

# Análisis del Riesgo de Caída de Cartera en Seguros: Metodologías de “Inteligencia Artificial” vs “Modelos Lineales Generalizados”

## *Analysis of the Risk of Insurance Portfolio Loss:*

## "Artificial Intelligence" Methodologies vs. "Generalized Linear Models"

María de Lourdes Gutiérrez Cordero\*

María Jesús Segovia-Vargas\*\*

María Ramos Escamilla \*\*\*

## Palabras clave

## Solvencia II, Riesgo de Caída de Cartera, Rough Set, GLM y C4.5

## Key words

PSolvency II, Lapse Risk, Rough Set, GLM y C4.5

JF

C52, D63, I3, O10, P36

\* Universidad Complutense de Madrid.

## Departamento de Economía Financiera y Contabilidad I

\*\* Universidad Complutense de Madrid.

Departamento de Economía Financiera y

[misegovia@ccee.ucm.es](mailto:misegovia@ccee.ucm.es)

\*\*\* Miembro del Sistema Nacional de  
Investigadores Instituto Tecnológico de

Pachuca. Departamento  
de Posgrado e Investigación  
ramos@itpachuca.edu.mx

56



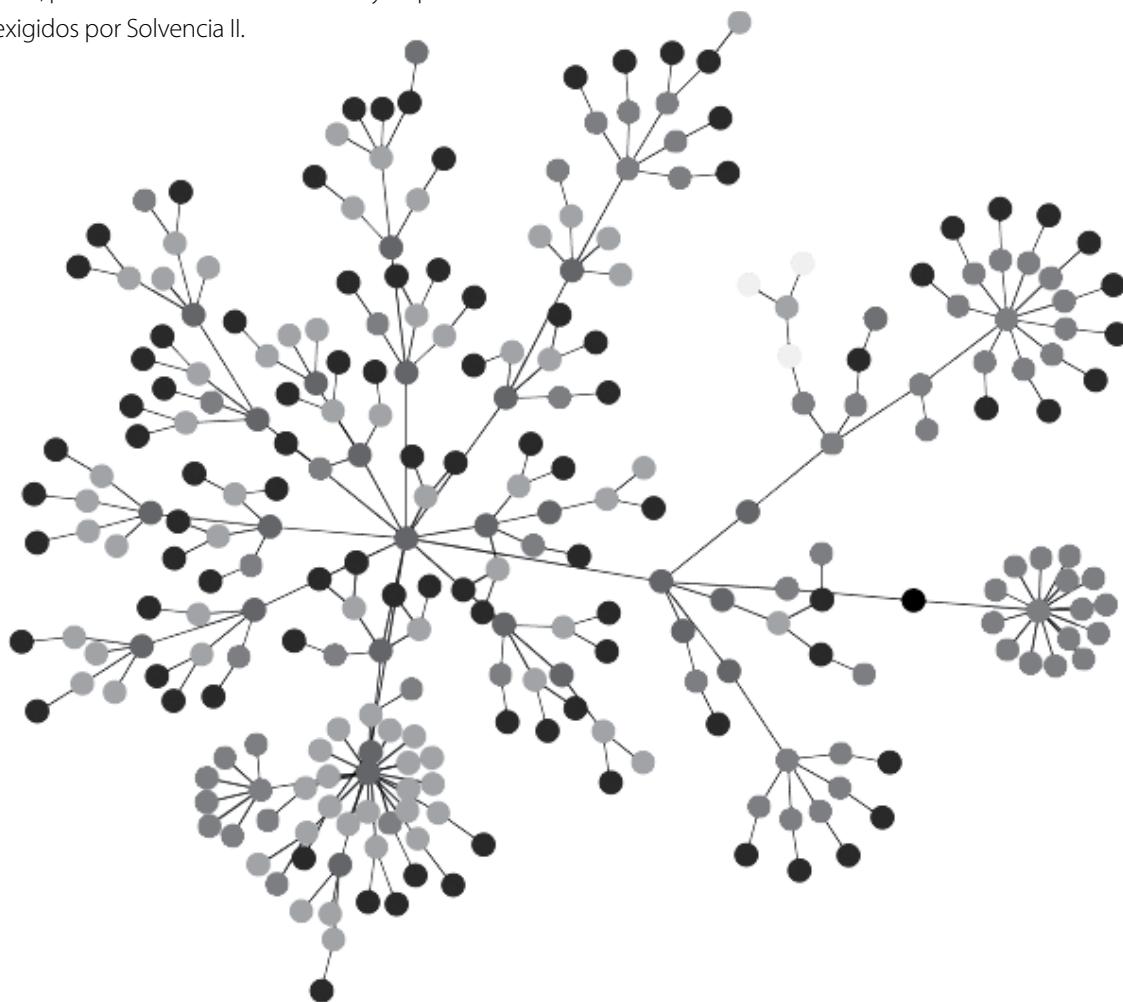
© 2017 Universidad Nacional Autónoma de México, Facultad de Economía. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Resumen

En este artículo se muestra la posibilidad de utilizar dos enfoques metodológicos distintos para la evaluación de riesgos: uno no paramétrico para lo cual se recurrirá a las técnicas de Inteligencia Artificial y, en contraste, la aplicación de los Modelos Lineales Generalizados provenientes de la estadística paramétrica. La aplicación práctica de ambas metodologías, se llevará a cabo con la finalidad de analizar el Riesgo de Caída de Cartera; haciendo referencia a uno de los tantos riesgos medibles que el sector asegurador ha de tener en cuenta de acuerdo a Solvencia II. Los resultados obtenidos y sus conclusiones muestran el nuevo enfoque y cómo dichas técnicas pueden ser utilizadas por las aseguradoras como una mejora en la Gestión de Riesgos; animando al sector a indagar en nuevas metodologías y técnicas, para cubrir con las necesidades y requerimientos exigidos por Solvencia II.

## Abstract

This paper describes the use of two different methodological approaches for risk assessment: non-parametric coming from Artificial Intelligence techniques and, in contrast, Generalized Linear Models from statistical parametric. Both practical applications will analyze Lapse Risk, one of the measurable risks that insurance sector must take into account according to Solvency II. Results and conclusions show a new approach and how these techniques can be used by insurance companies as an improvement in Risk Management; encouraging the insurance sector to investigate new methodologies and techniques to deal requirements demanded by Solvency II regulation.



## 1. Introducción

Después de varios años y a pesar de sus prórroga, Solvencia II (Comisión de las Comunidades Europeas, Directiva 2009/138/CE) es una realidad que ha puesto de manifiesto la adopción de medidas concretas para adaptarse al nuevo sistema; ya no es una práctica recomendable sino una materia exigible a las entidades. De esta forma, Solvencia II, sin lugar a dudas, es un desafío para la industria aseguradora.

Este proyecto es una verdadera transformación del modelo de gestión de riesgos y de la toma de decisiones en las entidades aseguradoras; por lo que el proceso de adaptación a esta nueva norma y filosofía exigirá un gran esfuerzo por parte de las entidades.

El objetivo de Solvencia II (González de Frutos, 2013) es el desarrollo y establecimiento de un nuevo sistema que permita determinar los recursos propios mínimos a requerir a cada aseguradora, en función de los riesgos asumidos y la gestión que se realice de ellos. Así mismo, engloba un conjunto de iniciativas para la revisión de la normativa existente, la valoración y supervisión de la situación financiera global de las entidades aseguradoras y modos de actuación interna de las mismas.

Existen varios motivos y necesidades que promueven el inicio de este proyecto, como son: el mantenimiento de la confianza en el sistema financiero-asegurador, el fortalecimiento de la gestión de riesgos, la armonización de las prácticas de valoración contables y regulatorias y el común acuerdo sobre las medidas de intervención. Todos estos conceptos e inquietudes han provocado que paulatinamente, el sector asegurador se cuestione sobre la innovación de los modelos y planteamientos en vigor; en otras palabras, se busca introducir al sector asegurador dentro de una nueva cultura del riesgo y su gestión. Ante la necesidad de

evolución del sector asegurador, surge el interés en temas como la gestión de riesgos y solvencia con la que es capaz de enfrentarse una entidad aseguradora (Romera, 2011).

Uno de los temas más controvertidos bajo esta regulación es cómo conseguir una adecuada evaluación de los riesgos asumidos por las entidades. Esto se traduce en lograr identificar las causas que puedan suponer una pérdida en sus recursos; así como en innovar en el campo técnico para lograr una correcta cuantificación de los riesgos posibles en los que podrían estar expuestas las entidades. Es así como las compañías aseguradoras están siendo sometidas a desarrollar nuevas técnicas para la cuantificación y control de los riesgos a los que se encuentran expuestas. Todo ello con el fin de lograr implementar una gestión integral del riesgo que contemple un adecuado nivel de solvencia (EIOPA, 2015).

Bajo otra perspectiva, en un competitivo mercado asegurador donde cada día toman relevancia temas como la guerra de precios, accesibilidad a múltiples cotizaciones, así como la constante innovación en el desarrollo de productos, surge la necesidad de retener y fidelizar a los clientes. De esta forma, ya no sólo se presta atención a niveles de primas altos sino a la capacidad de garantizar la rentabilidad de la entidad, lo cual no es una tarea fácil ante una situación de crisis financiera como la que se ha vivido recientemente. De aquí la importancia del riesgo de caída de cartera que exige la nueva regulación de Solvencia II, entendido como el riesgo que supone la pérdida de clientes ante la cancelación de cierto volumen de pólizas a su vencimiento por parte de los asegurados dentro de una entidad.

El objetivo de este trabajo es mostrar la posibilidad de utilizar dos enfoques metodológicos distintos para la evaluación de riesgos: uno no paramétrico para lo cual se recurirá a las

técnicas de Inteligencia Artificial (un árbol de decisión, el algoritmos C4.5, y, un algoritmo de inducción de reglas, la metodología Rough Set) y, en contraste, la aplicación de los Modelos Lineales Generalizados provenientes de la estadística paramétrica.

La aplicación práctica de ambas metodologías, se llevará a cabo con la finalidad de analizar el Riesgo de Caída de Cartera, el cual hace referencia a uno de los riesgos medibles que el sector habrá de tener en cuenta de acuerdo a Solvencia II. La relevancia de ambas aplicaciones empíricas, será poder tener una aproximación a la probabilidad de cancelación del cliente mediante dichos patrones que se traduciría en una mejora en la gestión del Riesgo de Caída de Cartera; contribuyendo al equilibrio y estabilidad de los niveles de solvencia de las entidades.

Para la realización de la aplicación empírica se utilizará una cartera española de pólizas del ramo de vida representativa de todas las regiones españolas. Los mercados avanzados, al que pertenece la muestra española de datos que utilizaremos en nuestro estudio, se han mantenido en un estancamiento desde la crisis económica del 2008, a pesar del crecimiento que experimentaron las primas en el año 2014 que se situó en el 2,9%, sin embargo, este porcentaje está todavía por debajo al 3,4% de la tasa de crecimiento a largo plazo de la situación anterior a la crisis (Swiss Reinsurance Company, *Informe Sigma 4/2015*, pp.7). Por tanto, en el contexto actual del mercado asegurador donde existe una disminución del volumen de negocio y tendencia creciente a la pérdida de la cartera de clientes; cobra importancia el tema de retención de clientes y con ello cobra importancia poder identificar el tipo de clientes propensos a causar baja (Guillen *et al.*, 2008). De esta forma, se podrán anticipar pérdidas mediante la implementación de estra-

tegias para la retención y atracción de nuevos clientes, es decir, lograr orientar la toma de decisiones por medio de la localización de algún patrón de comportamiento del tipo de cliente “cancelador” que permita establecer políticas comerciales atractivas para la captación y fidelización de su cartera. De aquí la relevancia de la presente aplicación empírica, para poder tener una aproximación a la probabilidad de cancelación del cliente mediante dichos patrones que se traduciría en una mejora en la gestión del riesgo de caída de cartera que a su vez, contribuiría al equilibrio y estabilidad de los niveles de solvencia que las compañías aseguradoras requieren.

El artículo se estructura como sigue: en la sección 2 se presenta Solvencia II; la sección 3 explica el problema objeto de nuestro estudio, es decir la caída de cartera; en la cuarta sección se explican las tres metodologías utilizadas, es decir, árbol de decisión, rough set y mínimos cuadrados generalizados; la quinta sección contiene el análisis empírico y finalmente la última sección incluye las conclusiones.

**E**n el contexto actual del mercado asegurador donde existe una disminución del volumen de negocio y tendencia creciente a la pérdida de la cartera de clientes; cobra importancia el tema de retención de clientes y con ello cobra importancia poder identificar el tipo de clientes propensos a causar baja.

## 2. El proyecto solvencia II: nuestro antecedente

Desde hace tiempo en algunos países surgió la inquietud de referenciar a la solidez financiera con los riesgos asumidos de manera implícita a la propia actividad de la entidad aseguradora.

En la década de los 50, los pioneros en la aplicación de esquemas basados en el riesgo fueron los finlandeses, quienes empezaron a utilizar un modelo de capital considerando el carácter estocástico de la actividad aseguradora mediante las “Reservas Especiales de Nivelación” (Pentikäinen y Rantala, 1982). Posteriormente, le secundó Canadá que a mediados de los 80 comienza a aplicar modelos que intentan englobar la totalidad de los riesgos mediante la generación de escenarios para el diseño de sus planes de negocio a través de las llamadas “Exigencias de Capital Mínimo para la Continuación” (Hardy y Panjer, 1998).

Bajo una línea similar, en los años 90, EE UU mediante la NAIC desarrolla el modelo RBC (Risk-based Capital) basado en un conjunto de normas haciendo una primera definición y basando los requerimientos de capital en una serie de riesgos independientes entre sí (Jacques y Nigro, 1997).

Se llega al año 2004 con el modelo suizo conocido como “Test Suizo de Solvencia” buscando un enfoque basado en el análisis de los riesgos reales que soporta una entidad aseguradora de una forma integrada (Keller *et al.*, 2005). Sin dejar de mencionar al modelo británico el cual funciona desde el año 2005 que también ha buscado relacionar los requerimientos de capital con los riesgos a los que están expuestos las entidades (Eling *et al.*, 2007).

Los antecedentes más directos de Solvencia II se sitúan en su predecesor, Solvencia I. Estaba basado en un conjunto de ratios que relacio-

nan el capital exigido con el volumen del negocio obtenido a partir del cálculo del Margen de Solvencia Obligatorio y el Fondo Mínimo de Garantía. Sólo se dirigía a los riesgos técnicos que surgen del pasivo de las entidades, sin tener en cuenta los riesgos asociados al activo como son las inversiones o la calidad crediticia de las operaciones. Sin dejar de mencionar que esta valoración y exigencias de capital no se hacen de acuerdo al mercado y sin considerar diversificación o transferencia de riesgos que implicase reducción de dichos requerimientos. Todas estas limitaciones dieron lugar a la necesidad de la creación de un nuevo modelo, materializado en la implementación de Solvencia II<sup>1</sup> (Ariza, 2013).

Se observa, por tanto, que el tema de la solvencia dentro del sector asegurador no es un tópico nuevo, así como su regulación, ya que existen varias directivas y normas que se perfeccionan y complementan entre sí. De ahí la necesidad de lograr establecer un conjunto de normas comunes que engloben la actual coyuntura, con el objetivo de adecuar la regulación a la situación actual, sin buscar cubrir una carencia sino completar las directrices ya existentes. Es así como el sector asegurador se encamina hacia a la propuesta por parte de la Comisión Europea de una nueva Directiva en materia de seguros y reaseguros, tanto del ramo de vida como de ramos distintos del de vida, bajo la denominación de “Solvencia II” (Solá, 2013).

---

<sup>1</sup> Comisión de las Comunidades Europeas. Directiva 2009/138/CE del Parlamento Europeo y del Consejo, de 25 de noviembre de 2009, sobre el seguro de vida, el acceso a la actividad de seguro y de reaseguro y su ejercicio (*Directiva de Solvencia II*). Disponible en: [https://ec.europa.eu/info/law/risk-management-and-supervision-insurance-companies-solvency-ii-directive-2009-138-ec\\_en](https://ec.europa.eu/info/law/risk-management-and-supervision-insurance-companies-solvency-ii-directive-2009-138-ec_en)

**Figura 1** Esquema conceptual del Proyecto Solvencia II

Solvencia II		
Pilar I	Pilar II	Pilar III
* Provisiones Técnicas	* Control y auditoría interna	* Difusión de la información
* Requerimientos de Capital	* Gobierno Corporativo	* Transparencia de la información
* Inversiones	* Administración de riesgos	* Revisión y competencias del mercado
* Reaseguro	* Proceso de supervisión	
<i>Requerimientos Cuantitativos</i>	<i>Requerimientos Cualitativos</i>	<i>Disciplina del Mercado</i>

Fuente: Propia de los autores.

61

El entorno normativo de Solvencia II, se compone de un conjunto de elementos que se ordena bajo una estructura basada en tres pilares (*Figura 1*), que se resumen de la siguiente forma (UNESPA, 2015):

- Pilar I – *Cuantitativo*: Se destina a los requerimientos cuantitativos. Este primer Pilar I se le ubica como el pilar técnico de Solvencia II;<sup>2</sup> ya que se busca la construcción de un primer Pilar matemáticamente sólido y, capaz de abarcar la totalidad de riesgos que se deben calcular en una adecuada valoración del perfil de riesgo de un negocio. Para cumplir con estos requisitos es necesario que primeramente las entidades aseguradoras establezcan y definan las reglas de valoración que seguirán para cuantificar todas las partidas relevantes del balance económico tanto del activo como del pasivo en conjunto, de tal forma que, se pueda obtener los niveles de capital adecuados al perfil de riesgos asumidos por la entidad.

2 Comisión de las Comunidades Europeas. *Comunicación de la Comisión*. Noviembre 2007. Revisión del proceso Lamfalussy. COM/2007/727 final – COD 55 de 28.2.2008. Disponible en: <http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=URISERV:l32056>

- Pilar II – *Cualitativo*: Se destina a los requerimientos cualitativos y las normas de supervisión. Busca una supervisión de alta calidad por parte de los organismos reguladores, con rigurosas exigencias en materia del gobierno en las entidades aseguradoras, que afectan a los órganos de gestión y dirección de la misma quienes son los principales responsables de los procesos de identificación, medición y gestión activa del riesgo. De esta forma, se ven obligadas a buscar mejoras en la gestión interna y así conseguir reforzar la estabilidad y solvencia del sector asegurador. Por otro lado, el Pilar II hace especial hincapié en la necesidad de preservar la coherencia entre las exigencias impuestas entre los distintos elementos que conforman el sector financiero. Se hace mención a fomentar una supervisión prudencial destinada a detectar aquellas entidades que presentan un riesgo elevado, por sus características financieras, organizativas o de cualquier otra índole; ya que ello podría tener graves consecuencias sobre la solidez financiera de las entidades.
- Pilar III – *Disciplina del mercado*: Se busca desarrollar la comunicación de la informa-

ción entre el supervisor y la entidad aseguradora con el fin de favorecer la disciplina y transparencia. Así se podrá lograr conseguir una mayor estabilidad financiera mediante una tendencia hacia la obtención de una contabilidad internacional homogénea. Mediante la implementación de este tercer Pilar, se verá reforzada la convergencia y transparencia de la actividad aseguradora, que se traducirá en el fortalecimiento de la supervisión del seguro.

### 3. El riesgo de caída de cartera: nuestro problema de estudio

62

Para cubrir con los requisitos exigidos por Solvencia II, las aseguradoras deberán cuantificar los niveles de capital adecuados al perfil de riesgos asumidos por la entidad, estableciendo y definiendo las reglas de valoración que seguirán para ello.

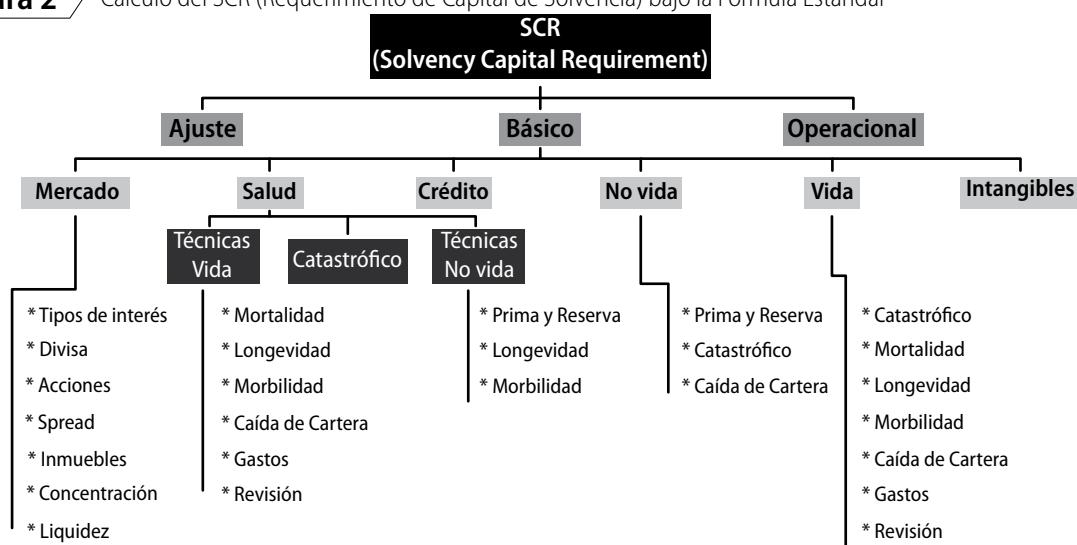
Es así como surge el término de Capital de Solvencia Obligatorio (*SCR – Solvency Capital Requirement*) que se define como el capital necesario para hacer frente a las posibles pérdidas económicas teniendo en cuenta to-

dos los riegos cuantificables a los que está expuesta, en un horizonte temporal de un año y con un nivel de confianza del 99.5% (*VaR al 99.5%*). Para la cuantificación de dicho Capital Requerido, Solvencia II facilita su propia metodología a través del denominado “modelo estándar”, o bien permite a cada entidad implementar un “modelo interno” basado en la experiencia propia de la compañía (Pozuelo de Gracia, 2008).

En términos generales, los “modelos internos” deberán construir sus propias hipótesis basadas en la experiencia de la compañía, justificando y documentando cada una de éstas, así como la estructura y calibración de cada riesgo considerado. Así mismo, si la entidad opta por utilizar su propio modelo, éste deberá ser presentado y aprobado por los órganos supervisores.

Por el contrario, el “modelo estándar” establece una fórmula general para el cálculo del SCR, la cual fue definida por el Comité Europeo de Supervisores de Seguros y de Pensiones de Jubilación (*CEIOPS*). Se resumen con un conjunto de normas que asume un enfoque ge-

**Figura 2** Cálculo del SCR (Requerimiento de Capital de Solvencia) bajo la Fórmula Estándar



Fuente: Propia de los autores.

neral de identificación y valoración de los riesgos que afectan a las entidades aseguradoras. Con base en dicha evaluación, se cuantifican cada uno de los riesgos y se calcula el capital necesario para cubrirlos. Dentro del planteamiento que sugiere la fórmula estándar, se identifican los riesgos más relevantes de la entidad aseguradora, tanto en la estimación de las provisiones técnicas como en la valoración de los activos bajo la generación de diferentes escenarios.

De esta forma, la valoración del requerimiento de capital se obtiene mediante el desglose de seis sub-módulos de cálculo correspondiente a la valoración de los riesgos asumidos por la entidad. Posteriormente mediante una matriz de varianza y covarianzas, se realiza la agregación de riesgos y se obtiene el SCR global, es decir, el Requerimiento de Capital de Solvencia bajo el enfoque de la Fórmula Estándar (*Figura 2*).

Uno de dichos riesgos contemplados es la caída de cartera que registra una entidad, entendiéndose como tal a la rotación o salida de asegurados, lo cual se ve directamente reflejado en el decrecimiento en el volumen de primas de la entidad.

De aquí la importancia del estudio del problema en cuestión, la cuantificación del riesgo de caída de cartera, que es exigido por Solvencia II, así como los principales agentes causantes del mismo y sus implicaciones, ya que se reflejan directamente sobre los márgenes de solvencia de la entidad (Ayuso *et al.*, 2011; 2012).

Siendo así que surge el concepto de “Caída de Cartera”, del cual no existe una definición precisa; por lo que puede precisar como el conjunto de pólizas que no optan por la renovación a su vencimiento por parte de los asegurados (Millán y Colomina, 2001).

En términos matemáticos, el número de pólizas que se anulan o cancelan durante un período determinado, se puede expresar de la siguiente manera:

$$Anul = Pol_i + Pol_{NP} - Pol_f$$

siendo:  $Pol_i$  = núm. Pólizas en Vigor al inicio del período

$Pol_{NP}$  = núm. Pólizas de Nueva Producción registradas durante el período

$Pol_f$  = núm. Pólizas en Vigor al final del período

63

De esta forma, se puede expresar el concepto de Caída de Cartera en términos de porcentajes de la siguiente manera:

$$TasaCaída = Anul/Pol_i$$

En la actualidad, cada entidad aseguradora ha ido desarrollando metodologías novedosas con el objetivo de estimar la caída de cartera que se registrará en un futuro (Brockett *et al.*, 2008). En la mayoría de los casos, basándose en su información histórica, utilizan modelos estadísticos medianamente complejos, y determinan los porcentajes de caída que definen mejor el riesgo al que está expuesto.

#### 4. Metodologías propuestas: marco teórico

En primer lugar, se debe situar a la “Inteligencia Artificial” dentro de las metodologías que se manejan en la rama de la disciplina de Aprendizaje Automático (*Machine Learning*, por sus siglas en inglés). Este enfoque utiliza algoritmos para analizar registros en bases de datos internas de los clientes de una empresa,

para descubrir ciertos patrones, interacciones o reglas que pueden describir o predecir las futuras tendencias que puedan indicar cualquier tipo de oportunidades competitivas (Mena, 1996) y, ayude a tomar decisiones o mejorar la comprensión o conocimiento que se pueda extraer a través de dichas bases de datos.

Ahora bien, dentro de esta disciplina se engloban las técnicas de Inteligencia Artificial, las cuales se basan en el aprendizaje a partir de los datos y de su semejanza con un pensamiento estructurado similar al comportamiento humano. Existen varias técnicas sugeridas dentro de esta rama de la Inteligencia Artificial como son las Redes Neuronales, los Vectores Soporte, los Algoritmos Genéticos, los Sistemas de Inducción de Reglas, los árboles de Decisión o la Teoría de *Rough Set*.

Por otro lado, los Modelos Lineales Generalizados (GLM –*Generalized Linear Models*) introducidos a comienzos de los años 70 (Nelder y Wedderburn, 1972), se han convertido en una de las principales herramientas de análisis estadístico en toda clase de áreas. De acuerdo con esta metodología, tras un análisis de regresión se obtiene la variable respuesta o dependiente (número de siniestros o importe reclamado), basándose en un conjunto de variables explicativas, es decir, una serie de factores relacionados con el evento que simula dicha variable respuesta (generalmente características propias del asegurado de la póliza) (Díaz *et al.*, 2010).

#### *4.1. Metodología no paramétrica. arbol de decisión c4.5*

A partir de un conjunto de datos se construyen diagramas de construcciones lógicas que hacen referencia a una clasificación óptima de los datos de acuerdo a sus características o atributos. De esta forma, se crean particiones recursivas que sirven para categorizar y representar una

serie de condiciones que ocurren de forma sucesivas, comúnmente llamadas reglas sobre la decisión que se debe tomar, para solucionar el problema planteado asignando un valor de salida a un determinado registro de entrada.

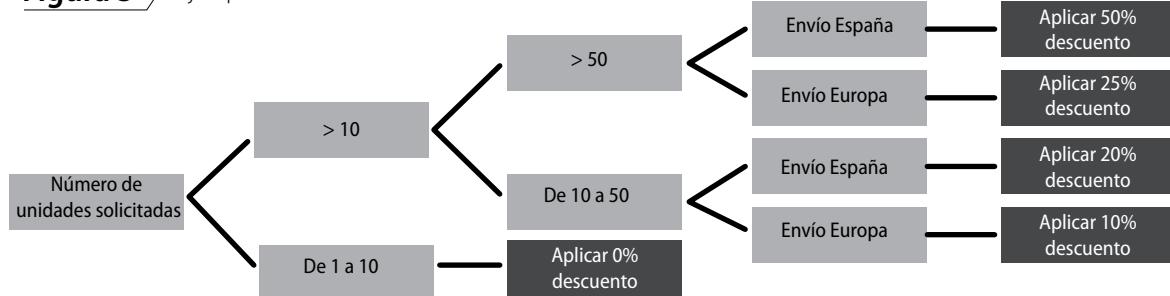
Dichas reglas, gráficamente, se representan en forma de árbol a través de hojas o ramas, de ahí el nombre de Árbol de Decisión. De esta forma, permite obtener de forma visual las reglas de decisión y, de aquí su principal ventaja que es la fácil interpretación de los resultados (Araya, 1994).

Existen varios algoritmos para la construcción de árboles de decisión: CLS (*Concept Learning Systems*), Método CHAID (Kass, 1980), Método CART (Breiman, *et al.* 1996), Algoritmo C4.5 (Quinlan ,1993). La diferencia entre estos algoritmos de aprendizaje radica en el criterio utilizado para realizar las particiones o Reglas.

El Algoritmo C4.5 (cuyo precedente está en ID3 (Quinlan, 1986) y cuyas mejoras se incorporan en la versión comercial C5.0) es uno de los algoritmos más utilizados en el ámbito de los árboles de decisión, habiendo sido aplicado a problemas de clasificación en general, y a problemas financieros, en particular (Díaz *et al.*, 2005 y 2009; Miranda *et al.*, 2013). Esta es la razón por la cual fue elegido este tipo de algoritmo para su uso en la aplicación empírica realizada.

El algoritmo C4.5 se basa en conceptos procedentes de la Teoría de la Información para hacer las particiones y fue desarrollado por Quinlan (Quinlan, 1993). Parte de la premisa de tomar en cada rama del árbol, para hacer la correspondiente partición, aquella variable que proporciona más información de cara a clasificar los elementos que constituyen el conjunto de entrenamiento o conjunto de datos usados para construir el árbol (Díaz *et al.*, 2004).

**Figura 3** Ejemplo de Árbol de Decisión



Fuente: Propia de los autores.

La información que proporciona un mensaje o la realización de una variable aleatoria  $x$  es inversamente proporcional a su probabilidad  $P_x$  (Reza, 1961). Con frecuencia en Ingeniería de Comunicaciones o en Estadística se mide esta cantidad en bits, que se obtienen como  $\log_2 1/P_x$ . El promedio de esta magnitud para todas las posibles ocurrencias de la variable aleatoria  $x$  recibe el nombre de entropía de  $x$ , es decir, el promedio se obtendría multiplicando los posibles estados que puede tomar la variable  $x$ ,  $\log_2 1/P_x$ , por su probabilidad de ocurrencia,  $p(x)$ . Luego la entropía de  $x$  será,  $H(x)$ :

$$H(x) = \sum_x p(x) \log_2 \frac{1}{p_x}$$

En consecuencia, la entropía es una medida de la aleatoriedad o incertidumbre de o de la cantidad de información que, en promedio, nos proporciona conocimiento de .

Originariamente (Quinlan, 1986 y 1993) seleccionaba para hacer cada partición aquella variable que proporcionaba la máxima información sobre  $x$ , es decir, maximizaba  $I(x; y_i)$  (magnitud denominada Gain). En las versiones posteriores del algoritmo, para establecer la variable que proporciona la mayor información, se selecciona aquella  $y_i$  que maximiza la magnitud  $I(x; y_i)/H(y_i)$ . En el caso del C4.5 se emplea este ratio de ganancia (denominado

Gain Ratio). Si el procedimiento descrito, se aplica de manera reiterada se va construyendo el árbol de decisión, hasta que se alcanza la pureza del nodo y con ello finaliza el proceso.

Se puede observar, mediante el ejemplo (Figura 3), que la interpretación de los resultados es sencilla y es fácil ver la lógica que se debe seguir para su aplicación a través del recorrido de sus reglas o ramas del árbol dibujado. De ahí su atractivo ya que puede ser analizado incluso por personas con poca experiencia en el tema.

#### 4.2 Metodología no paramétrica. Metodología Rough Set

La Inteligencia Artificial cuenta con numerosas técnicas dentro de las cuales se tienen los algoritmos de Inducción de Reglas. Es aquí donde se engloba la Teoría de Rough Sets (*Método de Conjuntos Aproximados*).

Esta metodología se ha aplicado con éxito en diversos problemas financieros (Witlox y Tindemans, 2004; Huang y Jane, 2009; Yao y Herbert, 2009; Shyng *et al.*, 2010 y Boudreau-Trudel y Kazimierz 2012). También hay estudios que lo han aplicado al sector asegurador (Sanchis, *et al.*, 2007; Shyng *et al.*, 2007 y Segovia-Vargas *et al.*, 2015), motivo que justifica su aplicación al problema que nos ocupa relativo a la caída de cartera al sector asegurador.

Esta teoría fue introducida en el año 1982 por Pawlak como una nueva técnica de gran utilidad para el análisis y contenido de tablas de información que describen a un conjunto de objetivos por medio de una serie de atributos (Pawlak, 1991).

Esta teoría utiliza la experiencia en eventos pasados acumulados sobre una serie de patrones de datos, para finalmente poder obtener una serie de reglas en forma de sentencias lógicas que nos ayuden en la toma de decisiones futuras. Es así como hace necesario referenciar que un problema de decisión implica un conjunto de *objetos* descritos por un conjunto de *atributos* y, el cual se puede representar mediante una *tabla de decisión* (Skowron y Grzymala la Busse, 1991).

De esta forma, se puede enumerar las siguientes ventajas que caracterizan a la Teoría *Rough Set* destacando:

- Utilización de variables tanto de tipo cuantitativo como cualitativo
- No necesita de ningún tipo de información preliminar o adicional de los datos como distribuciones de probabilidades estadísticas.
- Eliminación de variables redundantes y de esto modo enfocarnos en conjuntos mínimos de variables logrando una reducción del costo y tiempo del proceso asumido por el centro decisor
- Obtención de una serie de reglas de decisión de fácil comprensión. Así mismo dichas reglas están bien respaldadas por experiencia pasada lo cual argumenta las decisiones que se toman.

La Teoría *Rough Set* representa dicho conocimiento de los objetos en forma de una tabla de información: En la filas  $x$  se indican los objetos (acciones, empresas, etc.) y en las columnas  $q$  se representan los atributos. Los valores del

atributo son las entradas de la tabla tomando el valor  $f(x,q)$ . De esta forma, cada fila en la tabla representa la información sobre un objeto  $S$ ; siendo éste el sistema de información denominado como sistema de representación del conocimiento.

El conjunto de atributos se divide en un subconjunto de atributos de condición y otro subconjunto de atributos de decisión; en este caso, si se distingue entre ambos conjuntos, se obtiene la tabla de decisión.

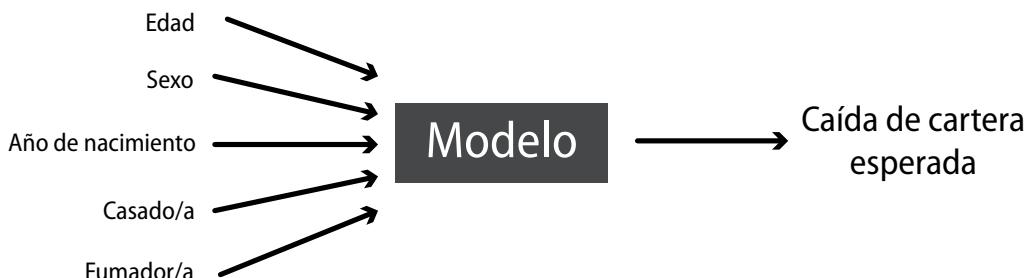
**Tabla 1** Ejemplo Tabla de Decisión

Criterios Candidatos	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	D
$X_1$	4	4	4	4	2	2	1	A
$X_2$	3	3	4	3	2	1	1	R
$X_3$	3	4	3	3	1	2	2	R
$X_4$	5	3	5	4	2	1	2	A
$X_5$	4	4	5	4	2	2	1	A
$X_6$	3	4	3	3	2	1	3	R
$X_7$	4	4	5	4	2	2	2	A
$X_8$	4	4	4	4	2	2	2	A
$X_9$	4	4	4	4	2	2	2	R
$X_{10}$	5	3	5	4	2	1	2	A
$X_{11}$	5	4	4	4	1	1	2	A
$X_{12}$	5	3	4	4	2	2	2	A
$X_{13}$	4	3	3	3	3	2	2	R
$X_{14}$	3	3	4	3	2	3	3	R
$X_{15}$	4	5	5	4	2	1	1	A

Fuente: Moscarola, 1978 y Slowiski, 1993

Ahora bien, otro de los conceptos tratados en este tipo de metodología son *las reglas de decisión*, siendo éstas el conjunto de datos que representan la experiencia. Se entiende que el conjunto de datos contiene información de un conjunto de *objetos* descritos por un conjunto de *atributos*. Por lo tanto, el tema consiste en encontrar reglas que determinen si un objeto pertenece a un subconjunto particular denominado *clase de decisión* o a un *concepto*.

**Figura 4** Traducción del Modelo Predictivo en Seguros de Vida



Fuente: Propia de los autores.

Como ya se ha mencionado, este tipo de reglas se presentan en forma de sentencias lógicas: *SI <condiciones> ENTONCES <clases de decisión>* (Sanchis et al. 2007; Segovia-Vargas, 2003; Segovia-Vargas et al., 2003).

Así, cada regla de decisión se caracteriza por el número de objetos que satisfacen la parte de la condición de la regla y pertenecen a la clase de la decisión sugerida, esto es, la denominada *fuerza* de la regla. Así se tiene que no todas las reglas son igual de importantes o fiables para el agente decisor, cuanto más *débil* la regla es, *menos fiable* es en la toma de decisión.

La técnica de *Rough Set* busca la mínima descripción posible en términos de atributos. Una descripción mínima permite un minucioso análisis de los conflictos. Además si los atributos son consecuencias de algunas decisiones, lo que en ciertas aplicaciones se puede interpretar como relaciones “causa-efecto”, la metodología *Rough Set* permite descubrir las mínimas dependencias elementales entre las consecuencias. Todo esto, en el contexto del presente trabajo sería de gran utilidad, ya que mediante mínimas interdependencias de las variables cualitativas que caracterizan a cada uno de los clientes dentro de una entidad aseguradora, se pueda definir el perfil de dichos clientes con mayor propensión a la anulación de su contrato de seguros (Shyng et al., 2007).

#### 4.3 Metodología paramétrica. modelos lineales generalizados (GLMs)

Existen pocos estudios que utilizan la metodología ofrecida por los GLM para el análisis del riesgo de caída de cartera al que está expuesta una entidad aseguradora (Cerchiara et al., 2008). Sin embargo, mediante la utilización de dicha metodología, se puede lograr identificar la información o características del asegurado que describan el tipo de clientes propensos a la anulación de su contrato de seguros. De esta forma, la metodología de los GLM podría ofrecer una herramienta que reconozca ciertas relaciones no lineales que podrían ayudar al análisis de los parámetros que afectan a este riesgo, y así permitir tener conocimiento sobre las correlaciones y dependencias de los factores que lo propician con el fin de lograr un control y gestión del riesgo en su globalidad.

Es así como se encuentran los modelos predictivos que trabajan relacionando un evento determinado (en este caso, la anulación o caída de cartera que puede presentar una entidad aseguradora), con un cierto número de factores (Figura 4):

Dentro de dichos modelos predictivos, se encuentran los Modelos Lineales Generalizados (GLM, por sus siglas en inglés), que constituyen una generalización de los tradiciona-

les Modelos Lineales (LM – *Linear Models* –), donde se asume que el valor esperado de la variable dependiente se encuentra condicionado a las variables independientes expresándose como una combinación lineal de los valores que dichas variables.

Los Modelos Lineales Generalizados, introducidos a comienzos de los años 70 (Nelder y Wedderburn, 1972), resumen un grupo homogéneo de métodos de regresión (logística, Poisson, gamma, etc.), previamente consideradas de forma independiente.

Los GLM cuentan con su propia estructura, elementos y método de análisis e interpretación de los resultados, que los hace más atractivos y ofrece mayor facilidad en su aplicación.

#### 4.3.1 Estructura

La estructura de un GLM presenta una relación lineal entre las variables explicativas y una transformación de la media de la variable respuesta. Esto es, que no existe una relación lineal entre ambas, sino entre una función de enlace (*función “link”*) y las variables explicativas:

#### 4.3.2 Componentes

Los Modelos Lineales Generalizados tienen tres componentes básicos, que se detallan a continuación:

- **Componente Aleatorio:**

Identifica la variable respuesta y su distribución de probabilidad. Este componente consiste en una variable aleatoria  $Y$  con observaciones independientes ( $y_1, \dots, y_N$ ).

En muchas aplicaciones las observaciones de  $Y$  son binarias y se identifican como éxito y fracaso. Aunque de modo más general, cada  $Y_i$  indicaría el número de éxitos de entre un nú-

mero fijo de ensayos; y se modelaría como una distribución binomial. En otras ocasiones cada observación es un recuento, con lo que se puede asignar a  $Y$  una distribución de Poisson o una distribución binomial negativa.

En otras ocasiones, las observaciones son continuas y se puede asumir para  $Y$  una distribución normal. Todos estos modelos se pueden incluir dentro de la llamada familia exponencial de distribuciones.

- **Componente Sistemático:**

Específica las variables explicativas, que entran en forma de efectos fijos en un modelo lineal, es decir las variables  $x_j$  que se relacionan como:

$$\alpha + \beta_1 x_1 + \dots + \beta_k x_k$$

Esta combinación lineal de variables explicativas se denomina predictor lineal, el cual se puede expresar como:

$$\eta = \sum_i \beta_j x_j$$

- **Función Link:**

Como ya se ha comentado, existe una relación entre la variable respuesta y las explicativas, la cual no siempre corresponde a una relación lineal entre ambas. Pues bien, de aquí surge el concepto de “*función vínculo*” o “*función enlace*” que se ocupa de linealizar la relación entre la variable dependiente y las variables explicativas mediante la transformación de la variable respuesta.

Se denota el valor esperado de  $Y$  como  $\mu = E[Y]$ , entonces la función link especifica una función que relaciona a  $\mu$  con el predictor lineal como:

$$g(\mu) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k$$

En otras palabras, la función link  $g(\mu)$  o función enlace que relaciona el componente aleatorio y el componente sistemático; de tal forma que:

$$\eta = g(\mu) = \sum_j \beta_j x_j$$

En la siguiente tabla, se resumen algunas de las funciones vínculos comúnmente más utilizados (Tabla 2):

**Tabla 2 Funciones Vínculo**

Función vínculo	$g(x)$	$g^{-1}(x)$
Identidad	$x$	$x$
Log	$\ln(x)$	$e^x$
Logit	$\ln(x/(1-x))$	$e^x/(1+e^x)$
Reciproca	$1/x$	$1/x$

**Fuente:** Propia de los autores a partir de Anderson *et al.* 2007.

### 4.3.3 Estructura de modelos comunes

Dependiendo de la naturaleza de los valores que toma la variable  $Y = (y_1, \dots, y_N)$ ; se deberá tomar la combinación idónea de la distribución de probabilidad y la función de enlace que mejor se adapte al objetivo que se plantea. De esta forma, algunos de los modelos más comunes con sus respectivas estructuras, se resumen a continuación (Tabla 3):

**Tabla 3 Estructura de los modelos más comunes**

$Y$	Frecuencias de Siniestros	Número de Siniestros	Coste medio de Siniestros	Probabilidad (renovación/cancelación)
Función enlace $g(x)$	$\ln(x)$	$\ln(x)$	$\ln(x)$	
Error	Poisson	Poisson	Gamma	Binomial
Parámetro Escalar	1	1	Estimado	1
Función Varianza $V(x)$	$x$	$x$	$x^2$	$x(1-x)$
Pesos $\omega$	Exposición	1	# siniestros	1
Offset $\zeta$	0	$\ln(\text{exposición})$	0	0

**Fuente:** Propia de los autores a partir de Anderson *et al.*, 2007.

### 4.3.4 GLM en la Práctica

Es importante tener en cuenta que no existe un único modelo válido que se pueda ajustar a la muestra de datos analizados. Es decir, la mayoría de las veces existe más de un modelo posible y es por ello, que el tema más complicado es saber y comprobar cuál ajusta mejor y por lo tanto, es el más adecuado de todos ellos.

Para ello, se pueden identificar cuatro fases que permiten estructurar la construcción de un GLM (Anderson *et al.*, 2007; McCullagh y Nelder, 1989), que serían:

**Análisis Preliminar:** El cual considera la etapa de preparación de los datos, así como un análisis exploratorio de las variables que serán consideradas dentro del modelo

- Distribución de los datos (frecuencias, n° de pólizas, importe de siniestros)
- Análisis Univariantes y Bivariantes
- Análisis de Correlaciones
- Categorización o Interacciones de Factores

**Iteración del Modelo:** En esta fase, se recurre a la selección adecuada de factores que mejor se ajustan a los datos, y por tanto, hacer uso recurrente del diagnóstico de las hipótesis y parámetros del modelo

- Selección de Factores
- Errores Estándar

- Pruebas de Devianza
- Consistencia en el tiempo
- Sentido Común

**Depuración del Modelo:** Implica refinar el modelo buscando la máxima simplificación posible, encontrando la interacción entre las variables y haciendo uso de la suavización de los resultados

- Residuos: Se pueden obtener varios tipos de residuos para analizar cómo los valores esperados varían de los observados; conviene analizarlos a través de los siguientes gráficos: histograma de los residuos, gráfico de residuos *vs* valores estimados, etc.
- Leverage: Esta prueba ayuda a identificar las observaciones que tienen influencia excesiva en el modelo; es decir, es útil en la identificación de observaciones puntuales con bastante influencia sobre los resultados del ajuste del modelo, esto es, valores atípicos.
- Suavizado: Se procede a la suavización de los parámetros en busca de una mejora en el poder predictivo del modelo; incorporando ciertos conocimientos o juicio del experto buscando un comportamiento natural del evento estimado o modelado.

**Interpretación de Resultados:** Finalmente, lograr traducir los resultados obtenidos para conseguir la mejor explicación del modelo e información obtenida del GLM planteado y, que de esta forma, se puedan entregar los resultados para una correcta toma de decisiones de una forma atractiva y que aporte valor.

## 5. Aplicación empírica: métodos paramétricos y no paramétricos en el análisis de la caída de cartera

### 5.1 Tratamiento de la información: inicio de la aplicación empírica

En esta sección se presentan los resultados obtenidos tras la aplicación de las metodologías propuestas a una cartera real de pólizas de Seguros de Vida Individual de una entidad que actualmente opera dentro del mercado asegurador español. Los datos no están disponibles a nivel general debido a la política de privacidad existente en España relativa a la protección de datos y a cuestiones de confidencialidad industrial. La muestra analizada consta de 19.784 pólizas para el estudio en cuestión. Dado que la compañía opera a nivel nacional y el territorio español está dividido en 17 comunidades autónomas y dos ciudades autónomas, la muestra incluye todas las regiones autónomas y es por tanto representativa a nivel nacional.

La selección de variables en modelos de estimación de tasas de anulación de pólizas dentro de una entidad aseguradora es un tema complicado, ya que pueden existir diversos factores influyentes en la caída de cartera. Por un lado, hay factores ligados al propio cliente como la edad o el sexo. Por otro, hay factores directamente ligados a las características de la póliza como son la antigüedad, el tipo de seguro. Finalmente hay factores ligados o bien al canal de venta o a niveles de competencia en que ha sido adquirido el contrato de seguros. Además, dicha selección se podrá ver limitada por la propia información disponible que se pueda considerar.

Con base en esta situación, y teniendo en cuenta la accesibilidad que se ha tenido a los datos utilizados, se ha seleccionado las siguientes variables que son susceptibles a explicar el

**Tabla 4** Variables seleccionadas para la aplicación empírica

Nombre	Descripción
SEXO	Sexo del asegurado
EDAD_ACTUARIAL	Edad del asegurado al cierre del ejercicio
ANTIGUEDAD	Años de antigüedad de la póliza desde su Fecha de Emisión hasta la Fecha de Cálculo
TIPO_PRODUCTO	Tipo de Producto contratado: Individual o Colectivo; Ahorro o Riesgo
TIPO_PRIMA	Tipo de Prima: Única o Periódica
RED	Tipo de Red: Propietaria o No Propietaria
FORMA_PAGO	Periodicidad del pago de la prima
ANO_EFECTO	Año de Emisión o Efecto de la póliza
EDO_CIVIL	Estado Civil del asegurado, si lo ha comunicado
HIJOS	Tiene (o no) hijos el asegurado, si lo ha comunicado
VALOR_CLIENTE	Valor "comercial" definido por la Aseguradora de acuerdo a metodologías internas
ICE	Índice de Capacidad Económica
NIV_INGRESOS	Nivel de Ingresos
NIV_ESTUDIOS	Nivel de Estudios
TIPO_PRESTACION	Indica si la póliza está en Vigor o Anulada

Fuente: Propia de los autores

comportamiento de la tasa de caída que presenta una entidad aseguradora, y las cuales han sido consideradas en las aplicaciones empíricas que se abordan (Tabla 4):

La única variable continua que se tiene en la muestra es la correspondiente a la EDAD. El empleo de este tipo de información implica una división del dominio original en algunos subintervalos; así como su correspondiente asignación de códigos cualitativos a dichos subintervalos (Segovia-Vargas, 2003). Esta manipulación o discretización hace que la interpretación de los resultados finales sea más sencilla. No existe una única forma para establecer los subintervalos; por lo que se toma la recomendación que se utiliza frecuentemente en los trabajos de investigación (Laitinen (1992), McKee, (2000) o Segovia-Vargas *et al.*, (2003)), que es el uso de percentiles que siguen las distribuciones en las variables continuas.

Por lo que la discretización y asignación de códigos para esta variable quedaría de la siguiente forma (Tabla 5):

**Tabla 5** Códigos asignados a la variable Edad

Cod	Rango de edad	Percentil	Observado
1	[ 0 - 37 )		
2	[ 37 - 43 )	P20	37
3	[ 43 - 49 )	P40	43
4	[ 49 - 58 )	P60	49
5	[ 58 - y más )	P80	58

Fuente: Propia de los autores

## 5.2 Resultados obtenidos y estudio comparativo

En esta sección, se analizan los resultados obtenidos tras la aplicación empírica de las metodologías propuestas sobre la cartera muestra. En dicha muestra se tiene la distinción de dos categorías que vienen dadas según el criterio que toma la variable de decisión TIPO DE PRESTACION. Esta variable toma el valor de 0 indicando que la póliza se encuentra en Vigor o por el contrario, toma el valor 1 si la póliza se encuentra Anulada (Tabla 6):

**Tabla 6 Códigos Variable Tipo de Prestación**

COD	TIPO_PRESTACION	Nº de pólizas	% Peso
0	VIGOR	16,568	83.74%
1	ANULADA	3,216	16.26%
TOTAL		19,784	100.00%

Fuente: Propia de los autores

Ambas categorías han sido evaluadas de acuerdo a los valores que tomas las 14 variables cualitativas y cuantitativas seleccionadas (Tabla 6) considerando una misma base de datos.

72

**Figura 5** Salida de Resultados- WEKA (Algoritmo C4.5)

```

Number of Leaves: 445
Size of the tree: 633
Time taken to build model: 0.58 seconds

==== Stratified cross-validation ====
==== Summary ====

Correctly Classified Instances      17076          86.3122 %
Incorrectly Classified Instances   2708           13.6878 %
Kappa statistic                   0.4111
Mean absolute error               0.1868
Root mean squared error          0.3221
Relative absolute error           68.5924 %
Root relative squared error     87.2969 %
Total Number of Instances        19784

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

          TP Rate  FP Rate  Precision Recall    F-Measure  ROC Area  Class
          0.954    0.603    0.891    0.954    0.921      0.818     0
          0.397    0.046    0.624    0.397    0.486      0.818     1
Weighted Avg.       0.863    0.512    0.847    0.863    0.85      0.818

==== Confusion Matrix ====

      a      b      <-- classified as
15798  770    |  a = 0
  1938  1278   |  b = 1
  
```

Fuente: Propia de los autores.

### 5.2.1 Árboles de Decisión

En primer lugar, la base de datos se ha introducido y programado en WEKA (Witten y Frank, 2005),<sup>3</sup> programa informático que ha desarrollado el análisis del algoritmo C4.5. Se puede observar que los resultados cuenta con un porcentaje de aciertos del 86.31% (*Correctly Classified Instances*) de acuerdo al Resumen de Validación de Resultados que arroja el programa; lo cual justifica su interpretación (Figura 5).

3 <http://www.cs.waikato.ac.nz/ml/weka/>

Además, si se observa la diagonal de la matriz de confusión, se tienen unos valores superiores a los elementos  $a_{21}$  y  $a_{12}$ . Esto es,  $a_{11}=15.798$  es mayor que  $a_{21}=1938$ ; y por otro lado,  $a_{22}=1.278$  es mayor que  $a_{12}=770$ . En concreto, se observa que un 39,7% de las pólizas anuladas son clasificadas correctamente y un 95,4% de las pólizas en vigor. Para analizar el modelo obtenido por la metodología C4.5 e interpretar dicho árbol, habría que ir descendiendo, hasta completar la totalidad de sus hojas (regla de decisión).

La variable ANTIGÜEDAD es una pieza clave en el modelo y suena lógico desde el punto de vista de la existencia de cierto nivel de fidelización por parte de los clientes. Por tanto, a mayor antigüedad dentro de la compañía de seguros, menor es la propensión o susceptibilidad que se tiene de cancelar su contrato de seguros. Sin embargo, existe cierto punto de inflexión a partir del cuarto -quinto año con la compañía, ya que a partir de que la variable ANTIGÜEDAD toma el valor de 4 y, el algoritmo de árboles no recoge alguna otra variable que nos indique cierto patrón de comportamiento de los clientes.

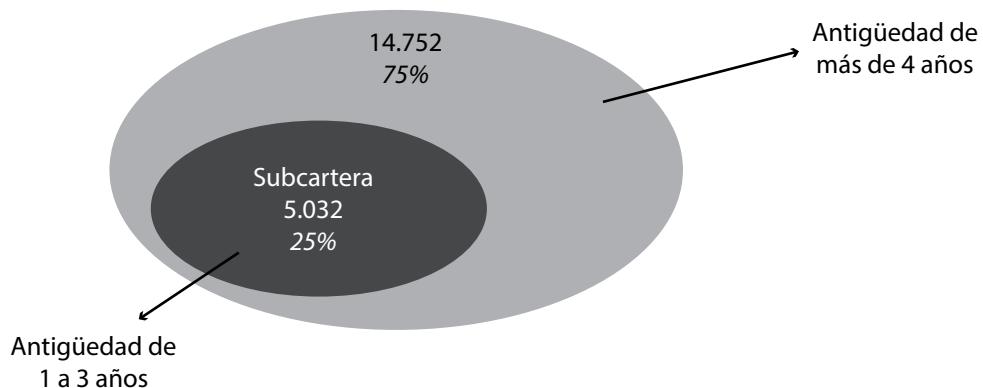
El problema para la compañía es que su cartera sobrepase los 4 años de antigüedad, es decir, el análisis de caída de cartera tendría que centrarse en los primeros años de vida de las

pólizas, que es donde realmente se presenta el riesgo de caída de cartera. Siendo así, ha sido necesario realizar el análisis del árbol, subdividiendo la muestra de acuerdo con la variable ANTIGÜEDAD, ya que dicha variable es un factor clave en el comportamiento de los clientes (es decir, de la cartera total de 19.784 pólizas, 14.752 pólizas son clasificadas por el árbol únicamente como pólizas en VIGOR mediante la variable ANTIGÜEDAD). Siendo así, se tiene que el 75% de la muestra, cuenta con más de 4 años de antigüedad dentro de la compañía, sin poder distinguir otro tipo de variable o característica del asegurado que determine específicamente el tipo de clientes que se retienen en cartera. Y es así como se debe substraer un tipo “*subcartera*” correspondiente a la cartera de clientes más recientes.

De esta forma se puede tener una fuerza global muy alta para la cartera total explicada principalmente por la variable ANTIGÜEDAD y, complementarla con una fuerza relativa determinada con respecto a la *subcartera*. Así, se centrará el análisis en los patrones de las antigüedades 1, 2 y 3 que corresponde al 25% de la muestra (5.032 pólizas), cuyo comportamiento de los asegurados dentro de esta *subcartera* se podría considerar independiente al resto de la muestra (Figura 6).

73

**Figura 6** Composición de la cartera para la subdivisión de acuerdo a la variable ANTIGÜEDAD



Fuente: Propia de los autores.

Por tanto, se centra el análisis de los patrones sobre la *subcartera* de pólizas más recientes de la compañía cuya ANTIGÜEDAD se encuentra entre 1 y 3 años de duración. A partir de esta *subcartera*, se obtienen las principales reglas de decisión sobre los 2 posibles comportamientos del asegurado:

- TIPO PRESTACIÓN = 1 correspondiente a los factores determinantes de los clientes susceptibles a la cancelación de su póliza; y
- TIPO PRESTACIÓN = 0 correspondiente a los patrones presentados en los clientes propensos a mantener su contrato de seguros en vigor

74

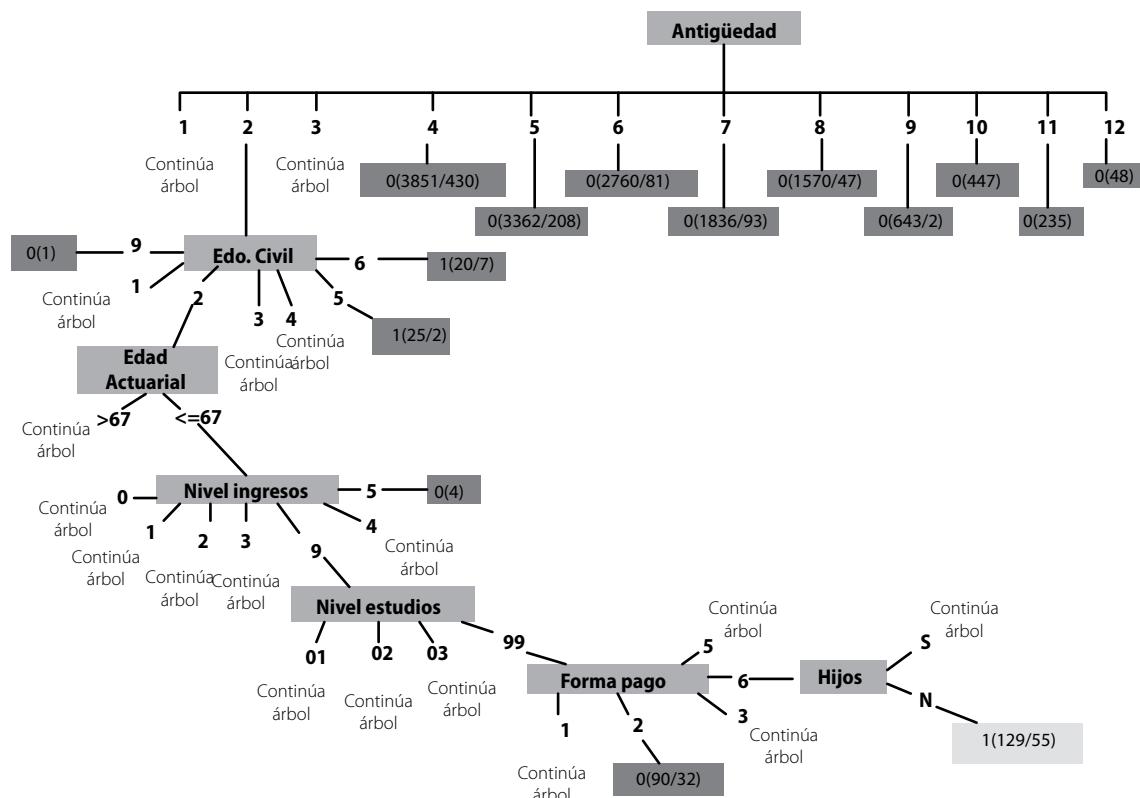
A continuación se presentan, a manera de ejemplo, una de estas ramas del árbol que dan lugar a las principales reglas detectadas para

la CLASE 1; esto es, las reglas de decisión que determinarían el tipo de clientes propensos a la cancelación de su póliza de seguros.

De tal forma, que descendiendo por cada uno de las ramas, se obtiene el perfil del cliente susceptible a anular su póliza de seguros o, para la categoría contraria, el perfil del cliente con mayor propensión a continuar con su contrato de seguros. Así finalmente se obtiene en términos matemáticos el cumplimiento del patrón: total de casos y el porcentaje de fuerza que tiene cada rama analizada.

El resumen de reglas para la Clase 1=Cancelación, sumando estas 3 reglas acumulan una fuerza del 2.6% sobre la cartera total de pólizas; y a su vez, representa un 10.3% como fuerza relativa sobre la subcartera (Tabla 7).

**Figura 7** Árbol de Decisión (Rama de Antigüedad=2\_Regla 3 de la Clase 1)



Fuente: Propia de los autores.

**Tabla 7** Resumen de Resultados Arboles de Decisión- Clase 1 CAIDA

Categoría: 1								Cartera	Subcart
Caída								19,784	5,032
								s / Cartera	s / Subcart
Regla 1	Prod-Ahorro	Ice-Medio	Edad-<=61				518	2.6%	10.3%
Regla 2	Prod-Riesgo	Sexo-Hombre					166	0.8%	3.3%
Regla 3	Edo-Casado	Edad- <=67	Ingrs-na	Estud-na	FP-Única	Hijos N	223	1.1%	4.4%
							129	0.7%	2.6%

---

**Fuente:** Propia de los autores

**Tabla 8 Resumen de Resultados Arboles de Decisión- Clase 0 RETENCIÓN**

**Fuente:** Propia de los autores

Por otro lado, el resumen general de reglas para la Clase 0=Retención engloba 4 reglas que acumulan una fuerza del 4.4% sobre la cartera total de pólizas, y a su vez, representa un 17.5% como fuerza relativa sobre la subcartera (Tabla 8).

A raíz de la aplicación de la técnica que ofrecen los Árboles de Decisión, se pueden resumir los principales resultados obtenidos con algunas de las variables que definen al perfil del cliente “cancelador” como son: primamente, la ANTIGÜEDAD de la póliza; otra de las variables que se encuentra relacionada con la duración del contrato es el TIPO PRODUCTO, el cual resulta ser un segundo patrón de comportamiento identificado; otras tres variables detectadas para ambas categorías analizadas, tanto Retención como Cancelación, son la EDAD y ESTADO CIVIL, con

lo cual esto puede de ser de vital importancia en la toma de decisión para la contención del riesgo de caída de cartera; la FORMA PAGO sugiere ser otra característica interesante que resulta del modelo y, finalmente, otra de las variables con menor fuerza, pero no por ello poco significativa, es la variable HIJOS, es decir, si se ha declarado tener o no Hijos.

### 5.2.2 Rough set

Uno de los primeros pasos que se debe realizar en la aplicación de este tipo de modelo, es la validación del mismo. Para ello, se efectúa un procedimiento de validación cruzada, la cual consiste en hacer ciertas particiones de igual tamaño en los datos dejando una muestra para estimar el modelo y otro conjunto de datos para su validación. Este proceso se repite

tantas veces como particiones se hacen. El resultado final es la media de los resultados obtenidos, con frecuencia se suele utilizar 10 particiones. Así se tiene que cuanto más alta es la tasa de validación cruzada, mayor fiabilidad del modelo obtenido.

Pues bien, siendo así la aplicación de la metodología *Rough Set* ha presentado una precisión satisfactoria utilizando dicha validación cruzada en 10 pliegues. Dando como resultado un 83.7%, indicando un porcentaje de pólizas correctamente clasificadas considerablemente bueno elevando el poder de predicción del modelo (Figura 8).

Al igual que se hiciese en la aplicación de Arboles de Decisión, se observa la diagonal de la matriz de confusión. Se tiene que el valor de =1.464,4 es mayor que =191,5; y por otro lado, =192,1 es mayor que =130. Por lo que se tiene que un 59,5% de las pólizas anuladas son clasificadas correctamente y un 88,4% de las pólizas en vigor; dejando evidencia la fiabilidad del poder predictivo de las reglas de decisión obtenidas del modelo.

A continuación, a manera de ejemplo, se presenta los principales resultados que arroja el modelo. Esto es, una selección de las sentencias o reglas con mayor fuerza que clasifican a la muestra de pólizas del estudio en la Categoría 1=Cancelación de la Póliza (Tabla 9).

Cabe observar que no existe regla alguna donde aparezcan todos los atributos, lo que significa que no existe una combinación global de todos las características de un cliente que distingan exactamente el perfil del cliente "cancelador". En otras palabras no existe alguna variable en especial que sobresalga sobre otras, sino por el contrario la interrelación entre los factores es lo que conforma cada una de las opciones que llevarían a definir y categorizar a los clientes. Sin embargo, sí existe la posibilidad ciertas condiciones o tipos de clientes determinen el nivel de riesgo de anulación o caída de cartera.

Por otro lado, se tiene el conjunto de reglas que clasifican a la Categoría 0=Retención de la Póliza (clientes susceptibles a conservar sus pólizas con la entidad aseguradora), se clasifican mejor, no sólo por la mayor fuerza que presentan sus reglas, sino por la cantidad de reglas de decisión que resultan.

A manera de resumen general de las variables más significativas para cada una de las categorías, se tiene (Tabla 10).

Así finalmente, se puede concluir que las variables que en ambas categorías están contenidas con mayor frecuencia y por tanto, pueden ser consideradas como los atributos que clasificarían el éxito o fracaso de la conserva-

**Figura 8** Salida de validación cruzada (*Rough Set*)

		0	1	No. Of. Obj.	Accuracy	Coverage
Actual	0	1464.4	192.1	1656.5	0.884	1
	1	130	191.5	321.5	0.595	1
True positive rate		0.92	0.5			

Total number of tested objects: 1,978

Total accuracy: 0.837

Total coverage: 1

**Fuente:** Propia de los autores.

**Tabla 9** Reglas con Mayor Fuerza- Rough Set (CATEGORIA 1= Cancelación)

	Sexo	Edad actuarial	Anti-güedad	Tipo producto	Forma pago	Edo. civil	Hijos	Valor cliente	Niv. ice	Niv. ingresos	Niv. estudios	Tipo prestación (clase)	Fortaleza
1	H		1	1	1	2					99	1	76
2	H		1	1	1		N					1	74
3			1	1	1			A				1	71
4	H		1	1	1				4			1	47
5		1	1	1	1	2	N					1	47
6			1	1	1	2	N		4			1	46
7		1	1	1	1	2				99		1	42
8	H	5	1					A				1	39
9	H		1	1				A	4			1	39
10		1	1	1		2	N	A				1	39
11			1	1	1	2	N		3	9		1	39
12			1	1	1	2	N		3		99	1	39
13		5	1	1				A	3			1	38
14		1	1	1			N	A			99	1	38
15	H	1	1	1	1						99	1	38
16		5	1	1			N	A				1	37
17		5	1	1	1	2				9		1	37
18		5	1	1	1	2				99		1	37
19		1	1	1	1	2			4			1	36
20		1	1	1		2		A	4			1	36
21		1	1	1				A	4		99	1	36
22		1	1		1	2	N	A				1	35
23		5	1		1	2			3	9		1	31
24		5	1		1	2			3		99	1	31
25		5	1		1		N		3	9		1	31
26		5	1		1		N		3		99	1	31
27	H	5	1		1	2						1	30
28	H	5	1		1		N					1	30

Fuente: Propia de los autores

77

**Tabla 10** Resumen de Resultados Rough Set por categoría (Caída\_1 y Retención\_0)

Categoría: 1 Caída	Categoría: 0 Retención
Variables significativas	Variables significativas
* Antigüedad	* Antigüedad
* Forma Pago	* Tipo Producto
* Tipo Producto	* Forma Pago
* Edo Civil	
* Hijos	

Fuente: Propia de los autores

ción o anulación de la cartera de pólizas de una compañía aseguradora son: ANTIGÜEDAD, TIPO PRODUCTO y FORMA PAGO.

### 5.2.3 Mínimos cuadrados generalizados (GLMs)

Se abordará la aplicación empírica de un modelo predictivo como lo son los GLM, para la identificación de variables o factores que indiquen el posible abandono de una póliza de vida en una entidad aseguradora.

Así mismo, dicha aplicación, más que buscar la exactitud de los resultados numéricos, busca ser una metodología contraste para las técnicas de Inteligencia Artificial. Es decir, mediante un modelo paramétrico como lo son los GLM se pretende complementar y fortalecer las conclusiones obtenidas del modelo no paramétrico que proporciona la Inteligencia Artificial.

Para la validación del modelo, así como de las variables, se trata de ir quitando los factores menos significativos, hasta conseguir que

todos o la mayoría de sus p-valor se encuentren por debajo de cierto intervalo de confianza. Inicialmente se puede considerar el 95% de confianza, aunque poco a poco seguramente se tenga que ser más conservador y bajar este intervalo de confianza.

Por otro lado, también se debe ir monitoreando como se mueve la medida de bondad de ajuste del Criterio de Información de Akaike (AIC), ya que nos proporcionará un dato de referencia para la selección del modelo óptimo. Así se busca tener que dicho indicador deje de descender, con lo cual se podría decir que se ha encontrado el modelo adecuado.

Siguiendo la técnica que se tomó de “prueba-error”, se han ido obteniendo resultados, logrando obtener el nivel de AIC más bajo. En consecuencia, se obtiene el siguiente cuadro resumen de resultados (Figura 9).

Para el diagnóstico del modelo se observa que todos los p-valores se encuentran por debajo del nivel objetivo, es decir, menor que el 0.05, lo que indicaría que se trata del mejor

**Figura 9** Salida de Resultados Modelo V- R (GLM, Binomial, Logit)

```
R Console

Call:
glm(formula = Prestacion ~ Antiguedad + Tipo + Forma + Hijos +
    EdoCivil, family = binomial, data = base1)

Deviance Residuals:
    Min      1Q   Median      3Q     Max 
-1.6605 -0.5196 -0.2834 -0.1092  3.2810 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 1.10486   0.15951   6.927 4.31e-12 ***
Antiguedad -0.77025   0.01493  -51.584 < 2e-16 ***
Tipo        0.29125   0.00707   3.779 0.000157 ***
Forma2     -0.22792   0.009023  -2.526 0.011539 *  
Forma3     -0.42071   0.09380  -4.485 7.29e-06 ***
Forma5     -0.14074   0.07815  -1.801 0.071728 .  
Forma6     -0.52526   0.08910  -5.895 3.75e-09 *** 
HijosS     -0.07650   0.04663  -1.641 0.100899    
EdoCivil2  0.17129   0.06874   2.492 0.012710 *  
EdoCivil3  0.02798   0.10505   0.266 0.789957    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

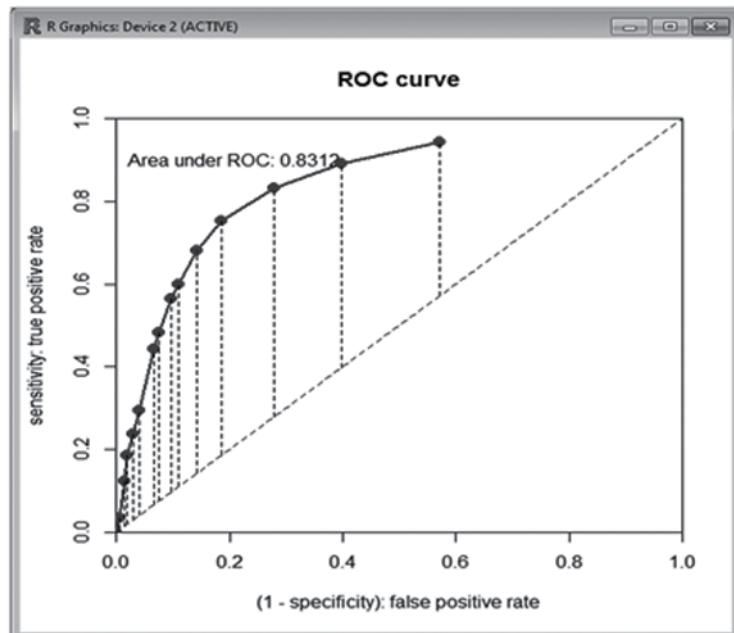
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 17564  on 19783  degrees of freedom
Residual deviance: 12940  on 19774  degrees of freedom
AIC: 12960

Number of Fisher Scoring iterations: 6
```

Fuente: Propia de los autores.

**Figura 10** Curva de ROC del Modelo VI- R (GLM, Binomial, Logit)



Fuente: Propia de los autores.

modelo. Este modelo no era el que obtiene el AIC más bajo (de hecho se obtiene un AIC un punto por arriba con respecto al menor AIC), sin embargo, se gana en ajuste, logrando algo por arriba del 95% inclusive.

Para hacer el diagnóstico del modelo, se suele hacer un análisis visual de los resultados, recurriendo a la Curva ROC (siglas en inglés *Receiver Operating Characteristic*). Se trata de una Teoría de detección de señales donde mediante la representación gráfica del ratio de “Verdaderos Positivos” frente al ratio de “Falsos Positivos”. El análisis de la curva ROC proporciona una herramienta para seleccionar los modelos posiblemente óptimos. Se suele utilizar como una medida para la elección entre pruebas diagnósticas distintas. A continuación la curva ROC resultante del modelo seleccionado sería (Figura 10).

Es decir, el resultado arroja un 83,12% lo cual significa que se trata de un test bueno, ya que supone que la probabilidad de “diagnosticar” a una póliza como candidata a ser

anulada sea correctamente clasificada es del 83,12%. Únicamente, a manera de contraste, se ha generado la curva ROC para los otros modelos anteriores al seleccionado, y en todos los casos, éste último es el que resulta con el mejor porcentaje, lo cual confirma que la elección del modelo ha sido acertada. Con lo cual, el modelo óptimo es el que arroja las variables ANTIGÜEDAD, TIPO PRODUCTO, FORMA PAGO y EDO CIVIL.

#### 5.2.4 Resumen de los resultados de las distintas metodologías

El objetivo básico de este trabajo es mostrar la complementariedad de las metodologías paramétricas (GLM) y no paramétricas (Árbol de decisión y *Rough Set*). En este sentido, a manera de resumen general, en la Tabla 11 se comparan los resultados obtenidos entre los distintos modelos en cuanto al porcentaje de clasificaciones correctas de las pólizas.

**Tabla 11 Comparativa de Resultados:  
Precisión de la clasificación**

	Metodologías		
	Árbol de Decisión	Rough Set	GLM
Precisión de la clasificación: porcentaje de clasificaciones correctas	86,31%	83,7%	83,12%

**Fuente:** Propia de los autores

Como puede observarse el porcentaje de aciertos en las clasificaciones es muy superior al 80% en las tres metodologías, lo cual valida los modelos obtenidos en nuestra aplicación práctica y justificaría su aplicación.

Por otro lado, las Tablas 12 y 13 muestran una comparativa del modelo paramétrico con respecto a las variables significativas que se han obtenido de las técnicas no paramétricas de la Inteligencia Artificial para cada una de las categorías o clases analizadas: Categoría 1- CAÍDA (Tabla 12) y Categoría 0 – RETENCIÓN (Tabla 13):

**Tabla 12 Comparativa de Resultados de Metodologías – CLASE 1\_Caída**

Categoría: 1	Caída	
	No paramétrico	Paramétrico
Arboles de decisión	Rough Set	GLM
Antigüedad	Antigüedad	Antigüedad
Tipo Producto	Forma Pago	Tipo Producto
Edad	Tipo Producto	Forma Pago
Edo Civil	Edo Civil	Edo Civil
Forma Pago	Hijos	
Sexo		
Hijos		
ICE		
Nivel Estudios		
Nivel Ingresos		

**Fuente:** Propia de los autores

**Tabla 13 Comparativa de Resultados de Metodologías – CLASE 0\_Retención**

Categoría: 0		
Retención		
No paramétrico		Paramétrico
Arboles de decisión	Rough Set	GLM
Antigüedad	Antigüedad	Antigüedad
Edo Civil	Tipo Producto	Tipo Producto
Tipo Producto	Forma Pago	Forma Pago
Edad		Edo Civil
Red		
Forma Pago		
Sexo		
Nivel Estudios		
Nivel Ingresos		

**Fuente:** Propia de los autores

Como puede observarse en la Tabla 12, para la CLASE 1 correspondiente a las variables o patrones de comportamiento que definen a los clientes susceptibles a la anulación de su contrato de seguros, se obtienen las mismas variables significativas que arroja el modelo GLM como factores con mayor significancia. Ahora bien, la Tabla 13 muestra que para la CLASE 0, correspondiente a los clientes propensos a quedarse en la compañía conservando su póliza contratada, sucede algo similar aunque en diferente orden de aparición, pero manteniéndose las variables más significativas para clasificar a los clientes.

## 6. Conclusiones

Como se ha comentado, existen varias causas y finalidades que han provocado el surgimiento, desarrollo y lanzamiento del marco normativo sobre el que descansará el sector asegurador, mejor conocido como Solvencia II. Solvencia II es un proyecto sumamente ambicioso ya que busca rediseñar la actual

metodología de cuantificación de la solvencia de las entidades aseguradoras y establecer los niveles de requisitos de capital que necesitan para hacer frente a los riesgos adquiridos frente a sus asegurados.

Además, a partir de su estructura de tres pilares, aporta una nueva cultura enfocada a optimizar la gestión de riesgos dentro del sector asegurador. Es más, dentro del Pilar I se busca analizar y determinar el perfil de riesgos que pretende administrar y soportar cada entidad. Esto se obtendrá en la medida en que se logre una mayor calidad en la gestión de riesgos a través de mejores técnicas de estrategia, planeación y administración de tal forma que, las compañías aseguradoras puedan ser capaces de mantener su posicionamiento frente a los diferentes riesgos que soportan mediante su identificación y control ante constante evolución.

Por lo tanto, es evidente que Solvencia II necesita entidades aseguradoras solventes cuyas decisiones estratégicas se tomen en función de esta nueva “cultura” sobre la gestión del riesgo, sin verse por ello amenazadas ante posibles debilidades de su patrimonio y prestigio. Todo esto se traducirá en un inminente fortalecimiento del sector asegurador.

Dentro de esta cultura de gestión de riesgos que propicia Solvencia II, nuestro estudio se ha centrado en el riesgo de caída de cartera. A través de la aplicación práctica realizada se han identificado una serie de variables o patrones de conducta que caracterizan a los tipos de clientes susceptibles a la cancelación de su contrato de seguros. Esto permitiría establecer estrategias comerciales de retención de clientes en aquellas pólizas con poca propensión a la anulación de su póliza o bien lograr una gestión eficiente de la caída de cartera y el riesgo que conlleva.

Recapitulando los resultados de la aplicación práctica a la muestra de datos, las metodologías sugieren que las variables que se deben tener en cuenta como posibles patrones de comportamiento son:

- Antigüedad de la Póliza
- Tipo de Producto a la que pertenece la Póliza
- Forma de Pago de la Póliza
- Estado Civil del Asegurado

Así mismo, mencionar que al tener una variable respuesta dicotómica (cancela o renueva), es posible estudiar el efecto contrario a la anulación del contrato de seguros. Es decir, se podría plantear el objetivo de manera inversa y analizar el patrón de comportamiento del perfil del asegurado fiel a la entidad. De aquí la propuesta de utilizar dichas metodologías para temas de retención o conservación de clientes.

Para lograr buscar estos patrones, en este contexto de exigencia que propone el entorno regulatorio, surge la necesidad de buscar metodologías de análisis novedosas. De esta forma se logrará incluir o detectar las características de los clientes con perfil de “anulador” que ayuden a complementar los métodos que hasta el día de hoy se han venido utilizando dentro del sector asegurador. Con base a ello, se ha procedido a aplicar una técnica estadística (paramétrica) y dos técnicas no paramétricas.

Con respecto a las técnicas no paramétricas que provienen de la Inteligencia Artificial (Teoría *Rough Set* y árbol de decisión C4.5) queremos resaltar las siguientes ventajas:

- En primer lugar su carácter descriptivo, que permite entender e interpretar fácilmente las decisiones tomadas por el modelo, ya que tenemos acceso a las reglas que se utilizan en la tarea predictiva. El modelo de

la metodología rough set es está formado por reglas con la forma de sentencias lógicas pero, es también posible derivar fácilmente reglas de decisión (para cada rama terminal) siguiendo las rutas marcadas en la estructura del árbol también en forma de sentencias lógicas. En consecuencia, tanto con la metodología Rough Set como con el C4.5 se pueden obtener sentencias lógicas fácilmente entendibles por el usuario final, sea o no experto.

- Además, las reglas de decisión proporcionadas por los modelos pueden usarse con fines predictivos y no sólo descriptivos. Esto es así porque se puede evaluar su precisión a partir de unos datos independientes (datos test o muestra de test que se puede utilizar para validar el modelo) a los utilizados en la construcción del modelo (datos o muestra de entrenamiento utilizada para estimar el modelo).

- Otra ventaja es que permiten trabajar con un número elevado de variables de entrada y no necesitan para trabajar un número elevado de datos.
- Su carácter estrictamente no paramétrico presenta una indudable ventaja para su aplicación al ámbito financiero ya que muchas de las variables económico-financieras, utilizadas habitualmente en los análisis, no suelen cumplir las hipótesis requeridas por las técnicas estadísticas.

Por otro lado, también se ha propuesto la aplicación de una técnica paramétrica con el fin de contrastar los resultados obtenidos por las técnicas no paramétricas que ofrece la Inteligencia Artificial. La técnica paramétrica utilizada son los Modelos Lineales Generalizados (GLMS), técnica muy utilizada en la actualidad en el sector asegurador aunque no tanto para el problema que nos ocupa. La ventaja de los GLMS es lograr un mayor ajuste al riesgo y poder al mismo tiempo ser más ecuánimes con los asegurados. Además, este tipo de modelos permite a las aseguradoras ser más dinámicas,

ya que la tarifa está en continuo movimiento, y va cambiando los coeficientes, en un breve periodo de tiempo.

En definitiva, la mayor aportación de este trabajo es la de incentivar al sector a buscar otro tipo de técnicas para la gestión de sus riesgos (Rodríguez-Pardo, 2012). Entre líneas, se debe leer que Solvencia II es una “cultura de cambio” y como parte de esa renovación, surge el planteamiento de este nuevo tipo de técnicas. Si bien, no buscando que éstas suplanten a las ya tradicionales técnicas estadísticas, sí podrían complementar la experiencia de los expertos dedicados a modelos de gestión de riesgos dentro de las entidades. De alguna manera darían un enfoque distinto y complementario a los tradicionales modelos de gestión de riesgos usados por el sector.

No es por demás comentar que, este estudio no está exento de limitaciones. Por un lado, el tema de la muestra se ha limitado a los productos con mayor volumen de producción de la entidad aseguradora y sólo se ha considerado un año de ejercicio contable de la entidad; con lo cual se trata de una muestra limitada en cuanto al número de casos analizados. Por otro lado, el número de variables cualitativas analizadas también se encuentra limitado, ya que las entidades aseguradoras han tenido poco interés en capturar en sus bases de datos demasiada información cualitativa del asegurado. No es hasta hace algunos años, cuando se empezó a hablar de la “Calidad de los Datos” a raíz de las exigencias marcadas por Solvencia II en el tema.

Finalmente, es necesario recalcar que aunque los resultados en términos de clasificación obtenidos por los modelos son bastante satisfactorios, lo que se perseguía con este estudio no era tanto la bonanza, fiabilidad y precisión de los resultados, sino presentar, aplicar y discutir la factibilidad y capacidad de aplicar las

metodologías ofrecidas por la teoría de Inteligencia Artificial en el campo de los Seguros de Vida. Lo mismo sucede con los resultados obtenidos del modelo GLM, ya que de hecho no se buscaba obtener la ecuación multivariada final con la que se podría predecir una tasa de caída o anulación.

En consecuencia, se puede afirmar que se ha cumplido con el objetivo marcado inicialmente: por un lado, logrando aplicar unas nuevas metodologías que han identificado perfiles o patrones de comportamiento de los clientes susceptibles a anular su póliza de seguros; y por el otro, se ha puesto de manifiesto la conveniencia del uso de los GLM en el campo de la estadística actuarial. Se ha pretendido, en definitiva, fomentar su uso como una alternativa novedosa o bien como un complemento a los actuales métodos utilizados en el sector, para que de manera conjunta, ofrezcan una gestión oportuna de los riesgos ante la propuesta de la nueva regulación de Solvencia II.

## Referencias

- Anderson, D., Feldblum, S., Modlin, C., Schirmacher, D., Schirmacher E. y Thandi, N.** "A practitioner's guide to generalized linear models. Casualty Actuarial Society (cas)", *Syllabus Year*, 1-116. 2007.
- Araya, R.** "Induction of decision trees when examples are described with noisy measurements and with fuzzy class membership." *INRIA Seminar, Projet CLOREC*. 1994.
- Ariza, F.** "El camino hasta Solvencia II en la industria europea del seguro y perspectiva de futuro." *Revista Universitaria Europea*, núm. 18. 2013. Disponible en: [http://www.aefin.es/articulos/pdf/C16\\_4\\_818995.pdf](http://www.aefin.es/articulos/pdf/C16_4_818995.pdf)
- Ayuso, M., Guillén, M. y Pérez-Marín, A. M.** "Modelos Internos en Solvencia II: Su aplicación al cálculo del coeficiente de caída de cartera." Septiembre-Diciembre. *Revista de Gerencia de Riesgos y Seguros* nº 112 de la Fundación Mapfre. 2012. Disponible en: <http://www.mapfre.com/fundacion/html/revistas/gerencia/n112/es/estudio2.html>
- Ayuso, M., Guillén, M. y Pérez-Marín, A. M.** "Metodología para el cálculo de caída de cartera en Solvencia II en presencia de contagio entre cancelaciones." *Anales del Instituto de Actuarios Españoles*, nº 22. 2011. Disponible en: [http://actuarios.org/wp-content/uploads/2017/02/anales2011\\_2.pdf](http://actuarios.org/wp-content/uploads/2017/02/anales2011_2.pdf)
- Boudreau-Trudel, B. y Kazimierz, Z.** "Comparison of Analytic Hierarchy Process and Dominance-Based Rough Set Approach as Multi-Criteria Decision Aid Methods for the Selection of Investment Projects." *American Journal of Industrial and Business Management*, 2, 7-12. 2012.
- Breiman, L., Friedman, J., Stone, C. J., y Olshen, R. A.** "Classification and Regression Trees." *Proceedings of the Thirteenth International Conference, Bari, Italy*. 1996.
- Brockett, P. L., Golden, L. L., Guillen, M., Nielsen, J. P., Parner, J. y Perez-Marin, A. M.** "Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection?" *Journal of Risk and Insurance* Nº 75.3. 2008.
- Cerchiara, R.R., Matthew E. y Gambini, E.** "Generalized linear models in life insurance: decrements and risk factor analysis under Solvency II." *18th International AFIR Colloquium*. 2008. Disponible en: [http://www.iaablog.org/AFIR/Colloquia/Rome2/Cerchiara\\_Edwards\\_Gambini.pdf](http://www.iaablog.org/AFIR/Colloquia/Rome2/Cerchiara_Edwards_Gambini.pdf)
- Comisión de las Comunidades Europeas.** *Directiva 2009/138/CE del Parlamento Europeo y del Consejo, de 25 de noviembre de 2009, sobre el seguro de vida, el acceso a la actividad de seguro y de reaseguro y su ejercicio (Directiva de Solvencia II)*. Disponible en: [https://ec.europa.eu/info/law/risk-management-and-supervision-insurance-companies-solvency-ii-directive-2009-138-ec\\_en](https://ec.europa.eu/info/law/risk-management-and-supervision-insurance-companies-solvency-ii-directive-2009-138-ec_en)

- Comisión de las Comunidades Europeas.** *Comunicación de la Comisión*. Noviembre 2007. Revisión del proceso Lamfalussy. COM/2007/727 final – COD 55 de 28.2.2008. Disponible en: <http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=URISERV:l32056>
- Díaz, Z., Fernández, J., y Segovia-Vargas, M. J.** *Sistemas de inducción de reglas y árboles de decisión aplicados a la predicción de insolvencias en empresas aseguradoras*. Universidad Complutense de Madrid. XII Jornadas de ASEPUA. 2004. Disponible en: [http://www.uv.es/asepuma/XII/comunica/diaz\\_fernandez\\_segovia.pdf](http://www.uv.es/asepuma/XII/comunica/diaz_fernandez_segovia.pdf)
- Díaz, Z., Fernández, J., Gil, J.A. y Pozo, E.M.** “Predicción de crisis empresariales en seguros no vida: una aplicación del algoritmo C4.5.” *Anales del Instituto de Actuarios*, 11, 47-88. 2005.
- Díaz, Z.; Sanchis, A.; Segovia, M.J.** “Analysis of Financial Instability by Means of Decision Trees and Lists.” in *Macroeconomics*, Baily, R.O. (ed.): *Emerging Topics*, Editorial Nova Publishers p.303-327. 2009.
- Díaz, Z., Fernández, J., Heras, A., Del Pozo, E. y Vilar, J.L.** “Modelos Aditivos Generalizados aplicados al análisis de la probabilidad de siniestro en el seguro del automóvil.” en *Investigaciones en seguros y gestión de riesgos*: RIESGO 2013 nº 194, Fundación Mapfre. 2010.
- EIOPA.** *Directrices sobre la evaluación interna de los riesgos y de la solvencia*. 14/259:1-9. Año 2015. Disponible en: [https://eiopa.europa.eu/GuidelinesSII/EIOPA\\_Guidelines\\_on\\_ORSA\\_ES.pdf](https://eiopa.europa.eu/GuidelinesSII/EIOPA_Guidelines_on_ORSA_ES.pdf)
- Eling, M., Schmeiser, H. y Schmit, J.** “The Solvency II process: Overview and critical analysis.” *Risk management and insurance review*, vol. 10, no 1, p. 69-85. 2007.
- González de Frutos, Pilar.** “El seguro español y Solvencia II: Tres conceptos básicos.” *Revista de Actuarios* nº 32 del Instituto de Actuarios Españoles. 2013. Disponible en: [https://www.fundacionmapfre.org/documentacion/publico/es/catalogo\\_imagenes/grupo.cmd?path=1075681](https://www.fundacionmapfre.org/documentacion/publico/es/catalogo_imagenes/grupo.cmd?path=1075681)
- Guillén Montserrat, Pérez Marín, A.M. y Nielsen, J.P.** “The need of monitoring customer loyalty and business risk in the European insurance industry.” *Geneva Papers on Risk and Insurance – Issues and Practice* N° 33. 2008.
- Hardy, M. y Panjer, H.** “A credibility approach to mortality risk.” *ASTIN Bulletin: The Journal of the IAA*, vol. 28, no 2, p. 269-283. 1998.
- Huang, K.Y. y Jane C.J.** “A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories.” *Expert Systems with Applications* 36 5387-5392. 2009.
- Jacques, K. y Nigro, P.** “Risk-based capital, portfolio risk, and bank capital: A simultaneous equations approach.” *Journal of Economics and business*, vol. 49, no 6, p. 533-547. 1997.
- Kass, G. V.** “An exploratory technique for investigating large quantities of categorical data.” *Applied statistics*, 119-127. 1980.
- Keller, P., Luder, T. y Stober, M.** “Swiss Solvency Test.” En *Solvency II & Risikomanagement*. Gabler Verlag, p. 569-593. 2005.
- Laitinen, E. K.** “Prediction of failure of a newly founded firm.” *Journal of Business Venturing*, 7(4), 323-340. 1992.
- McCullagh, Peter y John A. Nelder.** *Generalized linear models*. Vol. 37. Chapman & Hall Press. 2<sup>nd</sup> Edition. 1989.
- McKee, T. E.** “Developing a bankruptcy prediction model via rough sets theory.” *International Journal of Intelligent Systems in Accounting, Finance & Management*, 9(3), 159-173. 2000.
- Mena, J.** “Machine-learning the business: Using data mining for competitive intelligence.” *Competitive Intelligence Review* N° 4, Vol 7. 1996.
- Millán, A. y Muñoz, C. I.** (2001). “Indicadores de calidad en el sector asegurador.” En *Cruzando fronteras: tendencias de contabilidad directiva para el siglo XXI: actas VII Congreso Internacional de Costos y II Congreso de la Asociación Española de Contabilidad Directiva*. (p. 350). Servicio de Publicaciones.

- Miranda, M., Segovia-Vargas, M. J y Ramos, M.** "Técnicas de inteligencia artificial aplicadas a la resolución de problemas económico-financieros: Análisis de los factores determinantes del éxito exportador." *Revista Enlaces* 15, diciembre. 2013. Disponible en: [http://www.cesfelipecsegundo.com/revista/Articulos2013/Art%C3%ADculoMartaMirandaet.al.%20\\_2\\_.pdf](http://www.cesfelipecsegundo.com/revista/Articulos2013/Art%C3%ADculoMartaMirandaet.al.%20_2_.pdf)
- Moscarola, J.** *Multicriteria Decision Aid Two Applications in Education Management*. Springer, Berlin Heidelberg, 1978.
- Nelder, J.A. y Wedderburn, R.W.M.** "Generalized Linear Models." *Journal of the Royal Statistical Society Volº 135.3.* 1972.
- Pawlak, Z.** *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers. Dordrecht, Boston, London 1991.
- Pentikäinen, T. y Rantala, J.** (ed.). *Solvency of Insurers and Equalization Reserves: Report and general aspects*. Insurance Publishing Company. 1982.
- Pozuelo de Gracia, E.** "Solvencia II, capital económico en las aseguradoras." *Revista Economía Financiera*, nº 163. 2008.
- Quinlan, J.R.** "Induction of Decision Trees." *Revista Machine Learning* Volº 1-1. 1986. Disponible en: <http://link.springer.com/article/10.1023/A:1022643204877>
- Quinlan, J.R.** *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers, Inc. California 1993.
- Reza, F.M.** *An introduction to information theory*. Courier Corporation. 1961.
- Romera, S.** "Solvencia II, una oportunidad única que conviene aprovechar." *Revista Trébol*, nº 56. 2011.
- Rodríguez-Pardo del Castillo, J. M.** "Modelos predictivos aplicados al seguro de Vida." *Revista de Gerencia de Riesgos y Seguros* nº 114 de la Fundación Mapfre. 2012. Disponible en: <http://www.mapfre.com/fundacion/html/revistas/gerencia/n114/docs/Gerencia-de-Riesgos-y-Seguros-114.pdf>
- Sanchis, A., Segovia-Vargas, M. J., Gil, J. A., Heras, A., y Vilar, J. L.** "Rough sets and the role of the monetary policy in financial stability (macroeconomic problem) and the prediction of insolvency in insurance sector (microeconomic problem)." *European Journal of Operational Research* Nº 181(3). 2007.
- Segovia-Vargas, M.J.** *Predicción de Crisis Empresarial en Seguros No Vida mediante la Metodología Rough Set* (Tesis Doctoral). Universidad Complutense de Madrid. 2003. Disponible en: <http://eprints.ucm.es/tesis/cee/ucm-t26780.pdf>
- Segovia-Vargas, M.J., Gil, J. A., Heras, A. y Vilar, J. L.** *La metodología Rough Set frente al Análisis Discriminante en los problemas de clasificación multatributo*. Universidad Complutense de Madrid. Oviedo 2003. Disponible en: <http://www.uv.es/sala/malaga/XI/19.pdf>
- Segovia-Vargas, M. J., Camacho-Miñano, M.M. y Pascual-Ezama, D.** "Risk factors selection in automobile insurance policies: A way to improve the bottom line of insurance companies." *Revista Brasileira de Gestão de Negócios –RBGN*. vol. 17, nº 57, pp. 1228-1245. 2015.
- Shyng, J.H., Shieh, H.M., Tzeng G.H y Hsieh, S.H.** "Using FSBT technique with Rough Set Theory for personal investment portfolio analysis." *European Journal of Operational Research* 201, 601-607. 2010.
- Shyng, J.Y., Wang, F. K., Tzeng, G. H., y Wu, K. S.** "Rough set theory in analyzing the attributes of combination values for the insurance market." *Expert Systems with Applications* Nº 32.1. 2007.
- Skowron, A. y Grzymała-Busse, A.** *From the Rough Set Theory to the Evidence Theory*. Institute of Computer Science Reports 8. 1991.
- Slowiński, R.** *Rough set learning of preferential attitudes in multi-criteria decision making. Methodologies for Intelligent Systems*. Springer, Berlin Heidelberg, 1993.
- Swiss Reinsurance Company Ltd. (Economic Research & Consulting).** *Informe SIGMA: Word*

- Insurance in 2014: Back to life, nº 4/2015.* 2015 Disponible en: [http://www.swissre.com/library/042015\\_World\\_insurance\\_in\\_2014\\_back\\_to\\_life.html](http://www.swissre.com/library/042015_World_insurance_in_2014_back_to_life.html)
- Solá, F.** *El proyecto Solvencia II: Marco Conceptual, Normativo e Institucional. Working Paper IE Law School AJ8-198.* 2013. Disponible en : <https://catedraperezllorca.ie.edu/sites/default/files/AJ8-198.pdf>
- UNESPA.** *Solvencia II. De un vistazo.* 2015. Disponible en: <http://unespa-web.s3.amazonaws.com/main-files/uploads/2017/07/Solvencia-II.-De-un-vistazo.-FINAL.pdf>
- Witlox, F. y Tindemans, H.** "The application of rough sets analysis in activity-based modeling, opportunities and constraints." *Expert Systems with Application*, 27(2): 171–180. 2004.
- Witten, I. H. y Frank, E.** *Data Mining: Practical machine learning tools and techniques.* Morgan Kaufmann. 2nd Edition. San Francisco. 2005.
- Yao, J. y Herbert, J.** "Financial time-series analyses with rough sets." *Applied Soft Computing*, 9, 1000-1007. 2009.