

PREDICCIÓN DE CRISIS EMPRESARIALES EN SEGUROS NO VIDA. UNA APLICACIÓN DEL ALGORITMO C4.5

Zuleyka Díaz Martínez¹, José Fernández Menéndez²,
José Antonio Gil Fana³, Eva María del Pozo García⁴.

RESUMEN

El presente trabajo describe una investigación de carácter empírico consistente en la aplicación al sector asegurador del algoritmo de inducción de árboles y reglas de decisión C4.5, tomando como información de partida un conjunto de ratios de carácter financiero de una muestra de empresas españolas de seguros no-vida, con el objeto de comprobar su utilidad para la predicción de insolvencias en este sector. También se compara esta técnica, procedente del área del Aprendizaje Automático, con el tradicional Análisis Discriminante, observando que el método C4.5 proporciona mejores resultados y de interpretación más sencilla, además de salvar algunos de los inconvenientes de la mencionada técnica estadística.

Palabras clave: Insolvencia, Seguros No-vida, C4.5, Análisis Discriminante.

¹ *Profesora Ayudante.* Departamento de Economía Financiera y Contabilidad I (Economía Financiera y Actuarial). Universidad Complutense de Madrid. Facultad de Ciencias Económicas y Empresariales. Campus de Somosaguas, s/n. 28223 MADRID. Tfno.: 913942577. Fax: 913942570. E-mail: zuleyka@ccee.ucm.es

² *Profesor Asociado.* Departamento de Organización de Empresas. Universidad Complutense de Madrid. Facultad de Ciencias Económicas y Empresariales. Campus de Somosaguas, s/n. 28223 MADRID. Tfno.: 913942971. Fax: 913942371. E-mail: jfernand@ccee.ucm.es

³ *Catedrático de Universidad.* Director del Departamento de Economía Financiera y Contabilidad I (Economía Financiera y Actuarial). Universidad Complutense de Madrid. Facultad de Ciencias Económicas y Empresariales. Campus de Somosaguas, s/n. 28223 MADRID. Tfno.: 913942568. Fax: 913942570. E-mail: jagilfan@ccee.ucm.es

⁴ *Titular de Universidad.* Departamento de Economía Financiera y Contabilidad I (Economía Financiera y Actuarial). Universidad Complutense de Madrid. Facultad de Ciencias Económicas y Empresariales. Campus de Somosaguas, s/n. 28223 MADRID. Tfno.: 913942579. Fax: 913942570. E-mail: epozo@ccee.ucm.es

1. Introducción

La predicción de la insolvencia es uno de los temas centrales del análisis financiero que ha suscitado el interés no sólo del ámbito académico sino también de un amplio abanico de usuarios relacionados con el mundo empresarial. En el sector del seguro, dada su gran relevancia para el conjunto de la actividad económica, la detección precoz de insolvencias o de las condiciones que pueden llevar a que ésta acaezca es una cuestión de suma importancia e interés general.

Las investigaciones de tipo empírico acerca del fenómeno del fracaso empresarial se han llevado a cabo tradicionalmente a través de la aplicación de técnicas estadísticas de carácter paramétrico, como el Análisis Discriminante o la Regresión Logística, que utilizan ratios financieros como variables explicativas. Dadas las peculiaridades del sector asegurador, la mayoría de dichas investigaciones se centran en otros sectores de la economía. No obstante, en el sector de seguros español, cabe destacar los trabajos realizados por López *et al.* (1994), Mora (1994), Martín *et al.* (1999) y Sanchís *et al.* (2003), donde se pone de manifiesto la utilidad de estos métodos para valorar la situación financiera de este tipo de empresas.

Sin embargo, aunque los resultados obtenidos han sido satisfactorios, todas estas técnicas presentan el inconveniente de que parten de hipótesis más o menos restrictivas acerca de las propiedades distribucionales de las variables explicativas que, especialmente en el caso de la información contable, no se suelen cumplir. Además, dada su complejidad, puede resultar difícil extraer conclusiones de sus resultados para un usuario poco familiarizado con la técnica.

Por contra, las técnicas procedentes del campo de la Inteligencia Artificial, de carácter no paramétrico, no precisan de hipótesis preestablecidas sobre las variables de partida. Dentro de este tipo de técnicas, para el problema que nos ocupa son de gran utilidad las que se encuadran en el área del *Machine Learning* (Aprendizaje Automático). Un representante típico de esta categoría son las redes

neuronales, de las que se han desarrollado un buen número de aplicaciones en los más variados campos.

Otro importante conjunto de procedimientos de Aprendizaje Automático es el formado por los distintos sistemas de inducción de reglas y árboles de decisión. Dentro de los estudios realizados siguiendo este enfoque, destaca el trabajo, con una orientación marcadamente estadística, de Friedman (1977), que sirvió como base para la construcción del sistema CART (*Classification and Regression Trees*), descrito en Breiman *et al.* (1984). Más inspirados directamente en el campo de la Inteligencia Artificial son los algoritmos de inducción de árboles de decisión ID3 y C4.5 desarrollados por Quinlan (1979, 1983, 1986, 1988 y 1993), que han alcanzado una notable repercusión. Algunos estudios que implementan este tipo de metodología para la predicción de crisis empresariales son los de, entre otros, McKee (1995) y González *et al.* (1999), realizados a partir de una muestra de sociedades cotizadas en bolsa y una muestra de PYMES, respectivamente.

En la actualidad, como consecuencia fundamentalmente del proyecto comunitario denominado *Solvencia II*, la normativa reguladora del tratamiento de la solvencia de las entidades aseguradoras está sujeta a un proceso de revisión encaminado a conseguir un mayor ajuste a las circunstancias específicas de cada entidad de los requisitos en materia de solvencia a las que aquéllas se ven sometidas por parte de las autoridades reguladoras.

Aspecto esencial para la consecución de este objetivo es el logro de un mejor aprovechamiento de la información financiero-contable suministrada por las entidades sometidas a supervisión que permita extraer de dicha información toda su potencialidad latente en cuanto a caracterizar la situación específica de cada compañía, su grado de cobertura de los riesgos asumidos y su posibilidad de incurrir en una situación de insolvencia que le impida hacer frente a los compromisos adquiridos.

La utilización de nuevos y más ricos y sofisticados modelos analíticos para el tratamiento de la información suministrada por las entidades aseguradoras constituirá entonces un requisito inexcusable para

alcanzar el objetivo mencionado. Sin embargo, las cautelas previas exigibles a cualquier cambio de regulación en un sector de la importancia del del seguro hacen necesario que los modelos y los métodos de carácter estadístico-matemático que lo inspiren se vean sometidos a un estudio y deliberación prolongados.

En este sentido, se han propuesto recientemente distintos enfoques para la predicción del fracaso empresarial en el campo del seguro en España basados en técnicas procedentes de las áreas del Aprendizaje Automático y la Inteligencia Artificial, como Redes Neuronales (Martínez de Lejarza, 1999), Rough Set (Segovia, 2003), Algoritmos Genéticos y Support Vector Machines (Segovia *et al.*, 2004). Pero aunque todas estas técnicas salvan algunos inconvenientes de las técnicas estadísticas tradicionales, o bien requieren de un cierto nivel de conocimiento e implicación del decisor a la hora de establecer ciertos parámetros necesarios para su aplicación, o bien son modelos de “caja negra” que no permiten valorar la importancia relativa de las variables explicativas y, aunque proporcionen buenos resultados en términos de error de clasificación, no permiten establecer un modelo de predicción de insolvencias de interpretación sencilla.

Por estas razones planteamos este trabajo. Nuestra hipótesis de partida es la de que la utilización de técnicas de inducción de reglas y árboles de decisión, de carácter no paramétrico, que, aunque tradicionalmente han estado alejadas de la corriente principal de la investigación estadística, han demostrado su utilidad en los más diversos campos, pueden resultar de gran provecho para la valoración de la solvencia de las empresas de seguros.

Es sobre una de estas técnicas, el algoritmo C4.5, y su comparación con el Análisis Discriminante clásico, sobre lo que se basa nuestro trabajo, con el propósito de comprobar su grado de aplicabilidad a la valoración de las empresas aseguradoras, siendo el fin último de nuestra investigación el desarrollar un modelo sencillo basado en ratios financieros de previsión de insolvencias en empresas de seguros no-vida, de tal forma que, en función del valor que alcancen en estas entidades el pequeño conjunto de ratios derivados en nuestro modelo, se logre anticipar las insolvencias en este sector, orientando así los recursos limitados de la inspección hacia aquellas empresas

preseleccionadas como posibles candidatas al fracaso, lo que permitiría pensar en una normativa, en cuanto a requisitos de solvencia, menos estricta. De este modo, se podría profundizar en la combinación óptima de regulación e inspección.

El resto del trabajo se estructura de la siguiente forma: la sección 2 hace una breve referencia a las técnicas de clasificación. En la sección 3 se exponen brevemente los fundamentos teóricos del algoritmo C4.5. La sección 4 se refiere al estudio empírico realizado y engloba aspectos relativos a la selección de datos y variables que intervienen en el estudio, los resultados obtenidos con el algoritmo C4.5 y su comparación con el Análisis Discriminante. Finalmente, en la sección 5 exponemos nuestras conclusiones

2. Las técnicas de clasificación

La determinación de la solvencia futura de una empresa puede ser entendida como un problema de clasificación: dada una información inicial o conjunto de atributos asociados a una empresa, se pretende tomar la decisión de asignar esa empresa a una clase concreta de entre varias posibles.

Los sistemas clasificadores proceden fundamentalmente de la Estadística o de la Inteligencia Artificial y, entre otras, una manera de diferenciarlos radica en la forma en la que se construyen. Así, podemos entender la construcción de un sistema clasificador con n variables explicativas como la división del espacio n -dimensional que forman esas variables en una serie de regiones, cada una de las cuales se asigna a una de las clases previamente definidas. Esta división puede ser realizada de dos formas (De Andrés, 1998):

- Definiendo una o varias hipersuperficies separadoras. Algunas técnicas estadísticas muy empleadas en el análisis de la solvencia empresarial siguen este enfoque (Análisis Discriminante, Regresión Logística, etc.). Además, se han desarrollado en los últimos años modelos basados en redes neuronales artificiales, que permiten la

definición de hiper superficies separadoras muy complejas y tienen un carácter estrictamente no paramétrico.

- Realizando particiones sucesivas en el espacio de las variables explicativas, empleando una sola variable en cada partición. Dentro de este segundo tipo de sistemas clasificadores los Sistemas Expertos son la rama de la Inteligencia Artificial primeramente empleada en la gestión empresarial, si bien sus posibilidades se han visto enriquecidas con los nuevos enfoques que han aportado otras técnicas de aparición más reciente, como los algoritmos de inducción de reglas y árboles de decisión, entre los que se encuentra el algoritmo C4.5.

3. El algoritmo C4.5

Quizás de entre todos los métodos de aprendizaje automático sean los sistemas de aprendizaje basados en árboles de decisión los más fáciles de utilizar y de entender, y probablemente por ello constituyen uno de los modelos de clasificación más populares y utilizados.

Para el problema concreto que nos ocupa aquí, el de clasificación de empresas como sanas o fracasadas, la estructura de condición y ramificación de un árbol de decisión es idónea: éste puede utilizarse para clasificar un ejemplo concreto comenzando en su raíz y siguiendo el camino determinado por las respuestas a las preguntas de los nodos internos acerca de un atributo particular hasta que se llega a una hoja del árbol, que se refiere a una decisión.

Las sucesivas ramas de un árbol de decisión van realizando una serie de divisiones excluyentes y exhaustivas en el conjunto de elementos que se quieren clasificar. La diferencia básica entre los distintos algoritmos empleados radica en el criterio utilizado para realizar las particiones mencionadas.

La construcción automática de árboles de decisión se origina con los trabajos en el campo de las ciencias sociales de Morgan y Sonquist (1963) y Morgan y Messenger (1973). En estadística, resulta fundamental la aportación de Breiman *et al.* (1984), que proponen el algoritmo para construir árboles denominado CART (Classification

and Regression Trees). Más o menos al mismo tiempo los sistemas de inducción de árboles de decisión comienzan a usarse en el campo del aprendizaje automático, donde alcanzan una notable repercusión los algoritmos desarrollados por Quinlan (1979, 1983, 1986, 1988 y 1993), y en ingeniería (Henrichon y Fu, 1969; Sethi y Sarvarayudu, 1982).

El criterio utilizado para hacer las particiones en el algoritmo C4.5 (Quinlan, 1993) se apoya en una serie de conceptos procedentes de la Teoría de la Información y ha experimentado a lo largo de las dos últimas décadas una serie de notables mejoras. La idea central que comparte con otros algoritmos similares es la 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* (el conjunto de datos usados para construir el árbol).

La información que proporciona un mensaje o la realización de una variable aleatoria x es inversamente proporcional a su probabilidad (Reza, 1994). Resulta habitual medir esta cantidad en bits, que se obtienen como $\log_2 \frac{1}{p_x}$. El promedio de esta magnitud para todas las posibles ocurrencias de la variable aleatoria x recibe el nombre de *entropía* de x :

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

La entropía es una medida de la aleatoriedad o incertidumbre de x o de la cantidad de información que, en promedio, nos proporciona el conocimiento de x .

Análogamente, se puede definir la *entropía conjunta* de dos variables aleatorias x e y :

$H(x, y) = \sum_{x,y} p(x, y) \log_2 \frac{1}{p(x, y)}$, que es la cantidad de

información que, en promedio, nos proporciona el conocimiento de x e y . La *entropía condicional* de x dada y , $H(x/y)$, se define como

$H(x/y) = \sum_{x,y} p(x, y) \log_2 \frac{1}{p(x/y)}$, y es una medida de la

incertidumbre respecto a x cuando se conoce y . Es decir, es la cantidad de información que necesitamos para conocer plenamente x cuando ya tenemos la información suministrada por y . Se cumplirá siempre, lógicamente, que $H(x/y) \leq H(x)$, pues al conocer y tenemos más información que nos puede ayudar a reducir la incertidumbre sobre x . A esa reducción de incertidumbre se la denomina *información mutua* entre x e y : $I(x ; y) = H(x) - H(x/y)$, que es la información que una de las variables nos transmite sobre la otra. Se cumple además que $I(x ; y) = I(y ; x)$. La información mutua es una magnitud similar a la covarianza pero tiene algunas propiedades que la hacen superior a ella.

Podemos considerar que x es una variable aleatoria que nos muestra la clase a la que pertenece un elemento, mientras que y_i , $i = 1, 2, \dots, n$, son los atributos o variables que caracterizan a los elementos que queremos clasificar. Así, en nuestro caso, x indicará si la empresa pertenece al grupo de las sanas o al grupo de las fracasadas, mientras que las variables y_i serán los ratios financieros utilizados para la clasificación.

Inicialmente Quinlan seleccionaba para hacer cada partición aquella variable y_i que proporcionaba la máxima información sobre x , es decir, maximizaba $I(x ; y_i)$ (magnitud que denominaba *gain*). Aunque esto proporciona buenos resultados, introduce un sesgo en favor de las y_i con muchos valores distintos. Para corregirlo, las sucesivas versiones del algoritmo seleccionan aquella y_i que

maximiza la magnitud $\frac{I(x ; y_i)}{H(y_i)}$, a la que Quinlan denomina *gain ratio*. Ésta será el porcentaje de la información proporcionada por y_i que es útil para conocer x . Adicionalmente, para evitar seleccionar un atributo simplemente porque su entropía $H(y_i)$ sea pequeña, lo que aumentaría el valor del cociente anterior, se exige además que $I(x ; y_i)$ sea razonablemente grande.

El proceso de construcción del árbol de decisión mediante la aplicación reiterada del procedimiento descrito finaliza cuando se alcanza la pureza del nodo, entendiendo por “nodo puro” aquél al que sólo corresponden casos pertenecientes a una de las clases del problema, o cuando la ramificación del árbol ya no suponga ninguna mejora.

Generalmente, este método recursivo de construcción de árboles de decisión conducirá a la generación de árboles muy complejos y excesivamente ajustados a los datos del conjunto utilizado para dicha construcción. En consecuencia, harán una clasificación cuasi-perfecta. Esto, que en principio puede parecer óptimo, en realidad no lo es, ya que ajustarse demasiado a los datos de entrenamiento suele tener como consecuencia que el modelo sea muy específico y se comporte mal para nuevos elementos, especialmente si tenemos en cuenta que el conjunto de entrenamiento puede contener ruido, lo que hará que el modelo intente ajustarse a los errores, perjudicando su comportamiento global. Éste es un problema que en general presentan todas las técnicas de aprendizaje de un modelo a partir de un conjunto de datos de entrenamiento (Aprendizaje Automático), al que se conoce como “sobreajuste” (*overfitting*).

El modo más frecuente de limitar este problema en el contexto de los árboles de decisión y conjuntos de reglas consiste en eliminar condiciones de las ramas del árbol o de las reglas, consiguiendo con estas modificaciones la obtención de modelos más generales. En el caso de los árboles de decisión, este procedimiento puede verse como un proceso de “poda” del árbol. Esto aumentará el error de clasificación sobre el conjunto de casos de entrenamiento, pero cabe

esperar que lo disminuya sobre nuevos casos no usados en la construcción del árbol.

El algoritmo C4.5 permite combinar dos tipos de poda: prepoda y postpoda. El proceso de prepoda se realiza durante la construcción del árbol impidiendo la ramificación de nodos que contengan un número de ejemplos inferior a una cierta constante. Además, se implementa también un método de postpoda del árbol ajustado inicialmente que consiste en sustituir una rama del árbol por una hoja, en función de una tasa de error prevista o estimada. Consideremos que existe una hoja que cubre N casos clasificando incorrectamente E de ellos, lo que se puede interpretar suponiendo que existe una variable aleatoria que sigue una distribución binomial en la que el experimento se repite N veces obteniendo E errores. A partir de esto se estima la probabilidad de error p_e , que será la tasa de error prevista o estimada. Para ello se realiza una estimación de un intervalo de confianza para la probabilidad de error de la variable binomial y se toma como p_e el límite superior de ese intervalo (será una estimación pesimista). El nivel de confianza para dicho intervalo viene dado por $1-CF$, siendo CF un parámetro suministrado por el usuario que le permite controlar la intensidad de la poda (de forma que valores más pequeños de CF harán que ésta sea más acusada).

Entonces, para una hoja que cubra N casos, el número de errores previstos será $N \cdot p_e$. Si en lugar de una hoja tenemos una rama el número de errores previstos será la suma de los de cada una de sus hojas. De este modo, una rama será sustituida por una hoja (es decir, la rama será podada) cuando el número de errores previstos de ésta sea menor que el de aquélla.

Por otro lado, conforme el tamaño de los árboles de decisión aumenta, su inteligibilidad disminuye. Una característica importante del algoritmo C4.5 es que permite derivar, a partir del árbol, un conjunto de reglas de la forma *si* (condiciones) - *entonces* (decisión), un mecanismo éste de representación del conocimiento más inteligible que los árboles de decisión, puesto que cuando el problema es complejo, el árbol generado es tan grande que ni siquiera tras su poda resulta sencillo comprender el modelo de clasificación completo.

La representación de un árbol de decisión como un conjunto de reglas es trivial: de cada camino desde la raíz del árbol hasta una hoja se deriva una regla cuyo antecedente es una conjunción de condiciones relativas a los valores de los atributos situados en los nodos internos del árbol y cuyo consecuente es la decisión a la que hace referencia la hoja del árbol, esto es, la clasificación realizada. De este modo, se obtendrá una regla por cada hoja del árbol, pero estas reglas no son más que otra manera de describir el árbol y, por tanto, seguirán siendo mutuamente excluyentes.

Lo interesante sería poder simplificar estas reglas generadas, ya que pueden contener condiciones irrelevantes en su antecedente que podrían eliminarse sin que disminuya la precisión del clasificador, si bien entonces estas reglas más generales pueden dejar de ser mutuamente excluyentes.

Supongamos que R es una regla que cubre $Y_1 + E_1$ casos, de los cuales E_1 son errores, esto es, ejemplos mal clasificados. Si eliminamos una de las condiciones de la regla R , obtendremos una regla más general, llamémosla R^* , que además de los casos anteriores cubrirá otros $Y_2 + E_2$ casos adicionales de los cuales clasificará incorrectamente E_2 , de modo que R^* cubre en total $Y_1 + Y_2 + E_1 + E_2$ casos y clasifica incorrectamente $E_1 + E_2$.

El procedimiento implementado en el algoritmo C4.5 para decidir si simplificar o no la regla más estricta R en favor de la regla R^* más sencilla y general es el mismo que el empleado en la postpoda del árbol:

- Para la regla más estricta R se consideran $Y_1 + E_1$ y E_1 como parámetros procedentes de una distribución binomial en la que el experimento se repite $Y_1 + E_1$ veces obteniendo E_1 errores. Se realiza una estimación de un intervalo de confianza para la probabilidad de error de dicha distribución binomial y se toma el límite superior de ese intervalo como estimación de la probabilidad de error.
- Para la regla más sencilla y general R^* se consideran $Y_1 + Y_2 + E_1 + E_2$ y $E_1 + E_2$ también procedentes de una distribución binomial para la

cual se estima un intervalo de confianza para la probabilidad de error y se toma como estimación de esa probabilidad de error el límite superior del intervalo.

Si la probabilidad de error estimada es menor o igual para R^- que para R , se acepta R^- .

El proceso anterior se realiza para todas las condiciones que integran R y así se va eliminando sucesivamente la condición que proporcione con su eliminación la menor probabilidad de error estimada de entre todas las condiciones, siempre que tal probabilidad de error sea menor o igual que la que se obtiene sin eliminar la condición.

Este procedimiento se lleva a cabo para cada una de las reglas generadas a partir del árbol ajustado inicialmente, esto es, para cada una de sus hojas.

De este modo, se obtendrá un conjunto de reglas más simples, pero, generalmente, el número de ellas no variará. Por otro lado, no será habitual llegar a conjuntos de reglas exhaustivas y mutuamente excluyentes, sino que habrá casos cubiertos por varias reglas o bien por ninguna.

Para reducir el número de reglas se procede por grupos, conteniendo cada grupo todas las reglas que cubran una misma clase. Dentro de cada uno de esos grupos se selecciona un subgrupo de reglas basándose en el principio de la “longitud de descripción mínima” o principio MDL (*Minimum Description Length*) (Rissanen, 1983). Este principio es una versión formalizada del principio de economía de Occam (o “navaja de Occam”) de acuerdo con el cual cuando, en igualdad de condiciones, se pretende seleccionar entre diferentes teorías o diferentes conjuntos de reglas para explicar una determinada evidencia, se debe preferir la teoría más simple.

Formalmente, el principio MDL recomienda seleccionar la teoría h que minimice la expresión:

$$K(h) + K(D/h),$$

donde $K(h)$ es la complejidad en bits de describir la teoría h y $K(D/h)$ es la complejidad en bits de describir la evidencia D a partir de la teoría h (lo que incluye la descripción de los ejemplos que no son cubiertos, es decir, las excepciones).

No se considera adecuada una teoría simple con muchas excepciones, pero tampoco una teoría compleja sin excepciones. La idea del principio MDL es buscar un compromiso para cubrir bastantes ejemplos y capturar así las regularidades presentes en los datos (no subajustar) sin llegar a encontrar reglas demasiado específicas (no sobreajustar).

Una vez seleccionado un subgrupo de reglas dentro de cada grupo, para resolver los conflictos que se presenten cuando existan casos cubiertos por varias reglas correspondientes a clases distintas, se ordenan los subgrupos (dentro de cada subgrupo, el orden de las reglas no importa), comenzando por aquél que proporcione el menor número de falsos positivos, es decir, el menor número de asignaciones a la clase representada por el subgrupo de casos no pertenecientes a dicha clase.

Este primer subgrupo clasificará una serie de casos. El siguiente subgrupo será aquél que proporcione el menor número de falsos positivos con los casos restantes, y así sucesivamente.

Para contemplar la situación de que existan casos no cubiertos por ninguna de las reglas así construidas, se define además una regla o clase por defecto, que será la clase mayoritaria entre todos los casos del conjunto de entrenamiento no cubiertos por ninguna regla de las que constituyen los conjuntos de reglas representativos de cada clase. En caso de empate, se resuelve a favor de la clase con mayor frecuencia absoluta, es decir, la clase mayoritaria entre todos los casos que forman el conjunto de entrenamiento.

Una vez realizado el proceso de generalización, creados los conjuntos de reglas representativos de cada clase y definida la clase por defecto, se procede a la eliminación de reglas completas, analizando para cada uno de los conjuntos de reglas generalizadas que representan cada clase si la eliminación de alguna de las reglas que lo constituyen

disminuye el error producido en la clasificación y, de ser así, se elimina dicha regla, repitiéndose este proceso hasta que la eliminación de reglas ya no suponga ninguna mejora en la clasificación.

Este tipo de sistemas clasificadores presentan importantes ventajas para su aplicación al ámbito financiero, tales como su carácter estrictamente no paramétrico, lo que es importante teniendo en cuenta que la información que vamos a manejar en nuestro análisis, ratios contables, no siempre cumple las hipótesis requeridas por las técnicas estadísticas; al contrario que los Sistemas Expertos, no necesitan de la intervención del experto humano para la inferencia de los modelos clasificadores, ya que estos se elaboran automáticamente; además, frente a las redes neuronales y métodos similares, de indiscutible poder predictivo, sus resultados son mucho más fácilmente interpretables que los modelos de “caja negra” suministrados por aquéllos.

4. Estudio empírico

4.1. Aspectos metodológicos

En cuanto a la muestra de empresas seleccionada para el análisis, ésta es la utilizada para la aplicación de la metodología Rough Set a la predicción de crisis empresariales en seguros no-vida en el trabajo de Segovia (2003). Dicha muestra abarca datos del periodo comprendido entre 1983 y 1996. Consta de 36 empresas no fracasadas y 36 empresas fracasadas (en adelante, “buenas” y “malas”), emparejadas por tamaño - medido a través del volumen de primas - y año de procedencia de los datos, eliminando así el efecto de estas variables en el estudio, y escogiendo como criterio de selección de las empresas fracasadas el hecho de haber sido intervenidas por la Comisión Liquidadora de Entidades Aseguradoras (CLEA)⁵, por entender que se trata de una medida objetivamente determinable de las empresas que

⁵ La CLEA tenía por objeto asumir la liquidación de aquellas entidades aseguradoras que por encontrarse en una situación patrimonial irregular, de previsible insolvencia, les encomendase el Ministerio de Economía y Hacienda o el órgano competente de la respectiva Comunidad Autónoma. En la actualidad, la actividad de la CLEA ha pasado a ser desempeñada por el Consorcio de Compensación de Seguros.

fracasan. Los datos fueron extraídos de la publicación anual *Balances y cuentas. Seguros privados* de la Dirección General de Seguros.

Una vez tomada la muestra, nos situamos en periodos anteriores al de la insolvencia para tratar de determinar qué indicios de este suceso nos proporcionan los datos de las cuentas anuales en forma de ratios. El éxito o fracaso de una empresa será entendido como una variable dependiente que deberá ser explicada por un conjunto de ratios financieros que actuarán como variables independientes. Así que de cada una de las empresas se han obtenido las cuentas anuales de los dos años previos a la quiebra y, a partir de dicha información, se han calculado una serie de ratios, unos populares en la literatura contable para medir la solvencia empresarial y otros específicos del sector asegurador. Éstas son las mismas variables que las empleadas en el trabajo de Segovia (2003). En el **Anexo 1** se presentan los 21 ratios seleccionados. Dadas las peculiaridades de la muestra, los ratios 15 y 16 han sido eliminados en el análisis posterior por tomar valores carentes de sentido económico, utilizando finalmente como variables independientes los 19 ratios restantes.

Vamos a desarrollar dos modelos diferentes según que los datos procedan de uno o dos años previos a la quiebra. De este modo, se trata de predecir la crisis con uno o dos años de antelación, respectivamente. Llamaremos a estos modelos *Modelo 1* y *Modelo 2*.

Para desarrollar el *Modelo 1*, se utilizarán las 72 empresas disponibles. Sin embargo, no disponemos de los datos de la totalidad de estas empresas para el segundo año previo a la quiebra, únicamente de 70 de ellas. Al eliminar también las respectivas parejas de las dos empresas faltantes, contamos en total con 68 empresas para el desarrollo del *Modelo 2*.

Para verificar la capacidad predictiva de los modelos, llevaremos a cabo un proceso de validación *jackknife*.

Al disponer de pocos datos, reservar parte de ellos para el test supone utilizar todavía menos para la obtención de los modelos, lo que podría ocasionar que dichos modelos fueran de mala calidad. Además, el resultado sería demasiado dependiente del modo en el cual se hubiese

realizado la partición del conjunto completo en dos subconjuntos disjuntos de entrenamiento y test. Dado que, generalmente, esta partición se efectúa de manera aleatoria, podría ocurrir que dos experimentos distintos realizados con el mismo método sobre la misma muestra obtuvieran resultados muy dispares.

Un mecanismo que permite evitar la dependencia del resultado del experimento del modo en el cual se realice la partición es el método jackknife (también denominado *leave-one-out*). Siendo k el número de instancias que contenga el conjunto de entrenamiento (en nuestro caso, 72 para el *Modelo 1* y 68 para el *Modelo 2*), se elabora un modelo utilizando $k-1$ instancias y el caso restante se emplea para evaluar dicho modelo. Este procedimiento se repite k veces, utilizando siempre una instancia diferente para la evaluación del modelo. La estimación del error final se calcula como la media aritmética de los errores de los k modelos parciales.

Éste es un método muy atractivo por dos razones. En primer lugar, se utiliza la mayor cantidad posible de datos para el entrenamiento, lo que presumiblemente redundará de modo favorable en la calidad del modelo. En segundo lugar, el procedimiento es determinístico, los resultados obtenidos con el mismo método sobre la misma muestra siempre serán los mismos y no dependerán del modo en el que se realice la partición de la muestra. El inconveniente estaría en el elevado coste computacional por tantas ejecuciones, con lo que para bases de datos de gran tamaño no sería muy recomendable. Sin embargo, con pequeños conjuntos de datos como el nuestro, ofrece la oportunidad de conseguir la estimación más exacta que posiblemente pueda obtenerse.

Para aplicar el algoritmo C4.5 a nuestra muestra hemos utilizado la última versión publicada del mismo, denominada C4.5 Release 8, que se puede descargar gratuitamente desde la página de Ross Quinlan (<http://www.rulequest.com/Personal/>). Este programa, una vez compilado, puede ser ejecutado en sistemas operativos Unix. También se encuentran implementaciones suyas en algunas librerías. Además, existen nuevas versiones comerciales del algoritmo (C5.0 para Unix y See5 para Windows) que implementan mejoras y funcionalidades adicionales y se comercializan directamente por el

propio Quinlan (RULEQUEST RESEARCH) o a través de paquetes de minería de datos como Clementine, aunque también hay versiones de demostración gratuitas limitadas a bases de datos pequeñas (<http://www.rulequest.com/>). En cuanto al sistema de inducción de árboles de decisión, parece ser esencialmente el mismo que en C4.5. Sin embargo, la generación de reglas con las nuevas versiones es mucho más rápida y claramente se realiza de manera diferente. Como la forma concreta en que ésta se lleva a cabo no ha sido publicada, en el presente trabajo no hemos empleado las nuevas versiones del algoritmo.

4.2. Resultados

A continuación presentamos dos árboles de decisión derivados con los datos del primer año previo a la quiebra de las empresas (*Año 1*). A la izquierda, el árbol sin podar, y a la derecha, el árbol podado al *CF* por defecto del sistema (25%):

Año 1

Decision Tree:

```

R13 <= 0.68 :
| R1 <= 0.29 :
| | R2 > 0.06 : mala (3.0)
| | R2 <= 0.06 :
| | | R4 <= 0 : mala (3.0)
| | | R4 > 0 :
| | | | R2 > 0 : buena (6.0/1.0)
| | | | R2 <= 0 :
| | | | | R6 > 1.63 : buena (2.0)
| | | | | R6 <= 1.63 :
| | | | | | R3 <= 0.02 : buena (3.0/1.0)
| | | | | | R3 > 0.02 : mala (4.0)
| | R1 > 0.29 :
| | | R2 > 0 : buena (14.0)
| | | R2 <= 0 :
| | | | R9 <= 0.22 : buena (5.0)
| | | | R9 > 0.22 :
| | | | | R4 <= 0.12 : buena (4.0/1.0)
| | | | | R4 > 0.12 : mala (2.0)
R13 > 0.68 :
| R9 <= 0.36 : mala (12.0)
| R9 > 0.36 :
| | R17 <= 0.99 : mala (7.0)
| | R17 > 0.99 :
| | | R14 <= 2.02 : buena (5.0)
| | | R14 > 2.02 : mala (2.0)
    
```

Simplified Decision Tree:

```

R13 <= 0.68 :
| R1 > 0.29 : buena (25.0/4.9)
| R1 <= 0.29 :
| | R2 > 0.06 : mala (3.0/1.1)
| | R2 <= 0.06 :
| | | R4 <= 0 : mala (3.0/1.1)
| | | R4 > 0 :
| | | | R2 > 0 : buena (6.0/2.3)
| | | | R2 <= 0 :
| | | | | R6 > 1.63 : buena (2.0/1.0)
| | | | | R6 <= 1.63 :
| | | | | | R3 <= 0.02 : buena (3.0/2.1)
| | | | | | R3 > 0.02 : mala (4.0/1.2)
R13 > 0.68 :
| R9 <= 0.36 : mala (12.0/1.3)
| R9 > 0.36 :
| | R17 <= 0.99 : mala (7.0/1.3)
| | R17 > 0.99 :
| | | R14 <= 2.02 : buena (5.0/1.2)
| | | R14 > 2.02 : mala (2.0/1.0)
    
```

Evaluation on training data (72 items):

Before Pruning		After Pruning		
Size	Errors	Size	Errors	Estimate
27	3(4.2%)	21	5(6.9%)	(25.6%)

Como se puede observar, aún teniendo pocos datos, ninguno de los árboles se puede interpretar con facilidad. Por ello, más adelante, derivaremos también reglas de decisión.

El árbol de la izquierda, por ejemplo, se leería del siguiente modo:

- Si el ratio R13 es menor o igual que 0.68 y el ratio R1 es menor o igual que 0.29 y el ratio R2 es mayor que 0.06, la empresa será “mala” (fracasada).
- Si el ratio R13 es menor o igual que 0.68 y el ratio R1 es menor o igual que 0.29 y el ratio R2 es menor o igual que 0.06 y el ratio R4 es menor o igual que 0, la empresa será “mala”.

Y así continuaríamos descendiendo por el árbol, hasta completar la totalidad de sus hojas (recordemos que cada hoja del árbol se refiere a la decisión a tomar). En el árbol de la izquierda, al final de cada hoja aparece un valor n o n/m : n representa el número de empresas en la muestra que se clasifican de acuerdo a las condiciones que nos llevan hasta esa hoja y m el número de empresas mal clasificadas. En el árbol podado de la derecha, al final de las hojas aparece un valor n/e , donde el significado de n es el mismo que acabamos de comentar y e representa el número de errores previstos sobre n casos de prueba, es decir, $e = n \cdot p_e$, siendo p_e la estimación de la probabilidad de error que se realiza para llevar a cabo el proceso de postpoda.

Justo debajo de los árboles aparece la evaluación de los mismos sobre el conjunto de datos de entrenamiento (72 empresas). Antes de la poda, el tamaño del árbol es de 27 nodos (nodos internos y nodos hoja), y hay 3 errores en la clasificación (4.2%). Después de la poda, el árbol consta de 21 nodos y comete 5 errores (6.9%). Además, se proporciona una estimación del porcentaje de errores del árbol podado

obtenida como cociente de la suma de los errores estimados de todas sus hojas ($\sum e$) entre la suma de todos los casos cubiertos ($\sum n$).

El elevado porcentaje de clasificaciones correctas obtenidas tanto con el árbol sin podar como con el árbol podado (95.8% y 93.1%, respectivamente) no indica que sean un buen modelo de cara a la clasificación de nuevas empresas, puesto que, como ya hemos comentado, pueden estar sobreajustados a los datos de entrenamiento y comportarse mal para nuevos elementos. Para comprobar esto, llevamos a cabo el proceso de validación jackknife, que proporciona los siguientes resultados:

	Before Pruning		After Pruning		
	Size	Errors	Size	Errors	Estimate
train:	24.3	3.8(5.4%)	18.6	5.8(8.1%)	(25.3%)
test:	24.3	0.5(45.8%)	18.6	0.4(44.4%)	(25.3%)

Efectivamente se comprueba que, antes de la poda, la media de los errores de los 72 modelos parciales construidos es sólo del 5.4% mientras que el error medio al evaluar dichos modelos empleando una instancia distinta en cada iteración aumenta al 45.8%, y este mal comportamiento del árbol en el test es un indicador claro de la existencia de un problema de sobreajuste del mismo a los datos de entrenamiento. Después de la poda al *CF* del 25% ocurre algo similar. Esto sugiere realizar una postpoda más intensa (disminuyendo el valor de *CF*) o incluso una prepoda del árbol. De este modo, exigiendo que no se ramifiquen nodos que contengan un número de casos inferior a 6 (por defecto, el sistema establece este valor en 2) y realizando también una postpoda a un *CF* del 30% (ya que partiendo de un árbol prepodado quizás ya no sea necesario postpodarlo demasiado) obtenemos los siguientes árboles (a la izquierda, el árbol prepodado y no postpodado y a la derecha el árbol prepodado y postpodado con *CF* del 30%):

Decision Tree:

```

R13 <= 0.68 :
| R1 > 0.29 : buena (25.0/3.0)
| R1 <= 0.29 :
| | R4 > 0.08 : mala (7.0/1.0)
| | R4 <= 0.08 :
| | | R20 <= 0.43 : buena (6.0/1.0)
| | | R20 > 0.43 : mala (8.0/3.0)
R13 > 0.68 :
| R19 > 0 : mala (9.4)
| R19 <= 0 :
| | R17 <= 0.99 : mala (7.0)
| | R17 > 0.99 : buena (9.6/4.6)
    
```

Simplified Decision Tree:

```

R13 > 0.68 : mala (26.0/6.7)
R13 <= 0.68 :
| R1 > 0.29 : buena (25.0/4.6)
| R1 <= 0.29 :
| | R4 > 0.08 : mala (7.0/2.2)
| | R4 <= 0.08 :
| | | R20 <= 0.43 : buena (6.0/2.1)
| | | R20 > 0.43 : mala (8.0/4.3)
    
```

Evaluation on training data (72 items):

Before Pruning		After Pruning		
Size	Errors	Size	Errors	Estimate
13	12(16.7%)	9	13(18.1%)	(27.6%)

Así vemos que el tamaño de los árboles ha disminuido considerablemente y el porcentaje de error ha aumentado. No obstante, ambos árboles siguen clasificando correctamente un porcentaje elevado (más del 80%) de las empresas que constituyen el conjunto de entrenamiento, lo que indica que son capaces de capturar las regularidades presentes en los datos (no subajustan). Lo que interesa ahora es comprobar si estos modelos serían fiables a la hora de utilizarlos con fines predictivos, es decir, a la hora de clasificar nuevas empresas ajenas a la muestra empleada para la elaboración de los mismos. Por tanto, hemos de llevar a cabo la validación de los modelos. El método jackknife proporciona los siguientes resultados:

	Before Pruning		After Pruning		
	Size	Errors	Size	Errors	Estimate
train:	11.2	12.1(17.1%)	8.5	12.9(18.1%)	(27.4%)
test:	11.2	0.4(37.5%)	8.5	0.3(31.9%)	(27.4%)

Se observa que al haber realizado esta preoda el porcentaje de error ha disminuido en el test, sobre todo con el árbol también postpodado, que obtiene un error del 31.9% frente al 44.4% de antes. Podríamos intensificar la poda para conseguir un modelo más preciso. No obstante, aunque no ocurra aquí, si el problema es muy complejo el

árbol de decisión generado será muy grande y aún tras su poda no será fácil de interpretar. En estos casos, sería aconsejable derivar reglas de decisión, que pueden ser clasificadores tan fiables como los árboles y además de interpretación mucho más sencilla.

C4.5 transforma un árbol de decisión no postpodado en un conjunto de reglas mediante el procedimiento que hemos explicado anteriormente. Con los datos del primer año previo a la quiebra obtenemos el conjunto de reglas que se presenta a continuación (*Modelo 1*), partiendo del árbol prepodado impidiendo la ramificación de nodos que contengan un número de instancias inferior a 6 y fijando para el proceso de simplificación de reglas un *CF* del 30%:

Modelo 1

Rule 4:

R1 > 0.29
R13 <= 0.68
-> class buena [81.8%]

Rule 1:

R4 <= 0.08
R20 <= 0.43
-> class buena [77.8%]

Rule 5:

R13 > 0.68
R17 <= 0.99
-> class mala [87.5%]

Rule 2:

R1 <= 0.29
R20 > 0.43
-> class mala [79.8%]

Default class: buena

Evaluation on training data (72 items):

	Rule	Size	Error	Used	Wrong	Advantage
	----	----	-----	----	-----	-----
	4	2	18.2%	25	3 (12.0%)	0 (0 0)
buena	1	2	22.2%	7	1 (14.3%)	0 (0 0)
buena						
mala	5	2	12.5%	9	0 (0.0%)	6 (6 0)
mala						
mala	2	2	20.2%	25	4 (16.0%)	17 (21 4)

Tested 72, errors 10 (13.9%)

(a)	(b)	<-classified as
----	----	
32	4	(a): class buena
6	30	(b): class mala

Hemos obtenido cuatro reglas (dos para cada una de las clases) y la clase por defecto (en este caso “buena”). El número de cada regla es arbitrario, se deriva del orden de las hojas en el árbol original, y sólo sirve para identificar las reglas. El antecedente de cada regla se integra por una serie de condiciones relativas a los valores de ciertos atributos que han de ser todas satisfechas para que la regla sea aplicable, y el consecuente indica la decisión a tomar. Al lado, entre corchetes, se refleja la precisión estimada de la regla. Así, en la primera regla (Rule 4), se prevé que la clasificación será correcta para el 81.8% de casos de prueba que satisfagan las condiciones de la regla.

Después se presenta la evaluación de cada una de las reglas sobre los casos de entrenamiento. Por ejemplo, para la primera (Rule 4), se indica que el antecedente se compone de dos condiciones, la tasa de error prevista es del 18.2%, se usó 25 veces en la clasificación de los casos de entrenamiento y 3 fueron errores. La columna titulada “Advantage” muestra qué ocurriría si la regla se omitiese de la lista.

Cada entrada de la forma a (b/c) indica que, si la regla fuera omitida, b casos clasificados ahora correctamente por esta regla serían clasificados incorrectamente, y c casos ahora mal clasificados por esta regla serían clasificados correctamente por las reglas siguientes y la clase por defecto; el beneficio neto de no omitir la regla es entonces $a = b - c$.

Por último, se muestra el número total de errores (incluyendo las clasificaciones erróneas de la clase por defecto) así como una matriz de confusión que señala el tipo de errores cometidos.

Para evaluar la capacidad predictiva del modelo llevamos a cabo el procedimiento de validación jackknife, obteniendo los siguientes resultados:

train:	Tested 71.0, errors 11.1 (15.7%)
test:	Tested 1.0, errors 0.2 (25.0%)

Como se puede observar, la media de los errores de los 72 modelos parciales construidos es del 15.7 % (lo que representa un 84.3% de aciertos en la clasificación de los casos de entrenamiento) y el error medio al evaluar dichos modelos con los casos de test es del 25% (75% de aciertos). De este modo, el conjunto de reglas que representan el *Modelo 1*, además de ser un clasificador más simple y comprensible que los árboles de decisión anteriores, también se puede considerar más fiable a la hora de utilizarlo para clasificar otras empresas, en cuyo caso se aplicaría primero cualquiera de las reglas representativas de la clase ordenada en primer lugar (“buena”) y si ninguna de ellas fuera aplicable se pasaría a las reglas correspondientes a la clase ordenada en segundo lugar (“mala”); si éstas tampoco fuesen aplicables, a la empresa se le asignaría la clase por defecto (“buena”).

El antecedente de la primera de las reglas (Rule 4) refleja dos condiciones que se han de cumplir, relativas a la liquidez y solvencia de las empresas, para que éstas sean asignadas a la clase “buena” designada en el consecuente. El ratio R1 es un ratio de liquidez,

puesto que relaciona el fondo de maniobra con el activo total, y es en nuestro caso un ratio muy discriminante entre las dos clases de empresas. Lógicamente, en nuestra muestra las empresas “buenas” presentan una mejor posición en liquidez que las empresas “malas”. Una empresa de seguros en funcionamiento normal generará liquidez de forma continuada, debido a la inversión del proceso productivo que se da en este tipo de entidades, por lo que no será habitual que aparezcan problemas de esta índole. Si bien una de las cuestiones más importantes para asegurar el buen funcionamiento de cualquier tipo de empresa es la necesidad de un colchón de liquidez que le permita hacer frente a sus deudas a corto plazo sin tener que recurrir a la realización de sus activos fijos, en el caso de la empresa de seguros dicha necesidad reviste una mayor importancia, pues debe ser capaz de atender los siniestros sin estar condicionada al cobro de las primas o a la enajenación de activos no realizables. De este modo, aunque en las empresas aseguradoras no cabe esperar problemas de liquidez, su presencia sería un indicador de problemas financieros más graves que los que existirían en otro tipo de empresas. En nuestro caso, las empresas “malas” alcanzan un valor muy bajo en este ratio, incluso, en algunos casos, negativo, indicando que el fondo de maniobra también es negativo. Por otro lado, el ratio R13, como indican Martín *et al.* (1999), es un ratio de “solvencia en sentido estricto”, que recoge en el numerador la medida del riesgo anual basándose en la valoración de los riesgos que realmente han ocurrido, siniestros del ejercicio, registrados como Gastos Técnicos en la Cuenta de Pérdidas y Ganancias, y muestra en el denominador el soporte financiero real para el periodo (Capitales Propios + Provisiones Técnicas). De nuevo en nuestra muestra este ratio es un buen discriminante entre las empresas “buenas” y las empresas “malas”, alcanzando en estas últimas valores muy elevados, que indican una fuerte descapitalización y vaticinan la insolvencia de forma clara.

La segunda de las reglas (Rule 1), representativa también de la clase “buena”, consta en su antecedente de dos condiciones. Una de ellas informa a través del ratio R20 de la capacidad de afrontar los gastos anuales con las primas anuales correspondientes. Éste es un ratio de siniestralidad y es uno de los utilizados en Estados Unidos por la *National Association of Insurance Commissioners* (NAIC) dentro de su sistema IRIS (*Insurance Regulatory Information System*) de ratios

de alarma ante una posible situación de insolvencia. En las empresas “malas” de nuestra muestra este ratio presenta valores más elevados que en las “buenas”, que indican una merma en la capacidad mencionada. Por su parte, el ratio R4 relaciona el cash-flow, entendido como la capacidad de generar fondos por parte de la empresa, con el pasivo total. Ya en los clásicos estudios de Beaver (1966 y 1968) se mostraba este ratio como el de mayor valor predictivo. En principio, podría resultar un tanto contradictorio el sentido de la desigualdad de la condición de la regla referente a este ratio, ya que la lógica induce a pensar que este ratio no debería ser menor en las empresas “buenas” que en las “malas”, puesto que los recursos generados guardan una relación directa con el grado de autofinanciación de la empresa. Pero lo cierto es que muchas de las empresas “malas” de nuestra muestra alcanzan un valor de este ratio más elevado de lo que cabría esperar, sin embargo es necesario tener en cuenta que la mejoría relativa en la generación de fondos en años muy cercanos a la quiebra es un comportamiento habitual en las empresas con dificultades, que tratan de recuperar liquidez por todos los medios posibles realizando inmuebles, inversiones, anticipando cobros, postponiendo pagos... (Gabás, 1997). Esto explicaría el sentido a priori contrario a la intuición de la desigualdad de la regla relacionada con el ratio R4.

En la tercera regla (Rule 5), ahora representativa de la clase “mala”, de nuevo aparece el ratio de solvencia R13, aunque, lógicamente, con el sentido de la desigualdad contrario al que aparece en la regla para las “buenas”, porque valores mayores en el ratio indican una menor solvencia. Además hay una segunda condición relativa al ratio R17, que refleja la proporción que suponen los gastos de explotación respecto a las primas y recargos. Es otro de los ratios utilizados por la NAIC en su sistema IRIS de “alerta temprana”. Se diferencia del ratio R20 en el numerador, ya que a los gastos técnicos se añaden comisiones y otros gastos de explotación. El sentido que muestra la desigualdad asociada a este ratio es también contrario a la intuición, pero en este caso parece razonable buscar la explicación en el papel que juega el azar en los valores concretos que toman los ratios, papel más acusado en casos como el nuestro en los que el reducido tamaño de la muestra impide a la ley de los Grandes Números neutralizar el efecto de las pequeñas fluctuaciones aleatorias. Así, se puede observar

que los valores que toma R17 están muy concentrados en torno a la media. Al ser la dispersión tan pequeña variaciones poco o nada significativas pueden tener, bajo determinadas circunstancias, un papel relevante de cara a la discriminación. La primera parte de la regla ($R13 > 0.68$) es satisfecha por 26 empresas, 5 sanas y 21 fracasadas, para las cuales el valor que toma R17 es prácticamente idéntico. Sin embargo los valores para las 5 sanas están todos muy ligeramente por encima de 0.99, mientras que los valores para las fracasadas presentan una dispersión algo mayor, lo que unido al hecho de que sean más hace que existan 9 casos para los que R17 se sitúa muy ligeramente por debajo de 0.99. Tomando este valor como punto de corte se conseguirá entonces que todas las empresas cubiertas por la regla sean fracasadas con lo que la disminución de entropía y con ello la mejora en capacidad discriminante serán máximas.

Por último, en la cuarta regla (Rule 2), vuelven a aparecer en su antecedente los ratios R1 y R20. Pero como esta regla caracteriza a las empresas “malas”, el sentido de la desigualdad es contrario al de las reglas para las “buenas” donde intervienen estos mismos ratios.

A continuación, repetiremos el proceso con los datos del segundo año previo a la quiebra (*Año 2*). Debajo se presenta el árbol de decisión sin ningún tipo de poda (puesto que la poda por defecto, al *CF* del 25%, no afecta en absoluto, tal y como se puede comprobar en la evaluación sobre los 68 casos de entrenamiento):

Año 2

```
R1 <= 0.26 :
| R21 <= 0.41 : buena (5.0/1.0)
| R21 > 0.41 : mala (23.0/2.0)
R1 > 0.26 :
| R14 <= 0.92 :
| | R20 > 0.62 : buena (17.0)
| | R20 <= 0.62 :
| | | R3 > 0.11 : mala (2.0)
| | | R3 <= 0.11 :
| | | | R9 > 0.57 : buena (7.0)
| | | | R9 <= 0.57 :
| | | | | R17 <= 0.91 : buena (2.0)
| | | | | R17 > 0.91 : mala (2.0)
| R14 > 0.92 :
| | R17 <= 0.99 : mala (6.0)
| | R17 > 0.99 :
| | | R2 <= 0.05 : buena (2.0)
| | | R2 > 0.05 : mala (2.0)
```

Evaluation on training data (68 items):

Before Pruning			After Pruning		
Size	Errors	Size	Errors	Estimate	
19	3(4.4%)	19	3(4.4%)	(21.8%)	

Llevando a cabo el proceso de validación jackknife, se puede comprobar que se trata de un modelo sobreajustado, dado el elevado porcentaje de error obtenido en el test, tanto sin poda como con postpoda al *CF* por defecto (error del 41.2% y 39.7%, respectivamente):

Before Pruning				After Pruning		
Size	Errors		Size	Errors	Estimate	
train: 17.4	3.1(4.7%)		17.3	3.1(4.7%)	(21.2%)	
test: 17.4	0.4(41.2%)		17.3	0.4(39.7%)	(21.2%)	

Así que vamos a intentar obtener un árbol de decisión más general combinando prepoda y postpoda. Impidiendo la ramificación de nodos que contengan menos de 9 casos y con un *CF* para la postpoda del 25% obtenemos un árbol de decisión muy sencillo:

```

R1 <= 0.26 : mala (28.0/6.0)
R1 > 0.26 :
| R14 <= 0.92 : buena (30.0/4.0)
| R14 > 0.92 : mala (10.0/2.0)
    
```

Evaluation on training data (68 items):

Before Pruning			After Pruning		
Size	Errors	Size	Errors	Estimate	
5	12(17.6%)	5	12(17.6%)	(26.1%)	

Además de sencillo, los resultados del jackknife muestran su buena capacidad predictiva (26.5% de errores en el test, o lo que es lo mismo, 73.5% de clasificaciones correctas):

Before Pruning			After Pruning			
	Size	Errors		Size	Errors	Estimate
train:	6.0	11.7(17.5%)		5.9	11.8(17.6%)	(26.8%)
test:	6.0	0.3(27.9%)		5.9	0.3(26.5%)	(26.8%)

Por último, derivamos un conjunto de reglas de decisión (**Modelo 2**), partiendo del árbol prepodado impidiendo la ramificación de nodos que contengan un número de instancias inferior a 9 y fijando para el proceso de simplificación de reglas un *CF* del 15%:

Modelo 2

Rule 2:

R1 > 0.26
 R14 <= 0.92
 -> class buena [76.8%]

Rule 1:

R1 <= 0.26
 -> class mala [67.4%]

Rule 3:

R14 > 0.92
 -> class mala [58.6%]

Default class: buena

Evaluation on training data (68 items):

Rule	Size	Error	Used	Wrong	Advantage	
2	2	23.2%	30	4 (13.3%)	0 (0 0)	buena
1	1	32.6%	28	6 (21.4%)	15 (20 5)	mala
3	1	41.4%	10	2 (20.0%)	6 (8 2)	mala

Tested 68, errors 12 (18.4%)

(a)	(b)	<-classified as
26	8	(a): class buena
4	30	(b): class mala

jackknife:

train:	Tested 67.0, errors 12.4 (18.4%)
test:	Tested 1.0, errors 0.3 (27.9%)

En esta ocasión obtenemos tres reglas y la clase por defecto (“buena”), y los resultados del jackknife indican una capacidad predictiva prácticamente igual a la del árbol anterior.

Se puede observar que en el segundo año previo a la quiebra se manifiestan de nuevo como discriminantes entre empresas sanas y fracasadas el ratio de liquidez (R1) y la “solvencia en sentido estricto”, medida ahora a través del ratio R14 en lugar del ratio R13. Estos dos ratios sólo se diferencian en el numerador. El segundo recoge en dicho numerador Gastos Técnicos de Seguro Directo mientras que el primero recoge Gastos Técnicos de Negocio Neto, teniendo en cuenta por tanto la influencia del reaseguro. Pero, en realidad, la presencia en este modelo del ratio R14 en lugar del ratio R13 no obedece a otra razón más que al elevadísimo grado de correlación existente entre ambos ratios. Sobre todo en este segundo año previo a la quiebra, R14 y R13 serían virtualmente el mismo ratio, con lo que la elección de R14 es prácticamente arbitraria. De hecho, si eliminásemos de la muestra el ratio R14, los ratios discriminantes en el *Modelo 2* serían R1 y de nuevo R13.

4.3. Comparación con el Análisis Discriminante

Nuestro trabajo está orientado a poner de manifiesto la utilidad de los árboles y reglas de decisión construidos con el algoritmo C4.5 de cara a predecir el fracaso empresarial en las empresas de seguros. Sin embargo, esta tarea quedaría incompleta si no se realizase una comparación de los resultados alcanzados con los que se obtendrían aplicando al mismo problema alguna técnica alternativa y bien conocida que actuase como término de referencia con respecto al cual poder valorar de manera fundamentada la calidad real del método propuesto. Para este fin se ha elegido el Análisis Discriminante (AD).

El AD es la técnica estadística de clasificación por antonomasia y aunque desde su creación por R. Fisher en 1936 ha surgido un gran número de métodos de clasificación alternativos continúa siendo ampliamente utilizado, proporcionando además unos resultados notablemente satisfactorios. Es precisamente por su carácter de técnica estándar y por su habitual buen desempeño por lo que ha sido elegido aquí como método de clasificación con el que poder comparar los resultados alcanzados con el C4.5.

El AD supone la estimación de una serie de funciones lineales de las variables predictoras –las funciones discriminantes– que permiten dividir el espacio en el que se definen estas variables en un conjunto de regiones separadas por fronteras lineales. A cada una de las posibles clases o categorías a las que pertenezcan los datos analizados le corresponderá una de dichas regiones, de modo que la región en que se ubique un nuevo caso determinará la clase a la que sea asignado al hacer la predicción. Las fronteras lineales (rectas si hay 2 variables predictoras, planos si hay 3 e hiperplanos en dimensiones superiores) constituyen el tipo más sencillo posible de superficie separadora entre clases. La calidad de la clasificación alcanzada con ellas dependerá en gran medida del grado en que la frontera real entre cada par de clases pueda ser aproximada por una superficie “plana”. Los árboles y reglas de decisión, por el contrario, suponen superficies de separación entre clases con una estructura ramificada más compleja que la meramente lineal. Estas superficies tienen una estructura más flexible y pueden adaptarse mejor a la forma en que realmente se distribuyen los datos disponibles, pero tienen la restricción de que los distintos “trozos” de

superficie por los que están formados han de ser paralelos a los planos coordenados.

Las funciones discriminantes del AD se calculan admitiendo que las variables explicativas siguen una distribución normal multivariante y que las matrices de covarianza de las distintas clases son iguales (Peña, 2002). Bajo estas condiciones el AD constituye un clasificador óptimo en el sentido de que minimiza la probabilidad de clasificación errónea. De no verificarse la normalidad y la homoscedasticidad el discriminante perderá su carácter óptimo pero sin que ello necesariamente signifique, ni mucho menos, la desaparición de sus buenas condiciones como clasificador. Esta robustez puede ser vista como una consecuencia de que el AD admite una interpretación según la cual las funciones discriminantes permiten proyectar los vectores representativos de las distintas observaciones sobre un conjunto de direcciones que maximizan la separación, en el espacio de las variables proyectadas, entre los distintos grupos en que se pueden integrar los objetos clasificados. Esta interpretación, de naturaleza puramente geométrica, es independiente de cualquier hipótesis estadística acerca de la distribución de los datos utilizados y explica el buen comportamiento que suele exhibir el AD frente a violaciones de la normalidad y la homoscedasticidad.

En nuestro caso, como se indicará a continuación, ni los datos disponibles se distribuyen normalmente ni son iguales las matrices de covarianza correspondientes a empresas sanas y fracasadas. Esto significa que el AD no será un clasificador óptimo en el sentido de minimizar la probabilidad de error, pero sí se podrá afirmar que clasificará las nuevas observaciones asignándolas a aquel grupo al que se encuentren más próximas, midiendo esta proximidad por la distancia euclídea sobre un sistema de ejes que maximice la separación entre clases. Debido a ello, y a pesar de la ausencia de normalidad, la regla de clasificación obtenida será razonablemente buena.

En una situación en la que no se verifican la normalidad ni la homoscedasticidad podría ser razonable recurrir a la Regresión Logística para llevar a cabo la clasificación, sin embargo debe tenerse en cuenta que no estamos interesados en comparar el C4.5 con la

mejor alternativa posible (que por otra parte ni siquiera sería la Regresión Logística), sino en compararlo con una técnica que por su carácter absolutamente estándar constituya el “patrón de medida” más adecuado para su valoración. Por otra parte, se ha observado que el AD se comporta bien sobre una enorme variedad de conjuntos de datos (Hastie *et al.*, 2001) y que en los numerosos estudios realizados comparando el AD con la Regresión Logística no se han observado en general diferencias apreciables entre ambos métodos (McLachlan, 1992). Así, por ejemplo, en Michie *et al.* (1994) se compara una gran variedad de técnicas de clasificación sobre distintos conjuntos de datos encontrándose que tanto el AD como la Regresión Logística están entre las 5 mejores técnicas y que hay en la práctica poca diferencia entre ambas.

Para la realización del AD se ha utilizado el paquete estadístico SPSS 11.0 y el software R 2.0.0 distribuido por *CRAN Foundation* (R Development Core Team, 2004). Como paso previo al análisis, y debido a que el AD no acepta datos ausentes (*missing values*), se ha procedido a imputar esos valores faltantes. Lo más habitual para ello es usar la media o la mediana de los valores que sí están presentes, pero se ha elegido una alternativa algo más laboriosa y que se describe en Troyanskaya *et al.* (2001). En este artículo se comparan distintos métodos de imputación concluyendo que el que mejores resultados proporciona es el conocido como KNNimpute, que consiste en imputar un *missing value* de acuerdo con los valores que tome la variable correspondiente para las *k* observaciones más próximas a la que contiene el valor faltante que se desea imputar. El valor que se imputa será la media, ponderada de acuerdo a la distancia a la que se encuentren, de los valores correspondientes para los mencionados *k* vecinos más próximos. El valor concreto que se toma para *k* dependerá del conjunto de datos utilizado y se calcula tomando el subconjunto de observaciones que no presentan ningún *missing value*, eliminando al azar algunos valores, imputándolos con distintos valores de *k* y comparando la matriz de datos obtenida con la inicial. El proceso se repite un número de veces suficiente para garantizar la estabilidad del resultado. Los detalles concretos pueden consultarse en la referencia indicada.

Tras haber realizado la imputación de la manera descrita se ha contrastado la hipótesis de normalidad utilizando el test de Shapiro-Wilk multivariante. Los resultados se recogen en la siguiente tabla y obligan a rechazar la hipótesis nula, con lo que no se podrá mantener en modo alguno que los datos sigan una distribución normal multivariante.

<u>Año 1 empresas buenas</u>	<u>Año 1 empresas malas</u>
Shapiro-Wilk normality test	Shapiro-Wilk normality test
W = 0.2978, p-value = 5.969e-12	W = 0.1686, p-value = 4.875e-13
<u>Año 2 empresas buenas</u>	<u>Año 2 empresas malas</u>
Shapiro-Wilk normality test	Shapiro-Wilk normality test
W = 0.1705, p-value = 1.159e-12	W = 0.1676, p-value = 1.099e-12

A continuación se lleva a cabo una selección de las variables que se incluirán en los modelos discriminantes. Para ello se utiliza el procedimiento *stepwise* del SPSS. Con esta selección previa de las variables se consigue eliminar problemas de colinealidad que harían inestables los resultados, obtener modelos más sencillos y fáciles de interpretar y reducir el sobreajuste haciendo que mejore la capacidad de clasificar nuevos casos no usados para construir el modelo, aunque sea a costa de empeorar la clasificación para los casos de entrenamiento. Por el contrario, hay que indicar que los estadísticos que se utilizan para controlar el *stepwise* son muy sensibles frente a desviaciones de la normalidad con lo que los resultados obtenidos pueden ser inestables y deben ser acogidos con una cierta cautela.

En cualquier caso las variables seleccionadas para el año 1 son R1, R4, R10, R11, R13 y R19 y para el año 2 son R1, R6, R7, R14 y R21. Con estos dos conjuntos de ratios se construyen sendos modelos discriminantes. Los coeficientes estandarizados de las funciones discriminantes obtenidas proporcionan información acerca de la influencia de cada variable en el modelo correspondiente y se recogen en la siguiente tabla:

Año 1		Año 2	
Coeficientes estandarizados de las funciones discriminantes canónicas		Coeficientes estandarizados de las funciones discriminantes canónicas	
	Función		Función
	1		1
R1	-,686	R1	-,408
R4	,343	R6	,498
R10	-1,983	R7	-2,195
R11	1,918	R14	2,632
R13	,718	R21	-,406
R19	,497		

La hipótesis de que los modelos construidos discriminan realmente entre ambas poblaciones de forma que la clasificación obtenida no puede ser atribuida al azar se contrasta por medio del estadístico lambda de Wilks. El valor que éste adopta obliga a rechazar la hipótesis nula de no discriminación:

Año	Lambda de Wilks				
	Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
	1	,698	24,045	6	,001

Año	Lambda de Wilks				
	Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
	1	,769	16,664	5	,005

La homoscedasticidad se puede verificar con el estadístico M de Box. Los resultados obtenidos no dan soporte a esta hipótesis:

Año 1			Año 2		
Resultados de la prueba			Resultados de la prueba		
M de Box		234,315	M de Box		125,152
F	Aprox.	10,132	F	Aprox.	7,662
	gl1	21		gl1	15
	gl2	18022,192		gl2	17538,632
	Sig.	,000		Sig.	,000
Contrasta la hipótesis nula de que las matrices de covarianza poblacionales son iguales.			Contrasta la hipótesis nula de que las matrices de covarianza poblacionales son iguales.		

Por último se recogen los resultados que se obtienen al comprobar la capacidad clasificadora de los modelos, tanto medida para los casos de entrenamiento, indicador poco fiable, como estimada por medio del jackknife (denominado “validación cruzada” en la salida del SPSS), lo que sí proporciona una valoración realista de la verdadera capacidad predictiva del modelo:

Año 1

Resultados de la clasificación^c

		Mala-0 Buena-1	Grupo de pertenencia pronosticado		Total
			,00	1,00	
Original	Recuento	,00	22	14	36
		1,00	7	29	36
	%	,00	61,1	38,9	100,0
		1,00	19,4	80,6	100,0
Validación cruzada ^a	Recuento	,00	21	15	36
		1,00	9	27	36
	%	,00	58,3	41,7	100,0
		1,00	25,0	75,0	100,0

- a. La validación cruzada sólo se aplica a los casos del análisis. En la validación cruzada, cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.
- b. Clasificados correctamente el 70,8% de los casos agrupados originales.
- c. Clasificados correctamente el 66,7% de los casos agrupados validados mediante validación cruzada.

Año 2

Resultados de la clasificación^{b,c}

			Grupo de pertenencia pronosticado		Total
			,00	1,00	
Original	Recuento	0-Mala 1-Buena			
			,00	1,00	
	%	,00	1,00		
Validación cruzada ^a	Recuento	0-Mala 1-Buena			
			,00	1,00	
	%	,00	1,00		

- a. La validación cruzada sólo se aplica a los casos del análisis. En la validación cruzada, cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.
- b. Clasificados correctamente el 73,5% de los casos agrupados originales.
- c. Clasificados correctamente el 70,6% de los casos agrupados validados mediante validación cruzada.

Para finalizar, en la siguiente tabla se resumen los resultados obtenidos en cuanto a porcentajes de clasificaciones correctas estimados mediante el método jackknife con los dos enfoques de clasificación empleados: C4.5 (árboles y reglas de decisión) y Análisis Discriminante:

	C4.5 Árboles ⁶	C4.5 Reglas	A. Discriminante
Año 1	68.1%	75%	66.7%
Año 2	73.5%	72.1%	70.6%

Se observa que no existen grandes diferencias entre los distintos enfoques en cuanto a porcentaje de clasificaciones correctas. No obstante, si tuviésemos que establecer un ranking entre ellos, considerando la clasificación de forma global para los dos años situaríamos a C4.5 Reglas en primer lugar, C4.5 Árboles en segundo

⁶ Se han tomado los resultados alcanzados mediante la combinación de prepoda y postpoda.

lugar y por último Análisis Discriminante. También podríamos apuntar que los modelos basados en reglas parecen ser más estables, ya que obtienen un porcentaje de clasificaciones correctas similar en los dos años. Por otra parte, cabría preguntarse la razón por la cual C4.5 Árboles y Análisis Discriminante clasifican mejor en el año 2 que en el año 1, cuando la lógica normalmente indicaría que deberíamos esperar una disminución en la precisión clasificatoria ante el aumento del horizonte temporal de la predicción. El caso es que este fenómeno se observa en ocasiones en otros estudios sobre predicción de insolvencias (Martínez de Lejarza, 1999; Sanchís, 2003) y no parece haber ninguna razón clara que lo justifique, con lo que no puede ser achacado más que a las peculiaridades de los datos, el reducido tamaño de la muestra y la eventual mala calidad de la información contable.

5. Conclusiones

La detección precoz de la insolvencia empresarial es un problema que ha recibido constante atención por parte del mundo académico y profesional. El sector del seguro no ha sido ajeno a esta situación. La utilización de modelos eficientes de predicción de insolvencias facilitaría grandemente la labor de supervisión de las empresas aseguradoras permitiendo que los recursos limitados de la inspección se dirigiesen hacia aquellas preseleccionadas como potencialmente insolventes y, de forma paralela, se flexibilizase la normativa en cuanto a requisitos de solvencia.

En este trabajo hemos aplicado a una muestra de empresas españolas de seguros no-vida, partiendo de un conjunto de ratios de carácter financiero, un paradigma procedente del área de la Inteligencia Artificial conocida como Aprendizaje Automático, el algoritmo de inducción de árboles y reglas de decisión C4.5, con el objeto de comprobar su utilidad para la predicción de insolvencias en el mencionado sector. Al objeto de tener una referencia que pueda ser utilizada como término de comparación, hemos aplicado también Análisis Discriminante por ser ésta una técnica estadística estándar.

Los resultados obtenidos con C4.5 mejoran a los que se alcanzan mediante Análisis Discriminante, aunque la diferencia no es abismal. No obstante, el algoritmo C4.5 supera al Análisis Discriminante también en otros aspectos: se aplica con facilidad, proporciona modelos de interpretación más sencilla y es robusto ante el “ruido” introducido por valores faltantes y *outliers*, y por tanto se adecúa mejor a la información contable, que suele presentar datos interrelacionados, incompletos, adulterados o erróneos. Sin embargo, a pesar de sus múltiples ventajas, el algoritmo C4.5 presenta el inconveniente de que no permite considerar distintos costes de clasificación errónea, es decir, sólo computa el número global de errores sin distinguir si se trata de clasificar una empresa sana como fracasada o clasificar una fracasada como sana, error este último que resultaría mucho más grave. No obstante, esta desventaja ha sido subsanada en las nuevas versiones del algoritmo.

Por otro lado, aunque los resultados de la validación jackknife indican que ambos enfoques pueden considerarse adecuados, no debemos olvidar algunas de sus limitaciones a la hora de utilizarlos con fines predictivos, como son el hecho de que la muestra sea relativamente pequeña o el que el análisis se haya realizado al margen del factor tamaño, que puede ser importante a la hora de predecir la insolvencia. No obstante, si hubiésemos optado por no realizar el emparejamiento de las empresas de acuerdo a su tamaño, el peso tan determinante de dicha variable en la probabilidad de quiebra podría haber oscurecido el papel de las variables de carácter financiero en las que estamos especialmente interesados.

Sería interesante considerar un horizonte de predicción mayor del que se ha utilizado aquí y elaborar diferentes modelos para cada uno de los años previos a la crisis. Para paliar el problema derivado del desconocimiento del año concreto en que ésta se producirá, sería también conveniente elaborar un único modelo teniendo en cuenta los valores de los ratios de varios años consecutivos antes de la crisis tomados conjuntamente. En este sentido, nuestro trabajo podría ser considerado como un primer paso en ese camino.

Anexo 1: Ratios empleados

Ratio	Definición
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

Referencias bibliográficas

BEAVER, W.H. (1966): “Financial Ratios as Predictors of Failure”, *Journal of Accounting Research*, vol. 4, Empirical Research in Accounting: Selected Studies 1966, pp. 71-111.

BEAVER, W.H. (1968): “Market Prices, Financial Ratios, and the Prediction of Failure”, *Journal of Accounting Research*, vol. 6, nº 2, pp. 179-192.

- BREIMAN, L.; FRIEDMAN, J.H.; OLSHEN, R.A. y STONE, C.J. (1984): *Classification and Regression Trees*. Wadsworth International Group, Belmont, California.
- DE ANDRÉS SUÁREZ, J. (1998): *Caracterización económico-financiera de los sectores de la economía asturiana en función del nivel de rentabilidad de las empresas*. Tesis Doctoral, Universidad de Oviedo.
- FRIEDMAN, J.H. (1977): "A recursive partitioning decision rule for non-parametric classification". *IEEE Transactions on Computers*, vol. 26, nº 4, pp. 404-408.
- GABÁS TRIGO, F. (1997): "Predicción de la insolvencia empresarial", en A. CALVO-FLORES SEGURA y D. GARCÍA PÉREZ DE LEMA (eds.): *Predicción de la insolvencia empresarial*. AECA, Madrid, pp. 11-31.
- GONZÁLEZ PÉREZ, A.L.; CORREA RODRÍGUEZ, A. y BLÁZQUEZ MÚREZ, J.A. (1999): "Perfil del fracaso empresarial para una muestra de pequeñas y medianas empresas". Comunicación presentada al X Congreso AECA, Zaragoza.
- HASTIE, T.; TIBSHIRANI, R. y FRIEDMAN, J. (2001): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York.
- HENRICHON, Jr., E.G. y FU, K.S. (1969): "A nonparametric partitioning procedure for pattern classification", *IEEE Transactions on Computers*, 18, pp. 614-624.
- LÓPEZ HERRERA, D.; MORENO ROJAS, J. y RODRÍGUEZ RODRÍGUEZ, P. (1994): "Modelos de previsión del fracaso empresarial: Aplicación a entidades de seguros en España", *Esic Market*, 84, abril-junio, pp. 83-125.
- MARTÍN PEÑA, M.L.; LEGUEY GALÁN, S. y SÁNCHEZ LÓPEZ, 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, nº 49, Madrid.
- MARTÍNEZ DE LEJARZA ESPARDUCER, I. (1999): "Previsión del fracaso empresarial mediante redes neuronales: un estudio comparativo con el análisis discriminante", en E. BONSON PONTE (ed.): *Tecnologías Inteligentes para la Gestión Empresarial*. RA-MA Editorial, Madrid, pp. 53-70.

- McKEE, T.E. (1995): "Predicting bankruptcy via induction", *Journal of Information Technology*, 10, pp. 26-36.
- McLACHLAN, G.J. (1992): *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, New York.
- MICHIE, D.; SPIEGELHALTER, D. y TAYLOR (Eds.) (1994): *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, England.
- MORA ENGUÍDANOS, A. (1994): "Los modelos de predicción del fracaso empresarial: una aplicación empírica del logit", *Revista Española de Financiación y Contabilidad*, vol. 23, nº 78, pp. 203-233.
- MORGAN, J.N. y MESSENGER, R.C. (1973): *THAID: a Sequential Search Program for the Analysis of Nominal Scale Dependent Variables*. Survey Research Center, Institute for Social Research, University of Michigan.
- MORGAN, J.N. y SONQUIST, J.A. (1963): "Problems in the analysis of survey data, and a proposal", *Journal of the American Statistical Association*, 58, pp. 415-434.
- PEÑA, D. (2002): *Análisis de datos multivariantes*, McGraw-Hill, Madrid.
- QUINLAN, J.R. (1979): "Discovering rules by induction from large collections of examples", en D. MICHIE (ed.): *Expert systems in the microelectronic age*, Edimburgh University Press, Edimburgh.
- QUINLAN, J.R. (1983): "Learning efficient classification procedures", en R.S. MICHALSKI; J.G. CARBONELL y T.M. MITCHELL (eds.): *Machine learning: An Artificial Intelligence approach*, Tioga Press, Palo Alto, California.
- QUINLAN, J.R. (1986): "Induction of decision trees", *Machine Learning*, vol. 1, nº 1, pp. 81-106.
- QUINLAN, J.R. (1988): "Decision trees and multivalued attributes", *Machine Intelligence*, 11, pp. 305-318.
- QUINLAN, J.R. (1993): *C4.5: Programs for machine learning*. Morgan Kaufmann, San Mateo, California.
- R DEVELOPMENT CORE TEAM (2004): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. (URL <http://www.R-project.org>).
- REZA, F.M. (1994): *An introduction to Information Theory*, Dover Publications, New York.

- RISSANEN, J. (1983): "A universal prior for integers and estimation by minimum description length", *Annals of Statistics*, vol. 11, nº 2, pp. 416-431.
- SANCHÍS ARELLANO, A.; GIL, J.A. y HERAS MARTÍNEZ, A. (2003): "El Análisis Discriminante en la previsión de la insolvencia en las empresas de seguros de no vida", *Revista Española de Financiación y Contabilidad*, vol. 32, nº 116, pp. 183-233.
- SEGOVIA VARGAS, M.J. (2003): *Predicción de crisis empresariales en seguros no vida mediante la metodología Rough Set*. Tesis Doctoral, Universidad Complutense de Madrid.
- SEGOVIA VARGAS, M.J.; SALCEDO SANZ, S. y BOUSOÑO CALZÓN, C. (2004): "Prediction of Insolvency in Non-life Insurance Companies using Support Vector Machines, Genetic Algorithms and Simulated Annealing", *Fuzzy Economic Review*, vol. 9, nº 1, pp. 79-94.
- SETHI, I.K. y SARVARAYUDU, G.P.R. (1982): "Hierarchical classifier design using mutual information", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4, pp. 441-445.
- TROYANSKAYA, O.; CANTOR, M.; SHERLOCK, G.; BROWN, P.; HASTIE, T.; TIBSHIRANI, R.; BOTSTEIN, D. y ALTMAN, R.B. (2001): "Missing value estimation methods for DNA microarrays", *Bioinformatics*, vol. 17, nº 6, pp. 520-525.