

Modelo de predicción de insolvencias en empresas de seguros en base a Solvencia II. El algoritmo de inducción de árboles de decisión See5

ZULEYKA DÍAZ MARTÍNEZ

UNIVERSIDAD COMPLUTENSE DE MADRID

En el marco del proyecto comunitario denominado *Solvencia II*, el desarrollo y aplicación de nuevos criterios que permitan un aprovechamiento más eficiente de la información financiero-contable suministrada por las entidades aseguradoras se configura como una cuestión de carácter central.

De acuerdo con ello, el presente trabajo pretende examinar la aplicabilidad a la predicción de la insolvencia en el sector del seguro del algoritmo See5, técnica procedente del campo de la Inteligencia Artificial, tomando como información de partida un conjunto de ratios financieros obtenidos a partir de los estados contables de una muestra de empresas españolas de seguros no-vida.

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, ya que, debido al gran número de agentes e intereses afectados cuando se produce una insolvencia, la detección precoz de la misma o de las condiciones que pueden llevar a que ésta acaezca es una cuestión de suma importancia e interés general.

Las entidades aseguradoras han sido objeto desde hace tiempo de regulación y control con el propósito de comprobar que mantienen una situación de solvencia suficiente para cumplir su objeto social. Ahora bien, siendo la protección de asegura-

dos y beneficiarios amparados por un contrato de seguro la principal finalidad de la ordenación y supervisión de la actividad aseguradora, no puede el legislador dejar de considerar la enorme importancia del sector del seguro para el desarrollo de la economía en general, importancia que hace aconsejable, incluso imprescindible, una regulación y control estrictos que permitan evitar eventuales insolvencias que de producirse podrían ocasionar una pérdida de confianza en el funcionamiento del sector, de consecuencias nefastas para el conjunto de la actividad económica.

En la actualidad el marco legislativo en materia de solvencia para entidades aseguradoras está siendo objeto de un profundo estudio. El proyecto Solvencia II pretende establecer las directrices que reformen las reglas existentes en la Unión Europea en relación con la solvencia.

Previamente al proyecto *Solvencia II*, ha habido otros proyectos en el ámbito europeo relacionados con este tema. Tras la finalización en el año 1997 del denominado *Informe Müller* comienzan a concretarse las reflexiones sobre la solvencia de empresas de seguros. Sobre la base de este informe, que consideró el sistema de solvencia europeo globalmente satisfactorio, preconizando al mismo tiempo una serie de mejoras, se propuso, en primer lugar, la introducción de modificaciones en las normas sobre margen de solvencia de las Directivas existentes (proyecto *Solvencia I*), y, en segundo lugar, una revisión más profunda de todos los aspectos relacionados con la solvencia de las aseguradoras (proyecto *Solvencia II*).

El proyecto *Solvencia I* finalizó con la aprobación de dos Directivas, Directiva 2002/12/CE y Directiva 2002/13/CE, de 5 de marzo de 2002, que modifican, respectivamente, las primeras Directivas de vida y no-vida en lo que respecta a los requisitos del margen de solvencia para este tipo de entidades.

Por su parte, la problemática del proyecto *Solvencia II* se ha agrandado con el paso del tiempo. Este proyecto se ha dividido en dos fases, la primera de las cuales, cuyo objeto es determinar la forma general de un futuro sistema de solvencia de la UE, ha llegado a su fin. Para ello, los Estados miembros y la Comisión Europea han estudiado varias áreas, como el uso de sistemas de tipo «risk based capital» (RBC) (capital basado en riesgo), la experiencia obtenida del proceso de Basilea sobre el sistema bancario, el uso de modelos internos, los vínculos entre la presentación de la información financiera y las cuentas elaboradas a efectos de la supervisión, etc. Entre los trabajos realizados en esta primera fase, cabe destacar dos informes generales:

- El informe encargado por la Comisión Europea a la firma consultora KPMG, que fue

presentado en mayo de 2002 y lleva por título «Study into the methodologies to assess the overall financial position of an insurance undertaking from the perspective of prudential supervision» (Estudio de las metodologías para valorar la situación financiera global de una empresa aseguradora desde la perspectiva de una supervisión prudencial).

- El informe redactado por un grupo de trabajo de la Conferencia de Autoridades Supervisoras de Seguros de los Estados Miembros de la Unión Europea titulado «Prudential Supervision of Insurance Undertakings» (Supervisión Prudencial de las Empresas Aseguradoras), publicado en diciembre de 2002.

Asimismo, el Comité de Seguros de la Unión Europea realizó numerosos debates tomando como base las notas preparadas por la Comisión, y dedicó una de sus reuniones a las exposiciones de los Estados miembros sobre aspectos de su sistema prudencial o proyectos de reforma en curso. De manera paralela a estos debates, el Comité creó dos Grupos de Trabajo compuestos por expertos de varios Estados miembros y por un representante del Grupo Consultivo Actuarial Europeo, destinados a estudiar cuestiones más técnicas, un grupo para el seguro no-vida y otro para el seguro de vida.

Además, se invitó tanto a los Estados miembros como a los distintos participantes en el mercado (representantes de las compañías de seguros, ac-

tuarios, asesores) a formular su visión de la forma general del futuro sistema de solvencia.

Una vez definido el marco general, se han desarrollado actualmente los trabajos de la segunda fase, destinados a definir los detalles del sistema.

Los objetivos principales del proyecto *Solvencia II* son:

Además, se invitó tanto a los Estados miembros como a los distintos participantes en el mercado (representantes de las compañías de seguros, actuarios, asesores) a formular su visión de la forma general del futuro sistema de solvencia.

- El desarrollo y establecimiento de un nuevo sistema que permita determinar los recursos propios mínimos a requerir a cada aseguradora en función de los riesgos asumidos y la gestión que se realice de cada uno de ellos.
- El establecimiento de nuevas competencias y mecanismos de actuación de los supervisores.
- El establecimiento de la información que las entidades deberían proporcionar en relación, principalmente, a su política de gestión de riesgos, con el objeto de potenciar la disciplina de mercado.

Estos objetivos se integran en los denominados Tres Pilares, que son coincidentes con los acordados en el proyecto desarrollado, también en el seno de la Unión Europea, para el sector bancario (proyecto denominado *Basilea II*):

Pilar I: Exigencia de Recursos Propios.

Pilar II: Procesos de Supervisión.

Pilar III: Disciplina de mercado.

Por lo tanto, en línea con estos objetivos, desarrollar e implementar nuevos métodos que permitan analizar y/o anticipar el riesgo de insolvencia de una entidad aseguradora es una cuestión de gran actualidad e importancia, y a esta finalidad se orienta nuestra investigación, que trata de comprobar la utilidad de nuevas técnicas como herramientas para la toma de decisiones en este sector. El establecimiento de sistemas de «alerta temprana» de insolvencias supondría un ahorro en los costes, en tiempo y dinero, que conlleva la actividad supervisora. Así, se podrían salvar algunas de las limitaciones de la inspección material utilizando estos métodos para la preselección de empresas a investigar intensivamente, utilizando eficientemente los recursos limitados de la inspección. Además, se podría profundizar en la combinación óptima de regulación e inspección: orientando la inspección hacia las empresas pre-

seleccionadas, cabría pensar en una normativa menos restrictiva.

LAS TÉCNICAS DE CLASIFICACIÓN

Evidenciada la importancia y actualidad de la cuestión que vamos a abordar, trataremos de justificar el planteamiento de nuestra investigación.

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, extraídos en su mayor parte de los estados contables de la misma, lo que pretende el analista es tomar la decisión de clasificar a esa empresa dentro de una clase concreta de entre varias posibles.

Al no existir una teoría comúnmente aceptada que explique el fenómeno del fracaso empresarial, no es posible establecer a priori qué variables financieras ni qué valores en las mismas determinan la futura

solvencia o insolvencia de una firma. Por tanto, el estudio de la solvencia implica una investigación selectiva dentro de un conjunto inmenso de alternativas, debiendo estar basada dicha selección en reglas prácticas o heurísticas, fijando un criterio de suficiencia para determinar cuándo las soluciones encontradas son satisfactorias.

En el sector de seguros español, este análisis *heurístico* se ha implementado tradicionalmente a través de la aplicación de técnicas estadísticas de análisis multivariante tales como el análisis discriminante o los modelos de variable de respuesta cualitativa (logit, probit, etc.), que utilizan ratios financieros como variables explicativas. Cabe destacar, entre otros, los trabajos realizados por López *et al.* (1994), Mora (1994), Martín *et al.* (1999) y Sanchís *et al.* (2003), donde se pone de

El establecimiento de sistemas de «alerta temprana» de insolvencias supondría un ahorro en los costes, en tiempo y dinero, que conlleva la actividad supervisora.

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.

En un intento de superar estas limitaciones, surge el empleo de técnicas procedentes del campo de la Inteligencia Artificial, ya que, debido a su 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 de *Machine Learning* (Aprendizaje Automático). Un representante típico de esta categoría son las redes neuronales, de las que ya se han desarrollado un gran número de aplicaciones satisfactorias, algunas de las cuales están siendo utilizadas en el «mundo real» dentro de sistemas de gestión del riesgo de crédito. En cuanto a su aplicación a la predicción de crisis en el sector de seguros español. Martínez de Lejarza (1996) ha obtenido en su trabajo resultados claramente superiores a los del Análisis Discriminante. También el enfoque de los Conjuntos Aproximados «Rough Sets» en terminología anglosajona - pertenece a esta categoría. Una aplicación al sector de seguros español de la *Teoría Rough Set* (Pawlak, 1991) la podemos encontrar en el trabajo de Segovia (2003), donde se alcanzan resultados muy satisfactorios y similares a los del Análisis Discriminante, pero evitando con la aplicación del Rough Set los inconvenientes derivados del incumplimiento de las hipótesis requeridas por el Análisis Discriminante sobre las variables de partida.

Aplicación al sector asegurador del algoritmo See5 a fin de obtener un modelo de predicción de insolvencias en dicho sector basándonos en ratios financieros.

Otro gran bloque de procedimientos de Aprendizaje Automático es el constituido por los distintos sistemas de inducción de reglas y árboles de decisión. Dado que su desarrollo es más reciente que el de las redes neuronales, no se encuentran tan implantados en el mundo profesional como aquéllas, pero sí han despertado un gran interés en el mundo académico. 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 de clasificación 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 1D3, C4.5 y See5 desarrollados por Quinlan (1979, 1983, 1986, 1988, 1993 y 1997), que han alcanzado una notable repercusión. Algunas aplicaciones de dichos algoritmos a la predicción de crisis empresariales son las de, entre otros, McKee (1995) y González *et al.* (1999). En el sector de seguros español no se han realizado aún estudios aplicando este tipo de metodología, carencia que se pretende paliar con esta investigación, que consistirá, básicamente, en la aplicación al sector asegurador del algoritmo See a fin de obtener un modelo de predicción de insolvencias en dicho sector basándonos en ratios financieros.

Es necesario señalar que nuestro objetivo no consiste en sustituir la capacidad de juicio y experiencia del decisor humano en la evaluación de la solvencia, pero sí complementarla y añadir una sistemática o procedimientos objetivos comunes de análisis.

Como ya se ha mencionado, intentamos proveer de herramientas útiles para la toma de decisiones de los distintos usuarios de la información contable, pudiendo utilizarse estos modelos, en términos prácticos, como sistemas de «alerta temprana» por aquellos interesados en prevenir la crisis empresarial.

EL ALGORITMO DE INDUCCIÓN DE ÁRBOLES DE DECISIÓN See5

El algoritmo See5⁴ permite construir automáticamente a partir de un conjunto de datos de ejemplo o *entrenamiento* un árbol de clasificación que representa la relación que existe entre la decisión y sus atributos o variables. Para inferir el árbol, el algoritmo realiza particiones binarias sucesivas en el espacio de las variables explicativas, de forma que para realizar cada partición se escoge la variable que aporta más información en función de una medida de *entropía* o cantidad de información. Bonsón et al. (1999, p.117), indican que «la misión del algoritmo es la elaboración de un árbol de decisión bajo las siguientes premisas:

- 1) Cada nodo corresponde a un atributo y cada rama al valor posible de ese atributo. Una hoja del árbol especifica el valor esperado de la decisión de acuerdo con los ejemplos dados. La explicación de una determinada decisión viene dada por la trayectoria desde la raíz a la hoja representativa de esa decisión.
- 2) A cada nodo se le asocia aquel atributo más informativo que aún no haya sido considerado en la trayectoria desde la raíz.
- 3) Para medir el nivel informativo de un atributo se emplea el concepto de *entropía*. Cuanto menor sea el valor de la entropía, menor será la incertidumbre y más útil será el atributo para la clasificación».

Como se explica en Quinlan (1993, pp. 20-24), en las versiones iniciales de este algoritmo se usaba un criterio denominado *gain* (ganancia) para elegir el atributo (variable)

en base al cual hacer cada partición de las que forman el árbol. Muy brevemente, la idea es la siguiente: la información que proporciona un mensaje o la realización de una variable aleatoria depende de su probabilidad y puede ser medida en bits como menos el logaritmo en base 2 de esa probabilidad. Por ejemplo, si tenemos 8 mensajes equiprobables, la información contenida en cualquiera de ellos es $-\log_2\left(\frac{1}{8}\right)$ ó 3 bits.

Imaginemos que seleccionamos aleatoriamente un caso de un conjunto T de casos, y que éste resulta pertenecer a una clase C_j . La probabilidad del mensaje que nos indica la clase a la

que pertenece el elemento es $\frac{freq(C_j, T)}{|T|}$ y la información que proporciona dicho mensaje es $-\log_2\left(\frac{freq(C_j, T)}{|T|}\right)$ bits, donde $freq(C_j, T)$ repre-

senta el número de casos u observaciones en el conjunto T que pertenecen a la clase C_j (es decir, la frecuencia de C_j en T) y $|T|$ denota el número de casos u observaciones que contiene el conjunto T . Por tanto, tomando la esperanza matemática

de la cantidad $-\log_2\left(\frac{freq(C_j, T)}{|T|}\right)$ obtenemos la

cantidad media de información necesaria para identificar la clase, de entre k clases posibles, a la que pertenece un caso en el conjunto T (la denominada entropía del conjunto T). Esa cantidad media viene dada por la expresión:

$$info(T) = -\sum_{j=1}^k \frac{freq(C_j, T)}{|T|} \times \log_2\left(\frac{freq(C_j, T)}{|T|}\right) \text{ bits}$$

Si ahora conocemos el valor que toma un determinado atributo X (por ejemplo, un determinado ratio financiero) para cada elemento del conjunto T , entonces para clasificar cada elemento necesitaremos una cantidad de información menor que $info(T)$

⁴ Este algoritmo constituye una extensión de los algoritmos 1D3 y C4.5. Una descripción detallada puede verse en Quinlan (1993 y 1997).

(pues ya tenemos la información suministrada por el ratio anterior). Esta cantidad vendrá dada por:

$$info_x(T) = \sum_{i=1}^n \frac{|T_i|}{|T|} \times info(T_i)$$

donde T_i es cada una de las particiones hechas en el conjunto T de acuerdo con los distintos valores que tome el atributo X y $|T_i|$, el número de observaciones que contiene cada una de dichas particiones. La magnitud

$$gain(X) = info(T) - info_x(T)$$

mide la cantidad de información que se gana dividiendo el conjunto de datos T de acuerdo con el atributo X . Entonces, *el criterio de ganancia* selecciona para hacer la partición aquel atributo para el cual se maximiza la ganancia de información.

Posteriormente, Quinlan observó que este criterio favorecía a aquellos atributos con un número mayor de valores posibles, con lo que, por ejemplo, no clasificaba correctamente los ejemplos dados cuando alguno de los atributos era una variable continua. Para evitar este sesgo que favorece a los atributos con muchos valores posibles tomó la magnitud:

$$split\ info(X) = -\sum_{i=1}^n \frac{|T_i|}{|T|} \times \log_2 \left(\frac{|T_i|}{|T|} \right)$$

que representa la *entropía* del conjunto T cuando es dividido de acuerdo con los valores que toma el atributo X . Esta *entropía* será tanto mayor cuanto más elevado sea el número de dichos valores. De este modo, puede ser utilizado como divisor de $gain(X)$ para corregir los elevados valores que esta magnitud tomará para aquellos atributos que adopten un mayor número de valores posibles. Entonces, el atributo elegido para la partición será aquél para el cual el ratio de ganancia sea mayor, definiéndose esta medida como:

$$gain\ ratio(X) = \frac{gain(X)}{split\ info(X)}$$

El árbol construido por aplicación reiterada de este criterio consta del mínimo número de atributos (variables) que se requieren para la clasificación eficiente de los ejemplos dados, con lo que

es claro el gran poder explicativo de esta técnica. También se pueden elaborar, a partir del árbol, una serie de reglas de clasificación más simples y fácilmente interpretables, que definen las características que más diferencian a las distintas clases establecidas inicialmente.

Este tipo de sistemas clasificadores presentan importantes ventajas para su aplicación al ámbito financiero. Bonsón *et al.* (1997, pp. 198-199) señalan algunas:

- Superan a las técnicas estadísticas tradicionales en el sentido de que tienen un 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.
- Superan a los sistemas expertos (técnica de Inteligencia Artificial muy utilizada por entidades financieras en el análisis del riesgo de crédito) en que no necesitan de la intervención del experto humano para la inferencia de las reglas clasificadoras, ya que éstas se elaboran automáticamente.
- Frente a las redes neuronales presentan la ventaja de que las reglas son mucho más comprensibles por el usuario que la topología de una red, puesto que una red neuronal, por sencilla que sea, es un modelo de «caja negra», que no permite valorar la importancia relativa de cada una de las variables explicativas.

ANÁLISIS EMPÍRICO

1. Metodología

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). Consta de 36 empresas no fracasadas y 36 empresas fracasadas (en adelante, «buenas» y «malas»), empareja-

das por tamaño y tipo de negocio, 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 fracasan. Los datos fueron extraídos de la publicación *anual Balances y cuentas. Seguros privados* de la Dirección General de Seguros. A partir del año 1992 estos datos están disponibles en soporte informático. Dado que en el 82 se introdujo una nueva estructura en las cuentas anuales, no se utilizó información anterior a esa fecha para que la misma fuese lo más homogénea posible. En consecuencia, tenemos una muestra que abarca datos desde el 83 al 96.

Una vez tomada la muestra, nos situamos en el periodo anterior 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. De cada una de las empresas se han obtenido las cuentas anuales del año previo 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.

En la **Tabla I** 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.

De las 72 empresas de que consta la muestra original, hemos utilizado únicamente 27 empresas de cada una de las submuestras para la elaboración de nuestro modelo, reservando las 9 restantes para poder comprobar la validez del mismo aplicándolo a empresas cuyos datos no hubieran sido utilizados en dicha elaboración. En consecuencia, tendremos una muestra de *entrenamiento* para obtener el árbol y reglas de decisión formada por 54 empresas y una muestra de *validación* para verificar su capacidad predictiva formada por 18 empresas. La selección de la muestra de validación se ha efectuado aleatoriamente, tomando las empresas numeradas en Segovia (2003, pp. 162-164) de la 19 a la 27 para las «malas» y de la 119 a la 127 para las «buenas» como submuestras de validación.

El algoritmo se aplicó mediante el software *See5* de *Rulequest Research*.

Consta de 36 empresas no fracasadas y 36 empresas fracasadas (en adelante, «buenas» y «malas»), emparejadas por tamaño y tipo de negocio, Los datos fueron extraídos de la publicación *anual Balances y cuentas. Seguros privados* de la Dirección General de Seguros España.

De las 72 empresas de que consta la muestra original, hemos utilizado únicamente 27 empresas de cada una de las submuestras para la elaboración de nuestro modelo.

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.

TABLA 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

2. Resultados.

A continuación se presenta el árbol de decisión obtenido aplicando el algoritmo See5 a nuestra muestra, así como la evaluación del mismo tanto para la muestra utilizada en su elaboración como para la muestra de validación.

Como se puede observar, en el árbol aparecen únicamente 6 de los 19 ratios iniciales, lo que indica que 13 de los ratios empleados no aportan información relevante para clasificar las empresas como «buenas» o «malas». Como

```

R13 > 0.68:
...R9 <= 0.59: mala (14)
: R9 > 0.59:
: ...R17 <= 0.99: mala (3)
: R17 > 0.99: buena (3)
R13 <= 0.68:
...R1 > 0.29: buena (20/2)
R1 <= 0.29:
...R2 > 0.04: mala (3)
R2 <= 0.04:
...R6 > 0.64: buena (3)
R6 <= 0.64:
...R9 <= 0.85: mala (4)
R9 > 0.85: buena (4/1)
    
```

Evaluation on training data (54 cases):

```

Decision Tree
-----
Size      Errors
   8      3(5.6%)  «

(a)  (b)  <-classified as
-----
  27  ---  (a) : class buena
   3  24  (b) : class mala
    
```

Evaluation on test data (18 cases):

```

Decision Tree
-----
Size      Errors
   8      5(27.8%)  <<

(a)  (b)  <-classified as
-----
   7   2  (a): class buena
   3   6  (b): class mala
    
```

ya se ha mencionado, el árbol nos proporciona el menor número de atributos (ratios) necesarios para alcanzar el objetivo deseado. Nuestro árbol se leería del modo siguiente:

- Si el ratio R13 es mayor de 0,68 y además el ratio R9 es menor o igual de 0,59, la empresa será «mala», siendo 14 el número de empresas de la muestra que verifican este hecho.
- Si el ratio R13 es mayor de 0,68 y el ratio R9 es mayor de 0,59 y el ratio R17 menor o igual de 0,99, la empresa será «mala», cumpliendo estas condiciones 3 empresas.

Y así continuaríamos descendiendo por el árbol, hasta completar un total de 8 hojas. Obsérvese que 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.

La evaluación de este árbol de decisión construido con la muestra de entrenamiento (54 empresas) indica que el árbol consta de 8 ramas y comete un total de 3 errores (5,6%), lo que supone un porcentaje de aciertos del 94,4%.

También se muestra una *matriz de confusión* que señala el tipo de errores cometidos.

Por último, para comprobar la capacidad predictiva del árbol, se clasifican de acuerdo con éste las 18 empresas de la muestra de validación, obteniendo un porcentaje de clasificaciones correctas del 72,2%.

Aunque no en este caso, en ocasiones puede resultar difícil interpretar un árbol de decisión. Se permite solventar este problema derivando, a partir del árbol, un conjunto de reglas más simples de la forma *si* (condiciones) - *entonces* (decisión). Las reglas que se obtienen a partir del árbol anterior son las siguientes:

```

Rules:

Rule 1: (20/2, lift 1.7)
  R1 > 0.29
  R13 <= 0.68
  -> class buena [0.864]

Rule 2: (12/1, lift 1.7)
  R2 <= 0.04
  R6 > 0.64
  R13 <= 0.68
  -> class buena [0.857]

Rule 3: (7/1, lift 1.6)
  R9 > 0.85
  -> class buena [0.778]

Rule 4: (14, lift 1.9)
  R9 <= 0.59
  R13 > 0.68
  -> class mala [0.938]

Rule 5: (7, lift 1.8)
  R13 > 0.68
  R17 <= 0.99
  -> class mala [0.889]

Rule 6: (26/6, lift 1.5)
  R1 <= 0.29
  -> class mala [0.750]

Default class: buena
    
```

Cada regla consiste en:

- Una serie de estadísticas (un número n o n/m , seguido de otro número x); n y m representan lo mismo que en el árbol y x es el resultado de dividir la precisión estimada de la regla entre

la frecuencia relativa de la clase predicha en la muestra de entrenamiento. La precisión de la regla se estima mediante el denominado ratio de Laplace $(n - m + 1)/(n + 2)$ (Clark y Boswell, 1991 y Niblett, 1987).

- Una o más condiciones que deben ser satisfechas para que la regla sea aplicable.
- La clase predicha por la regla.
- Un valor entre 0 y 1 que indica el nivel de confianza con el que ha sido hecha la predicción.

También existe una clase por defecto (en este caso «buena») para cuando ninguna de las reglas sea aplicable.

El número de errores cometidos al clasificar mediante estas reglas y el tipo de los mismos coinciden, tanto con la muestra de entrenamiento como con la de validación, con los de las clasificaciones hechas con el árbol (aunque no siempre tiene por qué ser así).

A la luz de estos resultados, aquellas partes interesadas en evaluar la solvencia de una firma de seguros no-vida deberían tener en cuenta principalmente las siguientes cuestiones:

- La **liquidez**, medida a través del **ratio RI**. Si bien una de las cuestiones más importantes para asegurar el buen funcionamiento de cualquier tipo de empresa es la necesidad de una adecuada liquidez que le permita hacer frente a sus deudas a corto plazo sin tener que recurrir a la realización de sus activos, en el caso de la empresa de seguros dicha necesidad reviste una mayor importancia debido a la inversión del proceso productivo que se da en este tipo de entidades. Por otro lado, este ratio es también una medida del equilibrio financiero: si el ratio es positivo, implica que el capital circulante también lo es.
- La **rentabilidad de los recursos propios**, medida a través del **ratio R2**. Este ratio informa acerca de la eficiencia de la empresa en

la utilización de los fondos aportados por los accionistas.

- La **solvencia** en sentido estricto, medida a través de los ratios **R6** y **R13**. Estos ratios recogen en sus numeradores la medida del riesgo anual de la compañía, ya sea basándose en la cuantía de las primas destinadas a cubrir los riesgos asegurados en el año (R6) o basándose en la valoración de los riesgos que han ocurrido en el ejercicio (siniestros) y que se recogen como Gastos Técnicos en la cuenta de Pérdidas y Ganancias (R13). Los denominadores muestran el soporte financiero de las empresas. En R6 se contemplan

los capitales propios, lo que sugiere un soporte global del riesgo independiente de las características anuales, mientras que en R13 se suman a los capitales propios las provisiones técnicas, logrando

así el soporte financiero real para el periodo analizado.

- La **autonomía financiera**, medida a través del ratio **R9**. Este ratio muestra qué proporción de las obligaciones totales de la empresa representan los fondos propios (capital y reservas). Esto confirma la importancia, desde el punto de vista de la solvencia, de la adecuación de estos fondos porque puedan ser necesarios para hacer frente en última instancia a una elevada siniestralidad no prevista.
- La **tarificación correcta**, medida a través del ratio **R17**. Éste es uno de los ratios utilizados en Estados Unidos por la NAIC (National Association Insurance Commissioners) en su sistema de alarma basado en ratios financieros IRIS (Insurance Regulatory Information System) para detectar qué compañías aseguradoras pueden necesitar atención reguladora (este ratio es denominado por la NAIC *Expense Ratio* -ratio de gastos-. Informa de la capacidad de afrontar los gastos anuales con las primas anuales correspondientes, indicando si la

A la luz de estos resultados, aquellas partes interesadas en evaluar la solvencia de una firma de seguros No-Vida deberían tener en cuenta principalmente diversas cuestiones:

ESTUDIO

firma está llevando a cabo un proceso adecuado de tarificación que le permita hacer frente con los ingresos por primas a los gastos correspondientes.

A pesar de que los resultados que hemos obtenido son satisfactorios, es posible mejorarlos recurriendo a la opción que incorpora *See5* de *adaptive boosting*, basado en el trabajo de Freund y Schapire (1997). Muy brevemente, la idea consiste en generar varios clasificadores (árboles o conjuntos de reglas) en vez de sólo uno. Como primer paso, se construye un único árbol (o conjunto de reglas) del mismo modo que acabamos de ver, que cometerá algunos errores en la clasificación (3 en nuestro caso). Estos errores serán el foco de atención al construir el segundo clasificador en aras de corregirlos. En consecuencia, el segundo clasificador generalmente será diferente al primero y también cometerá errores que serán el foco de atención durante la construcción del tercer clasificador. Este proceso continúa para un número predeterminado de iteraciones o *trials*. Finalmente, cada caso será clasificado en la categoría a la que haya sido asignado por la mayoría de los clasificadores construidos. Los resultados obtenidos con este método son con frecuencia excepcionalmente buenos. Así, en nuestro caso, partiendo del primer árbol de decisión visto anteriormente los resultados que alcanzamos realizando 18 iteraciones son:

- Con la muestra de entrenamiento, el 100% de clasificaciones correctas, como podemos observar en la *matriz de confusión*:

(a)	(b)	<-classified as
----	----	
27	27	(a): class buena
		(b): class mala

- Con la muestra de validación, el 83,3% de clasificaciones correctas:

(a)	(b)	<-classified as
----	----	
7	2	(a): class buena
1	8	(b): class mala

En la siguiente tabla se exponen los porcentajes de clasificaciones correctas obtenidas en este último caso desagregadas por clase:

Clasificaciones correctas	Muestra de entrenamiento	Muestra de validación
Empresas "buenas"	100%	77,78%
Empresas "malas"	100%	88,89%
Total	100%	83,33%

Tal y como podemos observar, obtenemos un porcentaje de aciertos global elevado con la muestra de validación (83,33%), clasificando además correctamente el 88,89% de las empresas «malas», lo cual es importante teniendo en cuenta que precisamente lo que nos interesa captar es la insolvencia.

CONCLUSIONES

Los resultados que hemos obtenido en este estudio muestran la capacidad del algoritmo *See5* para responder de manera eficiente al problema de la predicción del fracaso empresarial, siendo una alternativa muy fiable a las técnicas estadísticas convencionales. Esta metodología acepta variables tanto de tipo cuantitativo como cualitativo, sin ninguna clase de limitación, así como la existencia de *missing values* (valores perdidos). Su carácter estrictamente no paramétrico le confiere una clara superioridad frente a las técnicas estadísticas en el sentido de que se adecúa más a la información contable, que suele presentar datos interrelacionados, incompletos, adulterados o

erróneos; proporciona modelos muy sencillos entendibles fácilmente por el analista humano, ya sea en forma de árboles o reglas de decisión, realizando una clasificación de las empresas entre solventes e insolventes que permite determinar la importancia de cada variable en el proceso de asignación. Además, da buenos resultados, incluso cuando se trabaja con escaso número de datos, aspecto este importante en las aplicaciones al ámbito financiero. Asimismo, aunque no se ha visto aquí, See5 posee una ventaja muy importante desde el punto de vista económico, ya que permite considerar distintos costes de clasificación errónea, distinguiendo si se trata de clasificar una empresa sana como fracasada o clasificar una fracasada como sana, error este último que resultaría mucho más grave.

Aunque los resultados de la validación externa indican que el modelo puede considerarse válido, no debemos olvidar alguna de sus limitaciones a la hora de utilizarlo con fines predictivos, como son el hecho de que la muestra sea relativamente pequeña o el que el análisis se haya realizado al margen de factores, tales como tamaño y tipo de negocio, que pueden ser importantes a la hora de predecir la insolvencia.

Además, es conveniente encontrar la manera de examinar la situación de una empresa con información no sólo del periodo anterior al que pretendemos evaluar, sino con varios periodos de desfase. Para comprobar el poder explicativo de los ratios con distintos desfases respecto al periodo que pretendemos examinar, se podrían elaborar diferentes modelos para cada uno de los años previos a la crisis. Asimismo, si nuestro objetivo es diseñar un modelo que pueda predecir la crisis *a priori*, para paliar el problema derivado del desconocimiento del año concreto en que ésta se producirá, sería conveniente elaborar un modelo teniendo en cuenta los valores de los ratios de varios años consecutivos antes de la crisis tomados conjuntamente.

Los resultados que hemos obtenido en este estudio muestran la capacidad de! algoritmo See5 para responder de manera eficiente al problema de la predicción del fracaso empresarial, siendo una alternativa muy fiable a las técnicas estadístico convencionales.

BIBLIOGRAFÍA

- BONSÓN, E., ESCOBAR, T. y MARTÍN, M. P. (1997). «Decision tree induction systems. Applications in Accounting and Finance», en Sierra, G. y Bonsón, E. (Eds.): *Intelligent Technologies in Accounting and Business*, Huelva, pp. 191-211.
- BONSÓN, E., ESCOBAR, T. y MARTÍN, M. P. (1999). «Aplicación de los sistemas de inducción de árboles de decisión a la gestión empresarial: toma de decisiones y control de tareas de decisión», en Bonsón, E. (Ed.): *Tecnologías Inteligentes para la Gestión Empresarial*, RA-MA Editorial, Madrid, pp. 115-132.
- BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. y STONE, C. J. (1984). *Classification and regression trees*. Wadsworth, Belmont.
- CLARK, P. Y BOS WELL, R. (1991). «Rule Induction with CN2: Some Recent Improvements», en Kodratoff, Y. (Ed.): *Machine Learning- Proceedings of the Fifth European Conference (EWSL-91)*, Springer-Verlag, Berlin, pp. 151-163.
- DÍAZ, Z., FERNÁNDEZ, J. y SEGOVIA, M. J. (2004). «Sistemas de inducción de reglas y árboles de decisión aplicados a la predicción de insolvencias en empresas aseguradoras». Comunicación aceptada en las XII Jornadas ASEPUMA, Murcia.
- EUROPEAN COMMISSION (1999). «The Review of the overall Financial Position of an Insurance Undertaking - Solvency II Review». Working paper. MARKT/2095/99 (en http://europa.eu.int/comm/internal_market/insurance/solvency_en.htm).
- EUROPEAN COMMISSION (2001). «Solvency II - Presentation of the proposed work». Working paper, MARKT/2027/0 I (en http://europa.eu.int/comm/internal_market/insurance/solvency_en.htm).
- EUROPEAN COMMISSION (2001). «Banking rules - Relevance for the Insurance Sector?». Working

- paper, MARKT/2056/0 1 (en http://europa.eu.mt/comm/internalmarket/insurance/solvency_en.htm).
- EUROPEAN COMMISSION (2001). «Risk-based Capital Systems». Working paper, MARK1 01 len http://europa.eu.int/comm/internalmarket/insurance/solvency_en.htm.
- EUROPEAN COMMISSION (2002). «Considerations on the links between financial statements and supervisory returns of insurance undertakings». Working paper, MARKT/25 14/02 (en http://europa.eu.int/comm/internal_market/insurance/solvency_en.htm).
- EUROPEAN COMMISSION (2002). «Risks models of insurance companies or groups». Working paper, MARKT/25 15/02 (en http://europa.eu.int/comm/internalmarket/insurance/solvency_en.htm).
- EUROPEAN COMMISSION (2002). «Current and future solvency work in the IAIS and within the actuarial profession from a Solvency II point of view». Working paper, MARKT/2520/02 (en http://europa.eu.int/comm/internalmarket/Insurance/solvency_en.htm).
- EUROPEAN COMMISSION (2002). «Report of the working group on life assurance to the IC Solvency Subcommittee». Working paper, MARKT/2528/02 (en http://europa.eu.int/comm/internal_market/insurance/solvency_en.htm).
- EUROPEAN COMMISSION (2002). «Report of the working group on non-life technical provisions to the IC Solvency Subcommittee». Working paper, MARKT/2 529/02 (en http://europa.eu.int/comm/internal_markeiiinsurance/solvency_en.htm).
- EUROPEAN COMMISSION (2002). «Solvency II: Review of work». Working paper, MARKT/25 36/02 (en http://europa.eu.int/comm/internal market/insurance/solvency_en.htm).
- EUROPEAN COMMISSION (2002). «Considerations on the design of a future prudential supervisory system». Working paper, MARKT/2535/02(en http://europa.eu.mt/comm/internal market/insurance/solvency_en.htm).
- EUROPEAN COMMISSION (2003). «Design of a future prudential supervisory system in the EU - Recommendations by the Commission Services». Working paper, MARKT/2509/03 (en http://europa.eu.int/comm/internalmarket/insurance/solvency_en.htm).
- FREUND, Y. Y SCHAPIRE, R. E. (1997). «A decision-theoretic generalization of on-line learning and an application to boosting». *Journal of Computer and System Sciences*, vol. 55(1), pp. 119-139.
- FRIEDMAN, J. H. (1977). «A recursive partitioning decision rule lbr non-parametric classification». *IEEE Transactions on Computers*, pp. 404-408.
- GONZÁLEZ, A. L., CORREA, A. y BLÁZQUEZ, J. A. (1999). «Perfil del fracase empresarial para una muestra de pequeñas y medianas empresas». Comunicación presentada al X Congreso AECA, Zaragoza.
- KPMG (2002): «Study into the methodologies to assess the overall financial position of an insurance undertaking from the perspective of prudential supervisión» (en http://europa.eu.int/comm/internal_market/insurance/solvency_en.htm).
- LÓPEZ, D., MORENO, J. y RODRÍGUEZ, P. (1994). «Modelos de previsión del fracaso empresarial: aplicación a entidades de seguros». *Revista española de Seguros*, 54, abril-junio, pp. 71-110.
- MARTÍN, M. L., LEGUEY, S. y 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, 49.
- MARTÍNEZ DE LEJARZA, I. (1996). «Forecasting company failure: neural approach versus discriminant analysis. An application to Spanish insurance companies of the 80's» en Sierra, G. y Bonsón, E. (Eds.): *Intelligent Systems in Accounting and Finance*, Huelva, pp. 169-186.
- McKEE, T. E. (1995). «Predicting bankruptcy via induction». *Journal of Information Technology*, Vol. 10, pp. 26-36.
- MORA, A. (1994). «Los modelos de predicción del fracaso empresarial: una aplicación empírica

- del logit». *Revista Española de Financiación y Contabilidad*, 78, enero-marzo, pp. 203-233.
- MULLER GROUP (1997): Müller Group Report. 1997. Solvency of insurance undertakings, Conference of Insurance Supervisory Authorities of The Member States of The European Union.
- NIBLETT, T. (1987). «Constructing decision trees in noisy domains», en Bratko, I. y Lavra, N. (Eds.): *Progress in Machine Learning (proceedings of the 2nd European Working Session on Learning)*, Sigma, Wilmslow, UK, pp. 67-78.
- PAWLAK, Z. (1991). *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, Boston, London.
- QUINLAN, J. R. (1979). «Discovering rules by induction from large collections of examples», en Michie, D. (Ed.): *Expert systems in the microelectronic age*, Edimburgh University Press, Edimburgh.
- QUINLAN, J. R. (1983). «Learning efficient classification procedures», en *Machine learning: an Artificial Intelligence approach*, Tioga Press, Palo Alto.
- QUINLAN, J. R. (1986). «Induction of decision trees». *Machine Learning*, Vol. 1. n° 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 Publishers, Inc., California.
- QUINLAN, J. R. (1997). See5 (en Internet, <http://www.rulequest.com/see5-info.html>).
- SANCHÍS, 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.
- SEGOVIA, M. J. (2003). *Predicción de crisis empresariales en seguros no vida mediante la metodología Rough Set*, Tesis Doctoral, Universidad Complutense de Madrid.
- SHARMA GROUP (2002): *Sharma Group Report. 2002. Prudential Supervision of Insurance Undertakings*, Conference of Insurance Supervisory Authorities of The Member States of The European Union (en <http://europa.eu.int/comm/internalmarket/insurance/solvencyen.htm>).