



Modelo de predicción de la solvencia empresarial mediante una red neuronal perceptrón multicapa

Autor: Ana Isabel Velasco Fernández

Institución: Universidad Alfonso X el Sabio

Otros autores: Marta Beatriz García García (Universidad Alfonso X el Sabio)

Resumen

En las últimas décadas se han realizado importantes contribuciones de mejora de las técnicas de predicción del fracaso empresarial. Estas mejoras se han concretado en la investigación del diseño muestral, la selección de variables y la elección de técnicas estadísticas. Salvo limitadas excepciones, la literatura financiera especializada se ha orientado, básicamente, al estudio de la predicción del fracaso en empresas grandes y cotizadas. La falta de datos adecuados, sin embargo, ha limitado las investigaciones en empresas menores o no cotizadas.

Todo ello, unido a la reciente recesión económica mundial, ha impulsado el desarrollo de nuevos y más sofisticados métodos de análisis de la solvencia, entre los cuales ocupan un papel destacado aquellos que están basados en técnicas de Inteligencia Artificial.

En este sentido proponemos una modelización de predicción de solvencia a partir de una muestra de empresas no cotizadas de los sectores de construcción e industria españoles, dos de los sectores más afectados por la reciente crisis económica. La técnica utilizada es una red neuronal perceptrón multicapa, del ámbito de la Inteligencia Artificial, que de forma similar a como el cerebro humano opera, aprende de los ejemplos-patrones que se le presentan y, a partir de ellos, es capaz de establecer relaciones y predecir nuevos casos. Para validar esta técnica la comparamos con una de las técnicas estadísticas tradicionalmente más usadas, el análisis discriminante.

Las investigaciones muestran mejores resultados, tanto en la clasificación como en la predicción, para el modelo neuronal, así como una mejor interpretación de los factores más importantes en la línea de evitar el fracaso de la empresa, con respecto al modelo tradicional.

Palabras clave: Modelos de Predicción de Solvencia. Ratios Financieros. Redes Neuronales Artificiales. Análisis Discriminante.

ABSTRACT

In this study, we have developed and compared two solvency prediction models for not quoted companies. The techniques used are the traditional statistics for a discriminant analysis model and intelligent techniques for the implementation of a model with an artificial neural network.

Research into this topic shows better results, both in the classification and in the prediction, for the neural model, as well as a better interpretation of the most important factors to avoid the failure of the company, in relation to the traditional model.

Keywords: Solvency Prediction Models. Financial Ratios. Artificial Neural Networks. Discriminant Analysis.

Introducción

Desde hace décadas estamos asistiendo a profundos cambios en el ámbito de la empresa. Independientemente de su actividad y de sus dimensiones, la globalización y el progreso tecnológico han supuesto para nuestras empresas españolas una mayor competitividad, al igual que la rápida evolución de los mercados ha implicado la revisión continua de la gestión empresarial, sus modelos económicos, operativos y organizativos. Además la globalización afecta tanto a las empresas que tienen presencia en el exterior como a las que no. La tecnología, el acceso a los clientes, la forma de hacer negocio, conforman la globalización y todos los mercados son susceptibles de verse modificados conforme a los condicionantes impuestos por los mercados exteriores.

Por otra parte, la reciente recesión económica iniciada en el año 2007, ha dañado sensiblemente los niveles de rentabilidad de las empresas, fundamentalmente las de menor tamaño, mucho más vulnerables a los vaivenes económicos, motivando la expulsión del mercado de un gran número de ellas.

Según el Informe Económico de la Escuela de Negocios Esade (1) de enero de 2013, la crisis económica en España, con un crecimiento negativo o prácticamente nulo del PIB otro año más, se prolonga ya seis años. En un contexto de debilitamiento del crecimiento global y de un práctico estancamiento en la zona euro, la economía española no consigue recuperar la actividad, y las empresas, fundamentalmente las PYMES, son las principales afectadas. Cada año sigue aumentando el número de empresas que se declaran en concurso de acreedores. De acuerdo a la Estadística del Procedimiento Concursal del INE (2), entre los años 2008 y 2012 el número de empresas concursadas fue de 27064, de las cuales el 32,5% pertenecían al sector de la construcción y un 20,6% al sector industria y energía.

Los factores que llevan a las empresas al fracaso varían. Muchos economistas atribuyen este fenómeno a las altas tasas de interés, a la disminución del margen de ganancias debida a la recesión, a la fuerte competitividad debida a los mercados abiertos entre los cuales hay que incluir mercados de otros países con bajos costes, etc. Además, el crédito de los bancos al sector privado ha disminuido más del 4% en el último año y más del 8% desde su punto máximo en 2008.

Como consecuencia de lo anterior la amenaza del fracaso es acuciante y los trabajos de predicción de riesgo financiero se intensifican.

La evidencia muestra que el valor de mercado de las empresas en riesgo disminuye sustancialmente, pero además, predecir la insolvencia empresarial es de interés para todos los agentes económicos implicados: accionistas, empleados, clientes, proveedores, auditores, acreedores y directivos, cuya riqueza y bienestar depende de la capacidad de las empresas para sobrevivir. A estos últimos, los modelos de predicción de la insolvencia pueden proporcionarles señales sobre la existencia de problemas, permitiéndoles introducir medidas correctoras que eviten el fracaso (3).

La forma tradicional de medir el riesgo financiero de las empresas ha sido mediante métodos estadísticos clásicos, que podían variar desde simples análisis univariantes a más complejos modelos multivariantes, como la puntuación Z-score de Altman (4), la puntuación O-score de Ohlson (5), o la de Zmijewski (6), y siempre utilizando como ejemplos grandes empresas que cotizan en Bolsa o entidades financieras, ya que tanto en uno como en otro caso el acceso a los datos financieros es mucho más sencillo.

Estudios más recientes han confirmado que estas puntuaciones siguen siendo válidas hoy en día como aproximaciones para la valoración del riesgo de fracaso en la empresa, pero también son muchos los estudios que las han modificado para adaptarlas a alguna situación particular obteniendo con ello mejores resultados (7).

El problema es que los análisis multivariantes de predicción de insolvencia, de los cuales los más empleados son el Análisis Discriminante Múltiple y Análisis Lógit (8), presentan el inconveniente de exigir a las variables el verificar unas fuertes condiciones estadísticas propias de las distribuciones paramétricas, que en el caso de los datos contables no siempre se cumplen.

Todos estos factores han dado como resultado la búsqueda de técnicas alternativas más eficientes y adaptadas a las nuevas exigencias de las empresas. En esta línea destacamos las redes neuronales artificiales que, sin exigir tantos requisitos estadísticos, proporcionan clasificaciones y predicciones con índices de aciertos comparables o sensiblemente mejores que los análisis multivariantes.

Las redes neuronales artificiales imitan la estructura del sistema nervioso, con la intención de construir sistemas de procesamiento de la información paralelos, distribuidos y adaptativos, que puedan presentar un cierto comportamiento inteligente.

Existen, no obstante, muchos detractores de las redes neuronales artificiales, que alegan que la mejora en la clasificación y predicción que proporcionan frente a la estadística tradicional no compensan la complejidad del funcionamiento y de la interpretación de los resultados.

La motivación de este trabajo ha sido la de, siguiendo la línea recientemente expuesta, desarrollar un modelo neuronal de predicción de insolvencia fiable para las empresas españolas no cotizadas, en particular para las empresas de los sectores de Construcción e inmobiliaria y Energía e industria, comparándolo con un modelo estadístico tradicional obtenido a partir de la misma base de datos.

Modelos de predicción de insolvencia

Técnicas estadísticas

El análisis financiero ha motivado el desarrollo de un gran número de técnicas con el propósito de ayudar a los decisores y a los analistas financieros en la predicción de la insolvencia, cuyos inicios se remontan a principios del siglo pasado. Están documentados trabajos de los años treinta que utilizan modelos univariantes básicos.

Pero es en la década de los sesenta cuando se produce un gran salto cualitativo con las aportaciones en 1966 de Beaver (9), quien incorpora a la investigación de la predicción de insolvencia en la empresa, el análisis discriminante univariante, consistente en la evaluación separada de la capacidad predictiva de cada uno de los ratios considerados.

El análisis univariante, que suele ser la primera fase de cualquier investigación estadística, supone, si es el único análisis utilizado, excesiva simplicidad. Estas técnicas tratan de determinar la significatividad de cada ratio para explicar la quiebra y la posible predicción de los ratios más significativos.

Sin embargo estos enfoques univariantes fueron pronto abandonados ya que no captaban la estructura multidimensional de la empresa. Bajo esta nueva orientación, las investigaciones basadas en el Análisis Discriminante Múltiple fueron imponiéndose, destacando los trabajos de Altman (4), cuyo modelo Z-score constituye el tipo de análisis más cualitativo de riesgo de fallido y por ello utilizado durante mucho tiempo por agencias de rating como Standard&Poor's o Moody's. Altman desarrolló una función discriminante resultado de la combinación de cinco ratios financieros cuya aplicación permitía una clasificación entre empresas sanas y fracasadas con un porcentaje de aciertos del 96% un año antes de la quiebra y de un 70% cinco años antes de la quiebra.

Otras técnicas igualmente muy utilizadas por los investigadores en la elaboración de modelos de predicción de solvencia son la Regresión Logística, o el Análisis Próbit (10).

Sin embargo la aplicación práctica de estos métodos se ve limitada por las estrictas condiciones estadísticas que deben cumplir, como son la normalidad multivariante, homogeneidad de las matrices de varianza-covarianza (homocedasticidad), linealidad y ausencia de multicolinealidad.

Técnicas inteligentes

Básicamente la investigación en Inteligencia Artificial (IA) se centra en el estudio del comportamiento inteligente desde un doble objetivo: aumentar la utilidad de las máquinas en general (ordenadores y procesos informáticos en particular) y analizar y comprobar los procesos de la inteligencia humana (11).

Desde hace décadas, numerosos estudios están aplicando técnicas de inteligencia artificial para la predicción financiera (12). Actualmente estas técnicas incluyen:

- i) árboles de decisión (ID3) y árboles de clasificación y regresión (CART);
- ii) diferentes arquitecturas de Redes Neuronales Artificiales (RNA), incluyendo la red Perceptrón MultiLínea (PML), los mapas autoorganizados (SOM, Self Organized Maps), o la red de *Learning Vector Quantization* (LVQ);
- iii) aproximaciones evolutivas como los algoritmos genéticos (GA) u optimización de enjambre de partículas (PSO, Particle Swarm Optimization);
- iv) nuevas técnicas utilizando máquinas de vectores soporte (SVM, Support Vector Machines).

Redes Neuronales Artificiales

Las Redes Neuronales Artificiales (RNA) siempre han llamado la atención a los estadísticos por su relativo buen comportamiento ante problemas no lineales o datos con mucho ruido, es decir datos incompletos o erróneos. Esto es lo que hace que se hable de ellas como técnicas de distribución libre o no paramétricas. Por ello las RNA han sido aplicadas a problemas de tradición estadística como son la predicción y la clasificación.

Las RNA trabajan sobre la base de reconocimiento de patrones, de manera que tienen la capacidad de adquirir conocimiento a partir de los ejemplos. Esta estructura otorga a la RNA gran capacidad para procesar datos y la habilidad para realizar procesos inteligentes como: aprender a partir de ejemplos, generalizar el conocimiento adquirido a nuevos casos y reconocer tendencias y patrones en los datos.

Su estructura se inspira en las primeras etapas de procesamiento del sistema nervioso biológico. Están formadas por numerosos elementos procesadores de información, que serían los equivalentes a las neuronas biológicas, interconectados entre sí y estructurados en niveles o capas. Existe un nivel de entrada, que introduce los datos a la red, y un nivel de salida que proporciona la respuesta a los datos de entrada, y uno o más niveles que procesan los datos. Aprenden la relación entre los datos de entrada y de salida, por lo cual, todo lo que se necesita para entrenar una RNA es un conjunto de datos que contengan la relación entrada-salida.

En realidad, las RNA son internamente modelos matemáticos multivariantes que utilizan procedimientos iterativos para minimizar unas funciones de error.

Los primeros trabajos sobre predicción de insolvencia utilizando redes neuronales son de los años 90. Sun et al (13) realizan una revisión de los trabajos más significativos sobre predicción de insolvencia aplicando diferentes arquitecturas de redes neuronales y comparándolas con técnicas estadísticas. De esta revisión se extrae la conclusión de que la ventaja fundamental de las redes neuronales es su fuerte capacidad de mapeo basada en la estructura de la red, además de que las relaciones estadísticas de las variables no requieren ser consideradas en el proceso de construcción del modelo neuronal. Como inconvenientes se destaca la necesidad de mayores muestras de datos y la dificultad de comprensión por parte del usuario.

En este sentido ya hay trabajos (13) para relacionar la forma de aprender de la red neuronal con unas reglas explicativas que permitan hacer uso de esta potencial ayuda en el mundo real.

Otra característica fundamental de las RNA es que son sistemas entrenables, es decir que son capaces de realizar un determinado tipo de procesamiento o cómputo aprendiéndolo a partir de un conjunto de ejemplos.

Cuando se construye una red neuronal artificial, se parte de una cierta arquitectura de red, con unos pesos sinápticos iniciales aleatorios. El proceso mediante el cual esos pesos se modifican hasta hacer la red operativa se denomina fase de aprendizaje y básicamente consiste en ir modificando los pesos de las sinapsis entre neuronas a partir de la optimización de una función error o coste, que mide la eficacia de la operación de la red. Este proceso es iterativo, actualizándose los pesos cada vez hasta que la red alcanza el rendimiento deseado.

El aprendizaje puede ser:

- supervisado, si le facilitamos ejemplos entrada-salida deseada para que, de forma iterativa, ajuste sus pesos de acuerdo a los mismos
- no supervisado, si se presentan patrones a la red sin adjuntar la respuesta deseada, de tal forma que es la red misma la que utilizando una regla de aprendizaje, es capaz de extraer rasgos de los datos, agrupar patrones según su similitud o reconocer regularidades en las entradas
- o híbrido si se basa en la coexistencia en la red de los dos tipos anteriores de aprendizaje, los cuales tienen lugar normalmente en distintas capas de neuronas.

Una vez que el sistema ha sido entrenado, los pesos y la estructura quedan fijos, y la red neuronal está preparada para procesar datos, dando lugar a la fase de recuerdo o ejecución.

Algunas redes neuronales son la red perceptrón multilínea (MLP, *MultiLayed Perceptron*), la red de función de base radial (RBF, *Radial Basis Function*), la red probabilística neural (PNN, *Probabilistic Neural Network*), la red neuronal de correlación en cascada (Cascor), la red de aprendizaje de vector de cuantización (LVQ, *Learning Vector Quantization*), la red de mapas autoorganizados (SOM, *Self Organized Maps*) son algunas de las

estructuras más comunes. Difieren en aspectos como el algoritmo de entrenamiento, las conexiones entre las neuronas, o el tipo de aprendizaje.

Investigación del fracaso en la empresa

La variable dependiente

Existen diferentes interpretaciones sobre cuándo una empresa fracasa o entra en crisis. Una de las más objetivas sería la de considerar una empresa fracasada cuando presenta concurso de acreedores (3), lo cual permite identificar fielmente el fracaso y el momento en el que se produce. Pero ello suele conducir a reducir la muestra de datos e incluso a falsearla, ya que el concurso supone, por un lado, unos costos fijos altos que algunas empresas no quieren asumir decidiendo optar por soluciones alternativas, como anexionarse con otras empresas, y por otro, antes de llegar a un concurso y con el objeto de evitarlo muchas veces las empresas falsean la información contable de los ejercicios previos.

Para alcanzar el objetivo pretendido en este estudio, y con el fin de ampliar lo más posible la muestra de empresas no solventes, hemos preferido adoptar el criterio de quiebra técnica, entendiendo por ésta aquella situación en la que la empresa no puede cumplir con sus obligaciones financieras con sus acreedores, proveedores y/o vendedores y que puede conducir a la empresa tanto a un proceso concursal como a una liquidación de la misma, es decir, en términos contables, la situación en la que el valor de sus deudas supere al total de sus inversiones (fondos propios negativos). Es cierto que este criterio también presenta problemas, ya que la existencia de fondos propios negativos puede no desembocar necesariamente en concurso o liquidación. Por ejemplo, el valor contable de algunos activos no afectos a la explotación pudiera ser inferior al valor de mercado, con lo que una posible venta reequilibraría la situación financiera de la empresa. Igualmente podría decirse de las empresas que manipulan los datos contables, a fin de presentar unos balances más saneados de los reales, máxime cuando son de un tamaño inferior al requerido para estar obligadas a ser examinadas por una auditoría externa. De cualquier forma serían casos muy puntuales que se dispersarían en muestras de un tamaño adecuado.

La variable dependiente es por tanto una variable nominal dicotómica que denominamos *Solvencia*, cuyos valores son 1 en caso de empresa solvente y 0 en caso de empresa en quiebra técnica.

La muestra

A pesar del decrecimiento en el ritmo de la construcción de los últimos años, es indudable la importancia del sector de la Construcción en la composición del PIB de España y en la generación de empleo. Antes del estallido de la burbuja inmobiliaria este sector representaba el 10% del PIB, daba empleo al 11,3 % de la población ocupada y constituía el 57% de la inversión que se realizaba en España. Fundamentalmente su importancia se debe al efecto arrastre que genera y a su fuerte relación con el crédito total. Su problema más relevante es que es muy dependiente del ciclo económico, razón por la cual las empresas constructoras grandes suelen diversificar su actividad a otros sectores o países más rentables, mientras que las de menor tamaño son más vulnerables a las situaciones adversas.

El sector de la Industria también tiene un efecto multiplicador sobre el resto de la economía, generando en media en forma indirecta 1,61 euros de PIB y 1,43 empleos por cada euro de PIB y empleo directo. Uno de los principales lastres para el crecimiento e internacionalización de la empresa industrial española es su reducido tamaño medio, que en media es más pequeña que la europea, lo que dificulta el aprovechamiento de economías de escala, el acceso al crédito, la inversión en I+D y la entrada en mercados internacionales, y merma la capacidad negociadora frente a proveedores y clientes.

La mayoría de los estudios de predicción de quiebra se han realizado sobre entidades financieras o sobre empresas que cotizan en Bolsa, ya que ambos fueron los primeros obligados a hacer públicos sus datos contables. Sin embargo el tejido empresarial español es básicamente de las PYMES y una parte importante del mismo, del sector de la Construcción y del sector de la Industria, que debido a la situación de crisis crediticia y recesión económica actual se enfrentan a dificultades y desafíos que ponen en riesgo su supervivencia.

Por esta razón, para la realización de este trabajo hemos utilizado una base de datos que, después de filtrar y eliminar los casos incompletos y erróneos, está formada por 318 empresas procedentes de los sectores *Construcción y promoción inmobiliaria e Industria y energía*, obtenidas de la agencia e-Infirma correspondientes a los años 2009 a 2011. Puesto que, afortunadamente, son más las empresas que sobreviven que las que no, la muestra refleja esa realidad. De esta forma la base consta de los datos de 205 empresas solventes y 113 empresas insolventes.

Para la aplicación de las dos técnicas objeto de nuestro estudio, la muestra completa se subdivide en dos de forma aleatoria. La primera es la muestra de entrenamiento y representa el 60% aproximadamente del total, mientras que el 40% restante es la muestra de validación o test, que se utiliza para estudiar la capacidad de predicción de los modelos obtenidos con la primera.

Las variables independientes

En cuanto a las variables independientes que permiten explicar el comportamiento de la variable Solvencia, tradicionalmente los modelos de predicción de quiebra se elaboran a partir de ratios económicos obtenidos de la información financiera de las empresas. A este respecto existe abundante literatura y diversidad de opiniones tanto a la hora de elegir los ratios más adecuados como la cantidad de los mismos, y si es conveniente o no añadir variables descriptivas de tipo edad de la empresa, tamaño, informes de auditoría, o incluso factores externos de tipo macroeconómico, como la tasa de crecimiento del PIB, etc.

Pero lo cierto es que no existe ninguna teoría económica que establezca un criterio único universal. En el estudio realizado por Ravi (10), se pueden consultar los ratios elegidos en una gran variedad de estudios y que varían desde los tres ratios, como en el caso del trabajo realizado por Zmijewski (6) hasta los 107 ratios en el caso del desarrollado por Ferrando y Blanco(14), por poner algunos ejemplos.

El modelo de predicción de insolvencia desarrollado por Altman(4) ha sido ampliamente aplicado en finanzas y en la investigación contable. Incluso muchos académicos lo proponen como un estándar de comparación con estudios de clasificación posteriores de insolvencia. Son muchos los trabajos en los que se han utilizado estas mismas variables para probar la eficacia de técnicas estadísticas y/u otros modelos de predicción, como los de Wilson(15), Sánchez-Laseheras (16) o Serrano Cinca (17). Algunos autores, como Begley (18) o Grice (19), han estudiado la aplicabilidad del modelo de Altman en distintos contextos temporales y geográficos, llegando a la conclusión de que con la modificación adecuada de los coeficientes, la ecuación de Altman sigue representando una aproximación válida para la predicción de la insolvencia.

Por otra parte, cuando las muestras no son suficientemente extensas, no es conveniente elegir una gran cantidad de variables porque es más difícil ajustar los parámetros perdiendo con ello fiabilidad y estabilidad.

Por último, los ratios del modelo de Altman, como veremos a continuación, corresponden a datos contables disponibles para cualquier empresa, lo cual permite no tener que desprestigiar casos por falta de la información financiera correspondiente.

Todas estas razones han sido las que nos han motivado a la hora de decidir cómo variables independientes de nuestro estudio los ratios propuestos por Altman y que aparecen en la Tabla 1.

Cada ratio refleja algún aspecto de la realidad financiera de la empresa y se encuadran dentro de las categorías más ampliamente utilizadas en los modelos de predicción de quiebra:

- R01 = capital de trabajo/total activos, como medida de los activos líquidos en relación con el tamaño de la empresa. Mide la relación del capital de trabajo, entendido como el efectivo que posee la empresa en caja, cuentas corrientes, cuentas por cobrar en un año, tras haber pagado sus deudas en el corto plazo con sus activos disponibles. Valores relativamente bajos en este ratio podrían indicar niveles de liquidez bajos.

- R02 = recursos propios/total activos, como medida del apalancamiento financiero de una compañía. Cuanto más elevado sea este ratio, y suponiendo que se comparan estructuras de inversión y riesgo semejantes, mayor es la solvencia pero menor el apalancamiento financiero. La pugna entre la rentabilidad y la solvencia se manifiesta por el grado de apalancamiento permitido, dejando a un lado la calidad de inversiones y recursos.
- R03 = ganancias antes de intereses y tasas /total activos, representa la eficiencia operativa, aparte de los impuestos y los factores de apalancamiento. Reconoce los beneficios de explotación como importantes para la viabilidad a largo plazo.
- R04 = Valor contable/pasivo exigible, también denominado como ratio de autonomía financiera. Cuanto mayor sea, menos dependencia tendrá la empresa de la financiación ajena. Sirve como indicador de la estabilidad financiera de que goza la empresa. En época de crisis interesa que el índice sea elevado, en épocas de crecimiento las empresas tienden a endeudarse más.
- R05 = total ventas /total activos, como medida estándar para el volumen de negocios total de los activos. Representa la rotación del activo total y mide el volumen de negocio en relación con el activo total y por tanto la eficiencia operativa.

Tabla 1: Variables independientes o Factores

R01	Capital de trabajo neto/total activos
R02	Reservas acumuladas/total activos
R03	Ganancias antes de intereses y tasas/total activos
R04	Valor contable/ pasivo exigible
R05	Total ventas/total activos

Metodologías aplicadas

Análisis Discriminante

Dentro de los métodos estadísticos que pueden aplicarse en los modelos de predicción de quiebra, el más extendido tradicionalmente es el Análisis Discriminante (AD).

El AD es una técnica estadística que permite construir un modelo predictivo para pronosticar el grupo al que pertenece una observación a partir de unas determinadas características que delimitan su perfil. De esta forma permite asignar o clasificar nuevos individuos u observaciones dentro de grupos previamente definidos.

Las dos grandes finalidades perseguidas en el uso del AD son la descripción de diferencias entre grupos y la predicción de pertenencia a grupos. La interpretación de las diferencias entre los grupos responde al objetivo de determinar en qué medida un

conjunto de características observadas en los individuos permite extraer dimensiones que diferencien a los grupos, y cuáles de estas características son las que en mayor medida contribuyen a tales dimensiones, es decir, cuáles presentan el mayor poder de discriminación. Las características usadas para diferenciar entre los grupos reciben el nombre de variables discriminantes, y a la ecuación matemática que las relaciona y que define el modelo se denomina función discriminante:

$$F_i = W_1 X_{1i} + W_2 X_{2i} + \dots + W_n X_{ni} + C$$

siendo: F_i la puntuación discriminante que alcanza la empresa i de la muestra,
 W_i los coeficientes de ponderación de las variables discriminantes o clasificadoras,
 X_{ji} el valor del ratio j en la empresa i ,
 C una constante

Una vez estimada la función discriminante, la capacidad predictiva del modelo se evalúa mediante el establecimiento de una puntuación de corte óptima, que permite asignar los casos a cada uno de los grupos definidos por la variable dependiente, obteniéndose la puntuación discriminante que corresponde a cada caso a partir de los valores que presenta el individuo en la combinación de variables explicativas que forman la función discriminante.

Estrictamente hablando, la función discriminante minimiza la probabilidad de equivocarse al clasificar los individuos en cada grupo. Para ello, la aplicación del AD se apoya en una serie de supuestos básicos como la normalidad multivariante, homogeneidad de matrices de varianza-covarianza (homocedasticidad), linealidad y ausencia de multicolinealidad, si bien algunos autores (14) han demostrado que el AD es una técnica bastante robusta que puede tolerar algunas desviaciones en estas hipótesis iniciales.

El supuesto de normalidad exige que cada grupo presente una muestra aleatoria extraída de una población con distribución normal multivariante sobre las variables discriminantes. La normalidad univariante no implica la multivariante, pero como esta última es difícil de comprobar, se contrasta la univariante con pruebas clásicas como la prueba de bondad de ajuste basada en Chi-Cuadrado, la prueba de Kolmogorov-Smirnov, el test de Shapiro Wilk o las pruebas de significación basadas en la asimetría y la curtosis.

El supuesto de homocedasticidad obliga a que las matrices de varianzas-covarianzas para las poblaciones de las que fueron extraídos los grupos sean iguales, lo cual suele probarse mediante la prueba M. de Box que se basa en los determinantes de las matrices de varianzas-covarianzas para cada grupo.

Por otro lado, la linealidad implica que existen relaciones lineales entre las variables dentro de cada grupo y suele comprobarse a partir de los diagramas de dispersión de las variables o mediante el cálculo de los coeficientes de correlación lineal de Pearson. La matriz de correlaciones de las variables también se utiliza para detectar la multicolinealidad, variables con correlación muy alta pueden ser redundantes (20).

Resultados del Análisis Discriminante

Realizamos con SPSS (21) Neural Networks 20, el Análisis Discriminante tomando como variable dependiente *solvencia* y como variables independientes los ratios *R01*, *R02*, *R03*, *R04* y *R05*.

Mediante el estadístico descriptivo previo verificamos la condición de normalidad en ambos grupos de la variable *solvencia*. Tanto con el estadístico de Kolmogorov-Smirnov como con el de Shapiro-Wilk se obtiene un alto grado de normalidad de acuerdo a la significación de ambos estadísticos.

Para los contrastes de significación, utilizamos el estadístico *Lambda de Wilks*, que es el cociente entre los determinantes de la matriz suma de cuadrados y productos cruzados intragrupos en el análisis de la varianza múltiple, y la matriz de la suma de cuadrados y de productos cruzados total. Según este contraste observamos que se rechaza la igualdad de medias de los ratios *R01*, *R02* y *R03* en los dos grupos (p-valores inferiores a 0,05) y se acepta la igualdad de medias para los ratios *R04* y *R05*, que son por tanto los que no se van a utilizar para la función discriminante (Tabla 1).

	Lambda de Wilks	F	gl1	gl2	Sig.
R01	0,932	13,988	1	192	0,000
R02	0,583	137,445	1	192	0,000
R03	0,949	10,224	1	192	0,002
R04	0,990	1,882	1	192	0,172
R05	0,999	0,104	1	192	0,748

M de Box		199,566
F	Aprox.	32,691
	gl1	6
	gl2	243786,744
	Sig.	0,000

Siguiendo con los requisitos iniciales, la prueba M. de Box contrasta con un alto grado de significación, la hipótesis nula de que las matrices de covarianzas poblacionales sean iguales (Tabla 2).

De las distintas opciones que ofrece el paquete estadístico SPSS a la hora de introducir las variables discriminadoras en el modelo, elegimos el procedimiento iterativo paso a paso, con el criterio para ir introduciendo los mejores ratios en la ecuación de la minimización de la *lambda de Wilks*. De esta forma en la primera iteración entra en la función el ratio que proporciona la mayor diferenciación entre los dos grupos de empresas, es decir, aquel cuya *lambda de Wilks* es menor. A continuación se va probando con el ratio elegido en primer lugar y el resto de los cinco ratios, y así sucesivamente hasta que la adición de nuevos ratios no mejore significativamente la capacidad discriminadora del modelo.

En ese proceso de búsqueda por pasos de la función discriminante a partir de las variables independientes para clasificar a las empresas según los valores de la variable *solvencia*, las variables introducidas para discriminar en el modelo son definitivamente los ratios *R02*, *R03* y *R01*, los cuales, de acuerdo al estadístico *F exacta*, tienen una buena capacidad explicativa según nos certifican los p-valores (*Sig.*) nulos obtenidos (Tabla 3).

Tabla 3. Variables introducidas/excluidas									
Paso	Introducidas	Lambda de Wilks							
		Estadístico	gl1	gl2	gl3	F exacta			
						Estadístico	gl1	gl2	Sig.
1	R02	0,583	1	1	192,000	137,445	1	192,000	0,000
2	R03	0,548	2	1	192,000	78,690	2	191,000	0,000
3	R01	0,533	3	1	192,000	55,478	3	190,000	0,000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

Los coeficientes de la función discriminante canónica estandarizados, donde las variables clasificadoras están tipificadas para evitar los problemas de escala que pudieran existir en los ratios, son los que aparecen a continuación, quedando la función de la siguiente forma:

$$f = 1,303 \times R02 - 0,318 \times R01 - 0,351 \times R03$$

Vemos que el ratio que mayor coeficiente tiene en la función es el *R02* (Reservas acumuladas/total activos). No en vano el ratio generalizado de solvencia (Patrimonio neto *100/activo total) es otra versión del ratio *R02*. Sin embargo en este sentido habría que ser cautos porque podría ocurrir que algunos de los activos fueran de dudoso cobro o bien que correspondieran a existencias de difícil realización en el mercado. Según la función discriminante, los siguientes ratios en importancia son el de rentabilidad de los activos (*R03*) y el relacionado con la liquidez (*R01*).

Finalmente, la bondad del modelo se evalúa a partir de la matriz de clasificación (Tabla 4) que muestra los casos en total que están correcta o incorrectamente clasificados. Vemos que, de la muestra de entrenamiento, hay un 93,2% de aciertos para la insolvencia, y un 90,6% para la solvencia, que da un total conjunto de 91,8% de aciertos en clasificación. En cuanto a la muestra de validación, o casos no clasificados inicialmente, tenemos un 92,9% de aciertos en insolvencia, y un 87,5% de aciertos en solvencia, lo cual significa un porcentaje total de aciertos en predicción del 90,2%.

Tabla 4. Resultados de la clasificación ^{a,b}						
			Solvencia	Grupo de pertenencia pronosticado		Total
				0	1	
Casos seleccionados	Original	Recuento	0	82	6	88
			1	10	96	106
		%	0	93,2	6,8	100,0
			1	9,4	90,6	100,0
Casos no seleccionados	Original	Recuento	0	26	2	28
			1	12	84	96
		%	0	92,9	7,1	100,0
			1	12,5	87,5	100,0
a. Clasificados correctamente el 91,8% de los casos agrupados originales seleccionados.						
b. Clasificados correctamente el 90,2% de casos agrupados originales no seleccionados.						

A partir de la muestra de validación calculamos también los errores Tipo I y Tipo II que comete el modelo. El error Tipo I es el correspondiente a aquellos casos en los que el modelo predice fracaso cuando debería predecir solvencia. Como vemos en la tabla 8 este error es del 12,5%. El error Tipo II es el que tiene lugar cuando el modelo predice solvencia para una empresa en fracaso, y que vemos que con el modelo discriminante es del 7,1%. Altman (4) realizó un estudio empírico del que resultó que el coste del error Tipo I es 35 veces mayor que el del Tipo II. Por esta razón vamos a intentar minimizar estos errores introduciendo técnicas inteligentes.

Red neuronal perceptrón multicapa

El uso del algoritmo de aprendizaje por retropropagación del error (*BackPropagation BP*) es lo que caracteriza un tipo específico de redes neuronales que se denominan redes perceptrón multicapa o MLP (*MultiLayer Perceptron*). La mayoría de las redes MLP están formadas por tres capas de unidades o neuronas (capa de entrada - capa oculta - capa de salida), aunque también se pueden encontrar topologías con dos o más capas ocultas. Cada unidad de la capa de entrada incorpora a la red el valor de una variable independiente, las unidades de la capa oculta se encargan del modelado y cada unidad de la capa de salida calcula el valor de una variable dependiente cuantitativa o de una categoría de una variable dependiente categórica.

Las unidades de cada capa están conectadas con todas las unidades de las capas adyacentes y cada conexión tiene una fuerza denominada peso de la conexión, que es determinada durante la fase de entrenamiento mediante el algoritmo de retropropagación del error. A su vez, cada unidad de las capas oculta y de salida tiene un peso de tipo umbral, que con funciones de activación de tipo sigmoideal debe ser fijado a un valor diferente de cero para garantizar que las zonas de decisión no se solapen innecesariamente. El peso umbral interviene en el cálculo del valor de activación como un término aditivo.

La información se transmite en el sentido de la capa de entrada hacia la de salida, por lo que se trata de redes con conexiones hacia adelante. Es importante no confundir el sentido en que fluye la información de los datos, con el sentido en que fluye el error, ya que es el primero y no el segundo, el que determina la clasificación de la red en las mencionadas redes con conexiones hacia adelante (*feedforward*) frente a las redes con conexiones hacia adelante y hacia atrás (*feedforward/feedback*) (22).

Como todas las redes neuronales, las redes MLP tienen dos modos de funcionamiento: entrenamiento y predicción. En el modo de predicción la red recibe información de las variables independientes, que es propagada a través de la red proporcionando un valor de la variable dependiente. En el modo de entrenamiento se realiza el procedimiento anterior y a continuación se compara el valor calculado por la red con el valor real, el error de predicción cometido se propaga hacia atrás (desde la capa de salida a la de entrada) para modificar los pesos de las conexiones con el objetivo de minimizar dicho error.

Una de las características más importantes de las redes MLP es que pueden aproximar cualquier función. La base matemática de esta afirmación se debe a Kolmogorov (23), quien demostró que una función continua de diferentes variables puede ser representada por la concatenación de varias funciones continuas de una misma variable. Este es el fundamento de la representación distribuida de la información de las redes neuronales artificiales, convirtiendo las redes MLP en particular, en herramientas de propósito general, flexibles y no lineales que, con suficientes datos y unidades ocultas pueden aproximar cualquier función con el nivel de precisión que se desee.

El objetivo del entrenamiento es hallar los pesos que proporcionan la mejor aproximación a la función que relaciona las entradas y las salidas del conjunto de datos de entrenamiento y consiste en la mayoría de los casos en un ejercicio de optimización de una función no lineal. Existen muchos métodos de optimización no lineales, pero no se puede decir que se haya encontrado un método universal mejor para todos los casos.

El problema del BP es que la red puede incurrir en el fenómeno de sobreaprendizaje o sobreajuste, que consiste en que la red en lugar de alcanzar el mínimo global de la función error, cae en un mínimo local debido a que en el propio proceso de aprendizaje la red *aprende demasiado*, es decir hasta los ruidos o errores de los patrones-ejemplos, dando lugar con ello a un error de generalización. Existen dos formas de luchar contra este fenómeno: la parada temprana o validación cruzada y limitar en lo posible siempre el tamaño de la red, eligiendo la arquitectura de tamaño mínimo que permita aprender los ejemplos disponibles.

Resultados de la Red Neuronal Perceptrón Multicapa

El software empleado de nuevo es el IBM SPSS 20 Neural Networks (21), que incluye la posibilidad de trabajar con algunas RNA, como la MLP o la RBF.

Elegimos una arquitectura personalizada con una única capa oculta, funciones de activación en la capa oculta de tangente hiperbólica y en la capa de salida de tipo sigmoide, que son funciones que cumplen con las condiciones de derivabilidad y continuidad requeridas para la aproximación de funciones (24).

Hacemos uso de la misma partición de la base de datos del método anterior, de manera que la muestra de entrenamiento nos sirve para estimar los parámetros de la red, y la muestra de prueba para prevenir el sobreajuste, así como para determinar de forma independiente la red final, la cual luego es aplicable a toda la base de datos y a cualquier base de datos nueva.

El procedimiento para entrenar mediante BP una arquitectura MLP dada es la siguiente:

- 1º) Establecer aleatoriamente los pesos y umbrales iniciales
- 2º) Para cada patrón μ del conjunto de aprendizaje:
 - a. Llevar a cabo una fase de ejecución para obtener la respuesta de la red ante el patrón μ -ésimo
 - b. Calcular las señales de error asociadas
 - c. Calcular el incremento parcial de los pesos y umbrales debidos a cada patrón μ .
- 3º) Calcular el incremento total (para todos los patrones) actual de los pesos y de los umbrales.
- 4º) Actualizar pesos y umbrales
- 5º) Calcular el error actual y volver a 2) si todavía no es satisfactorio.

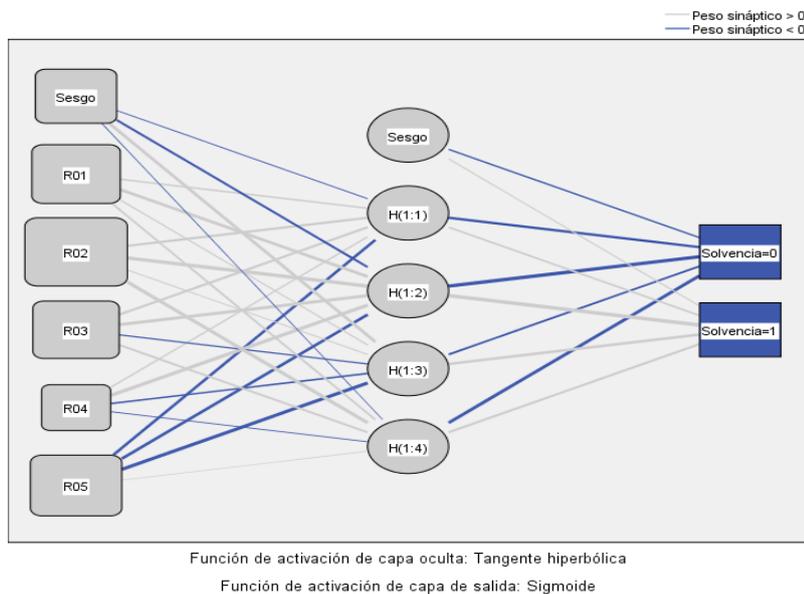


Figura 1: Red MLP 5:4:2. Fuente: IBM SPSS Neural Networks 20

Este proceso, que se realiza de forma automática, puede llevarse a cabo mediante *aprendizaje por lotes*: una fase de ejecución para todos y cada uno de los patrones del conjunto de entrenamiento, se calcula la variación en los pesos de cada patrón, se acumulan, y solamente entonces se procede a la actualización de los pesos, o mediante *aprendizaje en serie*, que consiste en actualizar los pesos sinápticos tras la presentación de cada patrón μ . Se ha demostrado que este último estima mejor el descenso del error, permite emplear tiempos de aprendizaje mayores y suele ser más rápido.

La red MLP obtenida (Figura 1) es una red 5:4:2, con los 5 ratios en la capa de entrada, 4 neuronas o unidades en la capa oculta y dos neuronas en la capa de salida correspondientes a los dos valores posibles de la variable dependiente solvencia.

El resumen de la clasificación realizada por la red neuronal se muestra en la Tabla 5. Vemos que para la muestra de entrenamiento, el modelo clasifica muy bien los casos de solvencia, con un 98,1% de aciertos, y los casos de insolvencia con un porcentaje de aciertos del 90,9 %.

La capacidad de una red entrenada para producir la respuesta correcta en registros de datos similares pero no idénticos a los de entrenamiento, es la conocida como *generalización*. En este caso, utilizando la muestra de prueba, vemos que la generalización de la red neuronal en conjunto es de 94,4% con una clara superioridad en la predicción de la insolvencia.

Tabla 5. Clasificación				
Muestra	Observado	Pronosticado		
		0	1	Porcentaje correcto
Entrenamiento	0	80	8	90,9%
	1	2	104	98,1%
	Porcentaje global	42,3%	57,7%	94,8%
Prueba	0	28	0	100,0%
	1	7	89	92,7%
	Porcentaje global	28,2%	71,8%	94,4%

Variable dependiente: Solvencia

En cuanto a los errores Tipo I y Tipo II, con la red neuronal observamos que teniendo en cuenta sólo la muestra de prueba, igual que hicimos con el modelo del Análisis Discriminante, los errores del Tipo I son del 0% ya que todos los casos de empresas fracasadas fueron correctamente clasificadas, y el porcentaje de errores del Tipo II es del 7%, lo cual mejora sensiblemente las predicciones realizadas con el estadístico discriminante.

La importancia de una variable independiente es una medida que indica cuánto cambia el valor pronosticado por el modelo de la red para diferentes valores de la variable independiente. El desconocimiento de este dato había sido hasta hace poco tiempo uno de los mayores inconvenientes de las redes neuronales, pero ya existen paquetes informáticos que proporcionan la importancia normalizada de las variables independientes, explicativas o predictoras. La importancia normalizada es el resultado de los valores de importancia divididos por el valor de la importancia mayor y expresados como porcentajes. Curiosamente vemos que, igual que obtuvimos con el modelo de AD, el ratio R^2 es el que mayor importancia cobraría en los pronósticos, pero en este caso el

siguiente ratio en importancia es el *R05* (*total ventas/total activos*), tal como se muestra en la Figura 2, poniendo de manifiesto que es la rotación del activo total, y por tanto el volumen de negocio en relación al activo total un factor trascendental en la viabilidad de las empresas.

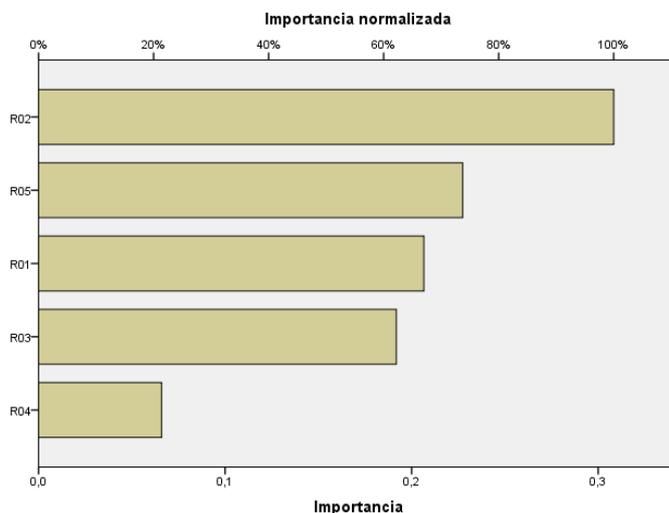


Figura 2: Importancia normalizada de las variables explicativas
Fuente: IBM SPSS Neural Networks 20

Conclusiones

El desarrollo de modelos de predicción de insolvencia ha sido considerado durante mucho tiempo una de las cuestiones más delicadas y complicadas en el ámbito de las finanzas y el mundo de los negocios. La capacidad de predecir con exactitud las dificultades financieras es fundamental para adoptar decisiones de forma adecuada, ya que una decisión incorrecta en este sentido puede conducir a una empresa a una situación de quiebra con el consiguiente perjuicio de todos los agentes implicados, desde trabajadores hasta inversores.

El objetivo de este trabajo era crear y comparar dos modelos de clasificación y predicción de solvencia o quiebra técnica en una muestra de empresas españolas no cotizadas.

Uno de ellos es el análisis discriminante multivariante utilizado tradicionalmente en modelos de predicción de fracaso en empresas, pero sujeto a unas condiciones estadísticas (normalidad, igualdad de matrices de varianzas y covarianzas y homocedasticidad), difíciles de cumplir en muchos casos, y el otro es una red neuronal perceptrón multicapa, que presenta la ventaja fundamental de evitar la verificación de esas condiciones iniciales impuestas al análisis discriminante.

Los modelos obtenidos con ambos métodos son significativos en cuanto a la clasificación realizada como han puesto de manifiesto los estadísticos calculados para valorar la significación de cada uno de ellos, proporcionando porcentajes de clasificación superior al 90% en ambos casos.

Sin embargo, hemos de concluir que el nivel global de acierto del modelo neuronal es superior al modelo discriminante, ya que considerando únicamente la muestra de

validación, es del 94,4% en la red neuronal frente a un 90,2% en el modelo de análisis discriminante.

En lo referente a los porcentajes de error de Tipo I, también el modelo neuronal es mejor, ya que hemos conseguido bajar del 12,5% del análisis discriminante al 0%, si bien en el error de Tipo II los porcentajes en ambos modelos son muy similares del orden del 7%. Teniendo en cuenta, que como demostró Altman (4) el coste del error Tipo I es 35 veces mayor que el del Tipo II, es clara la decisión, en este sentido, de cuál es el modelo más apropiado.

Por otro lado, es destacable la importancia de las variables explicativas otorgada por cada uno de los modelos. En el caso del análisis discriminante el ratio de mayor coeficiente es el *R02* (Reservas acumuladas/total activos). No en vano el ratio generalizado de solvencia (Patrimonio neto *100/activo total) es otra versión del ratio *R02*. Sin embargo en este sentido habría que ser cautos porque podría ocurrir que algunos de los activos fueran de dudoso cobro o bien que correspondieran a existencias de difícil realización en el mercado. Según la función discriminante, los siguientes ratios en importancia son el de rentabilidad de los activos (*R03*) y el relacionado con la liquidez (*R01*).

Análogamente, en el caso de la red neuronal MLP, el primer ratio en importancia es también el *R02*, pero seguido del ratio *R05* (total ventas/total activos), poniendo de manifiesto que es la rotación del activo total, y por tanto el volumen de negocio en relación al activo total un factor trascendental en la viabilidad de las empresas.

Podemos concluir que las redes neuronales artificiales constituyen un conjunto de técnicas más dentro de la amplia panoplia de métodos de tratamiento de datos, clasificación y predicción ya existentes, que no requieren una complejidad excesiva y que sin embargo proporcionan una mejora significativa en los resultados, por lo cual las recomendamos como herramienta en las tareas de gestión de las pequeñas y medianas empresas españolas.

BIBLIOGRAFIA

- (1) Disponible en <http://itemsweb.esade.edu/biblioteca/archivo/informeeconomicojunio2013.pdf> [Consultado el 28 de enero de 2014]
- (2) Instituto Nacional de Estadística. Información disponible en <http://www.ine.es> [Consultado el 15 de enero de 2014]
- (3) Gill de Albornoz, B.; Giner, B. "Predicción del fracaso empresarial en los sectores de construcción e inmobiliario: Modelos generales versus específicos". *Universia Business Review*. Tercer trimestre 2013. Madrid. España.
- (4) Altman, E. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", *The Journal of Finance*, 23 (4), (1968).
- (5) Ohlson, J.A. "Financial ratios and the probabilistic prediction of bankruptcy", *Journal of Accounting Research*, 18 (1), (1980).
- (6) Zmijewski, M.E. "Methodological issues related to the estimation of financial distress prediction models", *Journal of Accounting Research*, 22, (1984).
- (7) Kim, S.Y.; Upneja, A. "Predicting restaurant financial distress using decision tree and AdaBoosted tree models", *Econ. Model.* 36, (2014).
- (8) Dimitras, A.I.; Slowinski, R.; Susmaga, R. y Zopounidis, C. "Business Failure Prediction Using Rough Sets". European
- (9) Beaver, W. "Financial ratios as predictors of failure", *J. Account. Res.* 4 (1966).
- (10) Ravi Kumar, P., Ravi, V. "Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review", *European Journal of Operational Research*, 180, (2007).
- (11) Fortuna Lindo, J.M.; Busto Marroquín, B.; Sastre Centeno, J.M. (1991): "Los sistemas expertos: fundamentos y aplicaciones a la contabilidad". Partida Doble, Vol.17.
- (12) Chen M. –Y. "Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches", *Computers and Mathematics with Applications*, 62, (2011).
- (13) Sun, J.; Li, H.; Huang, Q. –H.; He, K. –Y. "Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modelling, sampling, and featuring approaches", *Knowl. Based. Syst.* 2014
- (14) Ferrando Bolado, M.; Blanco Ramos, F. "La Previsión del Fracaso Empresarial en la Comunidad Valenciana: Aplicación de los modelos discriminante y logit". *Revista Española de Financiación y Contabilidad*. Vol. XXVII, 95, (1998).
- (15) Wilson, R.L.; Sharda, R. "Bankruptcy prediction using neural networks", *Decision Support Systems*, (1994).
- (16) Sánchez-Lasheras, F.; de Andrés, J.; Lorca, P.; de Cos, F.J. "A hybrid device for the solution of sampling bias problems in the forecasting of firms' bankruptcy", *Expert Systems with Applications*, 38, (2012).
- (17) Serrano Cinca, C. "Self organizing neural networks for financial diagnosis". *Decision Support Systems*, 17 (3), (1996).

- (18) Begley, J.; Ming, J.; Watts, S. "Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models", *Review of Accounting Studies*, 1 (4), (1996).
- (19) Grice, F.S.; Ingram, R.W. "Tests of generalizability of Altman's bankruptcy prediction model", *Journal Business Research*, 54 (1), (2001).
- (20) Pérez López, C. "Modelos econométricos con SPSS". Ibergarceta publicaciones. Madrid, 2009.
- (21) IBM SPSS Neural Networks 20, Información disponible en <http://www.ibm.com/spss>. [Consultado el 6 de marzo de 2014]
- (22) Ripley, B. D. "Pattern recognition and neural networks". Cambridge Univ. Press. Cambridge, G.B. 1996.
- (23) Kolmogorov, A. N. "On the Representation of Continuous Functions of Several Variables by Superposition of Continuous Functions of one Variable and Addition". *Doklady Akademii. Nauk USSR*, 114, (1957).
- (24) Martín del Brío, B.; Sanz A. "Redes Neuronales y Sistemas Borrosos", 3ª Edición, RA-MA Editorial, Madrid, 2006.