

LA METODOLOGÍA DE LOS ROUGH SETS COMO TÉCNICA DE PREPROCESAMIENTO DE DATOS: UNA APLICACIÓN A LAS QUIEBRAS DE MICROEMPRESAS FAMILIARES

M. J. VÁZQUEZ CUETO

pepi@us.es

Universidad de Sevilla / Departamento Economía Aplicada III

Avenida Ramón y Cajal nº 1. 41018 Sevilla

A. I. IRIMIA DIÉGUEZ

airimia@us.es

Universidad de Sevilla / Departamento Economía Financiera y Dirección de Operaciones

Avenida Ramón y Cajal nº 1. 41018 Sevilla

A. J. BLANCO OLIVER

aj_blanco@us.es

Universidad de Sevilla / Departamento Economía Financiera y Dirección de Operaciones

Avenida Ramón y Cajal nº 1. 41018 Sevilla

Recibido (18/11/2014)

Revisado (13/07/2015)

Aceptado (21/07/2015)

RESUMEN: Las microempresas (MEs) representan más del 75% del tejido empresarial de la Unión Europea, acaparando más del 30% del empleo. No obstante, desde el inicio de la crisis económica en el año 2008, este segmento empresarial viene sufriendo elevadas tasas de quiebras y cierres empresariales, destruyéndose numerosos puestos de trabajo. La construcción de modelos que anticipen situaciones de insolvencia que permitan adoptar con suficiente antelación las medidas oportunas, es clave para evitar la quiebra de las MEs. A pesar de ello, es difícil disponer de información completa y relevante de las MEs, lo que hace muy difícil un buen ajuste de los modelos de predicción de la insolvencia empresarial para este tamaño de empresas. Aplicando la técnica de los Rough Sets como un método de pre-procesamiento de los datos, en el presente estudio, ordenamos las variables que mejor discriminan entre MEs solventes/insolventes en aras a incrementar la eficiencia en la predicción de su insolvencia. Además, ofrecemos una aplicación de la técnica a MEs de carácter familiar. En todo este proceso nuestros resultados realzan la relevancia de la consideración de variables no financieras para predecir la insolvencia de las MEs.

Palabras claves: Rough Sets, Modelos de quiebra empresarial, Ratios financieros, Microempresa.

ABSTRACT: Micro enterprises (MEs) represent over 75 % of all enterprises in the EU, accounting for over 30 % of employment. However, since the onset of the economic crisis in 2008, this business segment has suffered high rates of bankruptcies and business closures, destroying many jobs. The construction of models that anticipate insolvency to allow sufficient time to take appropriate action is important to avoid bankruptcy of the MEs. However, it is difficult to obtain complete and relevant information for MEs, making it very difficult to be a good fit of the models for predicting corporate failure for this size of company. Applying Rough Sets technique as a method for pre - processing of the data, in the present study, we order the variables that best discriminate between solvent / insolvent in order to increase efficiency in predicting insolvency MEs. Additionally, we provide an application of the technique to family- MEs. Throughout this process, our results highlight the importance of considering non-financial variables to predict insolvency of MEs.

Keywords: PYME, Micro enterprises, Rough Sets, Financial Ratios, Models of corporate bankruptcy.

1. Introducción

Las microempresas influyen decisivamente en la reducción del desempleo y el incremento de la innovación y competitividad de un país, siendo consideradas por muchos autores la columna vertebral de las principales economías desarrolladas (McPherson, 1996; Mead y Liedholm, 1998). Además, gracias a su sencilla estructura pueden responder rápidamente a los cambios tanto de las condiciones económicas como de las necesidades de sus clientes, creciendo en algunas ocasiones, a un ritmo superior al que lo hacen las grandes corporaciones (Altman y Sabato, 2007).

No obstante, en los últimos años, este segmento empresarial viene sufriendo elevadas tasas de quiebra, afectando negativamente al empleo y a la competitividad de las principales economías desarrolladas (Carter y Van Auken 2006). En este contexto, se hace necesario el desarrollo de modelos específicamente desarrollados para las microempresas que permitan anticipar situaciones de insolvencia y actuar en consecuencia para evitar las mismas. Sin embargo, existen una serie de características propias de este segmento empresarial, como son: (i) la limitación de la información financiera disponible, (Berger y Frame 2007), (ii) la más fácil manipulación contable ya que suelen coincidir en la misma persona las figuras de propietario y directivo (Claessens, Djankov, y Lang 2000; Fan y Wong 2001) y, (iii) su carácter familiar, que influye decisivamente en su performance (Morten, Nielsen, Perez-Gonzalez, y Wolfenzon 2007), que imposibilitan, limitan o distorsionan el desarrollo de adecuados análisis económico-financieros para predecir sus dificultades financieras. Los modelos de predicción de la insolvencia diseñados específicamente para este segmento empresarial deberán estar adaptados a la información disponible y considerar el carácter familiar de la microempresa como un factor especialmente relevante y diferenciador de su performance.

Por tanto, el principal objetivo del presente trabajo es mostrar cómo la técnica de los conjuntos Rough Sets (conjuntos vagos) puede ayudarnos a identificar la importancia que tienen determinadas variables en el desarrollo de modelos de predicción de la insolvencia empresarial, específicamente los desarrollados para las microempresas familiares.

Este trabajo se estructura como sigue. En el segundo epígrafe se expone la metodología a seguir, que está basada en la teoría de los “Rough Sets”. Expondremos brevemente los fundamentos de la técnica y cómo puede adecuarse a los objetivos fijados. En el tercer epígrafe mostramos la aplicación empírica. Comenzamos con la explicación de la base de datos y la forma de extracción de la muestra que posteriormente será utilizada. En los siguientes subepígrafes de este tercer epígrafe se definen las variables que vamos a utilizar en el estudio empírico: la variable dependiente y los conjuntos de variables independientes (financieras y no financieras). En el epígrafe cuatro se exponen los resultados haciendo especial hincapié en la importancia de las variables no financieras para incrementar la performance de nuestro modelo. Además, se cuantifica la contribución de cada variable para mejorar el modelo resultante. En el quinto y último epígrafe se exponen las principales conclusiones del presente trabajo.

2. Metodología

La técnica de los Rough Sets (también conocida como conjuntos vagos) fue desarrollada por Pawlak en 1982. La técnica parte de una matriz formada por un conjunto U de elementos u objetos en los que se han medido varios atributos de información, C , o de decisión (clasificación, asignación,...) D . Cada fila se corresponde con un elemento y cada columna con los valores que toma un determinado atributo en esos elementos. La submatriz (U, C) se denomina tabla de información y a la matriz completa (U, C, D) , tabla de información –decisión. En primer lugar debemos hacer notar que un conjunto es o no un rough set siempre en referencia a un subconjunto de atributos, ya sea de C , de D o de CUD .

Sea P un subconjunto de atributos. Toda la teoría descansa en la **relación de indiscernibilidad** o indistinguibilidad que definen dos elementos indistinguibles. Así, diremos que el elemento x es indistinguible del elemento y , $x, y \in U$, respecto al subconjunto P , si los valores que toman los atributos

de P en x coinciden con los valores que toman en y. Esta relación es una *relación de equivalencia* y, por tanto, determina una partición en el conjunto U de todos los elementos, [P]. Si ahora consideramos un subconjunto Q de P y la partición que origina Q en el conjunto U de todos los elementos coincide con la que define P, [Q]=[P], diremos que Q es una reducción de P. Este concepto nos ayudará a determinar los atributos de información de los que podremos prescindir.

Para un subconjunto X de elementos de U definimos

- *aproximación por debajo del conjunto P* como $\underline{P}X = \{x \in U / \text{la clase de equivalencia a la que pertenece } x, [P] \subseteq X\}$
- *aproximación por encima del conjunto P* como $\overline{P}X = \{x \in U / \text{la clase de equivalencia a la que pertenece } x, [P] \cap X \neq \emptyset\}$.

Cuando $\overline{P}X - \underline{P}X \neq \emptyset$ diremos que X es un *Rough Set* respecto a P, es decir, respecto a los atributos considerados en P.

La importancia de este concepto se pone de manifiesto cuando incluimos a la variable de decisión D, es decir, cuando trabajamos con la tabla de información-decisión. Supongamos que D toma solo dos valores. Sean D=0 y D=1. Podemos entonces descomponer el universo total de elementos en dos conjuntos, aquellos para los que D toma el valor 0, sea X₀, y aquellos para los que toma el valor 1, X₁. Si para el subconjunto de atributos de información P, X₀ es un *Rough Set* respecto a P, significará que, *en base a la información proporcionada por los atributos de P*, no podemos asegurar que los elementos de X₀ pertenezcan a la clase de decisión D = 0, de hecho los elementos pertenecientes a $\overline{P}X_0 - \underline{P}X_0$ son los elementos dudosos.

Al habernos restringido al caso en que D toma sólo dos valores, la calidad de la aproximación por P de la clase de decisión D es

$$(\text{card}\overline{P}X_0 + \text{card}\overline{P}X_1) / n^\circ \text{elementos}$$

y es el número de elementos correctamente clasificados.

La precisión que logramos para X_i por P viene definida por el cociente entre el número de elementos de su aproximación por debajo del conjunto P y el número de elementos de su aproximación por encima del conjunto P

$$(\text{card}\underline{P}X_i / \text{card}\overline{P}X_i) \text{ para } i=0,1.$$

Si Q es una reducción de P y conserva la “calidad de la aproximación”, diremos que Q es un **reducto** de P y podemos prescindir de aquellos elementos de P que no están en Q sin pérdida de información. La determinación de los reductos es pues un paso fundamental en la aplicación de la técnica al caso que nos ocupa¹.

Obviamente pueden existir muchos reductos², la intersección de todos ellos constituye el **núcleo** (core) de la información, es decir, aquellos atributos que son indispensables para la clasificación.

Una vez que hayamos detectado estos atributos, estos guardan relación con los valores de D, estas se expresan en la forma “si..., entonces...” y se denominan “**reglas de decisión**”. Una regla de decisión es pues **una aplicación** entre los atributos de información y los valores del atributo de clasificación. Las reglas elaboradas con cada uno de los reductos clasifican al conjunto de elementos de U con un 100% de aciertos. Una vez determinado el núcleo de la información, y medida la calidad de la aproximación que conseguimos con los atributos pertenecientes al mismo, aplicando reiteradamente la técnica,

¹ La determinación de los reductos se basa en el algebra boleana, y su explicación creemos que excede del objeto de este trabajo. Consultar Pawlak, 1982

² Si el número de atributos de información es m el número de reductos es $\sum_{k=1}^m \binom{m}{k} = (1+1)^m - \binom{m}{0} = 2^m - 1$

introduciendo cada vez uno de ellos, podremos determinar lo que aporta cada uno a la clasificación y, de esta forma determinar la importancia de los mismos.

3. Aplicación empírica

3.1. Base de datos

Aplicamos la técnica expuesta a una muestra de *micro-entities* familiares (microentidades³), extraída de una fuente de información secundaria proporcionada por el *Credit Management Research Centre* (CMRC) perteneciente a *Leeds University Business School* (Inglaterra), que contiene información (ratios económico-financieros, variables no financieras y macroeconómicas) de 4.813.391 pequeñas y medianas empresas (PYMEs) del Reino Unido entre los años 1999-2009 (ambos inclusive), un 98,32% consideradas como no insolventes y 1,68% insolventes. Limitamos nuestro trabajo al periodo temporal 2007 a 2009⁴. Para la extracción de la muestra de esta base de datos aplicamos, al igual que la mayoría de los trabajos empíricos sobre modelos de predicción de la insolvencia empresarial recogidos en la literatura previa (Beaver, 1966; Altman, 1968; Deakin, 1972; Edmister, 1972; Blum, 1974), la metodología del emparejamiento (*matched pairs methodology*)⁵. En la medida que se ha podido dada la base de datos, se ha procurado que los porcentajes de insolvencia dentro del año estén próximos al 50% y que todos los años estén igualmente representados (en torno cada uno al 33%). Finalmente, la muestra con la que se ha trabajado, por año y por situación de insolvencia, queda recogida en la Tabla 1.

Tabla 1. Composición de la muestra

Microentidad familiar		insolvencia		insolvencia		insolvencia	
		no	sí	no	sí	no	sí
		Recuento	Recuento	Porcentaje en el año	Porcentaje en el año	Porcentaje en la muestra	Porcentaje en la muestra
Año del análisis	2007	1144	1123	50,46%	49,54%	29,55%	24,86%
	2008	1318	1599	45,18%	54,82%	34,04%	35,40%
	2009	1410	1795	44,00%	56,00%	36,42%	39,74%

Fuente: Elaboración propia

La muestra está compuesta por un total de 8.389 empresas familiares, (2.269 del año 2007, 2.917 del año 2008 y 3.205 del año 2009), de las cuales el 46,16% eran no insolventes (aproximadamente el 14,8% del año 2007, el 15,4% del año 2008 y el 16% del año 2009)

3.2. Descripción de las variables

3.2.1. La variable dependiente: La insolvencia empresarial

Aunque la literatura especializada no suele distinguir entre los términos de quiebra, insolvencia, fracaso, dificultades financieras y problemas financieros empresariales, cuando trata de desarrollar modelos de predicción, estos no son conceptos similares. Por ello vamos a conceptualizar exactamente cuál es la variable dependiente a estudiar. En consonancia con nuestra base de datos las empresas consideradas insolventes son aquellas que estuvieron en alguna de las tres situaciones siguientes:

³ En la base de datos utilizadas se identifican las *micro-entidades* (MEs) usando la definición dada en Mayo de 2011 por el Consejo de Competitividad Europeo: aquellas empresas con una cifra de negocios anual de menos de 500.000 €, un activo total de menos de 250.000 € y un número de empleados menores a 10 personas (BIS, 2011).

⁴ Considerado 2008 como el año de estallido de la crisis, tomamos un año antes y un año después.

⁵ Esta técnica consiste en utilizar una muestra de casos en la que el número de empresas insolventes y solventes sean iguales. Como no queríamos perder la perspectiva temporal ni familiar y la proporción de casos insolventes es mucho menor que la de casos solventes, se procedió a realizar para cada uno de los años una selección aleatoria de casos no insolventes

- Liquidación empresarial motivada por problemas financieros.
- Administración concursal.
- Quiebra empresarial.

Por tanto, todas aquellas empresas que estuvieran en cualquiera de estas tres situaciones serán calificadas indistintamente en este trabajo como empresas insolventes, quebradas, fracasadas, o con problemas o dificultades financieras, y a la variable dependiente/decisión D, se le asignará el valor 1. En caso contrario tomará el valor 0.

3.2.2. Variables financieras

Existe un importante número de variables financieras, la mayoría de ellas ratios, que se han mostrado útiles en la clasificación/predicción de la quiebra de una empresa. Estudios empíricos previos han encontrado que una empresa es más probable que caiga en quiebra si es poco rentable, está muy endeudada y padece dificultades de liquidez (Myers, 1977). En este estudio hemos considerado catorce ratios financieros⁶ agrupados en cinco categorías de acuerdo con el aspecto financiero que miden: endeudamiento, liquidez, beneficio, actividad y tamaño. En la tabla 5 se describen los ratios, su forma de cálculo y la categoría en la que se ha encuadrado.

De acuerdo con la teoría de las finanzas corporativas las empresas con mayores obligaciones respecto a su capital son las que presentan mayores probabilidades de padecer problemas financieros. En este sentido, hemos considerado cuatro ratios de endeudamiento: *Capital Invertido/Recursos Ajenos l.p.*, *Recursos Ajenos a c.p./Activo*, *Recursos Ajenos l.p./Activo Corriente* y *Patrimonio Neto/Activo*.

La liquidez también es una categoría común en la mayoría de las decisiones acerca del crédito ya que representa la habilidad de convertir un activo en efectivo de manera rápida. En este trabajo hemos considerado cuatro ratios dentro de esta categoría: *Activos líquidos/Activo Corriente*, *Tesorería/Patrimonio Neto*, *Activo Corriente/Pasivo Corriente* y *Tesorería/Activo*. Respecto a la rentabilidad, hemos incluido el ratio *Reservas/Activo*, que mide la habilidad para acumular reservas y puede ser utilizado como una variable proxy de la rentabilidad futura.

Tabla 2: Variables financieras

Ratio	Cálculo	Categoría
Total Assets	Activo	Tamaño
Stlta	Recursos Ajenos a c.p. / Activo	Endeudamiento
HHI_Sales	Poder de mercado	
Celt	Capital Invertido/ Recursos Ajenos l.p.	Endeudamiento
Qaca	Activos líquidos / Activo Corriente	Liquidez
Prta	Reservas / Activo	Rentabilidad
Tlca	Recursos Ajenos l.p. / Activo Corriente	Endeudamiento
Ncnw	Tesorería / Patrimonio Neto	Liquidez
Cratio	Activo Corriente / Pasivo Corriente	Liquidez
Tctd	Proveedores / Clientes	Actividad
Tctl	Proveedores / Recursos Ajenos l.p.	Actividad
Tdta	Clientes / Activo	Actividad
Cashta	Tesorería / Activo	Liquidez
Nwta	Patrimonio Neto / Activo	Endeudamiento

Elaboración propia

⁶ Todos los ratios financieros que se utilizan en este estudio han sido previamente empleados por autores tales como Altman (1968), Altman et al. (2010), Ohlson (1980), Taffler (1984), and Zmijewski (1984).

Los tres ratios referentes a proveedores y clientes son importantes para las pequeñas empresas, que tienden a depender financieramente de la forma de pago a los proveedores⁷ y la forma de cobro a los clientes, utilizando también el crédito comercial como estrategia de ventas para luchar contra la competencia.

El tamaño de la empresa ha sido utilizado, entre otros, en los trabajos de Dietsch y Petey (2004) y Altman et al. (2010), que obtienen resultados contradictorios respecto a la influencia de esta variable.

3.2.3. Información no financiera

La literatura no es concluyente respecto a la contribución al incremento en la precisión de los modelos de la introducción de información no financiera. Así, Jacobson et al. (2008) sugieren que la información macroeconómica resulta útil para complementar la información proporcionada por los ratios financieros, tanto en empresas cotizadas como en las no cotizadas, y, en cambio, en el trabajo de Liou y Smith (2006) esta información tiene un poder explicativo relativamente bajo. Además, es extremadamente complejo obtener información no financiera sobre cualquier empresa, siendo especialmente difícil en el caso de las pequeñas, debido, fundamentalmente, a que no están obligadas a publicarla. Por ello, la gran mayoría de los estudios sobre modelos predicción de la quiebra empresarial no utilizan variables no financieras como predictores, siendo este uno de los principales aspectos a mejorar sugeridos por la literatura especializada. Siguiendo esta línea de pensamiento hemos considerado las variables no financieras que se muestran en la Tabla 3.

Tabla 3. Variables no financieras

Variables	Explicación	Categorías
Ln_Age	Logaritmo neperiano edad empresa	
Late Filing days_1	Retraso presentación cuentas	No (0) Si (1)
Late Filing days_2	Retraso presentación cuentas	No (0) Si (1)
No CCJs	Número acciones judiciales	Ninguna (0) Alguna (1)
CCJs_Value	Valor medio de las cantidades reclamadas	
Charge_Flag	Activos pignorados	No (0) Si (1)
Auditor_Changed	Cambio de auditor	No (0) Si (1)
Audited	Cuentas auditadas	No (0) Si (1)
AQ_Clean	Opinión favorable informe auditoria	No (0) Si (1)
AQ_No_Clean	Opinión no favorable informe auditoria	No (0) Si(1)

Elaboración propia

Según Hudson (1987), las empresas más jóvenes tienen mayores probabilidades de entrar en quiebra, por ello es por lo que se ha considerado la variable *Ln_Age* que expresa en días el tiempo transcurrido desde el comienzo de su actividad hasta la fecha en que presenta sus últimas cuentas.

⁷ Las pequeñas empresas suelen acudir a la financiación vía proveedores como sustituta de la financiación bancaria, representando dicha financiación un alto porcentaje de los recursos

La normativa del país del que proviene la muestra de empresas permite, a las que no cotizan, un período de 10 meses para presentar las cuentas a partir del final del ejercicio. Las variables *Late Filing days_1* y *Late Filing days_2* cuentan el número de días después de este periodo y, por tanto, se relacionan con el incumplimiento de la regulación que, a su vez, suele deberse a razones que afectan negativamente a la salud financiera de la empresa. La variable *Late Filing Days_1* se refiere al año previo y la variable *Late Filing Days_2* al año en cuestión.

La variable *No CCJs* cuenta el número de acciones judiciales llevadas a cabo por los acreedores para el cobro de las deudas que la empresa tiene contraídas con ellos en los últimos 12 meses, y se ha codificado según existan o no dichas reclamaciones y *CCJs_Value* es el valor medio de las cantidades reclamadas judicialmente.

También hemos considerado una variable relacionada con las garantías aportadas por los prestatarios al solicitar financiación, *Charge_Flag*, que toma el valor 1 si la empresa tiene activos pignorados y, 0, en caso contrario. Si una empresa tiene activos pignorados por un acreedor significa que dicha empresa ha incumplido con sus obligaciones de pago contraídas con aquel (Citron y Wright, 2008), por tanto, esta información es relevante desde la óptica del cumplimiento de sus obligaciones de pago.

Las siguientes cuatro variables están relacionadas con el proceso de auditoría. La primera de ellas es la variable *Auditor_Changed*. Generalmente, las empresas desean que el informe de auditoría no recoja aspectos u opiniones negativas acerca de su situación económica-financiera ya que, en ese caso, se verá con más dificultades para financiarse. Por ello, si el auditor emite una opinión desfavorable sobre la empresa en su informe de auditoría, dicha compañía deseara cambiar de auditor para el próximo año con el objetivo de tener una opinión diferente. Aunque el cambio de auditor puede ocurrir por diversas razones y no necesariamente por la emisión de informes de auditorías desfavorables, puede, a priori, ser una de las variables que alerten sobre la insolvencia empresarial. *Audited*, es una variable dummy que recoge el hecho de que la empresa haya sido o no auditada. Por lo general, la información financiera de las microentidades con las cuentas auditadas será más fiable que las de aquellas que no se auditan. Por otra parte, los auditores no sólo examinan la salud financiera de las empresas, sino que también aconsejan sobre cómo aumentar la eficiencia para conseguir sus objetivos. Para las empresas auditadas consideramos otras dos variables relativas al informe del auditor. Según la legislación británica (Altman et al., 2010) el auditor puede dar seis opiniones, ordenadas de peor a mejor, acerca de la situación de una empresa. No obstante, debido a la escasa proporción de empresas auditadas en nuestra muestra hemos procedido a reclasificar la opinión del auditor en dos variables *AQ_Clean* y *AQ_No_Clean* con la finalidad de formar grupos con más sentido estadístico: *AQ_Clean* toma el valor 1 cuando el informe del auditor es favorable, es decir, el auditor no detecta problemas financieros, y *AQ_No_Clean* toma el valor 1 cuando el informe del auditor no es favorable.

Añadiendo finalmente una variable macroeconómica indicativa de la situación económica-financiera del sector donde opera la empresa, *Indwoe*. Esta variable es la inversa de la probabilidad de quiebra del sector, y se ha calculado para cada empresa a partir de un modelo desarrollado por el *Credit Management Research Centre* de la Universidad de Leeds y ha sido utilizada en diversos artículos científicos de máximo nivel académico (Altman et al., 2010).

4. Resultados

4.1. Discretización de las variables continuas

Para la aplicación de la técnica propuesta se ha utilizado el software ROSE⁸ categorizando, en primer lugar, a las variables continuas en tres intervalos, en función de los valores que tomaba cada una, y posteriormente codificadas en los valores 0,1 y 2.

⁸ Desarrollado por el Laboratory of Intelligent Decision Support Systems of the Poznan University of Technology, y disponible en (<http://www-idss.cs.put.poznan.pl>)

Tabla 4: Codificación de las variables continuas.

Ln_Age: (-inf, 7.425) <7.425, 7.645) <7.645, +inf)
CCJ_Value medio: (-inf, 15.5) <15.5, 492.5) <492.5, +inf)
Indwoe: (-inf, -0.29) <-0.29, 0.225) <0.225, +inf)
Total_Assets: (-inf, 24500) <24500, 54500) <54500, +inf)
Stlta: (-inf, 0.00445437) <0.00445437, 0.0716855) <0.0716855, +inf)
HHI_Sales: (-inf, 175.06) <175.06, 570.59) <570.59, +inf)
Cetl: (-inf, -0.005) <-0.005, 0.495) <0.495, +inf)
Qaca: (-inf, 0.005) <0.005, 0.985) <0.985, +inf)
Prta: (-inf, -0.005) <-0.005, 0.235) <0.235, +inf)
Tlca: (-inf, 0.855) <0.855, 1.005) <1.005, +inf)
Ncnw: (-inf, 0.195) <0.195, 10.945) <10.945, +inf)
Cratio: (-inf, 1.175) <1.175, 2.165) <2.165, +inf)
Tctd: (-inf, 0.295) <0.295, 0.995) <0.995, +inf)
Tctl: (-inf, 0.295) <0.295, 0.995) <0.995, +inf)
Tdta: (-inf, 0.205) <0.205, 0.645) <0.645, +inf)
Cashta : (-inf, 0.175) <0.175, 0.405) <0.405, +inf)
Nwta: (-inf, 0.035) <0.035, 0.245) <0.245, +inf)

Fuente: Elaboración propia

4.2. Información de la que puede prescindirse

De las 25 variables consideradas, 21 han resultado imprescindibles para la clasificación. Las cuatro no considerados han resultado ser todas no financieras: *NoCCJ*, *Audited*, *AQ_CLEAN* y *AQ_NO_CLEAN*.

Utilizando solo las 21 variables imprescindibles obtenemos un porcentaje de clasificación correcta total del 81,12%. De hecho se obtienen tres reductos, cada uno de ellos con 23 variables. En ninguno de ellos está *No_CCJ*, y las variables *Audited*, *AQ_CLEAN* y *AQ_NO_CLEAN* son intercambiables⁹. Es decir, todos contienen a todas las variables financieras y son las no financieras las que marcan las diferencias entre ellos.

Hemos optado por considerar el reducto formado por

{*Ln_Age*, *Late Filing days_1*, *Late Filing days_2*, *CCJs_Value*, *Charge_Flag*, *AuditorChanged*, *AQ_Clean*, *AQ_No_Clean*, *Indwoe*, *Total_Assets*, *Stlta*, *HHI_Sales*, *Celt*, *Qaca*, *Prta*, *Tlca*, *Ncnw*, *Cratio*, *Tctd*, *Tctl*, *Tdta*, *Cashta*, *Nwta*},

con el que se obtiene un porcentaje de clasificación correcta del 81,57%.

4.3. Importancia de la información no financiera

Dada la dificultad de disponer de información, como se ha comentado anteriormente, vamos a analizar en este apartado si las variables no financieras, más difíciles de conseguir, aportan una mejora significativa a los modelos, en términos de porcentajes de clasificación correctos conseguidos.

Para ello, sea X_0 el conjunto formado por los elementos cuya variable de decisión es $D = 0$, es decir, por las microempresas no insolventes, y X_1 el conjunto formado por los elementos cuya variable de decisión es $D=1$, es decir, las microempresas insolventes. Y sea P el conjunto de atributos de información que incluye solo a las variables financieras

⁹ Puede eliminarse una y solo una de ellas.

{Total_Assets, Stlta, HHI_Sales, Celt, Qaca, Prta, Tlca, Ncnw, Cratio, Tctd, Tctl, Tdta, Cashta, Nwta}.

Bajo estas premisas se obtienen los resultados que se muestran en la tabla 5. En ella observamos que el conjunto X, formado por todas las microempresas es un rough-set en base a los atributos considerados en P. De 4.897 de sus 8.389 elementos no podemos asegurar, en base a la información proporcionada por las variables financieras, si pueden clasificarse o no como insolventes.

Si, por el contrario, consideramos a todas las variables del estudio, es decir, sin incluimos en P las **variables financieras y no financieras del reducto considerado**, obtenemos los resultados que se muestran en la tabla 6. En ella observamos que el número de elementos dudosos en la clasificación se ha reducido de 4.897 a 1.546, alcanzando, con la inclusión de estas variables un porcentaje total de clasificación correcta del 81,57%, mejorando considerablemente la precisión adquirida en el grupo de empresas insolventes¹⁰

Tabla 5: Calidad de la clasificación considerando solo variables financieras

	Total	\underline{PX}	\overline{PX}	Precisión	Calidad*
X=X ₀	3872	1723 elementos con toda seguridad son no insolventes	6620 elementos que no podemos asegurar que sean no insolventes	26,03%	
X=X ₁	4517	1769 elementos con toda seguridad son insolventes	6666 elementos que no podemos asegurar que sean insolventes	26,54%	
Total	8389	$\overline{PX} - \underline{PX} = 4897$		26,28%	41,63%

* Calidad de la aproximación (Elementos correctamente clasificados por P).
Fuente: Elaboración propia

Tabla 6: Calidad de la clasificación considerando todas las variables del estudio

	Total	\underline{PX}	\overline{PX}	Precisión	Calidad*
X= X ₀	3872	3100 elementos con toda seguridad son no insolventes	4646 elementos que no podemos asegurar que sean no insolventes	66,42%	
X=X ₁	4517	3743 elementos con toda seguridad son insolventes	5289 elementos que no podemos asegurar que sean insolventes	70,77%	
Total	8389	$\overline{PX} - \underline{PX} = 1546$		68,88%	81,57%

* Calidad de la aproximación (Elementos correctamente clasificados por P).
Fuente: Elaboración propia

4.4. Cuantificación de la importancia de cada una de las variables no financieras consideradas

Para ver lo que cada una ha aportado al aumento de la calidad del modelo, hemos tomado el análisis realizado con solo las variables financieras y lo hemos replicado nueve veces, incluyendo en cada una de ellas una y sola una de las variables no financieras consideradas. Después hemos cotejado los resultados obtenidos con los de partida. Presentamos en la tabla 7 los aumentos en la calidad de la clasificación, en puntos porcentuales, que se han conseguido con cada una y cómo ha afectado al grupo de mayor interés, cómo es el de las microempresas familiares insolventes.

Las dos variables que mayor calidad aportan en ambos criterios son Indowe y Late_Filing_Days_2, por este orden.

¹⁰ Puede resultar más interesante para el investigador la detección de las variables en el grupo de empresas insolventes.

El aumento de la calidad es mayor, si solo consideramos la clasificación de las insolventes, para las variables *CCJ_Value*, *Charge_Flag*, *Auditor_Changed* y *AQ_NO_CLEAN*, ocurriendo lo contrario para las cinco restantes.

Tabla 7: Aumento en la calidad de la clasificación total y en la precisión para las empresas no insolventes debido a la introducción de una variable no financiera.

Variable introducida	Aumento en la calidad	Aumento en la precisión para las insolventes
<i>Ln_Age</i>	6,44	5,58
<i>Late Filing_Days_1</i>	7,22	7,13
<i>Late Filing_Days_2</i>	10,55	9,85
<i>CCJ_Value</i> medio	5,73	6,72
<i>Charge_Flag</i>	2,15	2,48
<i>Auditor_Changed</i>	4,7	4,83
<i>AQ_CLEAN</i>	1,24	1,24
<i>AQ_NO_CLEAN</i>	0,97	0,99
<i>Indowe</i>	15,28	14,39

Fuente: Elaboración propia

5. Conclusiones

En este trabajo hemos analizado la importancia que tiene un conjunto de variables en la determinación de la quiebra empresarial, específicamente las variables no estrictamente financieras en la determinación de la quiebra de las microempresas. Para ello hemos utilizado la técnica de los Rough Sets, encuadrada dentro de los métodos de Inteligencia Artificial, que nos permite trabajar sin necesidad de que las variables de partida estén sometidas a ninguna hipótesis inicial.

De una base de datos que contiene información económico-financiera de 4.813.391 PYMEs del Reino Unido entre los años 1999-2009, hemos extraído una muestra de 8.389 microempresas de carácter familiar con datos relativos a los años 2007, 2008 y 2009, balanceada respecto al año y su situación de insolvencia o no, a la que hemos aplicado la metodología expuesta.

Para ello hemos considerado catorce variables financieras, ratios comúnmente usados como predictores de insolvencia, y once variables no financieras, tales como la edad de la empresa y si han sido o no auditadas favorablemente. Disponemos así de una tabla de veinticinco variables que hemos observado en 3.872 microempresas no insolventes y 4.517 insolventes, tratando de determinar cuál o cuáles de ellas distinguen mejor, en términos de aciertos en la clasificación.

En un primer análisis hemos determinado cuantas y cuáles de ellas eran prescindibles, en el sentido de que no proporcionaban información añadida, así hemos podido prescindir de la variable *N0_CCI*, que nos informaba acerca de si la empresa se había visto inmersa o no en proceso judiciales, y *AUDITED*, que nos informaba si la empresa había sido o no auditada, y nos hemos restringido a 23 variables. A ellas hemos aplicado la metodología desarrollada en el epígrafe dos, que es una adaptación de la técnica de clasificación basada en el procedimiento Rough Set de aprendizaje inductivo. De los resultados obtenidos destacamos la importancia que tiene la consideración de variables que no sean estrictamente financieras. Así, los porcentajes de clasificación correcta mejoran considerablemente cuando las incluimos, pasando estos de un 41,63% cuando solo consideramos variables financieras, a un 81,57% cuando las introducimos en el modelo.

La técnica, además, nos permite cuantificar la mejora que proporciona cada variable no financiera considerada individualmente, encontrando que las dos variables que más aportan en la clasificación correcta total son *Indowe* y *Late Filing_Days_2*, con aumentos respectivos de 15,28 y 10,55 puntos porcentuales. La primera de ellas es una variable macroeconómica que hace referencia a la situación económica-financiera del sector donde opera la empresa, y, que nos lleva a concluir la necesidad de

elaborar modelos para las microempresas específicos por sectores de actividad. La segunda, *Late Filing Days_2*, hace referencia al retraso en la presentación de las cuentas tal como exige la normativa, una variable fácilmente observable y que ha resultado determinante en la clasificación de la microempresa como fallida.

En resumen, la técnica de los Rough Sets puede ayudarnos a la preselección de las variables más importantes para una buena clasificación y, para el caso particular de las microempresas, nos ha mostrado la significatividad de variables que no son estrictamente financieras.

Referencias bibliográficas

1. M.A. McPherson. Growth of micro and small enterprises in southern Africa *Journal of Development Economics* 48 (1996) 253–277.
2. D.C. Mead, y C. Liedholm. The Dynamics of Micro and Small Enterprises in Developing Countries. *World Development* 26 (1998) 161–74. Elsevier.
3. E.I. Altman, y G. Sabato. Modeling Credit Risk for SMEs: Evidence from US Market *ABACUS* 43(3) (2007) 332–357.
4. R. Carter y H.V. Auken. Small Firm Bankruptcy. *Journal of Small Business Management* 44(2006) 493–512.
5. A.N. Berger y W.S. Frame. Small Business Credit Scoring and Credit Availability. *Journal of Small Business Management* 45 (2007) 5–22.
6. S. Claessens, S. Djankov y L. Lang. The separation of ownership and control in East Asian corporations. *Journal of Financial Economics* 58 (2000) 81–112.
7. J. Fan y T. Wong. Corporate ownership structure and the informativeness of accounting earnings in East Asia. *Journal of Accounting and Economics* 33 (2001) 401–426.
8. B. Morten, K.M. Nielsen, F. Perez-Gonzalez y D. Wolfenzon. Inside the Family Firm: The Role of Families in Succession Decisions and Performance. *The Quarterly Journal of Economics*, 122(2) (2007) 647–691.
9. Z. Pawlak. Rough sets. *International Journal of Information and computer Sciences* 11 (1982) 341–356.
10. W. Beaver. Financial Ratios as Predictors of Failure, Empirical Research in Accounting: Selected Studied. *Journal of Accounting Research* 4 (1966) 71–111.
11. E.I. Altman. Financial ratios, discriminant analysis and prediction of corporate bankruptcy. *Journal of Finance* 23(4) (1968) 589–609.
12. E. B. Deakin. A discriminant analysis of predictors of business failure. *Journal of Accounting Research* 10 (1) (1972) 167–179.
13. R. Edmister. An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *Journal of Financial and Quantitative Analysis* 7(2) (1972) 1477–1493.
14. M. Blum. Failing company discriminant analysis. *Journal of Accounting Research* 12 (1974) 1–25.
15. Comision Europea. Annual report on EU Small and Medium sized Enterprises 2010/2011.
16. BIS. (2011). *Simpler reporting for the smallest businesses*. Discussion paper for department of Business Innovations and Skills (BIS). Electronic document (<http://www.bis.gov.uk/assets/biscore/business-law/docs/s/11-1100x-simpler-reporting-for-smallest-businesses-discussion-paper>).
17. S. Myers. Determinants of corporate borrowing. *Journal of Financial Economics* 5(2) (1977) 147–175.
18. E.I. Altman, G. Sabato y N. Wilson. The value of non-financial information in small and medium-sized enterprise risk management. *Journal of Credit Risk* 6(2) (2010) 95-127.
19. J.A. Ohlson. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18(1) (1980) 109-131.
20. R.J. Taffler. Empirical Models for the Monitoring of UK Corporations. *Journal of Banking and Finance*, 8(2) (1984) 199–227.

21. M.E. Xin Zmijewsk. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research* 22 (1984) 59–82.
22. M. Dietsch y J. Petey. Should SME Exposures be treated as Retail or Corporate Exposures? A Comparative Analysis of Default Probabilities and Asset Correlation in French and German SMEs. *Journal of Banking and Finance* 28(4) (2004) 773–788.
23. T. Jacobson, R. Kindell, J. Linde y K. Roszbach. Firm Default and Aggregate Fluctuations. *Working Paper*, Sveriges Riskbank, no.226, September 2008 Available at SSRN: <http://ssrn.com/abstract=1471254>.
24. D.K. Liou y M. Smith. Macroeconomic Variables in the Identification of Financial Distress, *Working Paper*, May 2006, Available at SSRN: <http://ssrn.com/abstract=900284>.
25. J. Hudson. The Age, Regional and Industrial Structure of Company Liquidations. *Journal of Business Finance and Accounting* 14 (2) (1987) 199–213.
26. D. Citron, y M. Wright. Bankruptcy costs, leverage and multiple secured creditors: the case of managemnet buy-outs. *Accounting and Business Research*, 38(1) (2008) 781–799.