transformación final de los datos, esto con el fin de que se ajustaran al formato de entrada del algoritmo J4.8 de WEKA.

WEKA interpreta archivos de datos en formato ARFF (*Attribute-Relation File Format*).<sup>94</sup> Este formato consiste en un archivo de texto ASCII que describe una lista de instancias que comparten una serie de atributos.

Los archivos ARFF se componen de dos secciones: el encabezado y la sección de datos. El encabezado contiene el nombre de la relación, la lista de atributos y el tipo de atributo. La sección de datos contiene instancias cuya información va separada por comas. A continuación se muestra un ejemplo de un archivo ARFF:

```
@relation non-Hodgkin-lymphoma
@attribute AGE numeric
@attribute GENERAL-HEALTH-STATUS {poor, average, good}
@attribute BULKY-DISEASE {no, yes}
```

<sup>94</sup> WEKA también tiene la capacidad de aceptar otros formatos como CSV y TAB-Delimited, sin embargo, el formato ARFF es el que otorga mayor funcionalidad.

```

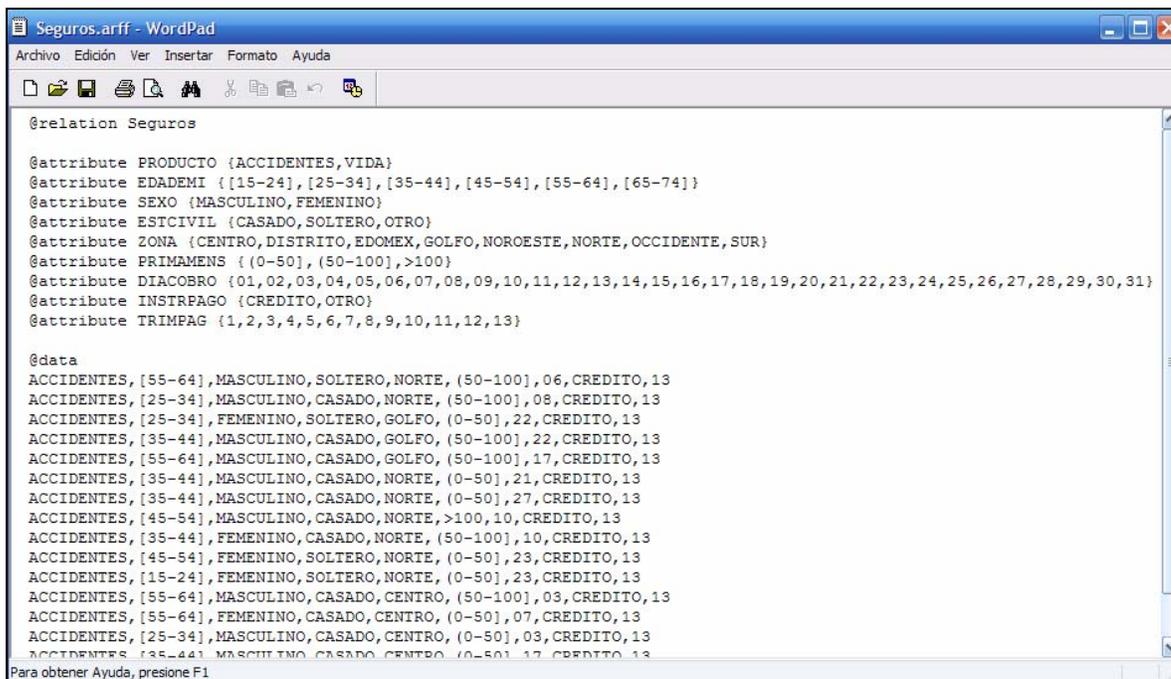
@attribute HISTOLOGICAL-CLASSIFICATION {low-grade, high-grade}
@attribute CLINICAL-STAGE {I, II1, II2, III, IV}
@attribute CLINICAL-PRESENTATION {none, hemorrhage,
perforation}
@attribute CT-RT-SCHEDULE {none, RT, CT, CT-next-RT}
@attribute SURGERY {none, palliative, curative}
@attribute 5-YEAR-RESULT {alive, death}
@data
10,average,no,high-grade,I,hemorrhage,RT,palliative,death
5,good,no,low-grade,I,hemorrhage,RT,none,alive
9,good,yes,high-grade,II2,none,RT,palliative,death
3,good,yes,high-grade,II2,none,RT,none,death
9,good,no,low-grade,I,none,RT,none,death
8,good,no,high-grade,II2,none,RT,palliative,death
11,good,yes,high-grade,II2,none,?,none,death
9,average,yes,high-grade,I,none,?,none,death
9,good,no,high-grade,I,none,RT,palliative,death
3,good,yes,high-grade,II2,none,RT,none,death
...

```

### 5.4.3 Seguros.arff

De esta forma, con la información final ya cargada en la Base de Datos Relacional, fue muy sencillo generar el archivo «Seguros.arff» (ver Gráfica 5.10).

**Gráfica 5.10 Archivo Seguros.arff**



Este archivo contiene 204,205 instancias con los nueve atributos finales. Su tamaño es de 11.8 MB y tarda menos de un minuto en ser cargado por WEKA.<sup>95</sup>

<sup>95</sup> Tiempo medido en una PC con sistema operativo “Windows 2000” y 256 MB en memoria.

## CAPÍTULO VI. MODELO DE MINERÍA DE DATOS

### 6.1 Minado de datos

#### 6.1.1 Introducción

El minado de datos es la etapa principal del «proceso de extracción de conocimiento en bases de datos». Durante esta fase, se aplican los diferentes algoritmos de extracción de conocimiento en datos que ya han sido preparados y transformados en etapas anteriores.

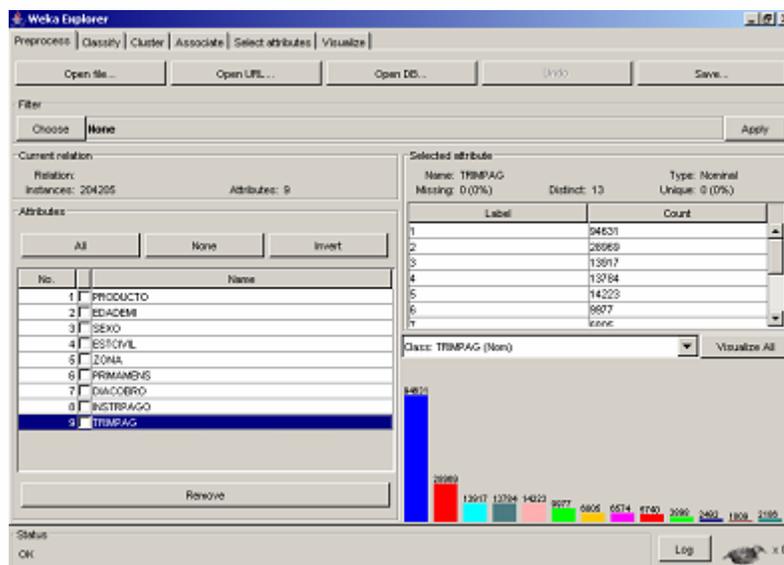
En el caso del presente trabajo, se ha elegido al algoritmo J4.8 de WEKA para la construcción de un modelo de minería de datos. El algoritmo J4.8 genera árboles de decisión los cuales clasifican los datos en clases o categorías. Esta clasificación sirve además para predecir la clase de instancias que aún no han sido clasificadas.

#### 6.1.2 Aplicación del algoritmo J4.8

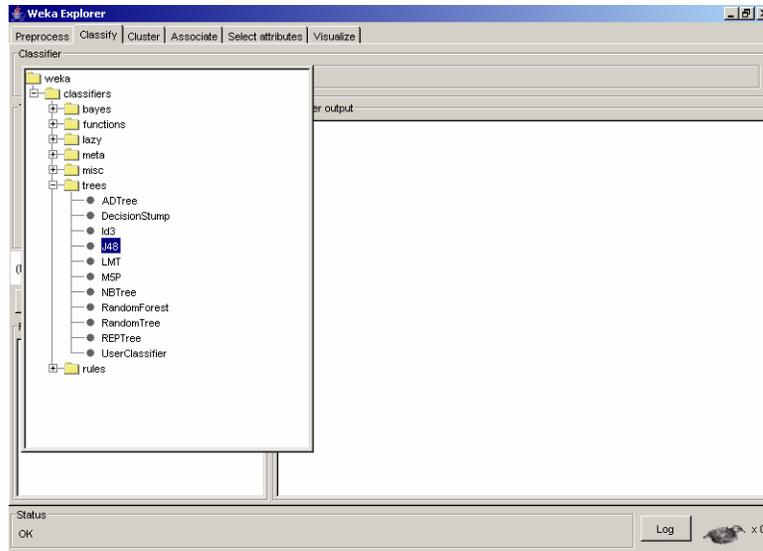
Una vez generado el archivo «Seguros.arff», se procedió a la carga del mismo en WEKA (ver Gráfica 6.1).

Posteriormente, se eligió dentro del módulo de exploración de datos (*WEKA Knowledge Explorer*) el clasificador J4.8 (ver Gráfica 6.2).

Gráfica 6.1 Carga de datos en WEKA



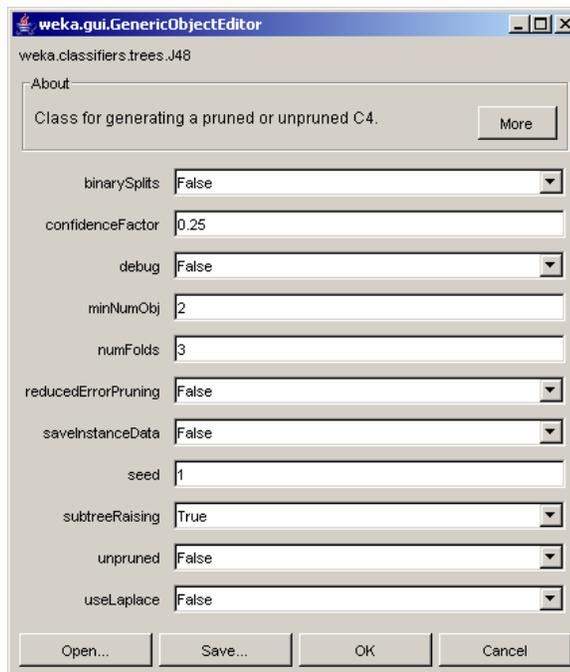
Gráfica 6.2 Clasificador J4.8



El algoritmo J4.8 maneja una serie de parámetros que permiten construir árboles de decisión de acuerdo a las necesidades del usuario (ver Gráfica 6.3).

Para el caso del presente modelo de *data mining*, se utilizaron los parámetros predeterminados por WEKA, los cuales se describen a detalle en las siguientes secciones.

Gráfica 6.3 Parámetros del algoritmo J4.8



#### **6.1.2.1 Parámetro «binarySplits»**

El parámetro «*binarySplits*» se utiliza para establecer si se utilizarán particiones binarias en atributos nominales. El valor por omisión es *False*.

#### **6.1.2.2 Parámetro «confidenceFactor»**

En este parámetro se indica el nivel de confianza para el proceso de poda, entre más pequeño sea este valor, mayor será el nivel de poda. El valor predeterminado por WEKA es *0.25*.

#### **6.1.2.3 Parámetro «debug»**

Si se requiere mayor información sobre el proceso de inducción, este parámetro debe tener el valor *True*. Para WEKA el valor por omisión es *False*.

#### **6.1.2.4 Parámetro «minNumObj»**

El parámetro «*minNumObj*» indica el mínimo de instancias que se deben considerar en cada hoja del árbol. El valor predeterminado es *2*.

#### **6.1.2.5 Parámetro «numFolds»**

Determina la cantidad de datos usados en la poda tipo “*reduced-error*”. La primera separación de datos es utilizada para la poda, y el resto se utiliza para desarrollar el árbol. El valor por omisión es *3*.

#### **6.1.2.6 Parámetro «reducedErrorPruning»**

En este parámetro se indica si el proceso de poda será de tipo “*reduced-error*”. El valor predeterminado es *False*, esto indica que la poda será del tipo C4.5.

#### **6.1.2.7 Parámetro «saveInstanceData»**

El parámetro «*saveInstanceData*» indica si los datos de entrenamiento serán salvados para su visualización. El valor predeterminado es *False*.

#### **6.1.2.8 Parámetro «seed»**

Este parámetro es utilizado en el proceso aleatorio de datos para la poda de tipo “*reduced-error*”. El valor por omisión es *1*.

#### **6.1.2.9 Parámetro «subtreeRaising»**

Determina si es utilizado el proceso de elevación de subárboles. El valor predeterminado es *True*.

#### **6.1.2.10 Parámetro «unpruned»**

Determina si es utilizado el proceso de poda. El valor predeterminado es *False*.

### 6.1.2.11 Parámetro «useLaplace»

En este parámetro se indica si el conteo en las salidas del modelo es suavizado mediante el criterio de Laplace. El valor por omisión es *False*.

De esta forma, y una vez establecidos los parámetros del algoritmo, se ejecutó la inducción del árbol de decisiones a través de WEKA.

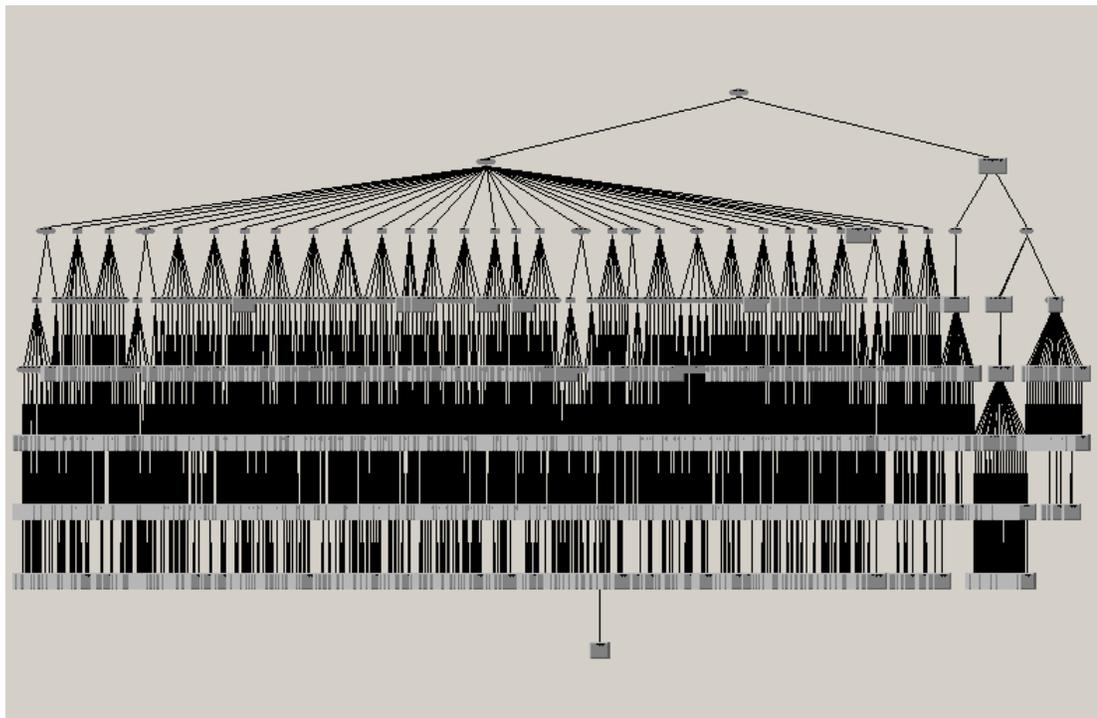
## 6.2 Interpretación de resultados

### 6.2.1 Árbol de decisión obtenido

Como resultado de la inducción, WEKA obtuvo un árbol de decisión de tamaño 6,421 (suma de nodos internos y "nodos hoja"), con 4,318 salidas o "nodos hoja" (ver Gráfica 6.4).

El modelo se construyó a partir de 134,775 instancias, las cuales fueron utilizadas como conjunto de entrenamiento (66% de los datos disponibles). El resto de los datos (34% de la muestra) sirvieron para probar el modelo obtenido.

Gráfica 6.4 Árbol de decisión obtenido



El proceso de *testing* arrojó que el 50.53% de las instancias fueron clasificadas correctamente, mientras que el 49.47% se clasificaron de manera incorrecta (ver Cuadro 6.1). En principio, estos resultados podrían parecer poco alentadores, sin embargo, en las siguientes secciones analizaremos el modelo a detalle para tratar de obtener algunos datos de interés.

**Cuadro 6.1 Evaluación en los datos de prueba**

```

==== Evaluation on test split ====
==== Summary ====

Correctly Classified Instances      35086      50.5344 %
Incorrectly Classified Instances    34344      49.4656 %
Kappa statistic                    0.2368
Mean absolute error                0.0978
Root mean squared error            0.226
Relative absolute error            85.2819 %
Root relative squared error        94.3598 %
Total Number of Instances         69430
  
```

### 6.2.2 Coeficiente de Kappa

WEKA calcula el «Coeficiente de Kappa» para mostrar la concordancia entre los datos de prueba y la clasificación hecha por el modelo. Cuando todas las instancias son clasificadas correctamente se obtiene la máxima concordancia, es decir, Kappa = 1. Por el contrario, cuando el valor del coeficiente es igual a cero, se dice que la concordancia observada es debida exclusivamente al azar.

Para el caso del árbol de decisión obtenido, el «Coeficiente de Kappa» resultó ser igual a 0.2368, lo cual indica un bajo nivel de concordancia entre los datos de prueba y la clasificación hecha por el modelo.

### 6.2.3 Detalle de precisión por clase

No obstante que el modelo general resultó ser poco preciso, se encontraron algunos datos de interés en el detalle de precisión por clase (ver Cuadro 6.2).

**Cuadro 6.2 Detalle de precisión por clase**

```

==== Detailed Accuracy By Class ====

TP Rate  FP Rate  Precision  Recall  F-Measure  Class
0.995  0.491  0.636  0.995  0.776  1
0.006    0.007    0.117    0.006    0.011    2
0.044    0.023    0.12     0.044    0.064    3
0.167    0.059    0.171    0.167    0.169    4
  
```

0.268	0.081	0.196	0.268	0.226	5
0.104	0.03	0.152	0.104	0.123	6
0.049	0.013	0.11	0.049	0.068	7
0.074	0.014	0.152	0.074	0.099	8
0.053	0.011	0.126	0.053	0.075	9
0.016	0.003	0.095	0.016	0.027	10
0.008	0.002	0.063	0.008	0.014	11
0.012	0	0.121	0.012	0.021	12
0.029	0.002	0.139	0.029	0.048	13

En el detalle de precisión por clase, se observa que la “Clase 1”<sup>96</sup> presenta una tasa de “verdaderos positivos” del 99.5%. Esto quiere decir que, dada una instancia perteneciente a la “Clase 1”, el árbol de decisión clasifica a dicha instancia como “Clase 1” el 99.5% de las veces.

En el mismo sentido, se tiene que la tasa de “falsos positivos” para las instancias clasificadas por el modelo es del 49.1%. Es decir, cerca de la mitad de las instancias que no pertenecen a la “Clase 1” son clasificadas dentro de esta clase.

Con las dos medidas anteriores, se obtiene la precisión del modelo para la “Clase 1”, la cual es del 63.3%.<sup>97</sup> Este porcentaje nos indica la proporción de aciertos que el modelo tiene al hacer clasificaciones en la “Clase 1”.

Una medida más que ofrece WEKA es la llamada «F-Measure»<sup>98</sup>, la cual representa la media armónica entre la precisión y el *recall*.<sup>99</sup> Entre más cercana sea a 1, mayor será la confiabilidad de modelo en la clase.

Para la “Clase 1” se tiene una confiabilidad del 77.6%. En contraste, para el resto de las clases, el modelo resulta ser poco confiable, alcanzando en el mejor de los casos, un grado de confiabilidad del 22.6% (ver Cuadro 6.2 y Cuadro 6.3).

De esta forma, se tiene que el árbol de decisión obtenido clasifica de manera aceptable a las instancias pertenecientes a la “Clase 1”. En otras palabras, se cuenta con un conjunto de reglas de decisión que clasifican con un buen grado de confiabilidad a las instancias enmarcadas en la “Clase 1”.

Se ha hecho mucho énfasis en la “Clase 1”, debido a que en esta clase es en donde se encuentra el mayor número de clientes que cancela su contrato de seguro; y a pesar de no haber obtenido un buen modelo en lo general, se puede contar un modelo de predicción para la

<sup>96</sup> Asegurados que cancelan su contrato de seguros durante el primer trimestre de vigencia.

$$^{97} \text{ Precisión} = \frac{(\text{Verdaderos Positivos})}{(\text{Verdaderos Positivos}) + (\text{Falsos Positivos})}$$

$$^{98} F - \text{Measure} = \frac{2 * (\text{Precisión} * \text{Recall})}{(\text{Precisión}) + (\text{Recall})}$$

$$^{99} \text{ Recall} = \frac{(\text{Verdaderos Positivos})}{(\text{Verdaderos Positivos}) + (\text{Falsos Negativos})}$$

“Clase 1”. Por ello, se decidió extraer dichas reglas del árbol de decisión, encontrándose un total de 67 (ver Anexo A).

**Cuadro 6.3 Matriz de confusión**

```

==== Confusion Matrix ====

  a  b  c  d  e  f  g  h  i  j  k  l  m <- classified as
31961 10 10 49 62 13 4 5 7 2 2 0 1 | a = 1
9026 57 111 193 257 97 38 49 32 10 2 1 8 | b = 2
2873 66 207 499 590 183 82 83 71 25 10 4 14 | c = 3
1697 75 287 785 1012 361 141 153 109 31 19 2 17 | d = 4
1355 76 275 788 1282 434 181 177 123 43 21 7 20 | e = 5
931 46 221 599 910 360 140 115 93 16 18 1 16 | f = 6
582 50 153 401 586 220 112 98 76 9 3 3 6 | g = 7
552 35 146 382 528 211 98 166 77 22 11 4 17 | h = 8
419 38 112 334 519 188 76 100 103 21 10 4 9 | i = 9
362 15 81 221 339 131 53 67 48 21 2 2 11 | j = 10
263 9 47 152 208 68 42 41 33 8 7 1 3 | k = 11
82 1 27 63 77 36 12 13 18 3 2 4 8 | l = 12
178 9 48 113 180 69 35 24 25 11 4 0 21 | m = 13

```

En resumen, no obstante haber obtenido un árbol de decisión que en general tiene un bajo nivel de concordancia, se pudo obtener, gracias a las bondades de los árboles de decisión, un conjunto de reglas de decisión que predicen con un buen grado de confiabilidad a los asegurados que tienden a cancelar su contrato de seguro durante el primer trimestre de vigencia.

## CONCLUSIONES

---

A lo largo de su historia, el sector asegurador mexicano se ha caracterizado por mantenerse en constante evolución. Su desarrollo y consolidación se ha venido dando desde principios del siglo XX a un ritmo importante, llevando incluso a los gobiernos posrevolucionarios a reglamentar cada vez más la actividad aseguradora. Sin embargo, no fue sino hasta la década de los 90's cuando se dio un gran impulso a la industria, aumentando el número de participantes, tanto de capital nacional como extranjero, en todas las operaciones y ramos del seguro. De entonces se ha observado un crecimiento consistente, que no sólo ha estado por arriba de los niveles de inflación, sino además, a un ritmo superior al del crecimiento económico.

Un claro ejemplo del crecimiento y la diversificación que ha tenido dicho sector en México lo representa la gran variedad de instituciones de Seguros que hoy en día existen. La entrada de los Grupos Financieros al negocio de los seguros, junto con sus socios extranjeros, representó no solo una importante inyección de capital al sector asegurador mexicano, sino al sistema financiero nacional. La industria de Seguros parecía ser un negocio con un gran potencial, y muy pronto, el capital extranjero encontró bastantes facilidades para su inserción. De tal forma que en la actualidad, se tiene dentro del sector asegurador mexicano a los más importantes proveedores de servicios financieros del mundo.

Dentro de este contexto de competencia, las empresas aseguradoras han buscado nuevas formas para la obtención de ventajas competitivas. Una de ellas gira en torno a la retención de clientes, la cual se basa en la premisa de que cuesta más conseguir un cliente nuevo que retener a un cliente actual.

Por tal motivo, muchas empresas aseguradoras están centrando sus esfuerzos en convertir a todos sus clientes hacia la lealtad.

La lealtad supone un profundo compromiso por volver a comprar un producto o servicio. Este compromiso de volver a comprar, se torna en un proceso complejo de decisión. Y es que en el mercado de los seguros, como en cualquier otro mercado, la decisión de compra se ve afectada por un sinnúmero de factores culturales, sociales, psicológicos y personales.

Las fórmulas que se tienen para llevar a un cliente hacia la lealtad son muy variadas, y cada día surgen nuevas ideas al respecto. Sin embargo, la parte verdaderamente crucial consiste en identificar a los clientes que tienden a ser desleales con la empresa. En cierta forma, se trata de categorizar a los clientes basándose en la información que la empresa posee, para posteriormente, tomar algún curso de acción que lleve a estos clientes a los terrenos de la lealtad.

Desde nuestra perspectiva, creemos que la minería de datos puede ser usada como herramienta para construir modelos predictivos que clasifiquen a los clientes de una institución Seguros de acuerdo a su grado de lealtad. Con lo anterior, dichas instituciones pueden canalizar esfuerzos específicos que fortalezcan los programas de retención de clientes en la empresa.

Durante el presente trabajo, nos hemos dado a la tarea de validar la tesis anterior. Para ello, buscamos construir un modelo de predicción para una de las principales empresas de Seguros

del país, el cual pudiera clasificar a los clientes en función al tiempo de permanencia con su contrato de seguros.

La técnica elegida fue la de Inducción de Árboles de Decisión, y aunque dicha técnica no nos arrojó un árbol de decisión con la suficiente precisión como para proponerlo como modelo predictivo, si obtuvimos, a partir del mismo árbol, un conjunto de reglas de decisión que predicen con un buen grado de confiabilidad a los asegurados que tienden a cancelar su contrato de seguro durante el primer trimestre de vigencia.

De esta forma, tenemos que las reglas de decisión clasifican a los clientes en dos tipos: los clientes que cancelan su contrato durante el primer trimestre, es decir, clientes que cumple con al menos una de las reglas; y los clientes que cancelan a partir del segundo trimestre, es decir, los clientes que no cumplen con ninguna de las reglas.

Este conjunto de reglas de decisión es el que se propone como modelo predictivo, ya que puede ayudar a la empresa de Seguros en estudio a identificar, desde la emisión de la póliza, a los clientes que son tendientes a cancelar su contrato dentro de los siguientes tres meses. Además, se pueden aportar elementos para fortalecer la actual estrategia de retención, ya que con el modelo es posible dar un seguimiento más personalizado a los asegurados que son más desleales con la empresa.

Hasta aquí hemos cubierto el objetivo general de nuestra investigación, así como los objetivos específicos de la misma, es decir, hemos obtenido un modelo predictivo, a partir de una técnica de minería de datos, que identifica a clientes que tienden a cancelar su contrato de seguros durante el primer trimestre de vigencia, así como a los que cancelan a partir del segundo trimestre.

Como hemos dicho, se cumplieron los objetivos del trabajo, pero talvez no como lo planteamos inicialmente, ya que hubiéramos deseado tener una clasificación más abierta, en donde se mostrara el número de trimestres de permanencia con la empresa.

Este último punto deja abierta una puerta a nuevas investigaciones para conseguir modelos que clasifiquen a los clientes basándose en el número de trimestres de lealtad. Consideramos que una posibilidad más de investigación podría ser el construir árboles más específicos, tal vez por producto, por región de venta, instrumento de pago, etc., o por algún cruce de estas variables. Y es que finalmente, la forma de minar datos es muy extensa; por ello, muchos investigadores consideran que dentro de todas estas técnicas hay mucho de ciencia, pero además, en el fondo, hay algo de arte.

Por último, mencionaremos que el presente trabajo es apenas un pequeño intento por mostrar la factibilidad de usar la minería de datos para modelar el comportamiento de los clientes en una institución de Seguros. Creemos que las bondades que pueden aportar estas técnicas al sector pueden ser enormes, solo es cuestión de explorar las áreas de oportunidad y emprender nuevos proyectos de minado de datos.

## BIBLIOGRAFÍA

---

- [1] Aguilera, M. (2003). Cambio estructural y desarrollo del sector asegurador mexicano. Trabajo presentado en la XIII Convención Nacional de Aseguradores, Mayo 19, Ciudad de México.
- [2] Carazo, L. M. (1997, Junio 11). Retención de clientes. Reforma, pp. 32.
- [3] Chen, M.S., Han, J. y Yu, P.S. (1996). Data Mining: An Overview from Database Perspective. IEEE Transactions on Knowledge and Data Engineering, 8 (6), 866-883.
- [4] Comisión Nacional de Seguros y Fianzas. (2003). Memoria de Actividades 2002 y Perspectivas para el 2003. México: Autor.
- [5] Comisión Nacional de Seguros y Fianzas. (2004a). Boletín de Análisis Sectorial, 3 (9).
- [6] Comisión Nacional de Seguros y Fianzas. (2004b). Boletín de Análisis Sectorial, 3 (11).
- [7] Comisión Nacional de Seguros y Fianzas. (2005). Boletín de Análisis Sectorial, 4 (13).
- [8] Comisión Nacional de Seguros y Fianzas. (2007). Boletín de Análisis Sectorial, 6 (21).
- [9] Comisión Nacional de Seguros y Fianzas. (2008). Boletín de Análisis Sectorial, 7 (24).
- [10] Cuesta, F. (2003). Fidelización. Un paso más allá de la retención. Madrid: McGraw-Hill.
- [11] Díaz, B. y Morillas, A. (2003). Minería de Datos y Lógica Difusa: Una aplicación al estudio de la rentabilidad económica de las empresas agroalimentarias en Andalucía. Trabajo presentado en el VI Encuentro de Economía Aplicada, Junio 5, Granada.
- [12] Elizondo, J. (1996). Vigilancia y Diagnóstico en Aseguradoras de Daños. México: IMCP.
- [13] Fayyad, U., Piatetsky-Shapiro, G. y Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. Artificial Intelligence Magazine, 17 (3), 37-54.
- [14] Febles, J. P. y González, A. (2002). Aplicación de la minería de datos en la bioinformática. ACIMED, 10 (2).
- [15] Federación Interamericana de Empresas de Seguros. (2003). Memoria FIDES 2002. Tegucigalpa: Graficentro Editores.
- [16] Fornell, C. y Wernerfelt, B. (1987). Defensive Marketing Strategy by Customer Complaint Management: A Theoretical Analysis. Journal of Marketing Research, 24 (4), 337-346.
- [17] Hardman, R. (2007). Decisions Grow on Trees. Oracle Magazine, 21 (2), 61-64.
- [18] Hernández, J., Ramírez, M.J. y Ferri, C. (2005). Introducción a la Minería de Datos. Madrid: Pearson Educación.

- [19] Hormozi, A. M. y Giles, S. (2004). Data Mining: A Competitive Weapon for Banking and Retail Industries. Information Systems Management, 21 (2), 62-71.
- [20] Han, J. y Kamber, M. (2000). Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufmann Publishers.
- [21] Huete, L. M. (1997). Revitalizando los servicios. Barcelona: IESE/Folio.
- [22] Islas, L. (2004, Agosto 31). Repunta Vida Individual y Crece 17.6% el Sector; Seguros de Personas Representa 56.9% del Total. Durante el Primer Semestre de 2004. El Asegurador, pp. E1, E6.
- [23] Ledesma, J. (1981). Panorama del Derecho Mexicano en el Siglo XIX. Jurídica, 13 (11), 625-652.
- [24] Loyola, J. y Pérez, A. (2002). Evolución del Sector Asegurador Mexicano de 1995 a 2001 (Documento Descriptivo No.15). México: CNSF, Dirección General de Desarrollo e Investigación.
- [25] Merlo, J. y Martín, J. L. (2000). El Marketing y el Seguro. Madrid: Comuniland.
- [26] Minzoni, A. (1994). Crónica de Doscientos Años del Seguro en México. México: CNSF.
- [27] Oliver, R.L. (1997). Satisfaction: A Behavioral Perspective on the Consumer. New York: Irwin/McGraw-Hill.
- [28] Palacios, H. E. (1996). Introducción al Cálculo Actuarial. Madrid: MAPFRE.
- [29] Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1 (1), 81-106.
- [30] Quinlan, J. R. (1996). Improved Use of Continuous Attributes in C4.5. Journal of Artificial Intelligence Research, 4 (1), 77-90.
- [31] Quiñones, C. E. (1999). La Individualización del Riesgo en el Contrato de Seguro Marítimo. Revista de Derecho, 11, 75-92.
- [32] Reichheld, F. F. (1996). The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value. Boston: Baie & Company.
- [33] Riegel, R. y Miller, J. (1980). Seguros Generales: Principios y Prácticas. México: Prentice-Hall.
- [34] Romero, H. (1993). El Seguro en el Mundo, 1990 - 1991. Actualidad en Seguros y Fianzas, 1 (9), 105-121.
- [35] Rosen, M. (1998). There's gold in that there data. Insurance & Technology, 23 (12), 28-32.
- [36] Sánchez, R. y Sánchez, A. (2000). Desregulación y Apertura del Sector Financiero Mexicano. Comercio Exterior, 50 (8), 686-697.

- [37] Smith, K. A., Willis, R. J. y Brooks, M. (2000). An analysis of customer retention and insurance claim patterns using data mining: a case study. Journal of the Operational Research Society, 51 (5), 532-541.
- [38] Turney, M. (2003). Intelligence behind data. Canadian Underwriter, 70 (5), 44-45.
- [39] Vega, A. (2000). Web-mining en el marco de los procesos de negocio y marketing de la empresa. Comunicaciones de Telefónica I+D, 19 (1), 125-130.
- [40] Venugopal, V. y Baets, W. (1995). Intelligent Support Systems for Organizational Learning. The Learning Organization, 2 (3), 22-34.
- [41] Wincuin, J. (1999). Tapping the mountain of consumer information. LIMRA'S MarketFacts, 18 (6), 26-29.
- [42] Witten, I. H. y Frank, E. (2000). Data Mining: Practical machine learning tools and techniques with java implementations. San Francisco: Morgan Kaufmann Publishers.

## ANEXO A. REGLAS DE DECISIÓN

---

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '02' AND  
Zona\_geográfica = 'CENTRO' AND  
Estado\_civil = 'CASADO' AND  
Edad\_de\_emisión = '[35-44]' AND  
Sexo\_del\_asegurado = 'FEMENINO' AND  
Tipo\_de\_cuenta = 'OTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '02' AND  
Zona\_geográfica = 'CENTRO' AND  
Estado\_civil = 'SOLTERO' AND  
Edad\_de\_emisión = '[35-44]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Sexo\_del\_asegurado = 'FEMENINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '02' AND  
Zona\_geográfica = 'NOROESTE' AND  
Estado\_civil = 'SOLTERO' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'OTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '04' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Zona\_geográfica = 'NORTE' AND  
Sexo\_del\_asegurado = 'FEMENINO' AND  
Edad\_de\_emisión = '[15-24]' AND  
Estado\_civil = 'CASADO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '04' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Zona\_geográfica = 'SUR' AND  
Estado\_civil = 'SOLTERO' AND  
Edad\_de\_emisión = '[35-44]' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '04' AND

Tipo\_de\_cuenta = 'OTRO' AND  
Zona\_geográfica = 'CENTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '04' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Zona\_geográfica = 'DISTRITO' AND  
Edad\_de\_emisión = '[45-54]' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '04' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Zona\_geográfica = 'EDOMEX' AND  
Edad\_de\_emisión = '[55-64]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '04' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Zona\_geográfica = 'GOLFO' AND  
Estado\_civil = 'CASADO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '04' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Zona\_geográfica = 'GOLFO' AND  
Estado\_civil = 'SOLTERO' AND  
Sexo\_del\_asegurado = 'FEMENINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '04' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Zona\_geográfica = 'SUR'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '05' AND  
Zona\_geográfica = 'CENTRO' AND  
Edad\_de\_emisión = '[35-44]' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Estado\_civil = 'OTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '06' AND  
Zona\_geográfica = 'SUR' AND

Edad\_de\_emisión = '[45-54]' AND  
Estado\_civil = 'OTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50)' AND  
Día\_de\_facturación = '07' AND  
Zona\_geográfica = 'OCCIDENTE' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Sexo\_del\_asegurado = 'MASCULINO' AND  
Estado\_civil = 'SOLTERO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50)' AND  
Día\_de\_facturación = '07' AND  
Zona\_geográfica = 'OCCIDENTE' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Sexo\_del\_asegurado = 'MASCULINO' AND  
Estado\_civil = 'OTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50)' AND  
Día\_de\_facturación = '07' AND  
Zona\_geográfica = 'OCCIDENTE' AND  
Edad\_de\_emisión = '[45-54]' AND  
Estado\_civil = 'OTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50)' AND  
Día\_de\_facturación = '08' AND  
Zona\_geográfica = 'GOLFO' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Edad\_de\_emisión = '[25-34]' AND  
Sexo\_del\_asegurado = 'FEMENINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50)' AND  
Día\_de\_facturación = '11' AND  
Zona\_geográfica = 'CENTRO' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Edad\_de\_emisión = '[25-34]' AND  
Estado\_civil = 'SOLTERO' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50)' AND  
Día\_de\_facturación = '14' AND  
Zona\_geográfica = 'SUR' AND

Estado\_civil = 'SOLTERO' AND  
Edad\_de\_emisión = '[35-44]' AND  
Sexo\_del\_asegurado = 'FEMENINO' AND  
Tipo\_de\_cuenta = 'OTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '19' AND  
Zona\_geográfica = 'GOLFO' AND  
Edad\_de\_emisión = '[15-24]' AND  
Estado\_civil = 'SOLTERO' AND  
Tipo\_de\_cuenta = 'OTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '20' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Zona\_geográfica = 'GOLFO' AND  
Edad\_de\_emisión = '[45-54]' AND  
Sexo\_del\_asegurado = 'FEMENINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '21' AND  
Zona\_geográfica = 'GOLFO' AND  
Estado\_civil = 'CASADO' AND  
Edad\_de\_emisión = '[45-54]' AND  
Tipo\_de\_cuenta = 'OTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '25' AND  
Zona\_geográfica = 'EDOMEX' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Estado\_civil = 'SOLTERO' AND  
Edad\_de\_emisión = '[55-64]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '29' AND  
Estado\_civil = 'SOLTERO' AND  
Zona\_geográfica = 'NORTE' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Edad\_de\_emisión = '[25-34]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50]' AND  
Día\_de\_facturación = '29' AND  
Estado\_civil = 'SOLTERO' AND

Zona\_geográfica = 'NORTE' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Edad\_de\_emisión = '[55-64]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50)' AND  
Día\_de\_facturación = '29' AND  
Estado\_civil = 'SOLTERO' AND  
Zona\_geográfica = 'NORTE' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Edad\_de\_emisión = '[65-74]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50)' AND  
Día\_de\_facturación = '29' AND  
Estado\_civil = 'OTRO' AND  
Zona\_geográfica = 'SUR' AND  
Edad\_de\_emisión = '[45-54]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50)' AND  
Día\_de\_facturación = '30' AND  
Zona\_geográfica = 'EDOMEX' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Edad\_de\_emisión = '[15-24]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(0-50)' AND  
Día\_de\_facturación = '31' AND  
Zona\_geográfica = 'EDOMEX' AND  
Tipo\_de\_cuenta = 'OTRO' AND  
Estado\_civil = 'CASADO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100)' AND  
Producto = 'ACCIDENTES' AND  
Edad\_de\_emisión = '[15-24]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100)' AND  
Producto = 'ACCIDENTES' AND  
Edad\_de\_emisión = '[25-34]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100)' AND  
Producto = 'ACCIDENTES' AND  
Edad\_de\_emisión = '[35-44]'

THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'ACCIDENTES' AND  
Edad\_de\_emisión = '[45-54]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'ACCIDENTES' AND  
Edad\_de\_emisión = '[55-64]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[15-24]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '05' AND  
Zona\_geográfica = 'EDOMEX' AND  
Sexo\_del\_asegurado = 'FEMENINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '06' AND  
Zona\_geográfica = 'NORTE' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '06' AND  
Zona\_geográfica = 'OCCIDENTE' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND

Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '06' AND  
Zona\_geográfica = 'SUR'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Zona\_geográfica = 'DISTRITO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '07' AND  
Zona\_geográfica = 'EDOMEX' AND  
Sexo\_del\_asegurado = 'FEMENINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '07' AND  
Zona\_geográfica = 'NORTE' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '11' AND  
Zona\_geográfica = 'GOLFO' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '11' AND  
Zona\_geográfica = 'NORTE' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND

Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '13' AND  
Zona\_geográfica = 'DISTRITO' AND  
Sexo\_del\_asegurado = 'FEMENINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '13' AND  
Zona\_geográfica = 'NORTE' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '17' AND  
Zona\_geográfica = 'EDOMEX'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '17' AND  
Zona\_geográfica = 'SUR' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '20' AND  
Zona\_geográfica = 'DISTRITO' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '20' AND  
Zona\_geográfica = 'EDOMEX'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '21' AND  
Zona\_geográfica = 'NOROESTE' AND  
Sexo\_del\_asegurado = 'FEMENINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '22' AND  
Zona\_geográfica = 'DISTRITO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '22' AND  
Zona\_geográfica = 'GOLFO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '22' AND  
Zona\_geográfica = 'SUR'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '25' AND  
Zona\_geográfica = 'GOLFO' AND  
Sexo\_del\_asegurado = 'FEMENINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '28' AND  
Zona\_geográfica = 'CENTRO' AND

Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '28' AND  
Zona\_geográfica = 'EDOMEX'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '28' AND  
Zona\_geográfica = 'NORTE'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '29' AND  
Zona\_geográfica = 'NORTE' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '30' AND  
Zona\_geográfica = 'CENTRO' AND  
Sexo\_del\_asegurado = 'MASCULINO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '30' AND  
Zona\_geográfica = 'EDOMEX'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND

Tipo\_de\_cuenta = 'CREDITO' AND  
Día\_de\_facturación = '31' AND  
Zona\_geográfica = 'NORTE'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[25-34]' AND  
Tipo\_de\_cuenta = 'OTRO'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[45-54]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[55-64]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '(50-100]' AND  
Producto = 'VIDA' AND  
Edad\_de\_emisión = '[65-74]'  
THEN Trimestre\_de\_cancelación = 1

IF Prima\_mensualizada = '>100'  
THEN Trimestre\_de\_cancelación = 1