

PREDICCIÓN DE LA INSOLVENCIA EMPRESARIAL: USO DE  
BÚSQUEDA TABÚ PARA SELECCIÓN DE RATIOS EXPLICATIVOS

IE Working Paper

DF8-121

15-12-2005

Joaquín Pacheco

Silvia Casado

Olga Gómez

Laura Núñez

Department of Applied  
Economics  
University of Burgos, Spain  
jpacheco@ubu.es /  
scasado@ubu.es /  
olga\_gomezpalacios@yahoo.es

Department of Finance  
Instituto de Empresa  
Business School  
Madrid, Spain  
laura.nunez@ie.edu

**Abstract**

En este trabajo se diseña un método de selección de variables para análisis discriminante que se aplica al problema de predicción de la insolvencia. El método diseñado se basa en la estrategia metaheurística Búsqueda Tabú (BT). La BT lee una solución inicial y a partir ella explora de forma sistemática e “inteligente” el conjunto de soluciones. Como se verá la BT consta de un procedimiento básico y de dos estrategias complementarias, denominadas Intensificación y Diversificación, que ayudan a hacer la búsqueda más robusta. Se estudia la eficacia de este método comparándolo con los algoritmos tradicionales de selección de variables empleados en el análisis discriminante (stepwise, backward y forward).

**Palabras clave**

Insolvencia empresarial, ratios contables, selección de variables, análisis discriminante, Búsqueda Tabú.

**Clasificación**

JEL: G33, C49.



## 1.- Introducción

El intento de extraer información relevante a partir de los datos financieros y contables de la empresa para predecir su comportamiento futuro (performance, cotización, etc.) ha constituido un objetivo de la investigación y la práctica económica desde hace ya muchos años. La resolución de este problema de extracción de información a partir de grandes volúmenes de datos se denomina “minería de datos” o “*data mining*” y su campo de estudio abarca prácticamente todas las áreas de la empresa, marketing, finanzas, planificación, producción, etc., así como todos los sectores empresariales, banca, seguros, comercio, industria, etc. (véase Brachman et al. 1996). Existen multitud de técnicas que pueden ser aplicadas a la resolución de problemas de minería de datos, entre las más comunes destacan las redes neuronales, los árboles de decisión, los algoritmos genéticos, el método del vecino más cercano y las reglas de inducción. Una de las aproximaciones más habituales al problema de la minería de datos es la extracción de la información relevante para realizar clasificaciones. Ejemplos típicos de problemas de clasificación en el ámbito de la empresa incluyen concesiones de créditos, seguridad en las operaciones comerciales, fraude con tarjetas de crédito, selección de productos, estimaciones de riesgos, selección de personal o predicción de insolvencia. Generalmente las herramientas de minería de datos extraen la información mediante la selección de aquellas variables que son relevantes para el problema en cuestión, obviando por tanto aquellas irrelevantes o redundantes. En este trabajo proponemos el uso de técnicas de minería de datos, y en concreto de búsqueda tabú, para seleccionar información relevante de los estados contables de las empresas con el objetivo de predecir la insolvencia empresarial. La predicción de la insolvencia empresarial se estructura como un problema de clasificación, en el que clasificar implica predecir la clase a la que pertenece cada empresa (solvente o insolvente).

La importancia de la insolvencia<sup>1</sup> empresarial, por sus efectos nocivos para accionistas, acreedores, empleados y para la economía en general, justifica los esfuerzos que desde los años 60 se vienen realizando por encontrar variables contables o financieras que permitan anticipar consistentemente la misma. Desde los pioneros trabajos de Beaver (1966) y Altman (1968), se ha desarrollado una considerable literatura en este campo. El análisis discriminante constituyó la técnica predominante durante la década de los 70 y aún hoy en

---

<sup>1</sup> A lo largo de este trabajo cuando hablamos de insolvencia nos referimos indistintamente a los estados de suspensión de pagos y quiebra.

día continua siendo una de las más empleadas (véase Altman 1968 y 1983; y Altman y otros 1994) a pesar de las restricciones que exige en cuanto a la estructura de los datos. Restricciones como las referidas al principio de normalidad de la distribución de los ratios, o a la igualdad de matrices de varianzas-covarianzas para los ratios de ambos tipos de compañías, solventes e insolventes, que son habitualmente violadas (véase Ezzamel y Molinero 1987; y Richardson and Davison, 1983 and 1984). Durante los años 80 comenzaron a usarse otras técnicas como la regresión logística (Ohlson 1980) o los modelos probit (Zmijewski), que junto con el análisis discriminante, constituyen las técnicas más ampliamente empleadas en la actualidad para modelizar la insolvencia empresarial y cuya capacidad de predicción resulta muy similar. En el ámbito español podemos citar como referencias los trabajos de (Gallego et al.,1997) en lo referido al estudio de la insolvencia en empresas no financieras, los de (Laffarga y otros 1985, 1986a, 1986b, 1987, 1990) sobre insolvencia en entidades bancarias, o los de (Sanchis y otros 2003) y (Gallego y otros., 2002) en el ámbito del sector asegurador.

En este trabajo se propone un método de selección de ratios contables basado en búsqueda tabú, que es incorporado en un modelo de análisis discriminante de predicción de insolvencia empresarial para una muestra de 198 empresas españolas clasificadas en solventes (131) e insolventes (67), para cada una de las cuales se tiene 141 ratios contables. La Búsqueda Tabú lee una solución inicial y a partir ella explora de forma sistemática e “inteligente” el conjunto de soluciones. Como se verá la búsqueda tabú consta de un procedimiento básico y de dos estrategias complementarias, denominadas Intensificación y Diversificación, que ayudan a hacer la búsqueda más robusta. Se estudia la eficacia de este algoritmo comparándolo con los métodos tradicionales de selección de variables que son empleados habitualmente en el análisis discriminante, como *stepwise*, *backward* y *forward*, y que son los que están incorporados en la mayoría de los paquetes estadísticos, como SPSS o BMDP. Como se va a comprobar, nuestro método va a ser más eficaz.

El trabajo se estructura de la siguiente forma. En la sección 2 se describe la muestra de empresas y ratios. La sección 3 presenta la formalización del problema de selección de variables para clasificación en el análisis discriminante. La sección 4 describe el algoritmo de Búsqueda Tabú que vamos a emplear para seleccionar variables en el análisis discriminante. La sección 5 compara los resultados obtenidos por el análisis discriminante en la predicción de la insolvencia empresarial con cada uno de los algoritmos (búsqueda

tabú, stepwise, forward y backward). Finalmente la sección 6 presenta las principales conclusiones del trabajo.

## 2.- Muestra de empresas y ratios

### 2.1 Empresas

La muestra empleada esta compuesta de 198 empresas españolas, de las que aproximadamente un tercio - 67 – suspendieron pagos o instaron la quiebra durante el año 2003. El resto – 131 – se encontraban activas en diciembre de dicho año. Las empresas fueron seleccionadas de la Base de Datos SABI, que comprende todas las empresas que depositan sus cuentas en el Registro Mercantil español. La selección se hizo de forma aleatoria para cada grupo (activas / en suspensión de pagos o quiebra), contemplando sólo las formas jurídicas de sociedad limitada y sociedad anónima, y aceptando sólo aquellas que tuvieran la información completa o a falta de muy pocos valores para los tres años anteriores. La tabla 1 muestra los datos de las empresas (solventes e insolventes) por sectores.

<b>CNAE-93 Rev. 1</b>	<b>Insolventes</b>	<b>Solventes</b>
01 - Agricultura y Ganadería	0	3
02 - Forestal	0	1
15 - industria Alimentación y Bebidas	5	6
17 - Ind. Textil	3	1
18 - Ind. Confección	1	1
19 - Fabr. Calzado	0	1
20 - Ind. Madera y Corcho	1	2
21 - Ind. del Papel	0	1
22 - Edición y Artes Gráficas	2	3
24 - Industria Química	0	4
25 - Fabric. Prod. Caucho y Plástico	1	2
26 - Fabr. Otros Prod. Minerales	0	1
27 - Metalúrgica	2	1
28 - Fabr. Pro. Metálicos	4	3
29 - Construcc. Maquinaria	5	3
31 - Fabr, Maquinaria Eléctrica	2	0
33 - Fabr. Equipo Médico	0	1
34 - Fabr,. Vehículos Motor	0	1
35 - Fabr. Otro Material Transporte	1	0
36 - Fabr. Muebles; Otras Industrias	4	3
41 - Captación, Depuración y Distrib. Agua	0	1
45 - Construcción	10	16
50 - Venta y Repara. Vehículos Motor	0	5
51 - Comercio al Por Mayor	12	16

52 - Comercio al Por Menor	7	11
55 - Hostelería	0	4
60 - Transporte Terrestre	0	2
61 - Transporte Marítimo	0	1
63 - Activ. Auxiliares de Transporte	1	1
65 - Intermed. Financiera (excepto seguros)	0	1
70 - Promoción Inmobiliaria	2	16
74 - Otras activ. Empresariales	2	12
80 - Educación	2	0
85 - Activ. Sanitarias y Veterinarias	0	2
92 - Act. Recreativas, culturales y deportivas	0	4
93 - Act. De Servicios Personales	0	1
<b>Total</b>	<b>67</b>	<b>131</b>

**Tabla 1.** *Distribución de la muestra de empresas por sectores*

## 2.2 Ratios

Se seleccionaron para cada empresa 36 ratios de los publicados en SABI – la base de datos empleada en el estudio – para cada uno de los tres años anteriores al 2003 (t, t-1 y t-2), lo que supone un total de 108 datos por empresa. Prácticamente se incluyeron todos los ratios publicados en SABI para las empresas españolas, a excepción de algunos para los que recurrentemente la información no estaba disponible, como por ejemplo el ratio de “plazo de pago a proveedores”, que desgraciadamente tuvo que excluirse. A su vez, se añadieron 11 nuevos ratios relativos a variaciones temporales para 11 de los 36 ratios previamente seleccionados – se calcularon tres variaciones para cada uno de estos ratios: entre el año t y t-1, entre t-1 y t-2 y entre t y t-2. Por tanto los datos totales seleccionados para cada empresa ascienden a 141. La introducción de las variaciones temporales de los ratios en el análisis de insolvencia no es una práctica habitual, sin embargo, puede resultar de gran interés, dado que es sabido que la distribución de los ratios de las empresas solventes permanece constante en el tiempo, mientras que la de las empresas insolventes varía significativamente, a causa del deterioro que sufren estos, véase (Beaver, 1966). Teniendo en cuenta este factor la variación en determinados ratios podría tener mayor poder predictivo que el propio nivel del ratio y por otra parte, parece a priori, que dicha variación pueda tener también mayor independencia con relación a los sectores de actividad de las empresas, que los niveles absolutos de los propios ratios. La tabla 2 presenta la definición del conjunto de ratios financieros empleados.

Ratios de Actividad	
Crecimiento de las ventas (%)	$[(Ventas_t - Ventas_{t-1}) / Ventas_{t-1}] \times 100\%$
Rotación de Activos	$Ventas / Activo\ Total$
Productividad	$[Ingresos\ Explot. - Consumos\ y\ Gastos\ Explot.] / Gastos\ de\ Personal$

Gastos de personal (%)	$[\text{Gastos de Personal} / \text{Ingresos de Explotación}] \times 100\%$
Crecimiento Valor Añadido (%)	$[(\text{Valor Añadido}_t - \text{Valor Añadido}_{t-1}) / \text{Valor Añadido}_{t-1}] \times 100\%$
Margen Operativo (%)	$[\text{Resultado antes de Impuestos} / \text{Ingresos de Explotación}] \times 100\%$
Rotación de activos netos	$\text{Ingresos de Explotación} / \text{Fondos Permanentes}$
<b>Ratios de Rentabilidad</b>	
ROCE	$[(\text{Resultado antes de Impuestos} + \text{Gastos Financieros}) / \text{Fondos Permanentes}] \times 100\%$
ROA	$[\text{Resultado del Ejercicio} / \text{Activo Total}] \times 100\%$
ROA antes de impuestos	$[\text{Resultado antes de Impuestos} / \text{Activo Total}] \times 100\%$
ROE	$[\text{Resultado del Ejercicio} / \text{Fondos Propios}] \times 100\%$
ROE antes de Impuestos	$[\text{Resultados antes de Impuestos} / \text{Fondos Propios}] \times 100\%$
Gastos financieros (%)	$[\text{Gastos Financieros} / \text{Ventas}] \times 100\%$
<b>Ratios de Equilibrio</b>	
Fondo maniobra (€)	$\text{Fondos Propios} + \text{Provisiones R y G} + \text{Acreedores LP} - \text{Inmovilizado}$
Necesidad fondo maniobra (€)	$[\text{ADNE} + \text{Gastos a Distribuir} + (\text{Activo Circulante} - \text{IFT-Tesorería})] - [\text{Ingresos a Distribuir} + (\text{Acreedores CP} - \text{Entidades de Crédito})]$
Tesorería (€)	$\text{IFT} + \text{Tesorería} - \text{Entidades de Crédito}$
Equilibrio	$[\text{Fondos Propios} + \text{Provisiones R y G} + \text{Acreedores LP}] / \text{Inmovilizado}$
<b>Ratios de Equilibrio Cinéticos</b>	
Fondo de maniobra (días)	$[\text{Fondo de Maniobra} / \text{Ventas}] \times 360$
Necesidad fondo de maniobra (días)	$[\text{Necesidad Fondo Maniobra} / \text{Ventas}] \times 360$
Tesorería (días)	$[\text{Tesorería} / \text{Ventas}] \times 360$
Crédito de Clientes (días)	$[\text{Deudores} / \text{Ingresos de Explotación}] \times 360$
Credito de Clientes por Ventas (días)	$[\text{Deudores} / \text{Ventas}] \times 360$
<b>Ratios de Solvencia</b>	
Endeudamiento (%)	$[(\text{Pasivo Total} - \text{Fondos Propios}) / \text{Pasivo Total}] \times 100\%$
Ratio de Solvencia (%)	$[\text{FP} / \text{Activo Total}] \times 100\%$
Fondos Propios sobre Permanentes (%)	$[[\text{FP} / (\text{FP} + \text{Acreedores LP} + \text{Provisiones R y G})] \times 100\%$
Capacidad devolución	$[\text{Acreedores LP y CP} / (\text{Ventas} + \text{Dotaciones} + \text{Var. Provisiones} + \text{PF})]$
<b>Ratios de Liquidez</b>	
Liquidez inmediata	$[\text{IFT} + \text{Tesorería}] / \text{Acreedores a CP}$
Liquidez corriente	$\text{Activo Circulante} / \text{Pasivo Líquido}$
Liquidez	$[\text{Activo Circulante} - \text{Existencias}] / \text{Pasivo Líquido}$
Cobertura de intereses	$\text{Resultado de Explotación} / \text{Gastos Financieros}$
<b>Ratios por empleado</b>	
Beneficio por empleado	$\text{Resultado antes de Impuestos} / \text{Número de Empleados}$
Ingreso por empleado	$\text{Ingresos de Explotación} / \text{Número de Empleados}$
Gastos de personal por empleado	$\text{Gastos de Personal} / \text{Número de Empleados}$
Fondos Propios por empleado	$\text{Fondos Propios} / \text{Número de Empleados}$
Fondo de Maniobra por empleado	$\text{Fondo de Maniobra} / \text{Número de Empleados}$
Activos Totales por empleado	$\text{Activo Total} / \text{Número de Empleados}$

**Tabla 2.** Ratios seleccionados (SABI).

\* abreviaturas: IFT (Inversiones Financieras Temporales); ADNE (Accionistas por desembolsos no exigidos); CP (Corto Plazo); LP (Largo Plazo); PF (Provisiones Financieras); Ry G (Riesgos y Gastos)

### 3.- Formalización del problema de selección de variables para clasificación usando análisis discriminante.

#### 3.1.- Revisión de la literatura

Aunque la investigación en el ámbito del problema de selección de variables para clasificación se inició a principios de la década de los sesenta (véase Lewis 1962; y Sebestyen 1962), ha sido en los últimos años cuando la literatura en este campo de conocimiento ha experimentado su mayor apogeo, generado posiblemente por el desarrollo de la capacidad de procesamiento y almacenamiento de los ordenadores y la generalización del acceso a volúmenes ingentes de información a través de Internet. Muchos de los trabajos de los últimos años sobre selección de variables están relacionados con la medicina, la biología y la genética (véase Lee y otros 2003; Shy y Suganthan 2003; y Tamoto y otros 2004) y en ellos se aplican metaheurísticas novedosas (algoritmos genéticos, búsqueda tabú, búsqueda dispersa, etc.) para la resolución del problema de clasificación. Sin embargo, es curioso como en el ámbito de la empresa, en el que disponemos cada vez de mayores volúmenes de datos e información apenas se encuentran trabajos que apliquen este tipo de técnicas.

Siguiendo a Liu y Motoda (1998) la selección de un subconjunto de variables para resolver problemas de clasificación conlleva diversas ventajas como la reducción del coste en la adquisición de los datos, la mejora en la comprensión del modelo final del clasificador o la eficiencia y eficacia del mismo. Desde un punto de vista computacional la búsqueda del subconjunto de variables relevantes es un problema NP-Hard, (véase Kohavi 1995; y Cotta y otros 2004), es decir que no existe ningún método que garantice la solución óptima en un tiempo polinomial en el tamaño de problema (NP = *No polinomic*). Los métodos aproximados o heurísticos han demostrado poder resolver este tipo de problemas en tiempos de computación razonables. Así destacan los trabajos de Bala y otros (1996), Jourdan y otros (2001), Oliveira y otros (2003), Inza y otros (2001a, 2001b) y Wong y Nandi (2004) que desarrollan algoritmos genéticos y el más reciente de García y otros (2003) que presenta un método basado en Búsqueda Dispersa.

Estos métodos buscan los subconjuntos de variables con mayor capacidad clasificatoria según diferentes criterios, pero ninguno de ellos está enfocado al uso posterior de las variables seleccionadas en análisis discriminante, una de las técnicas más empleadas para resolver problemas de clasificación. En la literatura para este específico propósito existe el conocido método *Stepwise* (Efroymsen, 1960) y variantes como O'Gorman (2004), así



como los métodos *Backward* y *Forward*. Estos son sencillos procedimientos de selección basados en criterios estadísticos (Landa de Wilks, F de Fisher, etc.) y que han sido incorporados en algunos de los paquetes estadísticos más conocidos como SPSS, BMDP, etc. Como destaca Huberty (1989) y Salvador (2000) estos métodos no son muy eficaces, y cuando las variables originales son muchas raramente alcanzan el óptimo.

### 3.2.- Modelización del problema

El problema de seleccionar el subconjunto de variables con más capacidad clasificatoria se puede formular de la forma siguiente: Considérese  $V$  un conjunto de  $m$  variables, simplificando  $V = \{1, 2, \dots, m\}$ , y  $A$  un conjunto de  $n$  casos. Para cada caso se conoce además la clase a la que pertenece. Considérese un valor predeterminado  $p \in N$ ,  $p < m$ . Hay que encontrar el subconjunto  $S \subset V$ , de tamaño  $p$  con la mayor capacidad clasificatoria en análisis discriminante,  $f(S)$ .

Más concretamente la función  $f(S)$  se define como el porcentaje de aciertos en  $A$  del clasificador de Fisher (Hair y otros 2001) obtenido a partir de las variables de  $S$ . Obsérvese que este clasificador se obtiene a partir de la distancia de Mahalanobis a los centroides de cada grupo. Por tanto la evaluación de  $f(S)$  se puede hacer en  $\theta(p^2) + \theta(n)$  operaciones: el primer sumando corresponde al cálculo de la inversa de la matriz de varianzas-covarianzas de  $S$ , y el segundo a la clasificación de los casos. En definitiva, que a la vez que se calcula  $f(S)$  se obtiene como paso previo el clasificador lineal de Fisher.

La solución inicial se genera mediante un método constructivo y un método de mejora, es decir cada solución  $S$  generada por el método constructivo es mejorada posteriormente por un procedimiento de búsqueda local.

#### a) método constructivo

La solución inicial se genera mediante un método constructivo. El funcionamiento del método constructivo es el siguiente: Se parte de solución inicial vacía ( $S = \emptyset$ ); en cada paso o iteración se elige un elemento que no esté en la solución y se añade; el proceso finaliza cuando la solución esta completa ( $|S| = p$ ). Para determinar que elemento se añade se hace uso de una función guía (función voraz)  $R_j$ ,  $j=1..m$ , que sirve como indicador de la bondad o

conveniencia de cada elemento candidato a entrar en la solución, eligiendo el elemento de mayor valor según esa función guía.

El funcionamiento del método constructivo en pseudocódigo es el siguiente:

*Iniciar*  $S = \emptyset$

*Repetir*

*Calcular*  $R_j, \forall j \in V - S$

*Determinar*  $R_{max} = \max \{ R_j / j^* \in V - S \}$

*Make*  $S = S \cup \{j^*\}$

*hasta*  $|S| = p$

En este trabajo se van a contrastar 3 funciones voraces o formas de calcular  $R_j$  diferentes, basadas en la F de Snedecor,  $\lambda$  de Wilks y la propia función objetivo  $f$ . Concretamente, sea una solución parcial  $S, \forall j \in V - S, R_j$  se puede calcular como sigue:

- a) Se calculan los residuos en un modelo de regresión lineal donde  $j$  es la variable dependiente y las independientes son los elementos de  $S$ . El valor  $R_j$  va a ser el cociente varianza entre grupos/varianza intragrupos de estos residuos. El objeto de la regresión es eliminar información redundante. En la iteración inicial ( $S = \emptyset$ )  $R_j$  va a ser el cociente varianza entre grupos/varianza intragrupos de la variable original  $j$ .
- b) Se determinan las matrices de varianzas-covarianzas entre grupos e intragrupos del conjunto  $S \cup \{j\}$ , que denotamos por  $B_j$  y  $W_j$ .  $R_j$  se calcula como el determinante de  $W_j^{-1} \cdot B_j$ .
- c)  $R_j$  se calcula como  $f(S \cup \{j\})$ .

Las dos primeras funciones se corresponden con la F de Snedecor y la inversa de la  $\lambda$  de Wilks. Estas coinciden con los criterios usados por diversos paquetes estadísticos como SPSS y BMDP en los métodos de selección para análisis discriminante que incorporan. El tercero evalúa directamente la capacidad clasificatoria tal y como se ha definido al comienzo de esta sección. En la próxima sección (la 3.3) mostramos como en efecto este último criterio es el más adecuado.

#### **b) mejora de la solución inicial**

Tal y como se ha comentado antes, a cada solución completa  $S$  generada por el método constructivo se la mejora posteriormente por un sencillo procedimiento de búsqueda local. En este caso, cada paso de la búsqueda local va a consistir en intercambiar una variable

que esté dentro de la solución por otra que este fuera. Más concretamente, sea  $S$  una solución se define

$$N(S) = \{ S' / S' = S \cup \{j\} - \{j'\}, \forall j \in S, j' \notin S \}$$

El procedimiento de búsqueda local se puede describir como sigue

*Leer Solución Inicial S*

*Repetir*

*Hacer valor\_ant = f(S)*

*Buscar  $f(S^*) = \max \{ f(S') / S' \in N(S) \}$*

*Si  $f(S^*) > f(S)$  entonces hacer  $S = S^*$*

*hasta  $f(S^*) \leq \text{valor\_ant}$*

Como se observa el procedimiento finaliza cuando ningún intercambio produce mejora.

### 3.3.- Resultados computacionales del análisis de las funciones guías o voraces

A continuación comparamos la eficacia de las 3 funciones guías que se proponen para el constructivo descrito en la sección anterior 2.2: F Snedecor,  $\lambda$  de Wilks y capacidad clasificatoria  $f$ . Para realizar dicha comparación usamos la tabla de 141 ratios financieros para el conjunto total de empresas en la muestra, 198. De esta tabla consideramos tablas de menores dimensiones con un número  $m$  de ratios financieros con el objetivo de tener múltiples tablas de datos para poder realizar las pruebas computacionales. Así se consideran los siguientes valores de  $m$ ,  $m = 40$  (correspondientes a los primeros 40 ratios financieros), 65, 90, 105 y 120. Para cada valor de  $m$  se consideran diferentes valores de  $p$  (número de ratios a seleccionar). Para cada problema se ejecuta el constructivo propuesto en la sección 3.2 usando cada una de las funciones guías. En la tabla 3 se muestra la capacidad clasificatoria  $f$  final obtenida.

$M$	$P$	F Snedecor	Inversa $\lambda$	$f$
40	4	0,71212121	0,71212121	0,77777778
40	5	0,72727273	0,72727273	0,77777778
40	6	0,73232323	0,73232323	0,78282828
40	7	0,73232323	0,73232323	0,76767677
40	8	0,72222222	0,72222222	0,76767677

<i>M</i>	<i>P</i>	F Snedecor	Inversa $\lambda$	<i>f</i>
65	6	0,74242424	0,74242424	0,78787879
65	7	0,76767677	0,76767677	0,78282828
65	8	0,77272727	0,75757576	0,77272727
65	9	0,78282828	0,76262626	0,76767677
65	10	0,77272727	0,78787879	0,76767677
90	8	0,75757576	0,75757576	0,83838384
90	9	0,77777778	0,77777778	0,79292929
90	10	0,77777778	0,77777778	0,80808081
90	11	0,78282828	0,78282828	0,7979798
90	12	0,77777778	0,77777778	0,84343434
105	10	0,76767677	0,76767677	0,81818182
105	11	0,76767677	0,76767677	0,79292929
105	12	0,77777778	0,77777778	0,82323232
105	13	0,78282828	0,78282828	0,81818182
105	14	0,78282828	0,78282828	0,85353535
120	12	0,81313131	0,78282828	0,82828283
120	13	0,81818182	0,81313131	0,82828283
120	14	0,7979798	0,79292929	0,84848485
120	15	0,81313131	0,78787879	0,82828283
120	16	0,81313131	0,78787879	0,82828283

**Tabla 3.** Comparación de funciones guía

En 22 de los 25 casos la propia capacidad clasificatoria *f* da lugar a mejores resultados que los otros 2 métodos más tradicionales. Estos obtienen resultados muy parecidos, aunque quizás los de la F de Snedecor sean ligeramente mejores. En este caso la propia función objetivo ha resultado ser la mejor función guía, pero hay que hacer notar que en muchos otros problemas esto no siempre tiene porque ser así.

#### **4.- Búsqueda Tabú como algoritmo de selección de variables para análisis discriminante.**

##### **4.1.- La búsqueda Tabú**

Los primeros trabajos sobre Búsqueda Tabú o Tabú Search (en adelante TS) con tal nombre se deben a Glover (1989 y 1990), que presentó sus principios básicos en un trabajo anterior, Glover (1986). La filosofía del TS se basa en la explotación de un conjunto de principios inteligentes, procedentes tanto del campo de la inteligencia artificial como de la optimización, para su aplicación a la resolución de problemas combinatorios complejos. Es un procedimiento iterativo que explora el espacio de búsqueda de todas las soluciones factibles a través de una secuencia de movimientos, utilizando el concepto de memoria (tomado del campo de la inteligencia artificial) con el fin de dirigir la búsqueda teniendo en cuenta la historia de esta. En este sentido se dice que la búsqueda es inteligente y que hay un cierto aprendizaje.

TS comienza como lo hacen los métodos habituales de búsqueda local, procediendo iterativamente de una solución  $S_i$  a otra  $S_j$  en el entorno o vecindario de la primera. Se permiten movimientos hacia soluciones del entorno aunque no sean tan buenas como la actual, de forma que se pueda escapar de óptimos locales y continuar la búsqueda de soluciones aún mejores. Pero este entorno es un entorno "reducido" que elimina o prohíbe, etiquetándolas como "tabú"<sup>2</sup>, las soluciones (o atributos de estas) visitadas por el algoritmo recientemente, con el objetivo de evitar que este cicle o se sitúe en óptimos locales, estimulando el descubrimiento de nuevas soluciones. A medida que la búsqueda avanza, las soluciones etiquetadas como "tabú" en iteraciones anteriores dejan de tener el estatus de "tabú" o prohibidas, ya que se considera que la búsqueda al cabo de un número determinado de iteraciones debe encontrarse en una región diferente del espacio de soluciones y por tanto ya no hay riesgo de que cicle. Este proceso se conoce como memoria a corto plazo. La condición de "tabú" puede ser ignorada bajo determinadas circunstancias dando lugar a los llamados "criterios de aspiración", que son precisamente los que fijan las condiciones en las que puede tomarse una solución aún teniendo esta estatus de "tabú" (por ejemplo si la solución "tabú" mejora la mejor solución guardada). Este mecanismo introduce

---

<sup>2</sup> El termino tabú hace referencia a *"un tipo de inhibición a algo debido a connotaciones culturales o históricas y que puede ser superada en determinadas condiciones..."*. (Glover, 1996).

cierta flexibilidad en la búsqueda. Por otro lado, TS se diferencia de otras metaheurísticas de búsqueda como los algoritmos genéticos porque limita en gran medida los procesos aleatorios, argumentando que una búsqueda inteligente debe estar basada en formas más sistemáticas de guía.

Para una revisión más amplia y detallada de los principios básicos del TS pueden encontrarse amplios tutoriales, que incluyen todo tipo de aplicaciones, en Glover y Laguna (1997 y 2002).

#### 4.2.- Algoritmo “básico” de Búsqueda Tabú

A continuación se describe un algoritmo básico de Búsqueda Tabú que usa los mismos movimientos vecinales que el procedimiento de búsqueda local descrito en el apartado 3.2. Estos movimientos consisten en intercambiar en cada paso un elemento que esta en la solución  $S$  por otro de fuera. Para evitar que el algoritmo cicle, cuando se lleva a cabo un movimiento, que consiste en intercambiar  $j$  de  $S$  por  $j'$  de  $V-S$ , se impide que el elemento  $j$  vuelva a  $S$  durante un número prefijado de iteraciones. Algunos movimientos “tabús” pueden ser permitidos en determinadas condiciones (“criterio de aspiración”), por ejemplo cuando mejora la mejor solución encontrada.

Concretamente, se define

$vector\_tabu(j)$  = número de la iteración en la que el elemento  $j$  salió de  $S$ .

El procedimiento de Búsqueda Tabú básico se describe a continuación en el algoritmo 1, donde  $S$  es la solución actual y  $S^*$  la mejor solución encontrada. El parámetro  $Tabu\_Tenure$  indica el número de iteraciones durante las cuales no se permite volver a la solución a un elemento que salió de la misma anteriormente. Tras realizar una serie de pruebas, el valor de  $Tabu\_Tenure$  más adecuado coincide con  $p$ , por lo que lo fijamos por tanto en  $p$ .

#### Algoritmo 1. Procedimiento básico de Búsqueda Tabú

*Leer solución inicial  $S$*

Hacer  $vector\_tabu(j) = - Tabu\_Tenure, j = 1..m; niter = 0, iter\_better = 0$  and  $S^* = S$

*Repetir*

$niter = niter + 1$

Calcular  $v_{jj'} = f(S \cup \{j'\} - \{j\})$

Determinar  $v_{j^*} = \max \{v_{j'} / \forall j' \in S, j' \notin S \text{ verificando:}$

$niter > \text{vector\_tabu}(j) + \text{Tabu\_Tenure}$  or

$v_{j'} > f(S^*)$  ('criterio de aspiración'))

Hacer  $S = S \cup \{j^*\} - \{j^*\}$  and  $\text{vector\_tabu}(j^*) = niter$

Si  $f(S) > f(S^*)$  entonces hacer:  $S^* = S$ ,  $f^* = f$  y  $\text{iter\_better} = niter$ ;

hasta  $niter > \text{iter\_better} + 2 \cdot m$

El procedimiento básico finaliza cuando trascurren  $2 \cdot m$  iteraciones sin mejora.

#### 4.3.- Algoritmo “completo” de Búsqueda Tabú: Intensificación y Diversificación

En muchas aplicaciones de búsqueda tabú el procedimiento básico descrito en el apartado anterior puede ser reforzado con estrategias de *Intensificación* y *Diversificación*. La *Intensificación* consiste en explorar con mas detalle las regiones donde se encuentran las mejores soluciones halladas hasta ese momento. La *Diversificación* se realiza después de la *Intensificación* y consiste en dirigir la búsqueda a regiones no exploradas. Un esquema del funcionamiento del algoritmo de Búsqueda Tabú completo se presenta en el algoritmo 2.

##### Algoritmo 2. Procedimiento completo de Búsqueda Tabú

*Construir solución inicial (constructivo + búsqueda local)*

*Repetir*

*Ejecutar Búsqueda Tabú Básico*

*Ejecutar Intensificación*

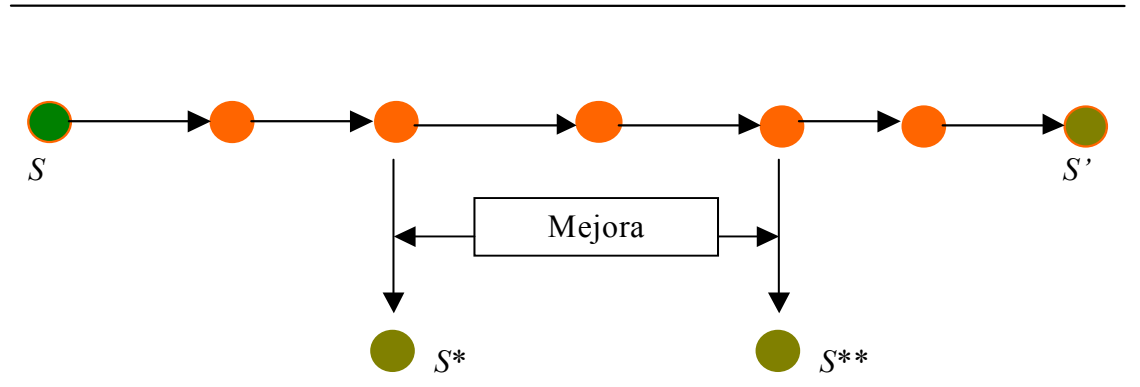
*Ejecutar Diversificación*

*Hasta alcanzar una condición de parada*

##### a) Intensificación

En nuestro caso la intensificación se realiza de la forma siguiente: se diseña una lista con los mejores óptimos locales encontrados hasta ese momento; cada par de soluciones diferentes de la lista se combina para generar una nueva solución. Para combinar las soluciones usamos una estrategia denominada Path Relinking. La idea básica es construir un “camino” de soluciones entre las dos soluciones iniciales uniéndolas. Un número de soluciones de ese “camino” o “cadena” son seleccionadas como nuevas soluciones. Estas soluciones intermedias se procura que estén

equidistantes entre sí. Posteriormente a esas soluciones intermedias se les aplica el método de mejora descrito en la sección 3.2.b basado en una búsqueda local. La figura 1 ilustra esta idea.



**Figura 1.-** Generación de nuevas soluciones usando Path Relinking

Desde cada par de soluciones, en la figura  $S$  y  $S'$ , se construye un camino que las une. Soluciones en posiciones intermedias en ese camino se seleccionan y se mejoran. De esta forma se generan las nuevas soluciones, (en la figura  $S^*$  y  $S^{**}$ ). Path Relinking es una estrategia que tradicionalmente se trató como un método de intensificación dentro de la Búsqueda Tabú aunque en los últimos años ha adquirido identidad propia. La idea que subyace es que en el camino entre dos buenas soluciones hay soluciones de igual o mejor calidad que las de partida. Ver Glover, Laguna and Martí (2000) para mayores detalles.

El número de óptimos locales de la lista se fija en 6. Inicialmente se crea con los 6 mejores óptimos locales obtenidos en la primera fase básica. Esta lista se actualiza después de cada fase básica y después de cada intensificación. Una vez actualizada tras cada fase de intensificación se eliminan los 3 peores óptimos locales de la lista. De esta manera en la siguiente actualización entrarán en la lista óptimos locales (al menos 3) de otras regiones del conjunto de soluciones. El objeto es asegurar en las siguientes intensificaciones combinar soluciones de diferentes regiones.

## b) Diversificación

La diversificación se realiza de la manera siguiente: se construye una nueva solución inicial. Concretamente se usa el método constructivo descrito en el apartado 3.2.a con la función guía modificada  $R'_j$  que se define como:



$$R'_j = R_j - \beta \cdot R_{\max} \frac{freq(j)}{freq_{\max}}$$

donde:

$freq(j)$  = numero de veces que el elemento  $j$  aparece en las soluciones visitadas en las fases básicas hasta el momento;

$freq_{\max} = \max \{ freq(j) / j \in V-S \}$ .

De esta manera se penaliza la elección de elementos que más han aparecido y se fuerza la elección de otros. En este trabajo se ha tomado el valor  $\beta = 1$ .

### 5.- Comparación de la búsqueda Tabú con los métodos tradicionales *Stepwise*, *Forward* y *Backward*

Finalmente, usando la tabla original de ratios contables entera ( $m = 141$  ratios), se compara el algoritmo de Búsqueda Tabú con los procedimientos tradicionales *Stepwise*, *Backward* y *Forward* que usan algunos de los paquetes estadísticos más conocidos como SPSS, BMDP, etc.

- El método *Forward* (o de *selección hacia delante*) comienzan eligiendo la variable que más discrimina según algún criterio. A continuación selecciona la segunda más discriminante y así sucesivamente. El algoritmo finaliza cuando entre las variables no seleccionadas ninguna discrimina de forma significativa.
- El método *Backward* (o de *eliminación hacia detrás*) actúa de forma inversa. Se comienza seleccionando todas las variables. En cada paso se elimina la menos discriminante. El algoritmo finaliza cuando todas las variables que permanecen discriminan significativamente.
- El método *Stepwise* o (*regresión por pasos*) utiliza una combinación de los dos algoritmos anteriores: en cada paso se introducen o eliminan variables dependiendo de la significación de su capacidad discriminatoria. Permite además la posibilidad de "arrepentirse" de decisiones tomadas en pasos anteriores, bien sea eliminando del conjunto seleccionado una variable

introducida en un paso anterior del algoritmo, bien sea seleccionando una variable previamente eliminada.

Como se ha comentado anteriormente, uno de los criterios más usado para la selección de variables es el de la  $\lambda$  de Wilks. La significación de la capacidad discriminante se mide con el estadístico F obtenido a partir de esta  $\lambda$ . Una explicación más detallada se puede encontrar en Salvador (2000). En la tabla 5 se muestran los resultados de ejecutar nuestro algoritmo de Búsqueda Tabú para valores de  $p$  desde 2 a 13. Así mismo también se han ejecutado los métodos *Backward*, *Forward* y *Stepwise*. También en la tabla 5 se muestran los resultados de las soluciones intermedias, correspondientes a esos mismos valores de  $p$ , obtenidos por la ejecución de los métodos *Backward* y *Forward*. El método *Stepwise*, en este caso, ha coincidido con los 6 primeros pasos del método *Forward*, es decir no ha habido eliminación de variables.

P	Forward	Backward	Constructivo + B. Local	Tabú Básico	Tabú completo
2	0,68181818	0,67676768	0,78787879	0,78787879	0,78787879
3	0,71212121	0,69191919	0,79292929	0,7979798	0,7979798
4	0,72222222	0,69191919	0,8030303	0,8030303	0,8030303
5	0,74242424	0,70707071	0,8030303	0,81313131	0,83333333
6	0,75757576*	0,71717172	0,8030303	0,81313131	0,84343434
7	0,77272727	0,72222222	0,8030303	0,84848485	0,84848485
8	0,76262626	0,71717172	0,8030303	0,82323232	0,85353535
9	0,77777778	0,72222222	0,8030303	0,82323232	0,85858586
10	0,78282828	0,72727273	0,8030303	0,84848485	0,85858586
11	0,79292929	0,72727273	0,8030303	0,86868687	0,86868687
12	0,76767677	0,73232323	0,82828283	0,85858586	0,85858586
13	0,7979798	0,72222222	0,84848485	0,87878788	0,87878788

\* solución final obtenida por el método *Stepwise*

**Tabla 4.** Comparación de búsqueda tabú con los métodos tradicionales *Forward* y *Backward*.

De la tabla 4 se pueden obtener las siguientes conclusiones:

- El método *Backward* parece funcionar claramente peor que el método *Forward* y *Stepwise*, al menos para este problema

- Nuestro método mejora significativamente las soluciones de los métodos tradicionales para cualquier valor de  $p$ . Obsérvese como con el método *Forward* apenas se consigue llegar al 79% de aciertos con 13 variables, mientras que con nuestro tabú se supera el 80% con 4 variables.

Considerando el doble objetivo de minimizar el número de variables y maximizar el porcentaje de aciertos es claro que las soluciones del modelo que proponemos dominan a las de los métodos tradicionales. Así se aprecia en el gráfico 1.

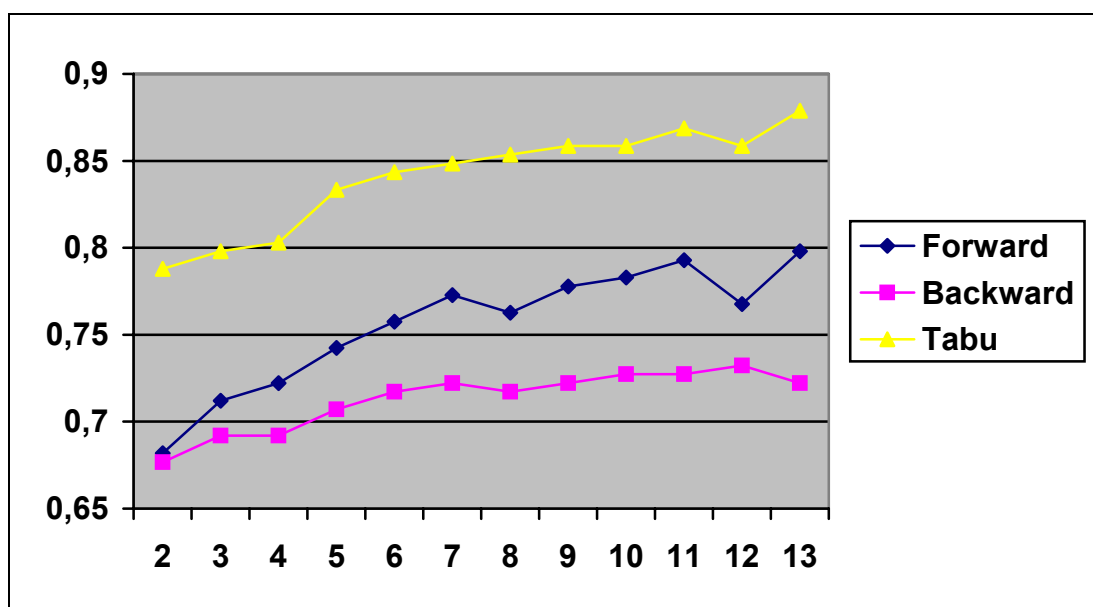


Gráfico 1: Comparación de Búsqueda Tabú con métodos tradicionales *Forward* y *Backward*.

En cuanto a las variables seleccionadas para predecir la insolvencia por nuestro algoritmo basado en TS, estas aparecen mostradas en la tabla 5 para los diferentes tamaños permitidos del subconjunto de variables a seleccionar. Destacan las siguientes observaciones:

- Con tan sólo dos variables (ROA y cobertura de intereses) se obtiene un porcentaje de aciertos cercano al 79%.
- La variable ROA aparece siempre en todos los subconjuntos (en total 12) de variables seleccionados, lo que parece indicar la importancia de este ratio en la predicción de la insolvencia empresarial.

- Otras tres variables, Crecimiento de las ventas, Endeudamiento y Productividad, parecen añadir también valor en la predicción de la insolvencia, ya que por un lado aparecen en 8, 7 y 6 de los subconjuntos seleccionados, y por otro lado se percibe un salto significativo en el porcentaje de aciertos cuando estas son incorporadas en el subconjunto de 5 y 6 ratios, pasando este al 83% y 84% respectivamente.
- Las restantes variables parecen tener una importancia limitada, ya que solo se incrementa el porcentaje de aciertos marginalmente con su incorporación. Con 13 variables el porcentaje de aciertos se sitúa en el entorno del 88%, un 4% por encima del alcanzado con menos de la mitad de las variables. Aunque este aumento de capacidad predictiva puede justificar el uso de este mayor número de variables en problemas reales, a nivel de comprensión del proceso de insolvencia resulta mucho más explicativo limitar el modelo a un máximo de 6 variables.

P	% aciertos	Variables seleccionadas
2	78,79 %	ROA_t Cobertura_intereses_t-2
3	79,80 %	ROA_t Productividad_t-1 Activos_totales_por_empleado(Mil)_t-1
4	80,30 %	ROA_t Cobertura_intereses_t-2 Activos_totales_por_empleado(Mil)_t-2 Gastos_de_personal_sobre_Ingresos_t-1
5	83,33 %	ROA_t Crecimiento de la cifra de ventas(%)_t Crecimiento de la cifra de ventas(%)_t-1 Endeudamiento(%)_t Capacidad de devolucion_t_vs_t-1
6	84,34 %	ROA_t Crecimiento de la cifra de ventas(%)_t Crecimiento de la cifra de ventas(%)_t-1 Productividad_t-2 Endeudamiento(%)_t Fondo de maniobra(dias)_t-1
7	84,85 %	ROA_t Crecimiento de la cifra de ventas(%)_t Crecimiento de la cifra de ventas(%)_t-1 Productividad_t-2 ROE_t Endeudamiento(%)_t Fondos Propios por empleado (Mil)_t
8	85,35 %	ROA_t Crecimiento de la cifra de ventas(%)_t Crecimiento de la cifra de ventas(%)_t-1 Productividad_t-2 ROE_t Endeudamiento(%)_t Fondo de maniobra(dias)_t-1 Liquidez inmediata_t_vs_t-1
	85,86 %	ROA_t

9		Crecimiento de la cifra de ventas(%)_t Crecimiento valor añadido(%)_t-1 Endeudamiento(%)_t Cobertura_intereses_t-2 Liquidez_t_vs_t-1 Necesidad fondo de maniobra(mil)t_vs_t-1 Periodo de cobro de clientes (dias)_t Fondos Propios sobre permanentes(%)_t_vs_t-2
10	85,86 %	ROA_t ROA_t-1 Crecimiento de la cifra de ventas(%)_t Activos Totales por empleado(Mil)_t-2 Crecimiento valor añadido(%)_t-1 Crecimiento valor añadido(%)_t-2 Endeudamiento (%)_t-2 Liquidez_corriente t_vs_t-1 Periodo de cobro de clientes (dias)_t Beneficio por empleado (Mil)_t-2
11	86,87 %	ROA_t Crecimiento de la cifra de ventas(%)_t Crecimiento de la cifra de ventas(%)_t-1 Productividad_t-2 ROE_t-1 Equilibrio_t Equilibrio_t-2 Capacidad de devolución_t_vs_t-1 Liquidez inmediata_t_vs_t-2 Liquidez inmediata_t-1 Fondos Propios por empleado (Mil)_t
12	85,86 %	ROA_t ROE_t-2 Gastos_de_personal_sobre_Ingresos_t Gastos_de_personal_sobre_Ingresos_t-1 Activos_totales_por_empleado(Mil)_t-1 Fondo_maniobra_por_empleado (Mil)_t-1 Necesidad fondo de maniobra(Mil)_t Necesidad fondo de maniobra(Mil)_t_vs_t-1 Tesoreria(mil)_t_vs_t-2 Tesoreria (dias)_t Ratio_solvenca_t_vs_t-1 Gasto medio por empleado (Mil)_t
13	87,88 %	ROA_t ROA_t-1 Crecimiento de cifra de ventas (%)_t Crecimiento de cifra de ventas (%)_t-1 Productividad_t-1 Endeudamiento_t Equilibrio_t Equilibrio_t-2 Necesidad fondo de maniobra (dias)_t-2 ROCE_t-1 ROCE_t-2 Fondo_maniobra_por_empleado (Mil)_t-1 Activos_totales_por_empleado(Mil)_t-1

**Tabla 5.** Variables seleccionadas por búsqueda tabú para diferentes valores de  $p$ .

## 6.- Conclusiones

En este trabajo se aborda el problema de predecir la insolvencia empresarial a partir de un conjunto de ratios financieros. Como en nuestro caso el número de ratios es alto un objetivo importante es seleccionar un subconjunto reducido de estos que permitan predecir la insolvencia empresarial de forma eficaz. Concretamente se propone un método de selección de ratios para su uso posterior en el análisis discriminante. Este método es un algoritmo basado en la estrategia metaheurística Búsqueda Tabú. Con la muestra de datos disponible, 198 empresas españolas y 141 ratios, este método ha resultado ser significativamente mejor que los usados tradicionalmente como *Stepwise*, *Backward* y *Forward*. Entendemos que este trabajo es interesante ya que una adecuada selección de ratios, entre otras ventajas, ayuda a obtener modelos más interpretables, eficaces y eficientes. Se ha de resaltar que no se han encontrado referencias en el campo financiero que estén enfocados principalmente a la selección de variables.

Por otro lado y en cuanto a la capacidad de discriminación de los diferentes ratios para predecir la insolvencia encontramos evidencia positiva a favor de la importancia en primer lugar del ROA y seguidamente de las variables Crecimiento de las Ventas, Endeudamiento y Productividad.

## Bibliografía

1. Altman, E. (1968): Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, September, pp. 189-209.
2. Altman, E. (1983): *Corporate financial distress-A complete guide to predicting avoiding and dealing with bankruptcy*. New York: Wiley.
3. Altman, E., Marco, G., and Varet, F. (1994). Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks. *Journal of Banking and Finance*, 18, 505-529.
4. Bala J., Dejong K., Huang J., Vafaie H. y Wechsler H. (1996): *Using Learning to Facilitate the Evolution of Features for Recognizing Visual Concepts*, *Evolutionary Computation*, 4, 3, 297-311.
5. Beaver, W., (1966). Financial ratios as predictors of failures, in *Empirical Research. Accounting, selected studies*, pp. 71-111.
6. Brachman R.J., Khabaza, T., Kloesgen, W, Piaterky-Shapiro, G., and Simoudis, E (1996). Mining business databases. *Communication of the ACM*, 39(11), 42-48. Cotta C., Sloper C. and Moscato P. (2004). Evolutionary search of thresholds for robust feature set selection: Application to the analysis of microarray data. *Lecture Notes In Computer Science 3005*: 21-30.
7. Chiang L.H. and Pell R.J.(2004): *Genetic Algorithms Combined with Discriminant Analysis for Key Variable Identification*, *Journal of Process Control*, 14, 2, pp.143-155.
8. Cotta C., Sloper C. y Moscato P. (2004): *Evolutionary Search of Thresholds for Robust Feature Set Selection: Application to the Analysis of Microarray Data*, *Lecture Notes In Computer Science 3005*: 21-30.
9. Efroymson, M.A.(1960): *Multiple Regression Analysis*, *Mathematical Methods for Digital Computers*(Ralston, A. and Wilf, H.S., ed.) Vol.1. Wiley, New York.
10. Ezzamel, M., and C. M. Molinero (1987): On the Distributional Properties of Financial Ratios, *Journal of Business Finance & Accounting*, 14, 463--481.
11. Gallego, A.M., Gómez, J.C. y Yáñez L. (1997): *Modelos de Predicción de Quiebras en Empresas no Financieras*. *Actualidad Financiera*, nueva época, nº 5: 3-14.
12. Gallego, A.M. y Gómez, M.A. (2002): *Análisis Integrado de la Absorción y Quiebra Empresarial Mediante la Estimación de un Modelo Multilogit.*. *Revista Española de Financiación y Contabilidad*, V. XXXI, nº111, pp. 111-144.

13. García, F.C., García-Torres, M., Moreno Pérez, J.M. and Moreno-Vega, J.M. (2003): *Búsqueda Dispersa para el Problema de la Selección de Variables, CAEPIA-2003*
14. Glover F. (1986): Future Paths for Integer Programming and Links to Artificial Intelligence, *Computers and Operations Research*, Vol. 13, pp. 533-549.
15. Glover F. (1989): *Tabu Search: Part I*, *ORSA Journal on Computing*. Vol. 1, pp. 190-206.
16. Glover F. (1990): *Tabu Search: Part II*, *ORSA Journal on Computing*. Vol. 2, pp. 4-32.
17. Glover F. y Laguna M. (1993): *Tabu Search in Modern Heuristic Techniques for Combinatorial Problems*, C. Reeves ed., Blackwell Scientific Publishing, pp. 70-141.
18. Glover F. Y Laguna M. (1997): *Tabu Search*, Kluwer Academic Publishers, Boston.
19. Glover, F. y Laguna, M. (2002): *Tabu Search*, in *Handbook of Applied Optimization*, P. M. Pardalos and M. G. C. Resende (Eds.), Oxford University Press, pp. 194-208.
20. Glover F., Laguna M. y Marti R. (2000): *Fundamentals of Scatter Search and Path Relinking*, *Control and Cybernetics*, vol. 29, pp.653-684.
21. Gómez O., Casado S., Nuñez L., y Pacheco J. (2004): *Resolución del Problema de Selección de Variables Cuantitativas mediante GRASP. Aplicación a Ratios Financieros*, Actas XII Congreso Asepuma celebrado en Septiembre de 2004 en la Universidad de Murcia.
22. Hair, J., Anderson, R., Tatham, R. y Black, W. (1999): *Análisis Multivariante*. 5ª Edición. Prentice Hall.
23. Huberty, C.J.(1994): *Applied Discriminant Analysis*. Wiley. Interscience.
24. Inza I., Merino M., Larranaga P., Quiroga J., Sierra B. y Giralá M.(2001a): *Feature Subset Selection by Genetic Algorithms and Estimation of Distribution Algorithms - A Case Study in the Survival of Cirrhotic Patients Treated with TIPS*, *Artificial Intelligence In Medicine* 23 (2): 187-205.
25. Inza I., Larranaga P. y Sierra B. (2001b): *Feature Subset Selection by Bayesian Networks: A Comparison with Genetic and Sequential Algorithms*, *International Journal of Approximate Reasoning* 27 (2): 143-164.
26. Jain A. And Zongker D. (1997): *Feature Selection: Evaluation, Application, and Small Sample Performance*, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153-158.



27. Jourdan L., Dhaenens C. y Talbi E. (2001): *A Genetic Algorithm for Feature Subset Selection in Data-Mining for Genetics*, MIC 2001 Proceedings, 4th Metaheuristics International Conference, 29-34.
28. Kohavi R. (1995): *Wrappers for Performance Enhancement and Oblivious Decision Graphs*, Stanford University, Computer Science Department.
29. Laffarga, J., Martín, J.L., Vázquez, M.J. (1985): *El Análisis de la Solvencia en las Instituciones Bancarias: Propuesta de una Metodología y Aplicaciones a la Banca Española*, Esic-Market, núm. 48, 2º trimestre, pp. 51-73.
30. Laffarga, J., , Martín, J.L., Vázquez, M.J. (1986a): *El Pronóstico a Corto Plazo del Fracaso en las Instituciones Bancarias: Metodología y Aplicaciones a la Banca Española*, Esic-Market, núm. 53, 3er trimestre, pp. 59-116.
31. Laffarga, J., , Martín, J.L., Vázquez, M.J. (1986b): *El Pronóstico a Largo Plazo del Fracaso en las Instituciones Bancarias: Metodología y Aplicaciones al Caso Español*, Esic-Market, núm. 54, 4º trimestre, pp. 113-167.
32. Laffarga, J., , Martín, J.L., Vázquez, M.J. (1987): *Predicción de la Crisis Bancaria Española: La Comparación entre el Análisis Logit y el Análisis Discriminante*, Cuadernos de Investigación Contable, núm. 1, otoño, pp. 103-110.
33. Laffarga, J., , Martín, J.L., Vázquez, M.J. (1990): *La Predicción de la Quiebra Bancaria: el Caso Español*. Revista Española de Financiación y Contabilidad, V. XX, nº 66, pp.151-166.
34. Lee S., Yang J. y Oh K.W. (2003): *Prediction of Molecular Bioactivity for Drug Design Using a Decision Tree Algorithm*, Lecture Notes In Artificial Intelligence 2843: 344-351.
35. Lewis P.M. (1962): *The Characteristic Selection Problem in Recognition Systems*, IEEE Trans. Information Theory, vol. 8: 171-178.
36. Liu H. Y Motoda H.(1998): *Feature Selection for Knowledge Discovery and Data Mining*, Boston: Kluwer Academic.
37. Narendra P.M. y Fukunaga K. (1977): *A Branch and Bound Algorithm for Feature Subset Selection*, IEEE Trans. Computers, vol. 26, no. 9: 917-922.
38. O’Gorman T.W. (2004): *Using adaptive Methods to Select Variables in Case-Control Studies*, Biometrical Journal 46,5, pp.595-605.
39. Ohlson, J.A. (1980): *Financial ratios and the probabilistic prediction of bankruptcy*. Journal of Accounting Research, 18 (1), pp. 109-111.

40. Oliveira L.S., Sabourin R., Bortolozzi F., y otros (2003): *A Methodology for Feature Selection Using Multiobjective Genetic Algorithms for Handwritten Digit String Recognition*, International Journal Of Pattern Recognition And Artificial Intelligence 17 (6): 903-929.
41. Richardson, F. M., and L. F. Davidson, (1984): On Linear Discrimination with Accounting Ratios, *Journal of Business Finance & Accounting*, 11 (4), 511--525.
42. Salvador Figueras M. (2000): *Análisis Discriminante*,[en línea] 5campus.com, Estadística <http://5campus.com/lección/discr>[10 de febrero de 2005]
43. Sanchis, A., Gil, J.A. y Heras, A. (2003): *El análisis discriminante en la previsión de la insolvencia en las empresas de seguros de no vida*. Revista Española de Financiación y Contabilidad, V. XXXII, nº 116, pp. 183-233.
44. Sebestyen, G. (1962): *Decision-Making Processes in Pattern Recognition*. New York: MacMillan.
45. Shy S. y Suganthan P.N. (2003): *Feature Analysis and Clasification of Protein Secondary Structure Data*, Lecture Notes in Computer Science 2714: 1151-1158.
46. Tamoto E., Tada M., Murakawa K., Takada M., Shindo G., Teramoto K., Matsunaga A., Komuro K., Kanai M., Kawakami A., Fujiwara Y., Kobayashi N., Shirata K., Nishimura N., Okushiba S.I., Kondo S., Hamada J., Yoshiki T., Moriuchi T. y Katoh H.(2004): *Gene expression Profile Changes Correlated with Tumor Progression and Lymph Node Metastasis in Esophageal Cancer*. Clinical Cancer Research 10(11):3629-3638.
47. Wong M.L.D. y Nandi A.K. (2004): *Automatic Digital Modulation Recognition Using Artificial Neural Network and Genetic Algorithm*. Signal Processing 84 (2): 351-365.
48. Zmijewski, M.E.(1984): *Methodological Issues Related to the Estimation of Financial Distress Prediction Models*. Journal of Accounting Research, 22(1), 59-82.

NOTAS

---

---

## NOTAS

---

---