

# Predicción de información contable con un modelo híbrido. Comparación con técnicas de regresión clásicas<sup>1</sup>

Antonio Arauzo  
Azofra  
[Arauzo\(a\)decsai.ugr.es](mailto:Arauzo(a)decsai.ugr.es)

Juan Luis Castro Peña  
[Castro\(a\)decsai.ugr.es](mailto:Castro(a)decsai.ugr.es)  
E.T.S. Ingeniería Informática  
Universidad de Granada  
18071 GRANADA

Miguel Delgado  
Calvo-Flores  
[Mdelgado\(a\)ugr.es](mailto:Mdelgado(a)ugr.es)

M<sup>a</sup> Angustias  
Navarro Ruíz

José M<sup>a</sup> de la Torre  
Martínez  
[Jdltorre\(a\)platon.ugr.es](mailto:Jdltorre(a)platon.ugr.es)

Facultad de Ciencias Económicas y  
Empresariales  
Universidad de Granada  
18011 GRANADA

## Resumen

En los últimos años se ha venido prestando una atención considerable al análisis y modelado de la información contable para toma de decisiones en el mercado de capitales. Para ello se ha empleado usualmente la regresión en sus múltiples variantes.

En este trabajo se presenta el empleo de una técnica híbrida (que combina la selección de variables con la aproximación con RNN's Artificiales) desarrollada para evitar las dificultades procedentes de la actuación local de los métodos de regresión. Trata de identificar ejemplos mal evaluados no desde un punto de vista local, sino global, es decir, teniendo en cuenta la repercusión de cualquier empresa de la muestra sobre el error global del ajuste.

Se realiza además una comparación de los resultados obtenidos mediante el empleo de la regresión logística y el antedicho método en el problema de clasificación de las empresas integrantes de la muestra, con objeto de identificar el método más adecuado para la determinación de las variables que discriminan en mayor grado la pertenencia de aquéllas al grupo de las entidades más valoradas en el mercado.

**Palabras Clave:** Métodos Híbridos, Redes Neuronales Artificiales, Información contable, mercado de capitales, regresión logística.

## 1 INTRODUCCIÓN

En los últimos tiempos, se ha dedicado una atención considerable, a determinar que información proporcionada por las empresas y disponible para los inversores resulta relevante para los mismos.

En estudios anteriores [13] [14] se trató de verificar la importancia que la información económico-financiera de las empresas pudiera tener en relación con la valoración que los inversores realizan de las mismas en el mercado de valores. En dichos trabajos, y en línea con otros anteriores [11] [17] [20], se llevó a cabo un análisis basado en información contable, empleando modelos de regresión logística que clasificaban a las empresas integrantes de la muestra seleccionada en dos grupos: aquéllas que habían permanecido de manera estable en el tiempo en un índice bursátil y aquéllas que no, en función de una serie de indicadores económico-financieros clásicos, elaborados de acuerdo con los criterios de devengo y caja, además de variables tendentes a informar acerca de algunos activos intangibles de las empresas, y otras de carácter cualitativo. Las conclusiones alcanzadas ponían de manifiesto la importancia que los inversores otorgaban a la política de dividendos seguida por las empresas a la hora de asignarles un determinado valor, así como su aparente preferencia por una estructura financiera apalancada [1] [22], a la vez que por una garantía suficiente que elimine el riesgo de que la unidad económica incurra en insolvencia financiera. Por otra parte, las conclusiones alcanzadas por numerosos estudios que han tratado de analizar la relevancia para la valoración de las variables de caja, aunque con diferentes conclusiones al respecto ([4], [6], [8], [9], [12], [18], [19], [21], o [24]), hacían pensar que dichas variables podrían tener un mayor poder discriminatorio a la hora de diferenciar las empresas más valoradas por los inversores. No obstante, los resultados obtenidos no corroboraron lo anterior.

---

<sup>1</sup> Este trabajo ha sido realizado con financiación del proyecto CICYT-TIC2000-1362-CO2-01.

El objetivo de la presente comunicación es presentar una técnica híbrida para identificar las variables y modelar las relaciones más relevantes para la toma de decisiones por parte de los inversores en el mercado de capitales. El mismo problema se resuelve empleando una técnica clásica, la regresión logística, para comparar los resultados.

## 2 METODOLOGÍA

### 2.1. SELECCIÓN DE LAS VARIABLES

Como variable dependiente se ha considerado la pertenencia o no de las empresas de la muestra al grupo de entidades que han permanecido en el Índice General de la Bolsa de Madrid desde 1993 hasta 1998; tomando el valor 1 en caso de pertenencia, y 0 en caso contrario. Tal como se ha indicado, dicha pertenencia viene determinada entre otros factores y de modo importante por la capitalización bursátil de los títulos, que vendría a indicar a su vez la valoración de los mismos por parte de los inversores participantes en el mercado. El uso de esta variable dependiente se ve apoyado por el empleo de los precios de las acciones en los trabajos de investigación en el área del mercado de capitales encuadrados dentro del paradigma utilitarista de la investigación contable [16] [15].

Las variables independientes seleccionadas han sido las que se relacionan en el Cuadro 1.

Cuadro 1: Relación de variables independientes

<input type="checkbox"/> NIVEL DE ENDEUDAMIENTO	<input type="checkbox"/> MARGEN
<input type="checkbox"/> COMPOSICIÓN DEL ENDEUDAMIENTO	<input type="checkbox"/> RENTABILIDAD ECONÓMICA TOTAL
<input type="checkbox"/> FONDO DE MANIOBRA	<input type="checkbox"/> RENTABILIDAD FINANCIERA
<input type="checkbox"/> DISPONIBILIDAD	<input type="checkbox"/> COSTE DEL ENDEUDAMIENTO
<input type="checkbox"/> TEST-ÁCIDO	<input type="checkbox"/> GASTOS FINANCIEROS
<input type="checkbox"/> SOLVENCIA	<input type="checkbox"/> I+D
<input type="checkbox"/> GARANTÍA	PUBLICIDAD
<input type="checkbox"/> TAMAÑO	DEMORA
<input type="checkbox"/> RECURSOS GENERADOS	DEMANDAS
<input type="checkbox"/> ROTACIÓN DEL ACTIVO	INCIDENCIAS
<input type="checkbox"/> ROTACIÓN DE EXISTENCIAS	AUDITORIA
<input type="checkbox"/> ROTACIÓN DE DERECHOS	SOLVENCIA CAJA (C/P)
<input type="checkbox"/> RENTABILIDAD ECONÓMICA DE LA EXPLOTACIÓN	SOLVENCIA CAJA (L/P)
	CALIDAD DEL BENEFICIO
	<input type="checkbox"/> RETORNO LÍQUIDO

### 2.2. SELECCIÓN DE LA MUESTRA

La muestra empleada está integrada por 64 empresas no financieras que cotizan en la Bolsa de Madrid, para las que se proporcionan datos consolidados relativos a balances y cuentas de resultados correspondientes a los ejercicios 1997 y 1998, obtenidos a partir de la base de datos de Intertell. Por otra parte, los datos relativos a los gastos en publicidad realizados por las firmas que conforman la muestra en los referidos ejercicios fueron suministrados por la empresa Infoadex. Para la selección de la muestra, se excluyeron las empresas inmobiliarias y constructoras, así como las concesionarias de autopistas, debido a que las características específicas de dichos sectores pueden distorsionar los resultados arrojados por los modelos resultantes del trabajo.

La relación de empresas fue obtenida a través de la página web de la Bolsa de Madrid, seleccionando, en primer lugar, aquellas que, desde 1993 hasta 1998, han formado parte de la composición del Índice General de la Bolsa de Madrid.

Las 64 empresas estaban divididas en dos grupos de 32. El primero formado por aquellas que presentan la característica de permanencia, y el de entidades que no reunían dicha característica.

### 2.3. REGRESIÓN LOGÍSTICA

Con el fin de alcanzar una explicación coherente acerca de la mayor o menor relevancia que los indicadores citados puedan tener en la capitalización bursátil de las unidades económicas de la muestra seleccionada, se aplicó en primer lugar el modelo de regresión logística.

La elección del mismo en detrimento de otros, tales como el análisis discriminante o la regresión lineal, se fundamenta en las características de las variables, tanto independientes como dependiente. [23]

En base a todo lo anterior, se aplicó la regresión logística a la muestra anteriormente especificada, utilizando para ello el programa SPSS 10.0, obteniéndose un modelo con un nivel de acierto global del 85% de los casos. Para su ajuste se siguió el método condicional de introducción por pasos de las variables consideradas.

### 2.4. UN MODELO HIBRIDO BASADO EN rrnaa Y ARBOLES DE DECISION

El segundo método de análisis aplicado se basa en el empleo de un método híbrido con objeto de seleccionar las características más relevantes, y ajustar el modelo QU se pretende. Se parte de la metodología propuesta en SEPARATE (SEmiglobal PARTition to design Authomatic sysTEms) [10], adaptándola para considerar tanto variables discretas como continuas, y combinándola con la metodología de selección de características del algoritmo FOCUS [5].

La filosofía de SEPARATE consiste en conjugar particiones globales y locales para aunar las ventajas de las Redes Neuronales (particiones globales, buena generalización) y los Árboles de Decisión (particiones locales, buen ajuste). En este caso, además, se pretende obtener las características más relevantes, con lo cual parecía apropiado utilizar también algún método para seleccionarlas. Se ha escogido el algoritmo de FOCUS por su alto grado de eficiencia y efectividad (lo que lo hace ser uno de los más empleados) aunque en principio esta limitado a problemas donde todas las variables fuesen de naturaleza discreta.

En un primer momento, se adaptó el algoritmo FOCUS, con el objetivo de considerar todas las variables tanto discretas como continuas, obteniendo de esta manera la variante C-FOCUS [7]. Finalmente, la adaptación de SEPARATE se realizó de la siguiente forma:

1. Obtener mediante C-FOCUS las variables más relevantes para el problema, ordenadas según su grado de relevancia.
2. Si en la primera posición hay variables discretas, utilizar un árbol de decisión para dividir el problema en casos, y volver a aplicar este algoritmo para cada uno de esos casos. Si, por el contrario, no hay variables discretas en la primera posición, ajustar el modelo mediante una red neuronal, utilizando solamente las variables obtenidas como relevantes.

En base a lo anterior, se aplicó el C-FOCUS, en el que se obtuvieron las variables relevantes. Dado que la primera era discreta, se aplicó de nuevo el algoritmo para los dos casos que definía esta variable. El paso posterior consistió en la aplicación de redes neuronales tanto para la muestra total como para los dos casos en los que se segmentó el problema. Para la aplicación de esta última técnica, se dividieron los casos en un 25% para validación y un 75% para entrenamiento, elegidos al azar. Además, se modificó ligeramente el procedimiento de entrenamiento BackPropagation, cambiando la función de error por una función truncada del mismo (se hace 0 si  $|\text{error}| < 0.2$ ), con lo que se consiguió elevar de forma notable el porcentaje de aciertos.

### 3 DISCUSIÓN DE LOS RESULTADOS OBTENIDOS Y CONCLUSIONES

Como hemos indicado anteriormente el principal objetivo del presente trabajo es presentar una técnica híbrida para selección y modelado de variables empleandola sobre el problema de la selección de las variables más relevantes para la toma de decisiones de los inversores en el mercado de capitales, comparando sus resultados con los obtenidos por regresión logística. Esta, al igual que el resto de técnicas clásicas de regresión para identificación de funciones a partir de datos, aunque en menor medida

que las demás de esta categoría, presenta el inconveniente de que los ajustes se realizan de manera local, de modo que el espacio de datos de entrada se divide en regiones, dentro de cada una de las cuales se estudia el ajuste de un modo relativamente independiente. Ello hace aconsejable la búsqueda de técnicas alternativas, que permitan superar estos inconvenientes, lo que ha llevado a la aplicación de técnicas basadas en inteligencia artificial.

La aplicación de la regresión logística a la muestra seleccionada ha llevado a la obtención de un modelo para el cual, como se observa en la Tabla 1, el test de significación global de los coeficientes presenta un nivel de significación de 0, lo cual hace rechazar la hipótesis nula, es decir, el modelo no ajusta bien sólo con el término independiente, lo que equivale a afirmar que los coeficientes obtenidos para las distintas variables integrantes del modelo son distintos de cero.

Tabla 1: Bondad del ajuste.

ESTADÍSTICO	VALOR	G	SIG.	H <sub>0</sub>	RESULTADO
TEST DE SIGNIFICACIÓN GLOBAL DE LOS COEFICIENTES	39,33	4	0,000	EL MODELO SÓLO CON EL TÉRMINO INDEPENDIENTE AJUSTA BIEN LOS DATOS	RECHAZAR
TEST DE BONDAD DEL AJUSTE DE HOSMER Y LEMESHOW	12,79	8	0,119	EL MODELO AJUSTA BIEN LOS DATOS	NO SE PUEDE RECHAZAR

El modelo clasifica correctamente las empresas que no pertenecen al grupo de permanentes al IGBM en un 89,7% de los casos, y las que pertenecen en un 81,3% de los casos, lo cual verifica de nuevo que el modelo es aceptable. Por último, se consigue asignar de forma correcta en su conjunto el 85,2% (Tabla 2).

Tabla 2: Clasificación de pertenencia.

VALORES OBSERVADOS	VALORES PRONOSTICADOS		PORCENTAJE DE ACIERTOS
	(0) NO PERTENENCIA	(1) PERTENENCIA	
(0) NO PERTENENCIA	26	3	89,7 %
(1) PERTENENCIA	6	26	81,3 %
TOTAL			85,2 %

La Tabla 3 muestra el modelo final obtenido, en el que aparecen las variables relativas a la tasa de reparto de dividendos, composición del endeudamiento, garantía y tamaño de la unidad económica, así como sus coeficientes, todos ellos significativos.

Tabla 3: Modelo final ajustado.

VARIABLE	SE	WALD	GL	SIG.
COMPOSICIÓN DEL ENDEUDAMIENTO	0,875	0,374	4,402	1 0,036
PAYOUT	6,872	3,300	4,798	1 0,028

GARANTÍA	1,284	0,512	6,294	1	0,012
TAMAÑO (1)	3,017	1,057	8,149	1	0,004
CONSTANTE	-6,576	1,938	11,510	1	0,001

Por otra parte, la aplicación del algoritmo C-FOCUS a la muestra considerada produjo los resultados que se muestran en la tabla 4. La misma proporciona una relación de las variables con mayor poder discriminatorio de entre las veintinueve incorporadas al estudio que resultan de la aplicación de esta técnica a la muestra total considerada, así como aquellas más relevantes en el caso en que la misma se subdivida en función de la dimensión de las unidades económicas que la integran.

Tabla 4: Resultados del algoritmo C-FOCUS.

VARIABLE	TODAS	MEDIANAS	GRANDES
TAMAÑO	8		
PAYOUT	8		8
GARANTÍA	8	8	
SOLVENCIA CAJA L/P	8	8	
ROTACIÓN ACTIVO	8		8
FONDO MANIOBRA	8		8
COSTE ENDEUDAMIENTO	8		
NIVEL ENDEUDAMIENTO			8
AUDITORÍA		8	

Gráficamente, los modelos obtenidos pueden representarse como se muestra en la Figura 1.

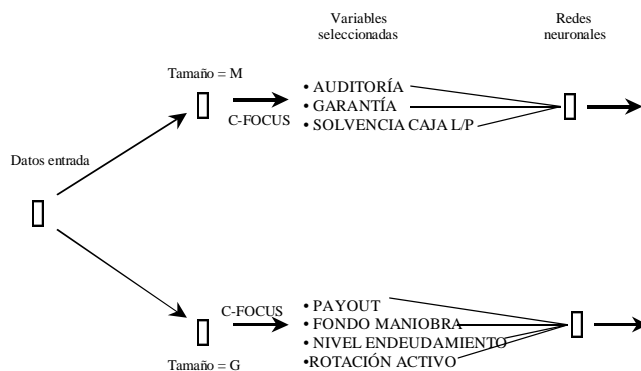


Figura 1: Aplicación del algoritmo C-FOCUS.

Los resultados alcanzados con los dos tipos de análisis realizados resultan coherentes entre sí, aunque las técnicas basadas en inteligencia artificial, en todos los casos, proporcionan porcentajes de acierto más elevados, como muestran las tablas 5 y 6, referidas, respectivamente, a la red neuronal aplicada sobre la muestra total una vez seleccionadas las variables con el C-FOCUS, y a las redes aplicadas a las dos submuestras.

Tabla 5: Clasificación de pertenencia, muestra total.

VALORES OBSERVADOS	VALORES PRONOSTICADOS		PORCENTAJE DE ACIERTOS
	(0) NO PERTENENCIA	(1) PERTENENCIA	
(0) NO PERTENENCIA	29	2	93'54 %
(1) PERTENENCIA	0	36	100 %
TOTAL			97'01 %

Tabla 6: Clasificación de pertenencia, submuestras.

VALORES OBSERVADOS	VALORES PRONOSTICADOS		PORCENTAJE DE ACIERTOS
	(0) NO PERTENENCIA	(1) PERTENENCIA	
TAMAÑO=M			
(0) NO PERTENENCIA	17	0	100 %
(1) PERTENENCIA	0	5	100 %
TAMAÑO=G			
(0) NO PERTENENCIA	12	2	85'7 %
(1) PERTENENCIA	0	31	100 %
TOTAL			97'01 %

Pese a la coincidencia, en líneas generales, de los resultados obtenidos, nuestro método proporciona niveles de acierto superiores a los correspondientes al análisis mediante regresión logística, lo que hace más aconsejable su uso para resolver el problema planteado, además de su capacidad de aprendizaje.

El tamaño resultó ser, en ambos casos, la variable con mayor poder discriminatorio a la hora de clasificar las empresas en el grupo de unidades económicas que han permanecido en el Índice General de la Bolsa de Madrid.

Es de destacar también que, al segmentar la muestra total en función de esta variable, los indicadores más valorados para las empresas grandes y para las de más reducida dimensión difieren. Mostrando de esta forma la idoneidad de la metodología SEPARATE en este problema.

## Referencias

- [1] ABAD, C., GARCÍA BORBOLLA, A., LAFFARGA, J., LARRÁN, M. y PIÑERO, J. (1997): "La información financiera de los grupos de sociedades frente a la información individual: evidencia empírica de la Bolsa de Madrid". *Comunicación presentada al X Congreso de AECA*. Salamanca, pp. 19-38.
- [2] ABARBANELL, J.S. y BUSHEE, B.J. (1997): "Fundamental Analysis, Future Earnings, and Stock Prices". *Journal of Accounting Research*, Vol. 35, nº 1, pp. 1-24.
- [3] ABARBANELL, J.S. y BUSHEE, B.J. (1998): "Abnormal Returns to a Fundamental Analysis Strategy". *Accounting Review*, Vol. 73, nº 1, January, pp. 19-45.

- [4] ALI, A. (1994): "The Incremental Information Content of Earnings Working Capital from Operations and Cash Flows". *Journal of Accounting Research*, Vol. 32, pp. 61-74.
- [5] ALMUALLIM, H. AND DIETERICH, T.G (1994). "Learning Boolean Concepts in the Presence of Many relevant Features". *Artificial Intelligence*, 69 (1-2): 279-305.
- [6] ANSÓN LAPEÑA, A. y PINA MARTÍNEZ, V. (1994): "Contenido Informativo de la Información Económico-Financiera para Evaluar la Rentabilidad Empresarial". *Revista Española de Financiación y Contabilidad*, Vol. XXIII, nº 78, pp. 143-157.
- [7] ARAUZO, A., CASTRO, JL (2002). "C-FOCUS: a continuous extension of FOCUS". DECSAI Technical Report 02-01-23.
- [8] BALL, R. y BROWN, P. (1968): "An Empirical Evaluation of Accounting Income Numbers". *Journal of Accounting Research*, Vol. 10, pp. 159-178.
- [9] BEAVER, W.H. y DUKES, R. (1972): "Interperiod Tax Allocation, Earning Expectations and the Behaviour of Security Prices". *The Accounting Review*, Vol. 47, pp. 320-332.
- [10] CASTRO, J.L., DELGADO, M. & MANTAS, C.M.(2000), "SEPARATE: a method of machine learning based on semiglobal partitions", *IEEE Transactions on Neural Networks*, vol. 11, nº 3, pp. 710-720, Mayo, 2000
- [11] CRESPO DOMINGUEZ, M. A. (1991): "El análisis contable y el mercado de capitales: un estudio empírico acerca de la relación entre la información contable y los precios de las acciones". *Revista Española de Financiación y Contabilidad*, Vol. 69, pp. 969-984.
- [12] CHARITOU, A. y PANAGIOTIDES, G. (1999): "Financial Analysis, Future Earnings and Cash Flows, and the Prediction of Stock Returns: Evidence for the UK". *Accounting and Business Research*, Vol. 29, nº 4, pp. 281-298.
- [13] DE LA TORRE MARTÍNEZ, J.M.; NAVARRO RUIZ, M.A. y SÁNCHEZ VIZCAÍNO, G. (2000): "Relevancia de la Información Contable para Discriminar las Empresas más Valoradas en el Mercado Español: Un Estudio Empírico". *IX Encuentro de Profesores Universitarios de Contabilidad*. Las Palmas de Gran Canaria. Vol. 2, pp. 669-680.
- [14] DE LA TORRE MARTÍNEZ, J.M.; NAVARRO RUIZ, M.A. y SÁNCHEZ VIZCAÍNO, G. (2001): "Incidencia de la Información Contable Basada en el Criterio de Devengo vs. Criterio de Caja en la Valoración de Acciones en el Mercado Español: Una Aproximación Empírica". *XI Congreso AECA*. Madrid.
- [15] GARCÍA BENAÚ, M.A. y GINER INCHAUSTI, B. (2000): "Perspectiva de Investigación de la Contabilidad Financiera en el Ámbito Internacional". *IX Encuentro de Profesores Universitarios de Contabilidad*. Las Palmas de Gran Canaria. Vol. 1, pp. 56-136.
- [16] JENNINGS, R.; ROBINSON, J.; THOMPSON, R.B. y DUVALL, L. (1996): "The Relation Between Accounting Goodwill Numbers and Equity Values". *Journal of Business, Finance and Accounting*, Vol. 23, nº 4, pp. 513-533.
- [17] LEV, B. y THIAGARAJAN (1993): "Fundamental Information Analysis". *Journal of Accounting Research*, Vol. 31, nº 2, pp. 190-215.
- [18] LIVNAT, J. y ZAROWIN, P. (1990): "The Incremental Information Content of Cash Flow Component". *Journal of Accounting of Economics*, Vol. 13, pp. 25-46.
- [19] McLEAY, S., KASSAB, J. y HELAN, M. (1997): "The Incremental Information Content of Accruals: Evidence Based on the Exponential Smoothing of Levels and Trends in Pre-Tax Earnings, Funds Flow and Cash Flow". *Journal of Business Finance & Accounting*, Vol. 24, nº 7 y 8, pp. 1147-1167.
- [20] MITCHELL, M. y MULHERIN, H. (1994): "The Impact of Public Information on the Stock Market". *Journal of Finance*, July, pp. 923-951.
- [21] PFEIFFER, R.J. y ELGERS, P.T. (1999): "Controlling for Lagged Stock Price Responses in Pricing Regressions: An Application to the Pricing of Cash Flows and Accruals". *Journal of Accounting Research*. Vol. 37, nº 1, pp. 239-247.
- [22] REES, W. (1997): "The Impact of Dividends, Debt and Investment on Valuation". *Journal of Business, Finance and Accounting*, vol. 24, pp. 1111-1140.
- [23] SÁNCHEZ VIZCAÍNO, G. (2000): "Regresión Logística", en LUQUE MARTÍNEZ, T. (ed.): *Técnicas de Análisis de Datos en Investigación de Mercados*. Ed. Pirámide. Madrid.
- [24] SLOAN, R.G. (1996): "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows About Future Earnings?". *The Accounting Review*, Vol. 71, nº 3, pp. 289-315.