

Predicción de insolvencias con el método Rough Set

M.J. Segovia Vargas, J.A. Gil Fana, A. Heras Martínez, J.L. Vilar Zanón*

Dpto. de Economía Financiera y Contabilidad I, Facultad de C. Económicas y Empresariales, Universidad Complutense de Madrid, Campus de Somosaguas, 28223 Madrid. Tfno. 913942564. Fax 913942570. E-mail mjsegovia@ccee.ucm.es.

Resumen

La detección precoz de la insolvencia de una empresa interesa tanto para proteger al público en general como para minimizar los costes económicos y sociales asociados a este problema.

Se han aplicado numerosos métodos estadísticos para afrontar este problema utilizando como variables explicativas los ratios financieros. Estas variables no suelen cumplir las hipótesis estadísticas que requieren estos métodos. En consecuencia, hemos aplicado la metodología Rough Set para la predicción de la insolvencia sobre una muestra de empresas españolas.

Esta metodología presenta, entre otras, estas ventajas: 1) es útil para analizar sistemas de información que representan el conocimiento adquirido por la experiencia, 2) elimina las variables redundantes reduciendo el coste, en tiempo y dinero, del proceso de decisión, 3) se obtienen unas reglas de decisión fácilmente comprensibles que no necesitan interpretación de ningún experto y, 4) las reglas están bien justificadas por extraerse de ejemplos reales lo que justificaría las decisiones que en base a ellas se tomen.

Los resultados muestran cómo esta metodología se adapta muy bien a los problemas de clasificación con atributos múltiples aplicándola a la detección de la insolvencia como problema de clasificación entre empresas sanas y fracasadas, y utilizando como atributos los ratios financieros.

Palabras clave: Fracaso empresarial, bancarrota, insolvencia, rough set

1.- INTRODUCCIÓN

Muchas decisiones financieras implican la clasificación de una observación (empresa, título de una cartera, ...) en una categoría o en un grupo, lo que ha motivado que muchos investigadores apliquen métodos de investigación operativa a los problemas financieros. Dentro de los problemas de clasificación nos encontramos el caso en que la clasificación es binaria, es decir cuando el número de grupos se limita a dos. Hay una gran cantidad de literatura financiera que se ha dedicado a este tipo de problemas de clasificación en dos grupos, entre los cuales destacamos: problemas de clasificación de créditos entre fallidos y no, fusiones y adquisiciones, clasificación de bonos, ..., o la predicción del fracaso empresarial.

Centrándonos en este último problema cabría preguntarse sobre por qué predecir una quiebra. La respuesta es clara, y es que a diferencia de otros problemas financieros, el

* Este trabajo ha sido en parte subvencionado por el proyecto de investigación BEC2001-1441.

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

Dentro de los posibles métodos aplicados a la predicción del fracaso empresarial hemos elegido un método multicriterio, y dentro de estos, la *Teoría Rough Set*. Es un enfoque que se encuadra dentro de las aplicaciones de la Inteligencia Artificial. Utiliza la experiencia de una forma objetiva, a través del estudio de la experiencia histórica de una manera cuantitativa, para así explicitar reglas que, resumiendo y objetivando esa experiencia acumulada, ayuden en las decisiones futuras.

El presente trabajo describe una investigación de carácter empírico consistente básicamente en la aplicación al sector de empresas de seguros no vida de la metodología Rough Set. Nuestro objetivo es comprobar su efectividad en los problemas de clasificación con atributos múltiples aplicándola a la detección de la insolvencia como problema de clasificación entre empresas sanas y fracasadas y utilizando como atributos los ratios financieros en este sector económico. Como resultado obtendremos un modelo de predicción de la solvencia empresarial tomando como referencia los datos contables.

El resto del trabajo se estructura en las siguientes secciones: en la sección 2 presentamos una breve descripción de los principales conceptos del método Rough Set, en la sección 3 describimos las variables que formarán parte del modelo, en la sección 4 se encuentran los resultados y, finalmente en la sección 5, explicamos las conclusiones.

El software utilizado es el programa ROSE facilitado por la Universidad de Poznan (Polonia) y desde aquí queremos expresar nuestro agradecimiento al Profesor Wilk por su inestimable colaboración. Queremos mencionar que cualquier ordenador personal conectado a internet puede acceder a la página web www-idss.cs.put.poznan.pl/rose. donde se puede descargar el programa ROSE, su manual e, informarse brevemente de algunas de sus características. Para una descripción más detallada de las posibilidades del programa nos remitimos a su manual y a la siguiente bibliografía Predki, B (1998, 1999).

2.- CONCEPTOS BÁSICOS DE LA TEORÍA ROUGH SET.

La teoría Rough Set fue originalmente desarrollada en los años ochenta por Z. Pawlak (Pawlak, 1991), entre otros, como herramienta formal para tratar con la incertidumbre o la vaguedad inherente a un proceso de decisión. Aunque existen en la actualidad extensiones de esta teoría nos referiremos al enfoque clásico.

Esta teoría implica el cálculo de particiones, divisiones o clases, según queramos. Es algo diferente tanto de la teoría estadística de la probabilidad como de la teoría *fuzzy set*. Se podría decir que hay tres categorías generales de imprecisión en el análisis científico. La primera categoría ocurre, cuando un acontecimiento es aleatorio por naturaleza y la imprecisión asociada con este tipo de acontecimientos puede describirse por la teoría estadística de la probabilidad. La segunda surge del hecho de que los objetos pueden no pertenecer exclusivamente a una única categoría sino que pueden pertenecer a varias

categorías aunque con diferentes grados; en este caso la imprecisión toma forma de pertenencia difusa a un conjunto y es el objeto de la lógica *fuzzy*. La tercera categoría es la teoría rough set que es útil cuando las clases en las que han de clasificarse los objetos son imprecisas, pero sin embargo pueden aproximarse mediante conjuntos precisos (Nurmi et al., 1996). Estas diferencias muestran una de las principales ventajas de la teoría rough set y es que no necesita ninguna información adicional acerca de los datos como puede ser una distribución de probabilidad en estadística o el grado o probabilidad de pertenencia en la teoría *fuzzy set*.

La filosofía del método se basa en la suposición de que a cada objeto del universo que estamos considerando se le puede asociar alguna información (datos, conocimiento). Los objetos caracterizados por la misma información no son discernibles a la vista de la información disponible. La relación de no-diferenciación generó de este modo la base matemática para esta teoría. Es, la información imprecisa, la causa de la no-diferenciación de los objetos en términos de datos disponibles y evita, en consecuencia, su asignación precisa a un conjunto. “Rough” podría traducirse por “vago, impreciso”; de aquí en adelante hablaremos de rough set. Por tanto, y de manera intuitiva, un rough set es un conjunto de objetos que, en general, no pueden ser caracterizados de manera precisa en términos de la información disponible. Si esta información consiste en un conjunto de objetos descrito por otro conjunto, en este caso, de atributos, diremos que un rough set es un conjunto de objetos que, en general, no pueden ser caracterizados de manera precisa en términos de valores de un conjunto de atributos.

La teoría rough set asume la representación del conocimiento de los objetos en forma de una *tabla de información*. En las filas de la tabla se indican los *objetos* (acciones, alternativas, candidatos, pacientes, empresas, etc.), mientras que las columnas se corresponden con los *atributos*. Las entradas en la tabla son los *valores del atributo*. Es decir, la entrada en columna q y en fila x tiene el valor $f(x, q)$. Por tanto, para cada par (objeto, atributo) se conoce un valor denominado *descriptor*. Cada fila de la tabla contiene descriptores que representan información correspondiente a un objeto. La *relación de no-diferenciación* ocurriría si dados dos objetos para todos los atributos tomasen los descriptores el mismo valor.

Aproximación de conjuntos, precisión y calidad de la clasificación

La no-diferenciación de objetos por medio de atributos impide generalmente su asignación precisa a un conjunto. Precisando aún más, un rough set es una colección de objetos que, en general, no pueden ser clasificados de manera precisa en términos de los valores del conjunto de atributos y, por tanto, tiene casos fronterizos (objetos que no pueden clasificarse con certeza como miembros del conjunto o de su complementario). Luego, un rough set puede ser reemplazado o representado por un par de conjuntos precisos, llamados la *aproximación por encima y por debajo*. La aproximación por debajo consiste en todos los objetos que seguramente pertenecen al conjunto y la aproximación por encima contiene los objetos que posiblemente pertenecen al conjunto. La frontera (o *región de duda*) es el conjunto de elementos que no pueden ser con certeza clasificados utilizando el conjunto de atributos.

Al cociente entre el número cardinal de la aproximación por debajo y el de la aproximación por encima se le denomina *precisión*. Si lo que tenemos es una clasificación en varios grupos, para cada clase se pueden calcular las aproximaciones por encima y por debajo. Al cociente de la suma de las aproximaciones por debajo de todas las clases dividido por el total de elementos que disponemos se denomina *calidad de clasificación*. Expresa el ratio de todos los objetos correctamente clasificados respecto a todos los objetos del sistema.

Reducción y dependencia de atributos

Una de las funciones más importantes de la investigación rough set es el descubrimiento de las dependencias entre atributos al analizar una tabla de información porque nos permite reducir el conjunto (eliminar aquellos atributos que sean redundantes), apareciendo el concepto de *reducto* o *conjunto mínimo*, que se define como el menor conjunto de atributos que mantiene la misma calidad de clasificación que el conjunto de todos los atributos. En una tabla de información puede haber más de un reducto y su intersección nos da el denominado *núcleo*, que es la colección de los atributos más relevantes en la tabla y que no pueden ser eliminados sin que disminuya la calidad de la aproximación de la clasificación. Uno de los modelos más utilizados para obtener los reductos y el núcleo es el de Skowron (1991) y se basa en el cálculo de la denominada *matriz de diferenciación*.

Reglas De Decisión

Un sistema de información reducido permite la obtención de reglas de decisión. De hecho ésta es la cuestión más importante del enfoque rough set.

Suponemos que el conjunto de datos contiene información de un conjunto de *objetos* descritos por un conjunto de *atributos*. Éstos a su vez se dividen en dos subconjuntos de *atributos de condición* y *atributos de decisión* que denotamos por C y D. El problema consiste en encontrar reglas que determinen si un objeto pertenece a un subconjunto particular denominado *clase de decisión*. La definición de esta clase es conocida porque, por ejemplo, lo ha definido el experto o un usuario.

Una regla de decisión puede expresarse como una sentencia lógica que relaciona la descripción de condiciones y las clases de decisión. Toma la siguiente forma:

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

Las reglas generadas pueden ser *determinísticas* o *no determinísticas*. Por determinística (consistente, precisa, exacta) entendemos si $C \rightarrow D$ y en otro caso, es no determinística (inconsistente, aproximada) que sería cuando las condiciones pueden conducir a varias posibles decisiones. Cada regla de decisión se caracteriza por su *fuerza*, esto es, el número de objetos que satisfacen la parte de la condición de la regla (en el argot se dice *cubiertos* por la regla) y que pertenecen a la clase de decisión sugerida por ella.

El conjunto de las reglas de decisión y la información sobre los atributos más significativos para la clasificación de los objetos puede considerarse como *una representación* del conocimiento adquirido por un *especialista* o *experto* sobre todos los casos/objetos contenidos en un sistema de información sin las redundancias típicas de las

bases de datos reales. Es más, los resultados obtenidos por el enfoque rough set se expresan de *forma similar al lenguaje natural humano*. Así, estas representaciones son muy fáciles de comprender por el usuario/analista y permiten justificar y explicar las conclusiones derivadas del análisis de los datos. Además, da la posibilidad al analista de controlar dicho análisis de manera simple. Tal posibilidad no es normalmente ofrecida por las técnicas tradicionales de análisis de datos, añadiendo además una baja carga computacional y posibilidad de interacción con un usuario en modo de diálogo.

El conjunto de reglas para todas las clases de decisión se denomina *algoritmo de decisión*. Los procedimientos para generar reglas de decisión a partir de una tabla de decisión operan sobre los principios del aprendizaje inductivo. Los algoritmos de inducción de reglas utilizan alguna de las siguientes estrategias: generación de un conjunto mínimo de reglas que cubran todos los objetos de la tabla, o generación de un conjunto exhaustivo de reglas consistentes en todas las reglas posibles de una tabla, o generación de un conjunto de reglas de decisión "fuertes" que cubran relativamente muchos objetos pero no necesariamente todos los objetos de la tabla.

Clasificación de *nuevos objetos* utilizando reglas de decisión

Dentro de los problemas de clasificación nos centraremos en los *problemas de clasificación multiatributo* en los que existe solo un atributo de decisión. Las reglas de decisión derivadas de una tabla con esta metodología pueden ser utilizadas para recomendaciones sobre la clasificación de *nuevos objetos* comparando su descripción con las condiciones contenidas en cada una de las reglas de decisión.

3.- SELECCIÓN DE LOS DATOS Y DE LAS VARIABLES

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

En cuanto a los datos, la muestra utilizada es la seleccionada para la aplicación del análisis discriminante en la predicción de la insolvencia en entidades aseguradoras españolas no vida (Sanchis, 2000). Consta de 36 empresas sanas y 36 empresas fracasadas, emparejadas, fundamentalmente, por tamaño (medido a través del volumen de primas) y sector, para minimizar el efecto de estas variables en el estudio. Para un mayor detalle sobre la selección de la muestra se puede consultar la bibliografía citada anteriormente.

Hemos de mencionar que para la construcción del modelo (reglas de clasificación) hemos utilizado únicamente 27 empresas de cada una de las submuestras dejando las 9 restantes para comprobar la validez de las reglas mediante su aplicación a datos que no hubieran sido utilizados para la elaboración de las mismas. La selección de la muestra de validación se ha efectuado aleatoriamente tomando las empresas numeradas en el trabajo mencionado de la 19 a la 27, para las fracasadas, y de la 119 a la 127 como submuestras de validación.

En cuanto a las variables, hemos partido de considerar que el éxito o fracaso de una empresa es función de una serie de variables. Siguiendo un paralelismo con conceptos matemáticos podríamos decir que el éxito o fracaso de una empresa sería la variable

dependiente y por variables independientes tomaríamos la información contable de las mismas reflejada a través del cálculo de determinados ratios financieros.

La definición de *empresa fracasada*, variable independiente, es aquella empresa que ha tenido que ser intervenida por la C.L.E.A (Comisión Liquidadora de Entidades Aseguradoras). Es una medida objetivamente determinable de las empresas que fracasan.

Los datos utilizados para el cálculo de los *ratios financieros*, variables dependientes, han sido los Balances y las Cuentas de Pérdidas y Ganancias de la publicación anual “Balances y Cuentas. Seguros Privados” de la Dirección General de Seguros. La utilización de documentos contables permite obtener una información cuantificable, objetiva y sistemática de la realidad económica y financiera de una empresa. Los datos contables empleados corresponden para las empresas fracasadas al año anterior a la quiebra, y por extensión, también para su pareja ese será el año base.

Para la selección de los ratios hemos acudido a los más utilizados en el análisis financiero para el pronóstico de la insolvencia de entidades aseguradoras, teniendo en cuenta nuestro criterio personal, dadas las propias características del sector seguros. Las particularidades sectoriales de las empresas de seguros, son tan significativas, que no resultan aplicables, al menos parcialmente, los modelos y contenidos de las Cuentas anuales elaboradas para empresas industriales y comerciales, por lo que junto con los ratios aplicables al conjunto de empresas industriales y comerciales hay que considerar todos aquellos específicos para el sector asegurador. Por tanto, dado que pretendemos con la metodología Rough Set, obtener el más pequeño subconjunto de ratios con el mayor poder discriminante en relación con el fracaso empresarial, hemos partido de una gran cantidad de ellos seleccionados, a nuestro criterio, por su aplicación en estudios anteriores y tradición financiera. La flexibilidad del método permite que las variables puedan también ser cualitativas aunque en este estudio no se hayan incluido. Los 21 ratios seleccionados son:

Ratios	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	$(\text{Gastos T\u00e9cnicos de seguro directo} + \text{Comisiones y otros gastos de Explotaci\u00f3n}) / \text{Total Primas adquiridas de seguro directo}$
R18	$(\text{Gastos T\u00e9cnicos de negocio neto} + \text{Comisiones y otros gastos de Explotaci\u00f3n}) / \text{Total Primas adquiridas de negocio neto}$
R19	$\text{Provisiones T\u00e9cnicas de reaseguro cedido} / \text{Provisiones T\u00e9cnicas}$
R20	$\text{RSD} = \text{Gastos T\u00e9cnicos de seguro directo} / \text{Total Primas adquiridas de seguro directo}$
R21	$\text{RSN} = \text{Gastos T\u00e9cnicos de negocio neto} / \text{Total Primas adquiridas de negocio neto}$

Aunque, se calcularon los ratios 15 y 16 no han sido utilizados en el an\u00e1lisis posterior porque las peculiaridades de la muestra hac\u00edan que tomaran valores que carec\u00edan de sentido econ\u00f3mico al no tener la mayor\u00eda de las empresas otros ingresos de explotaci\u00f3n. Con lo que solo hemos utilizado los 19 restantes.

De todos los ratios se han obtenido las principales medidas estad\u00edsticas. No hemos calculado la matriz de correlaci\u00f3n para el conjunto de ratios, aunque sabemos que algunos de ellos est\u00e1n muy correlacionados por dos motivos:

1.- No es un modelo estad\u00edstico con lo que los resultados no se ven afectados por las posibles correlaciones entre las variables. Es m\u00e1s, al poder incluir todas las variables no perdemos el contenido informativo que puede existir en ratios semiindependientes aumentando as\u00ed el poder predictivo, principal objetivo de la construcci\u00f3n del modelo.

2.- Dado que la metodolog\u00eda rough set proporciona el/los m\u00ednimo/s conjunto/s de variables que dan la misma capacidad de clasificaci\u00f3n que el conjunto de todas ellas, es decir, el/los reducto/s, en un mismo reducto no deber\u00eda haber ratios muy correlacionados y si los hubiera, al hacer un an\u00e1lisis de los mismos descartar\u00edamos aquellos que contuvieran ratios muy correlacionados.

4.-RESULTADO DE LA APLICACI\u00d3N DEL M\u00c9TODO ROUGH SET

Para la aplicaci\u00f3n de la investigaci\u00f3n rough set hemos formado un sistema de informaci\u00f3n en el que hemos transformado los valores continuos de los ratios en t\u00e9rminos cualitativos. De esta manera todos los atributos quedaban codificados. Esta transformaci\u00f3n implica una divisi\u00f3n del dominio original en subintervalos y una asignaci\u00f3n de valores nominales a ellos. Los valores frontera de los subintervalos son las denominadas *normas* y surgen de convencionalismos, h\u00e1bitos o de asignaci\u00f3n subjetiva. Esta codificaci\u00f3n no se impone por el rough set pero en el caso de utilizaci\u00f3n de ratios financieros se hace necesario para una interpretaci\u00f3n pr\u00e1ctica de los resultados y para extraer conclusiones generales de los datos en t\u00e9rminos de reglas de decisi\u00f3n. Adem\u00e1s, y dado que las normas se utilizan desde el principio del problema y hasta el final en la explicaci\u00f3n de las reglas de decisi\u00f3n, no falsean la imagen original de la situaci\u00f3n de decisi\u00f3n. Las empresas para las que los valores de estos ratios financieros est\u00e1n en los mismos subintervalos se considera que tienen el mismo comportamiento y caracter\u00edsticas. Los c\u00f3digos que se utilizan para referirse a un subintervalo no representan ning\u00fan orden de preferencias y su elecci\u00f3n no afecta a los resultados obtenidos.

Esta discretizaci\u00f3n de los ratios la hemos efectuado utilizando 4 subintervalos que han sido determinados a trav\u00e9s de los cuart\u00edles de la distribuci\u00f3n de cada ratio y asignando como c\u00f3digos los valores 1, 2, 3 y 4 a los intervalos obtenidos. Utilizando estos subintervalos y el principio “cu\u00e1nto m\u00e1s alto sea el c\u00f3digo, mejor subintervalo”, se obtuvo

una tabla de información codificada de manera automática ya que esta opción está incluida en el programa ROSE. Dado el principio que se aplicó, para los atributos decrecientes (aquellos en que los valores más bajos son mejores) se les dio un código en orden inverso y además hemos efectuado correcciones en la escala en el caso en que no estuviese nuestro criterio en concordancia con la secuencia de aumento o disminución de los subintervalos.

La tabla codificada utilizada para un posterior análisis, consta de las 54 empresas descritas por los 19 atributos codificados (ratios financieros), utilizando los datos del año previo a la quiebra y asignadas de forma binaria a una clase de decisión (sana o fracasada), representada por 1 ó 0 respectivamente. El análisis rough set de la información codificada de la tabla se desarrolló utilizando el programa ROSE. La precisión fue perfecta, esto es, igual a uno y la calidad de la clasificación fue también igual a uno. El resultado es muy significativo porque a pesar de un gran sistema de información, las empresas están muy bien discriminadas entre ellas. Es decir no hay casos frontera ni hay, por tanto, información imprecisa.

El siguiente paso del análisis fue la construcción del mínimo subconjunto de atributos independientes que asegurasen la misma calidad de clasificación que la totalidad del conjunto, esto es los reductos. Hemos obtenido 229 reductos de la tabla de información codificada, cada uno de los cuales contiene entre 4 y 7 atributos, lo que respecto a los 19 originales supone una reducción importante. Este resultado nos da una idea de la ayuda que supone la aplicación de este método en la eliminación de variables redundantes ya que, en cualquier caso al menos 12 atributos pueden ser eliminados sin ninguna consecuencia. Los ratios que aparecen con más frecuencia en los reductos (aparecen en más del 40%) son: R1, R3, R4, R9, R17, R19 lo cual indica que son variables muy discriminatorias en la muestra entre empresas solventes y no solventes. El núcleo de atributos está vacío luego ningún único atributo es absolutamente necesario para la aproximación de las dos clases de decisión.

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

- a) El reducto debía contener el menor número de atributos que fuese posible.
- b) Debía contener los atributos considerados a nuestro juicio más significativos para la evaluación de las empresas.
- c) No debía contener ratios que estuviesen correlacionados.

El reducto que hemos elegido es: R3, R4, R9, R14, R17. Elegido el reducto, el resto de atributos han sido eliminados de la tabla de información codificada y hemos pasado de un sistema inicial de información codificado de 19 columnas a otro de solo 5. El sistema de información reducido puede verse como una tabla de decisión de la cual obtendremos el conjunto de reglas. Hemos obtenido 27 reglas de decisión con la estrategia consistente en la obtención del mínimo conjunto de reglas que cubren todos los objetos de la tabla. Todas son determinísticas ya que la calidad de la clasificación es igual a uno. Son:

regla 1. (R3 = 2) & (R9 = 4) => (D = 0); empresas [{2, 14, 18, 23}]

regla 2. (R4 = 3) & (R14 = 2) => (D = 0); empresas: [17, 21, 25, 26]

regla 3. (R3 = 1) & (R4 = 4) => (D = 0); empresas: [7, 10, 22]

regla 4. (R4 = 1) & (R9 = 4) => (D = 0); empresas: [13, 19, 24]

regla 5. (R9 = 3) & (R14 = 3) => (D = 0); empresas: [1, 4]

regla 6. (R3 = 3) & (R9 = 1) => (D = 0); empresas:[6, 8]

regla 7. (R3 = 2) & (R9 = 3) & (R14 = 4) => (D = 0); empresas:[11, 15]

- regla 8.** (R3 = 4) & (R14 = 1) => (D = 0); empresas:[12, 16]
regla 9. (R3 = 4) & (R4 = 1) => (D = 0); empresas: [16, 20]
regla 10. (R3 = 2) & (R9 = 1) & (R14 = 3) => (D = 0); empresas: [9]
regla 11. (R9 = 3) & (R14 = 2) => (D = 0); empresas: [3]
regla 12. (R3 = 3) & (R9 = 2) & (R17 = 2) => (D = 0); empresas: [5]
regla 13. (R3 = 2) & (R4 = 2) & (R14 = 2) => (D = 0); empresas: [2, 18, 27]
regla 14. (R14 = 1) & (R17 = 3) => (D = 1); empresas: [29, 40, 50]
regla 15. (R9 = 2) & (R17 = 4) => (D = 1); empresas: [41, 44, 49, 51, 52]
regla 16. (R4 = 1) & (R9 = 3) => (D = 1); empresas: [28, 36, 38, 42]
regla 17. (R3 = 4) & (R4 = 2) => (D = 1); empresas: [33, 37]
regla 18. (R3 = 3) & (R4 = 4) & (R17 = 1) => (D = 1); empresas: [30, 32]
regla 19. (R3 = 1) & (R4 = 2) & (R9 = 4) => (D = 1); empresas: [39, 47]
regla 20. (R4 = 2) & (R17 = 2) => (D = 1); empresas: [53]
regla 21. (R4 = 3) & (R9 = 1) => (D = 1); empresas: [31]
regla 22. (R3 = 2) & (R14 = 1) => (D = 1); empresas: [43, 48]
regla 23. (R3 = 3) & (R14 = 4) => (D = 1); empresas: [34, 38, 42]
regla 24. (R3 = 4) & (R14 = 3) => (D = 1); empresas: [33, 35]
regla 25. (R3 = 1) & (R4 = 3) & (R9 = 2) => (D = 1); empresas: [46]
regla 26. (R3 = 1) & (R4 = 1) & (R17 = 2) => (D = 1); empresas: [54]
regla 27. (R4 = 4) & (R14 = 2) & (R17 = 4) => (D = 1); empresas: [45]

Las 27 reglas obtenidas y reunidas en un sistema de información codificada constituyen un algoritmo de clasificación, siendo una representación no redundante del conocimiento. La muestra de 54 firmas empleada en el estudio para derivar el algoritmo puede considerarse como una muestra de prueba utilizada para revelar las características financieras que discriminan las empresas viables de las que no lo son.

Para validar el algoritmo obtenido y ver la precisión predictiva hemos efectuado la clasificación de las 18 empresas que nos habíamos reservado como muestra de validación. Los porcentajes de clasificaciones correctas obtenidas con la aplicación del método rough set fueron:

Clases	Correcto	Incorrecto
0	77.78%	22.22%
1	77.78%	22.22%
Total	77.78%	22.22%

El programa ROSE utiliza unos parámetros para reclasificar aquellas empresas que no encajan con ninguna regla (ver bibliografía). Por otro lado, en este caso concreto, coincide el número de empresas clasificadas correctamente en ambos grupos por lo que los errores de tipo I y II son iguales.

5.- CONCLUSIONES

A lo largo de la exposición se han ido enumerando algunas ventajas de la metodología rough set por lo que podemos concluir que su utilización en el tratamiento de la información contable la convierte en una herramienta muy valiosa a la hora de analizar la situación económico-financiera de las empresas de seguros no vida.

Es un método rápido para procesar gran cantidad de información tanto cualitativa como cuantitativa. Lo hace, además, de manera sencilla porque las reglas de decisión no requieren para su interpretación la ayuda de expertos. Todas estas características hacen idónea su utilización por un gran número de usuarios (auditores, autoridades supervisoras, inversores y analistas) que pueden beneficiarse de una herramienta que les ayude a la hora de justificar sus decisiones con un gran ahorro en coste y en tiempo. Es decir, en términos prácticos, las reglas pueden usarse como sistemas de diagnóstico automático para preseleccionar por ejemplo aquellas empresas que necesiten una atención especial, de una manera rápida y a un coste bajo, y de este modo habría una gestión más eficiente del tiempo dedicado por el analista financiero o la autoridad supervisora dando, además, uniformidad a los juicios emitidos sobre una empresa. Es más, un análisis de las reglas de decisión obtenidas pueden mostrar algunos factores a tener en cuenta como posibles indicadores de una mala situación financiera. En cuanto a la validación de las reglas de decisión obtenidas, es evidente que no se pueden sacar conclusiones significativas dado el escaso número de empresas analizadas, y además no es este nuestro objetivo. Nuestra meta es señalar la idoneidad del método como herramienta que ayude a la hora de justificar o contrastar una decisión.

REFERENCIAS

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

Nurmi, H., Kacprzyk, J. y Fedrizzi, M. (1996). Probabilistic, fuzzy and rough concepts in social choice, *European journal of Operational Research* nº 95, pp. 264-277.

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

Predki, B., Slowinski, R., Stefanowski, J., Susmaga, R., Wilk, S. (1998). ROSE – Software Implementation of the Rough Set Theory, in L. Polkowski, A. Skowron, eds. *Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence*, vol. 1424. Springer-Verlag, Berlin, pp. 605-608.

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

Sanchis, A. (2000). *Una aplicación del Análisis Discriminante a la previsión de la Insolvencia en las empresas españolas de seguros no-vida*, Tesis Doctoral, Universidad Complutense de Madrid.

Skowron, A. (1991). *The Implementation of Algorithms based on Discernibility Matrix*, Manuscript.

Slowinski, R., ed. (1992). *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Zopounidis, C., y Dimitras A. (1998). *Multicriteria Decision Aid Methods for the Prediction of Business Failure*. Kluwer. Dordrecht.