

Análisis de información contable y financiera mediante técnicas de Inteligencia Artificial. Aplicación a la predicción de insolvencias en empresas españolas de seguros no-vida

M.J. Segovia-Vargas

Departamento de Economía Financiera y Contabilidad I, Facultad de Ciencias Económicas y Empresariales, Campus de Somosaguas, Universidad Complutense de Madrid, 28223 Pozuelo de Alarcón, Madrid. mjsegovia@ccee.ucm.es

Resumen

La detección precoz de insolvencias en el sector asegurador es una cuestión de gran actualidad. Los métodos estadísticos, tradicionalmente empleados en este sector para afrontar este problema, suelen utilizar como variables explicativas ratios financieros que no suelen cumplir las hipótesis estadísticas que requieren. Hemos aplicado varios métodos provenientes de la Inteligencia Artificial [Rough Set (*conjuntos imprecisos*), algoritmos genéticos, Support Vector Machines (*Máquinas de vectores soporte*) y el Templado Simulado (*Simulated Annealing*)] para la predicción de insolvencias sobre una muestra de empresas españolas de seguros no-vida a partir de ratios financieros. Nos hemos centrado en esta investigación en el método Rough Set por su carácter explicativo frente a otros métodos de la Inteligencia Artificial, aunque hemos comparado los resultados obtenidos con las otras metodologías citadas. La teoría Rough Set presenta estas ventajas: es útil para analizar sistemas de información que representan el conocimiento adquirido por la experiencia; elimina variables redundantes reduciendo el coste, en tiempo y dinero, del proceso de decisión, y, se obtienen reglas de decisión fácilmente comprensibles que se extraen de ejemplos reales lo que justificaría las decisiones que en base a ellas se tomen. Los resultados obtenidos son muy satisfactorios comparándolos con el análisis discriminante (técnica estadística tradicional aplicada en seguros) y muestran que las metodologías propuestas pueden ser útiles para evaluar la insolvencia de una entidad aseguradora.

Palabras clave: Insolvencia, inteligencia artificial, análisis discriminante, compañías de seguros

1.- INTRODUCCIÓN

A diferencia de otros problemas financieros, el gran número de agentes e intereses afectados cuando se produce una insolvencia, hacen que el estudio de la misma se mantenga siempre actual. Es más, la viabilidad y, en consecuencia, el riesgo de fracaso empresarial ha sido reconocido como un factor importante en el área de la macroeconomía o en los análisis industriales. Al mismo tiempo, tanto el número de empresas quebradas como la tendencia relativa, son considerados indicadores de la solidez de las industrias y del conjunto de la economía.

Además la actividad aseguradora, bancaria y de inversión constituyen, conjuntamente, una proporción muy significativa de las transacciones financieras. Por tanto, el sector asegurador tiene una gran importancia social y económica, luego, la detección precoz de insolvencias o de las condiciones que pueden llevar a que éstas acaezcan en una aseguradora, es una cuestión clave para las autoridades reguladoras, inversores, gestores, analistas financieros, bancos, auditores y asegurados. Se hace necesario minimizar los costes asociados al problema que estamos considerando como pueden ser los efectos sobre los fondos de garantía existentes en algunos países o las responsabilidades de gestores y auditores.

En definitiva, dada la importancia de este problema, la actividad aseguradora ha sido objeto desde hace tiempo de supervisión a través de la articulación de diversos sistemas de control: obligación de informar sobre su situación económico-financiera mediante la publicidad de las cuentas anuales y obligándolas a ser auditadas, imposición de normas encaminadas a garantizar su solvencia como pueden ser el cumplimiento de los coeficientes de garantía y de solvencia, provisiones obligatorias, inversión de las mismas, capital mínimo, regulación de los precios, remisión a la autoridad supervisora de los estados confidenciales, o, finalmente, realización de inspecciones in situ.

La solvencia en las entidades aseguradoras es una cuestión de interés mundial. De hecho, en la actualidad a nivel europeo, el marco legislativo en materia de solvencia para entidades aseguradoras

ha sido objeto de una profunda reforma. El proyecto *Solvencia II* marca las directrices de las normas existentes en la Unión Europea en relación con la solvencia. En consecuencia, desarrollar y aplicar nuevos métodos en relación con la supervisión de compañías aseguradoras o para analizar la información, tanto cuantitativa como cualitativa, que suministren las entidades aseguradoras relativa a su gestión de riesgos es una cuestión de gran actualidad e importancia.

En seguros muchas de las metodologías aplicadas para acometer este problema son de tipo estadístico (Martin *et al.* 1999, Mora, 1994; Sanchis *et al.*, 2003) y suelen emplear como variables explicativas ratios financieros. Este tipo de variables no suelen satisfacer las hipótesis estadísticas que requieren dichos métodos lo que dificulta su aplicación o cuestiona los resultados obtenidos. A veces en este tipo de técnicas la presencia de observaciones atípicas (*outliers*) afecta en gran medida a los resultados finales y éstos son difíciles de interpretar.

En la actualidad se ha dado un nuevo enfoque al problema aplicando al sector asegurador metodologías del ámbito de la Inteligencia Artificial (IA). Siguiendo a O' Leary (1995), los sistemas inteligentes pueden construirse a partir de dos enfoques.

El primer enfoque lo constituyen los denominados Sistemas Expertos: Consiste en introducir en el ordenador el conocimiento que los expertos humanos han ido acumulando a lo largo de su vida profesional; la mayor limitación de este enfoque radica en el proceso de captación de la información, que ha de hacerse mediante una serie de entrevistas a los expertos.

El segundo enfoque lo constituye el Aprendizaje Automático (*Machine Learning*). Este enfoque consiste en la elaboración de programas de ordenador que sean capaces de generar conocimiento a través del análisis de los datos y posteriormente utilizar dicho conocimiento para realizar *inferencias* sobre nuevos datos. Dentro de las técnicas aplicables de este enfoque encontramos: Redes Neuronales Artificiales, Algoritmos de Inducción de Reglas y Árboles de Decisión. Algunas de ellas tienen un carácter explicativo (inducción de reglas y árboles de decisión), otras se caracterizan por un enfoque de *caja negra- black box*, como las redes neuronales. Pero todas ellas, en general, no requieren que los datos satisfagan ningún tipo de requisito o hipótesis, no se ven afectadas por la presencia de observaciones atípicas.

Nos hemos centrado en la *Teoría Rough Set* por su carácter explicativo y compararemos sus resultados con otras metodologías de la Inteligencia Artificial. La teoría *Rough Set* utiliza la experiencia de una forma objetiva, a través del estudio de la experiencia histórica de una manera cuantitativa, para así explicitar reglas que, resumiendo y objetivando esa experiencia acumulada, ayuden en las decisiones futuras.

Esta metodología, al igual que el resto de las que utilizaremos y que provienen de la Inteligencia Artificial, no requiere que los datos satisfagan ningún tipo de requisito o hipótesis, no se ve afectada por la presencia de observaciones atípicas y el modelo final obtenido consiste en un conjunto de reglas de decisión en forma de sentencias lógicas que las hacen fácilmente comprensibles por cualquier usuario. Además dichas reglas están basadas en decisiones pasadas (experiencia del decisor), y por tanto, son casos reales lo que justificaría su utilización para futuras decisiones.

En definitiva, el presente trabajo describe una investigación de carácter empírico consistente básicamente en la aplicación al sector de empresas de seguros no-vida diversas metodologías de la IA. Nuestro objetivo es comprobar su efectividad en la detección de insolvencias como problema de clasificación entre empresas sanas y fracasadas y utilizando como variables explicativas los ratios financieros.

Hemos comparado los resultados obtenidos con los obtenidos aplicando el análisis discriminante, método estadístico tradicionalmente aplicado en seguros (Sanchis, *et al.*, 2003). Los resultados obtenidos son muy satisfactorios y similares a los conseguidos con el análisis discriminante pero evitando los inconvenientes que presenta dicho método. Este hecho muestra que las herramientas de la IA son de gran utilidad y un complemento y, en algunos casos, una alternativa a los métodos estadísticos ya que mejora en algunos aspectos algunos de sus problemas.

El resto del trabajo se estructura en las siguientes secciones: en la sección 2 presentamos una breve descripción de los principales conceptos del método Rough Set; en la sección 3 describimos las variables y datos que formarán parte de los modelos; en la sección 4 explicamos el desarrollo del análisis y los resultados obtenidos por el método Rough Set; la sección 5 contiene una comparación entre la metodología rough set y el análisis discriminante; en la sección 6 hemos comparado la metodología rough set con otros métodos de la IA; la sección 7 presenta otros problemas financieros que hemos analizados mediante técnicas de IA y, finalmente en la sección 8, explicamos las conclusiones.

2. PRINCIPALES CONCEPTOS DE LA TEORÍA ROUGH SET

2.1. Aproximación al concepto de rough set, tabla de información y relación de “no diferenciación”

En el ámbito de la predicción de insolvencias la teoría Rough Set ya ha sido aplicada a sectores distintos del asegurador (Slowinski y Zopounidis, 1995, Dimitras, *et al* 1999, y Ahn, *et al.*, 2000). Sin embargo, la aplicación de esta teoría al sector asegurador no se limita al problema que nos ocupa sino que también se ha aplicado a otros problemas (Shyng, Jhieh-Yu *et al.* 2007).

Fue originalmente desarrollada en los años ochenta por Z. Pawlak (Pawlak, 1991, 2002, 2007), entre otros, como herramienta formal para tratar con la incertidumbre inherente a un proceso de decisión. Aunque existen en la actualidad extensiones de esta teoría (Greco *et al.*, 1998) nos referiremos al enfoque clásico.

Para situar la metodología rough set cabría recordar que en todo proceso de decisión es necesario un *modelo de preferencias* que contenga “la mejor” solución para un problema de decisión dado. Un camino para construir dicho modelo es a través de la inteligencia artificial y constituye el modelo *vía aprendizaje de los ejemplos o adquisición inductiva del conocimiento* (también denominado *inducción de reglas, aprendizaje inductivo*). El modelo resultante es un conjunto de reglas de la forma Si/Entonces o un árbol de decisión. Este camino descansa sobre la hipótesis de que se puede obtener un modelo sobre las preferencias mediante el estudio de las valoraciones efectuadas por los decisores cuando se les presenta un conjunto de objetos representativos de un problema de decisión. Lo que subyace dentro de este enfoque es que ofrece más confianza la valoración efectivamente realizada por un decisor que la explicación que tuviera que dar sobre la misma dicho agente. El enfoque *rough set* es un representante típico de esta categoría al igual que las *redes neuronales*.

La filosofía del método se basa en la suposición de que con cada objeto del universo que estamos considerando se puede asociar alguna información (datos, conocimiento). Los objetos caracterizados por la misma información no son discernibles en vista de la información disponible. Esta “no diferenciación” de objetos generó de este modo la base matemática para esta teoría.

La información imprecisa es la que provoca la “no diferenciación” de los objetos en términos de datos disponibles y evita, en consecuencia, su asignación precisa a un conjunto. “Rough” podría traducirse por “vago, impreciso”; de aquí en adelante hablaremos de rough set. Por tanto, y de manera intuitiva, un rough set es un conjunto de objetos que, en general, no pueden ser caracterizados de manera precisa en términos de la información disponible. Si esta información

consiste en un conjunto de objetos descrito por otro conjunto, en este caso, de atributos, diremos que un rough set es un conjunto de objetos que, en general, no pueden ser caracterizados de manera precisa en términos de valores de un conjunto de atributos, y por lo tanto no los podemos clasificar en las diferentes categorías de manera precisa.

Por ejemplo, en nuestro caso, nuestros objetos son las empresas de seguros, la información disponible (atributos) son ratios financieros y las clases en que las vamos a clasificar son dos en función de si son insolventes o no. Tendríamos información imprecisa siempre y cuando con los ratios financieros disponibles no pudieramos clasificar una empresa en una de las dos categorías (solvente/insolvente) dado que tenemos empresas caracterizadas con valores similares de los ratios y una es insolvente y la otra no lo es. Es decir, en la realidad nos encontramos muchas veces con información de este tipo, por ejemplo por qué pacientes con los mismos síntomas desarrollan una enfermedad y otros no; o, nuestro caso por qué empresas con los mismos valores para determinadas variables financieras unas están intervenidas por ser incapaces de superar la crisis y otras no llegan a quebrar.

Por otro lado, la teoría del rough set asume la representación del conocimiento de los objetos en forma de una *tabla de información*, que es un caso especial de un sistema de información. En las filas de la tabla se indican los *objetos* (acciones, alternativas, candidatos, pacientes, empresas, etc.), mientras que las columnas se corresponden con los *atributos* (síntomas de una enfermedad, ratios financieros, etc.). Las entradas en la tabla son los *valores del atributo*. Es decir, la entrada en columna q y en fila x tiene el valor $f(x, q)$.

La *relación de no diferenciación* se expresaría de la siguiente forma dados dos objetos, x , y , decimos que x e y *no son discernibles* en base a un conjunto de atributos si, y solo si, $f(x, q) = f(y, q)$ para todos los atributos de la tabla.

En nuestro caso, dadas dos empresas, ambas no se pueden diferenciar si sus variables financieras (ratios) toman los mismos valores o varían muy ligeramente.

2.2. Aproximación de conjuntos, aproximación de la partición y calidad de la clasificación

Dado que nuestro problema será clasificar empresas entre solventes e insolventes, hemos de tener en cuenta que la clasificación de los objetos se basa en la información que hay accesible sobre ellos y no en los objetos en sí mismos.

Hasta ahora al hablar de la información sobre los objetos no hemos mencionado que suele ser imprecisa (inconsistente). Esta inconsistencia en la descripción de los objetos lleva a la ambigüedad en su clasificación, es decir, objetos descritos por los mismos valores de los atributos (atributos de condición) pero que sin embargo están asignados a diferentes clases (atributos de decisión) (Roy, 1989). Hay dos razones principales que justifican la existencia de imprecisiones: que la información de la que se dispone sea incompleta o que esta entre en conflicto. Desde el punto de vista de nuestro trabajo nos interesa el caso en que la evidencia entre en conflicto, esto es, que la información que tenemos sea inconsistente. Prácticamente eso significa que dados dos objetos, descritos por los mismos valores de los atributos (mismos descriptores), cada uno de ellos pertenece a conceptos o clases diferentes. (Es decir, recordemos, en nuestro caso, dadas dos empresas con los mismos valores para todos los ratios o con los valores de los ratios dentro de los mismos intervalos, una fracasa y la otra no).

Por tanto, si los objetos no se pueden distinguir mediante atributos, no podemos asignarlos de manera precisa a un conjunto con lo que se plantea la necesidad de una *aproximación* de las distintas clases en las que se asignarían los objetos.

Un rough set es una colección de objetos que, en general, no pueden ser clasificados de manera precisa en términos de los valores del conjunto de atributos. En consecuencia, cada rough set tiene

casos fronterizos, esto es objetos que no pueden clasificarse con certeza como miembros del conjunto o de su complementario y, por tanto, puede ser reemplazado o representado por un par de conjuntos precisos, llamados la aproximación por encima y por debajo. La *aproximación por debajo* consiste en todos los objetos que con seguridad pertenecen a la clase o categoría, y la *aproximación por encima* contiene los objetos que posiblemente pertenecen a la misma. La frontera (o *región de duda*) es el conjunto de elementos que no pueden clasificarse con certeza utilizando los atributos del sistema de información y es la diferencia entre ambas aproximaciones.

Al cociente entre la aproximación por debajo y la aproximación por encima se le denomina *precisión* de la aproximación. Este ratio indica los casos fronterizos o “dudosos” a la hora de clasificarlos que existen para una clase o categoría determinada. Este ratio toma el valor 1 para aquellas clases en los que no hay casos fronterizos (no hay por tanto información inconsistente).

Por otro lado, si en una tabla de información no solo existe una clase o categoría si no que existen varias (por ejemplo, en nuestro caso existen dos clases de empresas, la clase de las sanas y la de las fracasadas), para cada una de ellas se puede calcular la aproximación por arriba y por abajo. El cociente entre la suma de todas las aproximaciones por abajo y el número total de objetos del sistema, se denomina *calidad de la clasificación*. Expresa el ratio de todos los objetos correctamente clasificados respecto a todos los objetos del sistema (Slowinski y Stefanowski, 1994). Si toma el valor 1 significa que no hay objetos “dudosos” a la hora de ser clasificados.

2.3. Reducción y dependencia de atributos

Una de las funciones más importantes de la metodología rough set es descubrir las dependencias entre atributos al analizar una tabla de información. Descubrir dichas dependencias permite que el conjunto de atributos se pueda reducir (eliminar aquellos atributos que sean redundantes o innecesarios), apareciendo el concepto de *reducto* o *conjunto mínimo*, que se define como el menor conjunto de atributos que mantiene la misma calidad de clasificación (ver sección anterior) que el conjunto de todos los atributos. En una tabla de información puede haber más de un reducto. La intersección de todos los reductos nos da el denominado *núcleo*, que es la colección de los atributos más relevantes en la tabla y que no pueden ser eliminados sin que disminuya la calidad de la clasificación.

2.4. Reglas de decisión

Un sistema de información reducido (ya sin redundancias) permite la obtención de reglas de decisión. Estas reglas serán las que determinen si un objeto pertenece a una determinada clase.

Una regla de decisión puede expresarse como una sentencia lógica que relaciona la descripción mediante atributos de un objeto (condiciones) y las clases de decisión. Toma la siguiente forma:

SI <se cumplen condiciones> **ENTONCES** <el objeto pertenece a una clase de decisión dada>

Cada regla de decisión se caracteriza por su *fuerza*, esto es el número de objetos que satisfacen la parte de la condición de la regla (en el argot se dice *cubiertos* por la regla) y que pertenecen a la clase de decisión sugerida.

El conjunto de las reglas de decisión y la información sobre los atributos más significativos para la clasificación de los objetos puede considerarse como *una representación* del conocimiento adquirido por un *especialista* sobre todos los casos/objetos contenidos en un sistema de información, sin las redundancias, tan típicas en las bases de datos reales. Es más, los resultados obtenidos por el enfoque rough set se expresan de *forma similar al lenguaje natural humano*. Así, las reglas son muy fáciles de comprender por el usuario/analista y permiten justificar y explicar las conclusiones derivadas de análisis de los datos. Además, da la posibilidad para el analista de controlar dicho análisis de manera simple. Tal posibilidad no es normalmente ofrecida por las técnicas tradicionales de análisis de datos.

El conjunto de reglas para todas las clases de decisión se denomina *algoritmo de decisión*. Los procedimientos para generar reglas de decisión a partir de una tabla de decisión operan sobre los principios del aprendizaje inductivo. No vamos a entrar en los algoritmos de inducción de reglas simplemente indicar que utilizan alguna de las siguientes estrategias: 1) generación de un conjunto mínimo de reglas que cubran todos los objetos de una tabla de decisión, 2) generación de un conjunto exhaustivo de reglas consistentes en todas las reglas posibles de una tabla de decisión, 3) generación de un conjunto de reglas de decisión "fuertes", incluso en parte discriminantes, cubriendo relativamente muchos objetos pero no necesariamente todos los objetos de la tabla de decisión.

Utilizaremos las reglas de decisión derivadas de una tabla con esta metodología para clasificar *nuevos objetos* (en nuestro caso empresas nuevas). La clasificación de cualquier nuevo objeto puede hacerse comparando su descripción con las condiciones de cada una de las reglas de decisión.

3.- SELECCIÓN DE LOS DATOS Y DE LAS VARIABLES

En esta fase de nuestra investigación hemos procedido a seleccionar los datos y las variables que serán objeto de estudio, y que posteriormente formarán parte del modelo.

En cuanto a los *datos*, la muestra utilizada en nuestro trabajo es la empleada para la aplicación del análisis discriminante en la predicción de la insolvencia en entidades aseguradoras españolas no vida (Sanchis *et al.*, 2003). Consta de 36 empresas fracasadas y 36 empresas sanas, emparejadas, fundamentalmente, por tamaño (medido a través del volumen de primas) y sector, para minimizar el efecto de estas variables en el estudio.

En cuanto a las *variables*, también hemos decidido utilizar las mismas al objeto de poder efectuar un análisis comparativo entre ambas metodologías. Se considera que el éxito o fracaso de una empresa es función de una serie de variables. Siguiendo un paralelismo con conceptos matemáticos podríamos decir que la solvencia o insolvencia de una empresa sería la variable dependiente y por variables independientes tomaríamos la información contable de las empresas sistematizada a través del cálculo de determinados ratios financieros.

La definición utilizada de *empresa insolvente*, variable dependiente, es aquella empresa que ha tenido que ser intervenida para su liquidación por el organismo oficial existente a dicha fecha en España para la liquidación de entidades aseguradoras (la CLEA- Comisión Liquidadora de Entidades Aseguradoras). Es una medida objetivamente determinable de las empresas que fracasan.

Los datos utilizados para el cálculo de los *ratios financieros*, variables dependientes, han sido los Balances y las Cuentas de Pérdidas y Ganancias de la publicación anual "Balances y Cuentas. Seguros Privados" de la Dirección General de Seguros (organismo que depende del Ministerio de Economía y Hacienda español). La utilización de documentos contables permite obtener una información cuantificable, objetiva y sistemática de la realidad económica y financiera de una empresa. Los datos contables empleados corresponden para las empresas fracasadas a los cinco años anteriores a que las mismas quebraran, y por extensión, ese mismo periodo será el que se aplique a su pareja. Para las empresas sanas se comprobó que siguiesen funcionando varios años después de su período de análisis para no afectar los resultados del trabajo. Hemos de mencionar que dado que en 1982 se introdujo una nueva estructura en las cuentas anuales en este sector, no se ha utilizado información anterior a esa fecha para que la misma fuese lo más homogénea precisa. La serie temporal abarca desde el 83 al 94.

Los ratios que formaran parte del análisis rough set se han obtenido agregando todos los ratios que forman parte de las funciones discriminantes para cada horizonte que fueron estimadas en el trabajo anteriormente referenciado. Por tanto, partimos de un total de 17 ratios.

Tabla 1: lista de ratios

Ratios	DEFINICIÓN
A1	Capitales Propios/ Pasivo Total (Ratio de autonomía financiera)
A5	Fondo de Maniobra/ Activo Total (Ratio de Liquidez)
A6	Activo Circulante/ Activo Total (Ratio de Estabilidad)
B3	Negocio neto/ Activo Total
B6	Provisiones para prestaciones/ Siniestralidad
B7	Primas y recargos netos de anulaciones/ Fondos propios (Ratio de solvencia)
B8	(Capitales propios + Provisiones Técnicas) / Primas y recargos (Ratio de solvencia)
C1	Beneficio antes de impuestos/ Pasivo exigible
C4	Cash-flow / Fondos propios
C5	Cash-flow / Pasivo exigible
C6	Resultado acumulado / (Capital suscrito-resultado acumulado)
C7	Beneficio antes de impuestos/ Fondos propios
D1	Provisiones para prestaciones/ Primas y recargos
D4	Provisiones técnicas/ Total primas
D6	Provisiones Técnicas/ Fondos propios
D7	Provisiones Técnicas del reaseguro cedido/ Capitales propios
D8	Provisiones Técnicas al cierre para riesgos en curso y pp/ Primas y recargos

4.-APLICACIÓN DEL MÉTODO ROUGH SET Y RESULTADOS. COMPARACIÓN CON OTRAS HERRAMIENTAS DE LA INTELIGENCIA ARTIFICIAL.

El software utilizado es el programa ROSE (web www-idss.cs.put.poznan.pl/rose. Predki *et al.*1998 y Predki y Wilk, 1999).

El problema que nos ocupa, predicción de insolvencias, es un problema de clasificación multiatributo consistente en la asignación de cada objeto (empresa), descrito por un conjunto de atributos múltiples (ratios financieros), a una clase o categoría (solvente o insolvente).

4.1. Tabla de información

Para la aplicación de esta metodología lo primero que hemos construido es una tabla de información con las 72 empresas, 36 solventes y 36 insolventes. Las empresas han sido evaluadas de acuerdo con los valores que toman los 17 ratios (nuestros atributos) para cada una de ellas. Estas empresas han sido clasificadas en una de las dos categorías citadas para el año que denominaremos *base* (año 1), que es el año anterior a la quiebra. Por tanto estamos considerando un atributo de decisión, *d*, que efectúa en esta tabla de información (a partir de ahora tabla de decisión) una partición dicotómica: si $d=0$ la firma pertenece al grupo de las insolventes y si $d=1$, la empresa pertenece al grupo de las sanas.

4.2. Discretización

Antes de comenzar el análisis, hemos de adecuar la información que contiene nuestro sistema de información. Para ello, hemos formado un sistema de información en el que hemos transformado los valores continuos de los ratios en términos cualitativos. De esta manera todos los atributos quedan codificados. Esta transformación implica una división del dominio original en subintervalos y una asignación de valores nominales o discretos a los mismos. Los valores frontera de los subintervalos son las denominadas *normas*.

Esta codificación no se impone por el método rough set pero en el caso de utilización de ratios financieros se hace necesario para una interpretación práctica de los resultados y para extraer conclusiones generales de los datos en términos de reglas de decisión. Además, y dado que las normas se utilizan desde el principio del problema y hasta el final, en la explicación de las reglas de decisión, no falsean la imagen original de la situación de decisión. Las empresas para las que los valores de estos ratios financieros están en los mismos subintervalos se considera que tienen el

mismo comportamiento y características. Los códigos que se utilizan para referirse a un subintervalo no representan ningún orden de preferencias y su elección no afecta a los resultados obtenidos.

Esta discretización de los ratios la hemos efectuado utilizando 4 subintervalos que han sido determinados a través de los cuartiles de la distribución de cada ratio (**Tabla 2**) y asignando como códigos los valores 1, 2, 3 y 4 a los intervalos obtenidos (**Tabla 3**). La utilización de percentiles para establecer intervalos en los ratios es frecuente en los trabajos de investigación (Laitinen, 1992; García *et al.*, 1997 o McKee, 2000).

Tabla 2: lista de subintervalos (cuartiles)

Ratio	1°	2°	3°	4°
A1	(-∞, 0.155)	(0.155, 0.385)	(0.385, 0.68)	(0.68, +∞)
A5	(-∞, -0.29)	(-0.29, -0.005)	(-0.005, 0.195)	(0.195, +∞)
A6	(-∞, 0.52)	(0.52, 0.705)	(0.705, 0.875)	(0.875, +∞)
B3	(-∞, 0.325)	(0.325, 0.55)	(0.55, 0.96)	(0.96, +∞)
B6	(-∞, 0.07)	(0.07, 0.495)	(0.495, 1.35)	(1.35, +∞)
B7	(-∞, 0.635)	(0.635, 1.435)	(1.435, 3.185)	(3.185, +∞)
B8	(-∞, 0.775)	(0.775, 1.465)	(1.465, 2.485)	(2.485, +∞)
C1	(-∞, -0.04)	(-0.04, 0)	(0, 0.04)	(0.04, +∞)
C4	(-∞, 0.005)	(0.005, 0.13)	(0.13, 0.41)	(0.41, +∞)
C5	(-∞, 0.005)	(0.005, 0.095)	(0.095, 0.33)	(0.33, +∞)
C6	(-∞, -0.245)	(-0.245, -0.025)	(-0.025, 0.05)	(0.05, +∞)
C7	(-∞, -0.03)	(-0.03, 0.01)	(0.01, 0.06)	(0.06, +∞)
D1	(-∞, 0.04)	(0.04, 0.295)	(0.295, 0.965)	(0.965, +∞)
D4	(-∞, 0.08)	(0.08, 0.785)	(0.785, 1.63)	(1.63, +∞)
D6	(-∞, 0.07)	(0.07, 0.565)	(0.565, 2.82)	(2.82, +∞)
D7	(-∞, 0)	(0, 0.01)	(0.01, 0.46)	(0.46, +∞)
D8	(-∞, 0)	(0, 0.355)	(0.355, 0.435)	(0.435, +∞)

Utilizando estos subintervalos y el principio “cuánto más alto sea el código, mejor subintervalo”, se obtuvo una tabla de información codificada. Dado el principio que se aplicó, para los atributos decrecientes (aquellos en que los valores más bajos son mejores) se les dio un código en orden inverso y además hemos efectuado correcciones, como puede comprobarse, en la escala en el caso en que no estuviese nuestro criterio en concordancia con la secuencia de aumento o disminución de los subintervalos.

Tabla 3: códigos asignados a los intervalos

Ratio	Intervalos			
	1°	2°	3°	4°
A1	1	3	4	2
A5	1	2	3	4
A6	3	4	2	1
B3	1	3	4	2
B6	1	4	3	2
B7	1	3	4	2
B8	1	3	4	2
C1	1	2	3	4
C4	1	2	3	4
C5	1	2	3	4
C6	1	2	3	4
C7	1	2	3	4
D1	1	3	4	2
D4	1	3	4	2
D6	1	3	4	2
D7	1	3	4	2
D8	1	3	4	2

En definitiva, la tabla codificada utilizada para un posterior análisis, consta de las 72 empresas descritas por los 17 ratios financieros codificados, utilizando los datos del año previo a la quiebra y asignadas de forma binaria a una clase de decisión (solvente-1- o insolvente-0).

4.3. Análisis rough set

El análisis rough set de la información codificada de la tabla se desarrolló utilizando el programa ROSE. La *precisión* fue perfecta, esto es, igual a uno y la *calidad de la clasificación* fue también igual a uno. Esto significa que las empresas están muy bien discriminadas entre ellas y no hay casos “dudosos”. La explicación de este hecho es que la tabla de información utilizada está formada por empresas insolventes en el año anterior a ser intervenidas y por lo tanto, con una situación financiera extremadamente deteriorada, y en consecuencia, muy diferente a la situación financiera de las empresas sanas

El siguiente paso del análisis fue la construcción de los *reductos*. Hemos obtenido 452 reductos de la tabla de información codificada, cada uno de los cuales contiene entre 4 y 8 atributos, lo que respecto a los 17 originales supone una reducción importante.

Este resultado nos da una idea de la ayuda que supone la aplicación de este método en la eliminación de variables redundantes ya que, en cualquier caso al menos 9 atributos pueden ser eliminados sin ninguna consecuencia. Los ratios que aparecen con más frecuencia en los reductos son: B8, D4, A5, C6, A1 lo cual indica que son variables muy discriminatorias en la muestra entre empresas solventes y no solventes y revelan las siguientes cuestiones a tener en cuenta a la hora de evaluar la solvencia de una entidad de seguros:

- a) La *liquidez* medida a través del ratio A5: Una de las cuestiones más importantes para asegurar el adecuado funcionamiento de cualquier empresa es la necesidad de tener suficiente liquidez. Pero en el caso de una empresa de seguros, este problema no debería aparecer, y si aparece reviste una mayor importancia que en otro tipo de empresas, ya que la inversión del proceso productivo implica que las primas se cobran antes de que acaezcan los eventuales siniestros. Si una aseguradora no puede pagar los siniestros que ocurran, sus clientes y el público en general perderían la confianza en dicha compañía.
- b) Una política *apropiada* de *financiación* medida a través del ratio C6. Una financiación escasa debida a que no se generan recursos o ingresos suficientes puede ser la causa del comienzo de un proceso de quiebra.
- c) Unos adecuados *capitales propios* medidos a través del ratio A1. Se confirma la importancia, desde el punto de vista de la solvencia, de estos fondos destinados a satisfacer, en caso necesario, futuros siniestros.
- d) La necesidad de cumplir con las exigencias en relación con el cálculo de las *provisiones técnicas* (D4), dado que las aseguradoras deben tener suficientes provisiones para hacer frente a los eventuales siniestros. Una infradotación de las mismas conduciría a un resultado superior al real, y en el caso, de que no fueran suficientes para pagar los siniestros acaecidos, llevaría a recurrir a los fondos propios para hacer frente a las reclamaciones lo que provocaría un debilitamiento financiero de la empresa.
- e) El ratio B8 es un ratio de solvencia específico para el sector asegurador. Dado que relaciona los fondos propios con el volumen de negocio podríamos considerarlo como una medida de la solvencia dinámica de la empresa. Se hace necesario, en consecuencia, una conveniente *tarificación* en el cálculo de las primas. Una mala tarificación lleva a una insuficiencia de las primas en relación con los costes tanto de los siniestros como del resto.

Por otro lado, el *núcleo* de atributos está vacío luego ningún único atributo es absolutamente necesario.

El reducto se ha seleccionado teniendo en cuenta los siguientes criterios:

- a) El reducto debía contener el menor número de atributos que fuese posible.

b) Debía contener los atributos considerados a nuestro juicio más significativos para la evaluación de las empresas. Entre otros, el mayor número posible de ratios de los que más frecuentemente aparecen en los reductos.

El reducto que hemos elegido es el formado por los ratios A5, A6, B6, B8, C6, D8. Elegido el reducto, el resto de atributos han sido eliminados de la tabla de información codificada y hemos pasado de un sistema inicial de información codificado de 17 columnas a otro de solo 6.

El sistema de información reducido puede verse como una tabla de decisión de la cual obtendremos el conjunto de reglas. De las posibles estrategias que incorpora ROSE para la generación de reglas de decisión, hemos escogido la estrategia consistente en la obtención del mínimo conjunto de reglas que cubren todos los objetos de la tabla de decisión.

Hemos obtenido 30 reglas de decisión (**Tabla 4**). Dichas reglas son una representación sin redundancias del conocimiento. Si observamos en el algoritmo obtenido sólo se han utilizado 74 descriptores, lo que supone un 5% de los que aparecían en el sistema de información inicial (1.368). En general, cuanto más corta es una regla más general es.

Es necesario recordar que las reglas son sentencias lógicas. Así por ejemplo la regla número 2 se leería: *Si el ratio A5 toma el valor 2 (-0.29, -0.005) (ver **tabla 2**) y el ratio B6 toma el valor 4 (>1.35), entonces la empresa pertenece al grupo de las fracasadas*. La fuerza de esta regla es de 5 porque son 5 las empresas que verifican este hecho en nuestra muestra.

Tabla 4: Reglas de decisión

Nº Regla	A5	A6	B6	B8	C6	D8	Decisión	Fuerza
1	1		3		1		0	5
2	2		4				0	5
3	1	4	3				0	3
4					1	3	0	2
5				3	2		0	5
6	1			2	2		0	1
7		1		1	4	1	0	2
8		3			4		0	2
9	1			1			0	4
10	2	4					0	5
11				2		3	0	3
12		4			2		0	5
13	4				1		0	1
14	1		1				0	2
15	3			1			0	2
16	4					3	1	6
17	3			4			1	6
18		2		2			1	4
19	4			3			1	4
20	4				3		1	7
21		1			4	2	1	2
22	1		2			3	1	2
23		2	2				1	4
24	2		1			1	1	2
25	3	1		2			1	1
26			2	2	1		1	1
27	4	2			4		1	3
28	2			3	1		1	2
29				2	3	1	1	1
30			3	4	2		1	1

Nuestro algoritmo se compone de 30 reglas porque hemos escogido la estrategia que consiste en que todos los objetos de la tabla fueran cubiertos. Otro tipo de estrategias podrían llevarnos a sistemas más reducidos y por ende más fáciles de manejar e interpretar en la práctica. Así, si no tenemos en cuenta las reglas más débiles, aquellas que están verificadas por un objeto (6, 13, 25, 26, 29, 30), por ejemplo, tendríamos 24 reglas que cubrirían casi un 93% de las empresas.

En definitiva, la muestra de 72 empresas utilizada en el estudio para derivar los algoritmos de decisión puede considerarse como una muestra de prueba utilizada para revelar las características financieras que discriminan las empresas solventes de las insolventes.

Para validar el algoritmo obtenido y ver la precisión predictiva hemos efectuado la clasificación de los datos de las 72 empresas para 2, 3, 4 y 5 años previos a la declaración de la quiebra (Dimitras *et al.*, 1999). Los porcentajes de clasificaciones correctas obtenidas con la aplicación del método rough set fueron:

Clasificaciones correctas	Año 1	Año 2	Año 3	Año 4	Año 5
Clase 0 (empresas quebradas)	100%	83.33%	80.77%	76%	63.16%
Clase 1 (empresas sanas)	100%	77.78%	72.41%	75%	68.15%
Total	100%	80.56%	76.36%	75.50%	65.85%

En general los resultados obtenidos son bastante satisfactorios. Se puede observar como lógicamente el porcentaje de clasificaciones correctas va disminuyendo conforme avanzamos hacia atrás en el tiempo y nos alejamos del año 1 o año previo a la quiebra.

5.- COMPARACIÓN ENTRE LA METODOLOGÍA ROUGH SET Y EL ANÁLISIS DISCRIMINANTE

Dado que el análisis discriminante constituye una de las técnicas estadísticas clásicas más utilizadas en el problema que nos ocupa (predicción de insolvencias) y en el sector asegurador, hemos procedido a la comparación en cuanto de la misma con la metodología rough set.

Aunque la filosofía de las dos técnicas es muy diferente, ambas metodologías se pueden aplicar a la predicción del fracaso empresarial sobre un conjunto de datos comunes.

Brevemente indicaremos que, el análisis discriminante es una técnica estadística que se utiliza para clasificar objetos en diferentes grupos basándose en la observación de algunas características de dichos objetos. Básicamente consiste en estimar una función lineal discriminante que calcula un resultado para cada objeto. Esta función es una combinación lineal ponderada de los valores que toman las características observadas del objeto. Las ponderaciones representan, en esencia, la importancia relativa y el impacto de las características que forman parte de la función discriminante. El objeto se clasifica en un determinado grupo o clase basándose en la puntuación obtenida con la función discriminante.

El análisis discriminante está sujeto a un gran número de hipótesis tales como: la muestra debe ser estadísticamente representativa; el número de atributos discretos que acepta es limitado; los atributos continuos deben seguir una distribución normal; las matrices de covarianzas de cada grupo han de ser idénticas; y, las matrices de covarianzas, las probabilidades a priori y los errores de clasificación han de ser conocidos. Además el número de objetos que debe haber en cada clase debe ser comparable (de ahí la frecuente utilización de muestras emparejadas)

En la práctica, cuando se utilizan datos reales estas hipótesis no se suelen cumplir lo que puede conducir a resultados erróneos o a cuestionar los mismos. Sin embargo, ha sido el método más utilizado, en primer lugar porque fue el primer método que se aplicó y en segundo lugar porque ha dado muy buenos resultados empíricos en problemas financieros en los que intervienen como variables ratios contables pese a que estos no satisfagan las hipótesis estadísticas.

Como ya se ha mencionado, el método rough set no necesita que los datos verifiquen a priori ningún tipo de hipótesis. Además, se puede aplicar cuando existen únicamente atributos cualitativos

o coexisten cualitativos con cuantitativos y cuando la muestra no es, desde el punto de vista estadístico, representativa

La primera comparación que hemos efectuado entre ambas metodologías es en relación al número de ratios que aparecen en los reductos y en las funciones discriminantes estimadas en el trabajo realizado por Sanchis *et al.* (2003, p. 220). Las funciones discriminantes tienen para tres de los cinco horizontes, 8 ratios y, para los dos restantes, 7 y 6, dando lugar a un total de 17 ratios distintos para poder realizar predicciones. Los reductos producidos por la metodología rough set contienen entre 4 y 8 ratios; el escogido contiene 6 ratios, que serían los únicos necesarios para poder realizar predicciones con es método rough set. Este resultado demuestra la potencia de esta metodología en cuanto a selección de características y variables.

Para comparar los resultados de ambas metodologías hemos extraído los resultados de la aplicación del análisis discriminante correspondientes a la clasificación externa del modelo lineal sin ponderación de *outliers* (Sanchis *et al.*, 2003, p. 221). No hemos escogido los datos ponderados de outliers porque la metodología rough set incluye todos los datos tal cual se presentan. Hemos considerado el modelo lineal en lugar del cuadrático porque, aunque no se cumple la hipótesis que requiere el análisis discriminante en cuanto a la normalidad, es el que da mejores resultados. Los resultados de ambas metodologías son:

Clasificaciones correctas	Año 1	Año 2	Año 3	Año 4	Año 5
Rough Set	100%	80.56%	76.36%	75.50%	65.85%
Análisis Discriminante	81.86%	81.27%	76.79%	75.34%	77.78%

Del análisis de ambos resultados se puede concluir como, excepto para el año 5, los resultados son muy parecidos. El año 1 no es comparable porque la estrategia seguida, tal y como ya hemos indicado, en la aplicación del ROSE es la generación de reglas que cubran todos los objetos de la tabla.

Por tanto, a la luz de estos datos, podemos concluir que la metodología rough set es un complemento y, en algunos casos, una alternativa al análisis discriminante. Incorpora, entre otras, las siguientes ventajas: toma los datos tal cual se presentan y no se les requiere que verifiquen ningún tipo de requisito y las reglas de decisión son muy sencillas de entender y de aplicar con lo que el usuario final no requiere de la ayuda de un experto para interpretar los resultados o para aplicarlos.

6.- COMPARACIÓN DE LOS RESULTADOS DEL MÉTODO ROUGH SET CON OTRAS TÉCNICAS DE INTELIGENCIA ARTIFICIAL (MÁQUINAS DE VECTORES SOPORTE, ALGORITMOS GENÉTICOS, TEMPLADO SIMULADO Y ÁRBOLES DE DECISIÓN)

En análisis posteriores elaboramos una nueva tabla de ratios, dado que observamos que en algunos estudios se incorporaban nuevas variables y que algunas de las utilizadas en todos los algoritmos salían como redundantes. (tabla 5).

Tabla 5- Nuevos ratios financieros

R1	Fondo de Maniobra / Activo Total
R2	Beneficio antes de Impuestos(BAI)/ Capitales propios
R3	Ingresos Financieros/ Total Inversiones
R4	BAI*/ Pasivo Total BAI* = BAI+ Amortizaciones + Provisiones + Resultados Extraordinarios
R5	Total Primas adquiridas de seguro directo / Capitales propios
R6	Total Primas adquiridas de negocio neto / Capitales propios
R7	Total Primas adquiridas de seguro directo / Capitales propios + Provisiones Técnicas
R8	Total Primas adquiridas de negocio neto /Capitales propios + Provisiones Técnicas
R9	Capitales Propios / Pasivo Total
R10	Provisiones Técnicas / Capitales Propios
R11	Gastos Técnicos de seguro directo / Capitales propios
R12	Gastos Técnicos de negocio neto / Capitales propios
R13	Gastos Técnicos de seguro directo / Capitales propios + Prov. Técnicas
R14	Gastos Técnicos de negocio neto / Capitales propios + Provisiones Técnicas
R15	Ratio Combinado 1 = Ratio Siniestralidad de seguro directo (RSD)+ Ratio de Gastos (RG) RSD = Gastos Técnicos de seguro directo/ Total Primas adquiridas de seguro directo RG = Comisiones y otros gastos de explotación/ Otros ingresos explotación
R16	Ratio Combinado 2 = Ratio Siniestralidad de negocio neto (RSN)+ Ratio de Gastos (RG) RSN = Gastos Técnicos de negocio neto/ Total Primas adquiridas de negocio neto RG = Comisiones y otros gastos de explotación/ Otros ingresos explotación
R17	(Gastos Técnicos de seguro directo + Comisiones y otros gastos de Explotación)/ Total Primas adquiridas de seguro directo
R18	(Gastos Técnicos de negocio neto + Comisiones y otros gastos de Explotación)/ Total Primas adquiridas de negocio neto
R19	Provisiones Técnicas de reaseguro cedido / Provisiones Técnicas
R20	RSD = Gastos Técnicos de seguro directo/ Total Primas adquiridas de seguro directo
R21	RSN = Gastos Técnicos de negocio neto/ Total Primas adquiridas de negocio neto

Las metodologías que vamos ahora brevemente a comentar se han demostrado ser tanto excelentes clasificadores [las Máquinas de Vectores Soporte (SVM- en inglés, Support Vector Machines] y/o excelentes “selectores de características [Algoritmos Genéticos-AG; Árboles de Decisión-AD y el Templado Simulado- SA (en inglés, Simulated Annealing)], y se han aplicado con éxito en multitud de tareas en el ámbito de la Ingeniería, de la Medicina y de la Economía, entre otras.

Sin embargo, dentro de las herramientas de la Inteligencia Artificial y, en concreto, dentro de las técnicas de Aprendizaje Máquina nos interesan aquéllas que, siendo buenos clasificadores (y, por lo tanto, ser buenos predictores), sirvan para analizar el conocimiento (las relaciones) existente entre los datos. Es decir, nos interesa analizar todo el conocimiento que existe en los datos para que le sea útil al decisor, en nuestro caso las entidades aseguradoras. Por lo tanto dentro de las Técnicas de Aprendizaje Máquina escogeremos aquellas que no tengan un enfoque denominado de “caja negra”. Éste es el motivo por el que la utilización de estas técnicas solo se ha hecho a efectos comparativos tanto en términos de clasificación como de características o variables que resulten más relevantes para el problema que estamos considerando. Pensamos que el método Rough Set es un método más completo para un problema financiero real porque no pertenece a este “enfoque de caja negra” sino que analiza todo el conocimiento existente en los datos de una manera explícita para el decisor o usuario del modelo.

Describiremos brevemente estas técnicas así como los resultados obtenidos con ellas:

La SVM (Salcedo-Sanz, *et al.* 2005, Segovia *et al.* 2004) es una máquina de aprendizaje que se basa en la siguiente idea: cuando no sea posible separar los datos en el espacio de entrada con un hiperplano lineal, trasladar, mediante aplicación no lineal, los vectores de entrada a un espacio de dimensión muy alta. En este nuevo espacio se construirá una superficie de decisión lineal. Esta idea

se puede aplicar tanto en el caso en que los datos sean separables sin errores como para el caso de no separable. La SVM la hemos utilizado para clasificar los datos, pero previamente hemos efectuado una eliminación de las variables (ratios) redundantes (lo que se denomina problema de “selección de características”) mediante la utilización de AG y del SA. De esta forma mejoraríamos la tasa de clasificación de la SVM.

Los AG (Salcedo-Sanz, *et al.* 2004, Segovia *et al.* 2004) son un logro más de la Inteligencia Artificial en su intento de replicar comportamientos biológicos mediante la computación. Se trata de algoritmos de búsqueda basados en la mecánica de la selección natural y de la genética. Utilizan la información histórica para encontrar nuevos puntos de búsqueda de una solución óptima del problema planteado, con esperanzas de mejorar los resultados. Inspirados en el proceso de evolución biológica, un Algoritmo Genético funciona del siguiente modo: en cada generación, se crea un conjunto nuevo de “criaturas artificiales” (cadenas) utilizando bits y las partes más adecuadas del progenitor. Esto involucra un proceso aleatorio que no es, en absoluto, simple. La novedad que introducen los Algoritmos Genéticos es que explotan eficientemente la información histórica para especular sobre nuevos puntos de búsqueda, esperando un funcionamiento mejorado.

A diferencia del anterior el SA (Salcedo-Sanz, *et al.* 2004, Segovia *et al.* 2004) está inspirado en el proceso físico de calentar una sustancia y, posteriormente, enfriarla lentamente, hasta obtener una estructura cristalina fuerte. Este proceso comienza “simulando” una temperatura inicial (es decir, partimos de una configuración inicial, la que sea) más baja que se va enfriando hasta que el sistema alcanza un equilibrio (obtenemos una primera configuración), y se van comenzando nuevas fases en cada una de las cuales se va enfriando la temperatura (vamos obteniendo nuevas configuraciones de tal forma que si la configuración posterior obtenida es mejor, entonces sustituye a la anterior, y así, sucesivamente). El resultado final es la configuración que se obtenga en la última fase.

Por tanto tenemos un algoritmo que se denomina híbrido: los AG y el SA los utilizamos para quedarnos con los ratios más significativos y, la SVM la utilizamos para clasificar.

Los mejores ratios obtenidos con el AG es el conjunto formado por {R1, R9, R13} con una tasa de error $Pe = 0.23$. Los ratios más relevantes seleccionados por el SA son dos conjuntos {R1, R9, R13} y {R3, R9, R19}, ambos con una probabilidad de error $Pe = 0.23$ (Segovia-Vargas *et al.* 2004). Debemos hacer notar que los ratios {R1, R9, R13} son comunes a ambos algoritmos y que los resultados de clasificación (logrados utilizando la SVM) son para ambos algoritmos iguales, esto es, $Pe = 0.23$.

Estos resultados, en cuanto a clasificación, son similares a los obtenidos con la metodología Rough Set (RS). En Segovia-Vargas *et al.* 2004 se comenta otro “experimento” en el que procedimos a seleccionar con los datos del año 1 dos muestras, una de estimación y otra independiente para validar el algoritmo obtenido. De esta forma podíamos comparar con el SA, GA y la SVM. En dicho experimento obtuvimos para la teoría RS un porcentaje de aciertos del 77,78, lo que implica una tasa de error de $Pe = 0.22$, por eso los resultados son similares.

Sin embargo, los reductos obtenidos utilizando la teoría RS contenían de 4 a 6 ratios. Los ratios más significativos en los reductos son R1, R3, R4, R9, R17, R18 y R19. Si observamos, hay varias coincidencias con los seleccionados tanto por el AG como con el SA. Esto indicaría cuales son en nuestra muestra las variables más discriminatorias entre empresas solventes e insolventes.

Hay, además que destacar que para la misma tasa de clasificación tanto el SA como el AG utilizan solo 3 ratios (menos que la metodología RS) lo cual muestra que son herramientas más poderosas en cuanto a selección de las variables más significativas. Sin embargo, estos resultados, a diferencia de los obtenidos por el método RS (reglas de decisión), son difíciles de justificar y no se pueden analizar claramente debido a su carácter de “caja negra”.

Por otro lado, hemos comparado los ratios obtenidos tanto por el método RS, como por el AG y el SA con los ratios obtenidos mediante la generación de árboles de decisión basados en la programación genética. (Salcedo-Sanz *et al.* 2005). De esta forma confirmaríamos las conclusiones previas sobre cual es el mejor conjunto de ratios a utilizar para el problema que nos ocupa.

Hemos de mencionar que un árbol de decisión básicamente consiste en una estructura compuesta de una secuencia de preguntas sencillas, cuyas respuestas trazan un camino que conduce hacia abajo en el árbol. El punto final alcanzado (hojas) determina la clasificación o predicción hecha por el modelo, que puede constituir tanto una respuesta cualitativa como cuantitativa.

Hay muchos procedimientos para construir árboles de decisión, nosotros hemos generados árboles basados en la programación genética (en Díaz *et al.* 2005 se comentan otros análisis que hemos efectuado basados en árboles de decisión para el problema que estamos considerando). Los ratios seleccionados como más relevantes por los árboles de decisión son R1, R4, R5, R7, R10, R11, R18, R20, R21. Por lo tanto el ratio R1 sería la variable más discriminatoria entre empresas solventes e insolventes ya que aparece en todos los métodos aplicados

Toda esta información sería la que tendríamos que facilitar a cualquier decisor interesado en evaluar la solvencia en entidades aseguradoras especializadas en el ramo de “no-vida”.

7.- OTROS PROBLEMAS FINANCIEROS ANALIZADOS CON TÉCNICAS DE IA

Hemos extendido la aplicación de estas metodologías a otros problemas financieros. Concretamente en Sanchis *et al* 2007 y en , hemos profundizado en el papel en el papel de la política monetaria en la ocurrencia de crisis bancarias sistémicas. Hemos utilizado una amplia muestra de 79 países (datos anuales) durante el periodo comprendido entre 1981 y 1999 y hemos aplicado la metodología Rough Set para analizar el papel de la política monetaria a la hora de explicar las crisis sistémicas controlando un conjunto de variables macroeconómicas, tanto de tipo cualitativo como de tipo cuantitativo.

Concretamente, la variable que estudiamos (variable dependiente) es el hecho de que un país sufra inestabilidad financiera, es decir, el acaecimiento en dicho país de crisis financiera, principalmente crisis bancarias (definidas en el sentido de que todo el capital bancario o prácticamente todo se ha agotado). En cuanto a las variables independientes las hemos clasificado en dos categorías. Hemos de resaltar, de nuevo, que se han empleado tanto variables de tipo cualitativo como de tipo cuantitativo. La posibilidad de utilizar ambos tipos de variables es una de las ventajas que ofrece la utilización de la metodología rough set.

Variables Objetivo

- *Estrategias de política monetaria*: estas variables (objetivo de tipo de cambio, objetivo de política monetaria) son variables artificiales. El objetivo de tipo de cambio toma cuatro valores dependiendo del régimen de tipo de cambio: libre flotación, flotación administrada, tipo de cambio fijo y “*currency board*” o caja de conversión (que limita la emisión de dinero a la cantidad de reservas en dólares, asegurando la permanente convertibilidad de la moneda nacional). El objetivo de política monetaria toma el valor uno en los periodos en que los objetivos se basan en agregados monetarios, dos cuando el objetivo es la inflación, tres cuando las dos variables entran en la función objetivo y cero en cualquier otro caso.

• *Independencia respecto al Banco Central*: Esta variable mide en que medida el banco central es legalmente independiente conforme a sus estatutos. Va desde el valor 0 (una independencia mínima) a un valor 1 (la mayor independencia).

VARIABLES DE CONTROL

a) Variables macroeconómicas

- *Inflación*
- *Tipo de Interés Real*
- *Flujos Netos de Capitales respecto al PIB*
- *PIB real per cápita expresado en dólares de EEUU (base 1995)*
- *Crecimiento Real del PIB*
- *Crecimiento Real del PIB mundial*

b) Variables Financieras

- *Crecimiento del Crédito Interior*
- *Liquidez Bancaria respecto al activo total*
- *Pasivos exteriores bancarios respecto al activo exterior*
- *Crisis Previas*: Esta variable es igual a cero si el país ha sufrido un episodio previo de crisis, dos en el caso de dos o tres crisis previas y tres en cualquier otro caso.

8.- CONCLUSIONES

A lo largo de la exposición se han ido enumerando algunas ventajas de la aplicación de técnicas de la Inteligencia Artificial a un problema financiero real como es el que nos ocupa. Podemos concluir que su utilización en el tratamiento de la información contable las convierte en herramientas muy valiosas a la hora de analizar la situación económico-financiera de las empresas. Recapitulando dichas ventajas tendríamos, entre otras:

- analizan los hechos escondidos en los datos, sin necesitar información adicional o preliminar sobre los datos. No necesitan que los datos verifiquen algún tipo de hipótesis,
- permiten utilizar variables cualitativas y cuantitativas,
- obtienen la mínima representación del conocimiento
- eliminan, con gran eficacia, aquellos atributos que son redundantes en el sistema de información

Fruto de estas ventajas son las numerosas aplicaciones de estas metodologías a campos tan diversos como son la medicina, farmacología, industria, ingeniería, finanzas, geología, etc.

En este trabajo se ha contrastado dicha idoneidad mediante su aplicación a un problema real en el sector de seguros no-vida. Obteniendo resultados muy satisfactorios y revelando algunas variables que, entre otras, deberían tenerse en cuenta para la evaluación de la solvencia en dicho sector.

Hemos, además, comparado una de las técnicas de la Inteligencia Artificial (la metodología Rough Set) con el análisis discriminante. Hemos obtenido resultados similares pero creemos que en el casos en que no se cumplan las hipótesis que necesita el discriminante, el método rough set puede ser una alternativa muy útil.

En términos prácticos, estas técnicas pueden usarse como sistemas de diagnóstico automático para preseleccionar por ejemplo aquellas empresas que necesiten una atención especial, de una manera

rápida y a un coste bajo, y de este modo habría una gestión más eficiente del tiempo dedicado por el analista financiero o la autoridad supervisora dando, además, uniformidad a los juicios emitidos sobre una empresa.

Con nuestro trabajo, aplicando técnicas de IA a un problema financiero real, hemos señalado la idoneidad de los métodos de la Inteligencia Artificial como herramientas que ayuden a la hora de justificar o contrastar una decisión.

BIBLIOGRAFÍA

AHN, B.S., CHO, S.S. y KIM, C. Y. The Integrated Methodology Rough Set Theory and Artificial Neural Network for Business Failure Prediction. *Expert Systems with Applications*, 2000, 18, 65- 74.

DÍAZ-MARTÍNEZ, Z., FERNÁNDEZ-MENÉNDEZ, J., SEGOVIA-VARGAS, M.J. y del POZO-GARCÍA, E.M. See5 Algorithm versus Discriminant Analysis. An Application to the Prediction of Insolvency in Spanish Non-life Insurance Companies, *Investment Management and Financial Innovations*, 2004, nº 4, 100-112.

DIMITRAS, A.I., SLOWINSKI, R., SUSMAGA, R. Y ZOPOUNIDIS, C. (1999): “Business failure prediction using Rough Sets”, *European Journal of Operational Research*, 114, 263-280.

GARCÍA, D., CALVO-FLORES, A. Y ARQUES, A. (1997): “Factores discriminantes del riesgo financiero en la industria manufacturera española”, en Calvo-Flores, A. y García, D. (coord.), *Predicción de la Insolvencia Empresarial*, Asociación Española de Contabilidad y Administración de Empresas, pp. 125-156.

GRECO, S., MATARAZZO, B., Y SLOWINSKI, R. (1998): “A new rough set approach to evaluation of bankruptcy risk”, in C. Zopounidis (ed.), *New Operational Tools in the Management of Financial Risks*, Kluwer Academic Publishers, Dordrecht, pp. 121-136.

GUTIÉRREZ, P.A., SEGOVIA-VARGAS, M.J., SALCEDO-SANZ, S., HERVÁS-MARTÍNEZ, C., SANCHÍS, A,

PORTILLA-FIGUERAS, J.A. y FERNÁNDEZ-NAVARRO, F. (2010): *Hybridizing Logistic Regression with Product Unit and RBFnetworks for accurate detection and prediction of banking crises*, *Omega*, The International Journal of Management Science nº38, 333-344.

LAITINEN, E. K. (1992): “Prediction of failure of a newly founded firm”, *Journal of Business Venturing*, julio, 323-340.

MARTÍN, M.L., LEGUEY, S., SÁNCHEZ, J. M. (1999): *Solvencia y estabilidad financiera en la empresa de seguros: Metodología y evaluación empírica mediante análisis multivariante*, Cuadernos de la Fundación Mapfre Estudios, vol. 49.

MCKEE, T. (2000): “Developing a Bankruptcy Prediction Model via Rough Sets Theory”, *International Journal of Intelligent Systems in Accounting, Finance and Management*, nº 9, pp. 159-173.

MORA, A. (1994): “Los modelos de predicción del fracaso empresarial: una aplicación empírica del logia”, *Revista Española de Financiación y Contabilidad*, 78, enero-marzo, pp. 203-233.

O’ LEARY, D.E. Using Neural Networks to Predict Corporate Failfure” , *International Journal of Intelligent Systems in Accounting Finance and Management*, 1998, Vol.7, pp. 187-197.

PAWLAK, Z. *Rough Sets. Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht/ Boston/ London, 1991.

PAWLAK, Z. Rough sets and intelligent data analysis”, *Information Sciences*, 2002, Vol. 147, pp. 1-12.

PAWLAK, Z. y SKOWRON, A. Rudiments of rough sets”, *Decision Sciences*, 2007, Vol. 117, pp. 3-27.

PREDKI, B., SLOWINSKI, R., STEFANOWSKI, J., SUSMAGA, R. Y WILK, S. (1998): “ROSE – Software Implementation of the Rough Set Theory”, in L. Polkowski, A. Skowron, eds. *Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence*, vol. 1424. Springer-Verlag, Berlin, pp. 605-608.

PREDKI, B. Y WILK, S. (1999): “Rough Set Based Data Exploration Using ROSE System”, in: Z.W. Ras, A. Skowron eds. *Foundations of Intelligent Systems, Lecture Notes in Artificial Intelligence*, vol. 1609, Springer-Verlag, Berlin, pp. 172-180.

ROY, B. (1989): "Main sources of inaccurate determination, uncertainty and imprecision in decision models", *Math. Comput. Modell.*, 12, pp. 1245-1254.

SANCHIS, A., GIL, J.A. Y HERAS, A. (2003): "El análisis discriminante en la previsión de la insolvencia en las empresas de seguros no vida", *Revista Española de Financiación y Contabilidad*, 116, enero-marzo, pp.183-233.

SANCHIS, A., SEGOVIA, 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*, 2007, 181, issue 3, pp. 1554-1573.

SALCEDO-SANZ, S., FERNÁNDEZ-VILLACAÑAS, J. L., SEGOVIA-VARGAS M.J. Y BOUSOÑO-CALZÓN, C. (2005).- Genetic programming for the prediction of insolvency in non-life insurance companies", *Computers and Operations Research*, vol.32, nº 4, pp. 749-765, April.

SALCEDO-SANZ, S., PRADO-CUMPLIDO, M., SEGOVIA-VARGAS, M.J., PEREZ-CRUZ, F. y BOUSOÑO-CALZÓN, C. (2004): "Feature selection methods involving Support Vector Machines for prediction of insolvency in non-life insurance companies", *Intelligent Systems in Accounting, Finance and Management*, .12, 261-281.

SEGOVIA-VARGAS M.J., SALCEDO-SANZ, S. Y BOUSOÑO-CALZÓN, (2004).-Prediction of Insolvency in non-life insurance companies using support vector machines and genetic algorithms, *Fuzzy Economic Review*, Volume IX, number 1, may, pp. 79-94.

SLOWINSKI R. y STEFANOWSKI J. (1994): *RoughDas: Rough Set Based Data Analysis System-Version 2.0- User's Guide book*, Poznan, Poland.

SLOWINSKI, R. Y ZOPOUNIDIS, C. Application of the rough set approach to evaluation of bankruptcy risk, *International Journal of Intelligent Systems in Accounting, Finance and Management* 1995, 4, 1, 27-41.