Romero Martínez, Mariano; Carmona Ibáñez, Pedro; Pozuelo Campillo, José

Article

Utilidad del Deep Learning en la predicción del fracaso empresarial en el ámbito europeo

Revista de Métodos Cuantitativos para la Economía y la Empresa

Provided in Cooperation with:
Universidad Pablo de Olavide, Sevilla

Suggested Citation: Romero Martínez, Mariano; Carmona Ibáñez, Pedro; Pozuelo Campillo, José (2021) : Utilidad del Deep Learning en la predicción del fracaso empresarial en el ámbito europeo, Revista de Métodos Cuantitativos para la Economía y la Empresa, ISSN 1886-516X, Universidad Pablo de Olavide, Sevilla, Vol. 32, pp. 392-414, https://doi.org/10.46661/revmetodoscuanteconempresa.5172

This Version is available at:
https://hdl.handle.net/10419/286257

REVISTA DE MÉTODOS CUANTITATIVOS PARA LA ECONOMÍA Y LA EMPRESA (32). Páginas 392-414.
Diciembre de 2021. ISSN: 1886-516X. D.L: SE-2927-06.
www.upo.es/revistas/index.php/RevMetCuant/article/view/5172

Utilidad del Deep Learning en la predicción del fracaso empresarial en el ámbito europeo

ROMERO MARTÍNEZ, MARIANO
Universidad de Valencia
Correo electrónico: mromero@uv.es

CARMONA IBÁÑEZ, PEDRO
Universidad de Valencia
Correo electrónico: pedro.carmona@uv.es

POZUELO CAMPILLO, JOSÉ
Universidad de Valencia
Correo electrónico: jose.pozuelo@uv.es

RESUMEN

En este trabajo pretendemos constatar la utilidad de las redes neuronales del Deep Learning en la predicción del fracaso empresarial, en particular, de las redes de alimentación hacia adelante (feedforward neuronal networks, en inglés). Se trata de una metodología caracterizada por proporcionar muy buenos resultados en términos de capacidad predictiva cuando se dispone de grandes tamaños muestrales. Para ello hemos desarrollado un modelo de predicción empresarial en empresas europeas, basado en dicho algoritmo, sobre una muestra formada por 61.624 empresas de las cuales 12.128 fueron declaradas en concurso en 2016. Como variables independientes se han considerado ratios y magnitudes económico-financieras obtenidas de las cuentas anuales del año anterior a la fecha del fracaso. Deep Learning logra una capacidad predictiva del 94%, de manera que presentan una mayor propensión al fracaso aquellas que tienen un mayor tamaño y una solvencia menor. Los resultados que se presentan se han contrastado en una submuestra de comprobación independiente y diferente a la empleada para estimar el modelo.

Palabras clave: fracaso empresarial, Deep Learning, aprendizaje automático, ratios financieros, modelo de predicción.
Clasificación JEL: C53.
MSC2010: 62M45; 62P20; 91B99.

Artículo recibido el 24 de julio de 2020 y aceptado el 23 de junio de 2021.

The usefulness of Deep Learning in the prediction of business failure at the European level

ABSTRACT

In this paper we intend to substantiate the usefulness of Deep Learning, especially feedforward neuronal networks, in the prediction of business failure. This methodology provides very good results in terms of predictive performance when large sample sizes are available. Therefore, we have developed a business failure prediction model for European companies, based on this algorithm on a sample of 61,624 companies, of which 12,128 were declared bankrupt in 2016. As independent variables were considered ratios, and economic and financial data obtained from the financial statements for the year preceding the date of failure. Deep Learning achieves a predictive performance of 94%, where companies with larger size and lower solvency are more prone to failure. The obtained results have been tested on an independent test sample, different from that used to estimate and train the model.

Keywords: business failure, Deep Learning, machine learning, financial ratios, prediction model.
JEL classification: C53.
MSC2010: 62M45; 62P20; 91B99.

1. Introducción.

La reciente crisis económica vivida y la consecuente desaparición de empresas que provocó ha estimulado el interés de investigadores en el análisis del fracaso empresarial. Ello ha supuesto la aparición de numerosos estudios que, desde diferentes perspectivas y sirviéndose a menudo de nuevas metodologías, analizan sus causas e identifican los indicadores que anticipan las situaciones de dificultad financiera.

Entre las metodologías más novedosas utilizadas recientemente en este ámbito de investigación destacan el *Gradient Boosting Machine* (GBM) y las basadas en el *Machine Learning* o Inteligencia Artificial, como la *Deep Learnig*, cuya idoneidad viene avalada por proporcionar muy buenos resultados cuando se dispone de grandes tamaños muestrales (Big Data).

En este trabajo se propone un modelo estadístico de predicción del fracaso empresarial basado en la utilización de la metodología *Deep Learning* o Aprendizaje Profundo (AP). Con él se identificarán las variables más relevantes que detectan las situaciones de dificultad financiera, en nuestro caso escenarios concursales.

Además, nos parece una interesante aportación ya que no hay prácticamente trabajos que lo hayan abordado utilizando la técnica del Aprendizaje Profundo.

El estudio se estructurará revisando brevemente, en primer lugar, las aportaciones de los últimos años con relación al objeto de nuestro trabajo. Posteriormente, se describirán los datos de la muestra de trabajo y las variables empleadas. A continuación, se describirán los fundamentos de la metodología AP, el estudio estadístico y los resultados obtenidos. Finalmente, se muestran las conclusiones del trabajo, junto a posibles líneas futuras de investigación.

2. Antecedentes históricos.

Si bien existen revisiones completas de la literatura sobre fracaso empresarial que abordan esta línea de investigación desde sus comienzos (Tseng & Hu, 2010; Tascón & Castaño, 2012; Romero, 2013), haremos referencia solamente a las últimas metodologías incorporadas, comenzando por las redes neuronales (*neuronal networks*). Las redes neuronales, técnica que forma parte del "Machine Learning", tienen como base un sistema de neuronas dispuestas en tres niveles (entrada, oculta y salida) que realizan ciertos cálculos o tareas en función de la arquitectura de las conexiones que utilice. Sus primeras aplicaciones a este campo datan de la década de los noventa del siglo pasado (Bell, Ribar & Verchio, 1990; Odom & Sharda 1992; Wilson & Sharda, 1994; Boritz & Kennedy, 1995), prolongándose hasta la actualidad (Ahn, Cho & Kim, 2000; López & Pastor, 2015; Popescu, Andreica & Popescu, 2017; Jayanthi, Kaur & Suresh, 2017). Diversos estudios avalan que las metodologías apoyadas en redes neuronales presentan ventajas de robustez respecto a otras anteriores (Tam, 1991; Tam & Kiang, 1992; Wilson & Sharda, 1994; Zhang, Tomczak & Tomczak 1999), mostrando una interesante capacidad de adaptación a la nueva información (Tam, 1991; Tam & Kiang, 1992) lo que contribuye a obtener unos resultados significativamente mejores respecto a los obtenidos con el uso de técnicas estadísticas basadas en análisis discriminante, logit o probit (Fletcher & Gross, 1999; Ravi & Ravi, 2007; Zhang et al., 1999).

Otro conjunto de metodologías actuales de "*Machine Learnnig*" es el "*Boosting*" que usa principalmente árboles de decisión como clasificadores. Su utilidad y capacidad predictiva queda de manifiesto en diferentes trabajos tanto en el ámbito internacional (West, Dellana & Qian, 2005; Kim & Kang, 2010; Kim & Upneja, 2014; Wang, Ma & Yang, 2014; Kim, Kang & Kim, 2015; Zieba, Tomczak & Tomczak, 2016) como nacional (Díaz, Fernández & Segovia, 2004; Alfaro, Gámez & García, 2007a, 2007b, 2008, 2013; Alfaro, Gámez, García & Elizondo, 2008; Momparler et al., 2016; Pozuelo, Martínez & Carmona, 2018).

Un paso más evolucionado dentro del *Machine Learnnig* es la metodología denominada Aprendizaje Profundo (*Deep Learning*), que supone un importante avance sobre los modelos de redes neuronales al ampliar las capas a las que se ven sometidos los datos muestrales sobre los que se aplica el modelo. Esta técnica permite realizar operaciones complejas y obtener resultados fiables cuando se dispone de una gran cantidad de datos. Su uso está expandiéndose en los últimos años en diferentes ámbitos de investigación, habiéndose aplicado con éxito principalmente en áreas como la medicina, procesamiento de audios, imágenes, vídeos y datos, en diversidad de juegos, procesamiento del lenguaje, áreas de interés comercial, etc. (Deng & Yu, 2013; Lecun, Benglio & Hinton, 2015; Patel, Armstrong & Ganguli, 2016; Goodfellow, Benglio & Aaron, 2016; Abbot, Deshowitz, Murray & Larson, 2018). Sin embargo, se ha utilizado muy poco en el campo del análisis y predicción del fracaso empresarial. Apenas encontramos el reciente trabajo de Chaudhuri y Ghosh (2017), donde identifica las señales que evidencian la proximidad del fracaso de la empresa aplicando esta metodología a una muestra de empresas de construcción coreanas y empresas no financieras estadounidenses y europeas. Estos autores concluyen que el modelo utilizado de aprendizaje profundo tiene una capacidad predictiva superior frente a otros modelos previos considerados como tradicionales.

Esta relativa ausencia de trabajos previos sobre predicción del fracaso empresarial apoyados en esta novedosa metodología del *Deep Learning* añade un motivo adicional para la realización de este estudio, en el que se pretende constatar la utilidad de este algoritmo como herramienta en la predicción en este campo.

3. Desarrollo metodológico.

A continuación, se expone el desarrollo metodológico llevado a cabo en el este estudio, abordando la definición de fracaso empresarial, la selección de la muestra y las variables explicativas.

3.1. Definición de fracaso empresarial utilizada en el estudio.

Es común encontrar en la literatura sobre el fracaso empresarial diferentes definiciones del fracaso pudiendo distinguir tres grupos: fracaso económico, fracaso financiero y fracaso jurídico. Todas ellas tienen como denominador común que contemplan diferentes estados que influyen negativamente en la empresa y que pueden suponer su desaparición. En la mayoría de los estudios realizados se identifica el fracaso empresarial con una situación concursal o con las antiguas figuras jurídicas de quiebra o suspensión de pagos, a las que se podría añadir otras situaciones como la morosidad en la atención de obligaciones financieras, incumplimiento en los compromisos de pago o las reducciones de capital por pérdidas.

En nuestro trabajo se ha optado por una definición de fracaso empresarial que permita distinguir con claridad las empresas sanas de las que no lo son, equiparándose la situación de fracaso exclusivamente a la calificación de concurso. Con la aplicación de este criterio se ha pretendido homogeneizar al ámbito estudiado la diversidad de figuras jurídicas que contemplaban diversas situaciones de dificultad financiera por la que podrían pasar las empresas consideradas.

3.2. Selección y definición de las variables explicativas.

Uno de los aspectos más relevantes en la elaboración de modelos de predicción de fracaso empresarial es determinar las variables independientes que lo integrarán, en nuestro caso serán fundamentalmente ratios económico-financieras.

Ante la ausencia de una teoría general que guíe el proceso de selección de ratios, lo que constituye una fuerte limitación a la hora de modelizar el fracaso empresarial, en principio hemos procurado conciliar la experiencia aportada por otros autores (véase el estudio de Tascón & Castaño (2012) que

analiza detalladamente las variables empleadas en los principales trabajos sobre predicción del fracaso empresarial) con los objetivos propuestos.

Las ratios consideradas se han extraído de las definidas en la base de datos utilizada. Sobre este conjunto, dado que nuestro objetivo es la formulación de modelos de predicción de fracaso empresarial hemos seleccionado aquellas que informan sobre aspectos de la solvencia, rentabilidad y endeudamiento. A continuación, hemos incluido dos variables más, el país de origen de la empresa y el tamaño, representado por la cifra de activo de cada empresa.

Por último, se ha realizado una depuración eliminando aquellas que presentaban un elevado porcentaje de valores perdidos.

La lista y descripción de las variables consideradas inicialmente y separadas por categorías la mostramos en la Tabla 1.

Tabla 1. Variables explicativas.

CLAVE	VARIABLES
V1	INGRESOS EXPLOTACIÓN
V2	RESULTADO DEL PERIODO / FONDOS DE LOS ACCIONISTAS
V3	RESULTADO DEL PERIODO / ACTIVO
V4	INGRESOS NETOS /FONDOS DE LOS ACCIONISTAS
V5	INGRESOS NETOS / ACTIVO
V6	BENEFICIO MARGINAL BENEFICIO ANTES DE IMPUESTOS / INGRESOS OPERACIONALES
V7	EBIT *Earnings Before Interest and Taxes* (beneficio antes de intereses e impuestos)
V8	ROTACIÓN DE ACTIVOS NETOS INGRESOS OPERACIONALES / FONDOS DE LOS ACCIONISTAS + PASIVOS NO CORRIENTES
V9	PERIODO DE COBRO DEUDORES / INGRESOS OPERACIONALES
V10	PERIODO DE CRÉDITO ACREEDORES / INGRESOS OPERACIONALES
V11	RATIO ACTUAL (ACTIVO CORRIENTE / PASIVO CORRIENTE)
V12	ACTIVO CORRIENTE–ACTIVO CORRIENTE FRO / PASIVO CORRIENTE
V13	RATIO DE SOLVENCIA (FONDOS DE LOS ACCIONISTAS / ACTIVO TOTAL)
V14	VOLUMEN DE ACTIVO
V15	PAÍS (*)

(*): Bélgica, España, Francia, Luxemburgo, Italia, Holanda, Finlandia.

Fuente: Elaboración propia.

Todas las partidas integrantes de las ratios han sido derivadas del balance de situación y de la cuenta de pérdidas y ganancias de las empresas que componen la muestra.

3.3. Selección de la muestra de empresas.

En el proceso de selección y obtención de la muestra de empresas se ha recurrido a la base de datos financieros ORBIS que dispone de información financiera sobre empresas del continente europeo en formatos estándar, lo que ha permitido realizar comparaciones entre empresas de distintos países.

Se ha trabajado con empresas de Italia, Bélgica, Luxemburgo, Holanda, Finlandia, Francia y España. La elección de estos países se ha basado en la homogeneización de las diversas figuras jurídicas que contemplaban situaciones de dificultad financiera.

Para seleccionar las empresas fracasadas de la muestra de estimación se han considerado aquellas firmas que presentaban situación de concurso en el año 2016 pertenecientes a prácticamente todos los sectores productivos.

Sobre la población inicial se realizaron dos filtrados para su depuración. Uno por el que se descartaron aquellas empresas de reciente creación (hasta tres años) y otro, que permitió eliminar las que no contenían datos contables completos de los tres ejercicios anteriores a la fecha del fracaso y las que presentaban valores extremos entre sus ratios que pudieran modificar sustancialmente los valores medios para los mismos.

Particularmente, se han considerado valores extremos o outliers aquellas observaciones que están por debajo de (Q1 – 1,5 × RIQ) o por encima de (Q3 + 1,5 × RIQ). Donde, Q1 es el primer cuartil, Q3 es el tercer cuartil y RIQ es el rango inter-cuartil (diferencia entre el tercer y el primer cuartil).

También, por sus singularidades, se ha prescindido de las empresas de seguros y financieras. Tras estos procesos de selección y filtrado, el número de empresas se redujo a 12.126, que son las que definitivamente se integrarán como parte de la muestra de estimación.

Para completar esta muestra y poder aplicar las herramientas estadísticas de clasificación que permitan apreciar sus diferencias se completó la muestra con empresas sanas. Para ello se exigió que la firma estuviera activa (condición coincidente con la categoría "empresa activa" de la base de datos ORBIS) y de manera similar a las empresas fracasadas, que tuviese información contable completa de al menos tres ejercicios anteriores a 2016. Tras este proceso se añadieron 49.498 empresas sanas a la muestra, quedando la muestra de estimación formada por 61.624 empresas de las que 49.498 son sanas y el resto fracasadas.

Como último paso, para contrastar la capacidad predictiva del modelo y su grado de generalización, se ha reservado un 20% de las observaciones para comprobar los resultados obtenidos con la muestra de entrenamiento o de estimación. Todos los indicadores de rendimiento y de precisión del modelo ajustado de aprendizaje profundo incluidos en este estudio se han obtenido con la muestra independiente de comprobación.

4. Metodología.

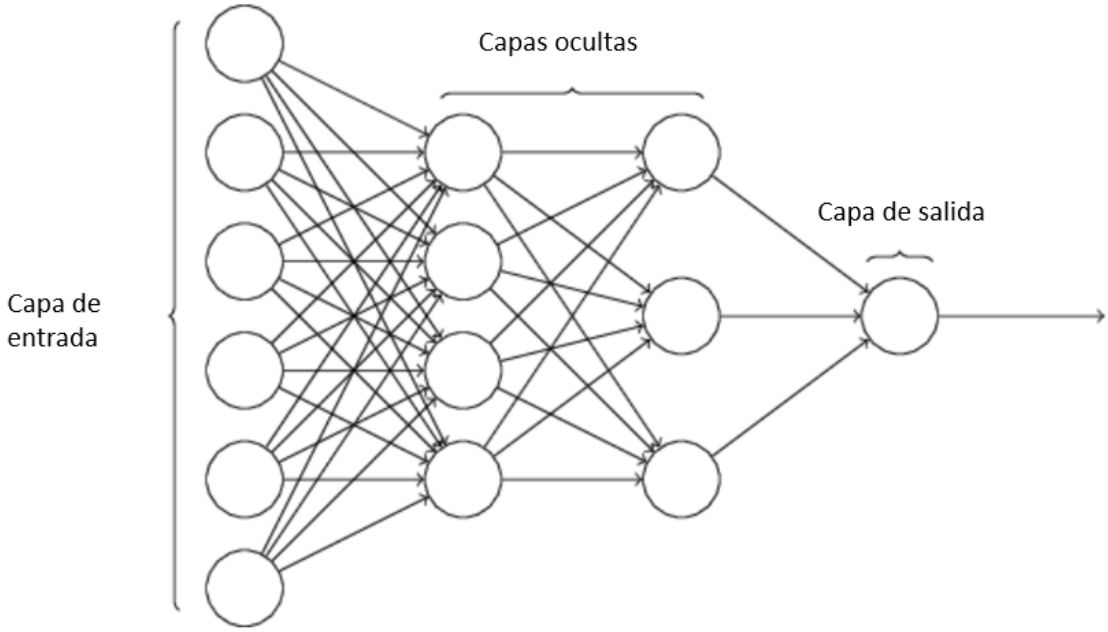
En este trabajo vamos a utilizar el Aprendizaje Profundo (AP) (*Deep Learning*, en inglés), una metodología relativamente reciente que hace referencia a un tipo particular de algoritmos dentro del Aprendizaje Automático (*Machine Learning*, en inglés). Esta técnica se caracteriza porque proporciona muy buenos resultados en términos de capacidad predictiva cuando se dispone de tamaños muestrales ciertamente grandes. El AP está formado por un conjunto de metodologías basadas en redes neuronales artificiales, que en cierto modo están inspiradas en la estructura de las neuronas del cerebro humano. La palabra *profundo* se refiere a la presencia de muchas capas en la red neuronal artificial. Hace menos de 5 años, 10 capas ya eran suficientes para considerar una red como profunda, sin embargo, hoy en día los modelos pueden llegar a tener cientos de capas (Gulli & Pal, 2017).

Estos últimos años el AP se ha aplicado con gran éxito en diferentes dominios, en particular para el reconocimiento de imágenes, texto, vídeo o audio. Los resultados han sido muy sorprendentes, en el sentido de que esta técnica ha mejorado considerablemente los niveles de precisión alcanzados con técnicas más tradicionales y que eran consideradas hasta hace muy poco como referentes del estado del arte de estos problemas de reconocimiento. Como señalan Gulli y Pal (2017), en gran medida, el éxito del AP reside en la cada vez mayor disponibilidad de datos para entrenar a los algoritmos (como ImageNet para imágenes) y la posibilidad de disponer de ordenadores más potentes con tarjetas GPU que permiten realizar cálculos numéricos de modo muy eficiente. Google, Microsoft, Amazon, Apple,

Facebook y muchos otros utilizan esas técnicas de aprendizaje profundo todos los días para analizar cantidades masivas de datos.

Una red de AP es una red neuronal con múltiples capas y se puede entender como una red neuronal potenciada. Las redes neuronales tuvieron cierta popularidad entre el mundo académico hasta los años 80; sin embargo, otros algoritmos se evidenciaron como más precisos y, por tanto, se abandonó su uso en favor de estos últimos. Las limitaciones que había entonces, en cuanto a los costes que suponía incrementar la capacidad numérica de los ordenadores y disponer de grandes cantidades de datos, impedía que pudieran construirse redes neuronales con más de una capa. De hecho, en sus inicios las redes neuronales esencialmente consistían en variaciones de metodologías basadas en la regresión logística (Schmidhuber, 2015). Sin embargo, a principios de la década del 2010 la presencia de ordenadores más potentes preparados para el cálculo masivo y la mejora del algoritmo *backpropagation* (Wang, Zeng & Chen, 2015), hizo que despertase de nuevo el interés hacia esta metodología. Estos cambios, darían lugar al aprendizaje profundo moderno, una clase de redes neuronales caracterizadas por un número significativo de capas de neuronas (arquitecturas neuronales con múltiples capas ocultas), que son capaces de aprender relaciones bastante complejas mediante niveles de abstracción progresivos (Gulli & Pal, 2017). En definitiva, el AP es la implementación de redes neuronales con más de una capa de neuronas ocultas.

Figura 1. Ejemplo de arquitectura del perceptrón multicapa.



Fuente: Elaboración propia.

En este trabajo emplearemos un esquema del AP basado en el perceptrón multicapas (Figura 1), pues es una de las clases más extendidas de este tipo de redes neuronales. También se conocen como redes de alimentación hacia delante (*feedforward neuronal networks*, en inglés) y se consideran la quintaesencia de los modelos de AP, tal y como indica Ledell (2018). Estos modelos se denominan redes de alimentación hacia adelante porque la información fluye desde las neuronas de la capa de entrada hacia las neuronas de la capa de salida, pasando por las capas ocultas, y siempre en este sentido. No hay ningún tipo de conexión que se retroalimente de información que venga de capas posteriores. Cuando las redes de alimentación hacia adelante incluyen retroalimentación que proviene de capas posteriores se denominan redes neuronales recurrentes (*recurrent neuronal networks*).

Como se aprecia en la Figura 1, las redes son densas (*densely connected*, en inglés) en el sentido de que cada neurona de una capa está conectada con todas las neuronas situadas en la capa anterior y con todas las neuronas de la capa posterior. La primera capa es la capa de entrada, cada nodo se alimenta de una entrada y transmite su salida como entrada de cada nodo o neurona de la siguiente capa. No existen conexiones entre las neuronas de una misma capa y la última capa produce el resultado de la red neuronal. La parte central está formada por las capas ocultas, se caracterizan porque las neuronas no tienen ninguna conexión con el exterior (ya sea con la capa de entrada o con la capa de salida) y se activan sólo por medio de neuronas de la capa anterior. Estas redes neuronales si tienen más de una capa oculta ya se consideran como profundas; en cambio, si sólo tienen una capa oculta se denominan superficiales.

La capa de entrada está formada por un número de neuronas equivalente al número de variables que tienen los datos. El número de capas ocultas y neuronas en cada una de ellas lo especifica el diseñador de la red neuronal, de modo que tendrá que escoger una combinación adecuada de acuerdo con las características de los datos. En un problema de clasificación binario, como el que nos ocupa (empresas sanas frente a empresas fracasadas), la capa de salida está formada por una sola neurona; y el algoritmo proporciona la probabilidad de pertenencia a cada uno de los dos grupos.

La importancia del uso de los algoritmos del AP radica en el hecho de que, en comparación con los algoritmos tradicionales, para un tamaño de datos elevado los resultados del AP son mucho mejores. Y esta mayor capacidad predictiva se acentúa más a medida que aumenta el tamaño de las bases de datos. Por ejemplo, y como recoge Crawford (2016), Facebook ha demostrado la alta capacidad predictiva de las redes neuronales que utilizan el AP en la identificación de una cara en una fotografía. Cuando se pregunta si dos fotos de caras desconocidas muestran a la misma persona, un ser humano acertará el 97,53% de las veces. El nuevo software desarrollado por investigadores de Facebook puede obtener un 97,50% de aciertos, independientemente de las variaciones en la iluminación o si la persona en la imagen está mirando directamente a la cámara.

En las redes del AP cada capa de nodos o neuronas recibe información de un conjunto distinto de características basadas en las salidas de la capa precedente. Cuanto más se avanza hacia las capas más profunda, más complejas son las particularidades que los nodos pueden llegar a reconocer, ya que agregan y recombinan características de la capa anterior. Esto se conoce como jerarquía de las particularidades y se trata de una jerarquía con una complejidad y abstracción crecientes. Todo ello permite que las redes del AP sean capaces de manejar conjuntos de datos muy grandes y de gran dimensión con miles de millones de parámetros que se procesan mediante funciones no lineales.

Las características principales de un modelo de redes AP se pueden esquematizar en los siguientes pasos (adaptado de https://www.kdnuggets.com/2016/12/artificial-neural-networks-intro-part-1.html, visitado en septiembre de 2019):

- Paso 1. Cuando se alimenta el nodo de entrada con una serie de variables, se activa un conjunto único de neuronas en la primera capa, comenzando una reacción en cadena que se dirige hacia el nodo de salida.
- Paso 2. Las neuronas activadas de la primera capa envían señales a cada neurona conectada en la siguiente capa. Esto determinará qué neuronas se activarán en la siguiente capa.
- Paso 3. En la siguiente capa, se activa otro conjunto de neuronas de acuerdo con las señales recibidas de las neuronas anteriores.
- Paso 4. Los pasos 2-3 se repiten para todas las capas restantes (siempre que el modelo tenga más de 2 capas), hasta llegar al nodo último o de salida.
- Paso 5. En un problema de clasificación binario: el nodo de salida clasifica la salida como 0 o 1, teniendo en cuenta las señales recibidas directamente de las neuronas de la capa precedente. Cada combinación de neuronas activadas en la capa precedente conduce a una solución, aunque cada solución puede representarse mediante diferentes combinaciones de neuronas activadas.

Para que las redes neuronales sean capaces de producir modelos con una alta capacidad predictiva es necesario identificar los valores óptimos de una serie de hiperparámetros que determinan su configuración. En las siguientes líneas vamos a hacer referencia a los hiperparámetros más importantes de las redes de AP basadas en el perceptrón multicapas o redes de alimentación hacia delante.

La ratio de aprendizaje del algoritmo del AP. Cuanto más alta sea la ratio el algoritmo convergerá antes, pero los resultados serán peores. En cambio, un valor demasiado bajo puede ocasionar que el tiempo necesario para que el algoritmo encuentre una solución óptima sea excesivo. Como señala Cook (2017), un valor adecuado puede ser 0,005.

Uno de los factores más cruciales en las redes de AP es la función de activación de los nodos o neuronas, pues permiten la construcción de modelos a partir de datos que no tienen una relación lineal. Dependiendo de la función de activación con la que se construya la red neuronal los resultados serán diferentes. En realidad, lo que hacen estas funciones es una transformación de los datos. De ahí la importancia de identificar la función de activación más apropiada. Entre las más populares se encuentran las siguientes: Maxout, Rectifier, Tanh y Sigmoid. La función no lineal Sigmoid transforma un número real en otro dentro de un rango comprendido entre 0 y 1. La función no lineal Tahn transforma un número real en otro, pero dentro del intervalo [-1, 1]. En la función Rectifier, si el número real es negativo se convierte en 0 y si es positivo se mantiene. La función Maxout se asemeja a la Rectifier: cualquier número positivo no se transforma y se mantiene, pero los números negativos se convierten en otros también negativos pero muy próximos al cero.

El diseño de una red neuronal de AP basado en el perceptrón multicapas también requiere especificar el número de capas ocultas y el número de neuronas en cada capa. Dependiendo de las características de los datos los valores óptimos de estos dos elementos serán diferentes. Un número demasiado bajo daría lugar a modelos con una capacidad predictiva muy baja y, en cambio, un número demasiado alto produciría una red neuronal que no resultaría generalizable (Khashman, 2010).

La tasa de exclusión (*dropout*, en inglés) de las neuronas se refiere al porcentaje de neuronas de la capa de entrada o de las capas ocultas que se excluyen durante el proceso de entrenamiento del algoritmo. Es muy importante identificar los porcentajes óptimos con el fin de no ajustar una red que proporcione resultados insuficientes en un conjunto de datos independientes o nuevas observaciones. Las neuronas que se excluyen se desactivan y no envían ninguna señal, de modo que no intervienen en la activación de las neuronas de la siguiente capa.

También con el fin de evitar que una red neuronal no resulte generalizable hay que especificar dos parámetros de regularización: L1 y L2. El primero también se conoce como regularización *lasso* y el segundo como regularización *ridge*. Sus valores se encuentran dentro del intervalo [0, 1]. En general, suelen tomar valores muy bajos y próximos a 0. La regularización penaliza el peso o importancia de los predictores durante el entrenamiento de una red neuronal.

Otro hiperparámetro que hay que determinar es el número óptimo de *epochs*. Un *epoch* está formado por el conjunto de observaciones que constituyen la muestra de entrenamiento. Se puede entender también como una pasada completa del algoritmo por todos los datos, con lo que se asemejaría con el número de pasadas que efectúa el algoritmo.

Es sumamente importante especificar correctamente todos estos hiperparámetros a los que acabamos de referirnos, de lo contrario la red neuronal podría memorizar los datos con los que se ha entrenado y proporcionar unos resultados muy pobres en una muestra de comprobación, lo cual sería indicativo de un problema de sobre-ajuste (*over-fitting*, en inglés). Por tanto, la red neuronal obtenida no resultaría generalizable y carecería de interés.

Cada neurona tiene un peso asignado (w_1, w_2, w_3… w_n) a cada una de sus entradas (Figura 2), que se van modificando hasta alcanzar los valores óptimos durante el proceso de aprendizaje de la red neuronal. También hay una entrada de sesgo (o w_0) en cada neurona, que puede considerarse como un

valor constante que también hay que determinar. Los impulsos o información de cada neurona que pasan a la siguiente capa se obtienen mediante las funciones de activación (Ledell, 2018). La entrada que le llega a cada neurona se puede representar matemáticamente como una relación lineal:

$$y = w_0 + \sum_{i=1}^{n} w_i x_i$$

Los sesgos y los pesos se especifican con w_0 y w_i, y las variables con x_i. A continuación, sobre el valor de *y* obtenido con la agregación se aplica la función de activación *g(y)*, dando lugar al *output* que pasa a las neuronas de la siguiente capa:

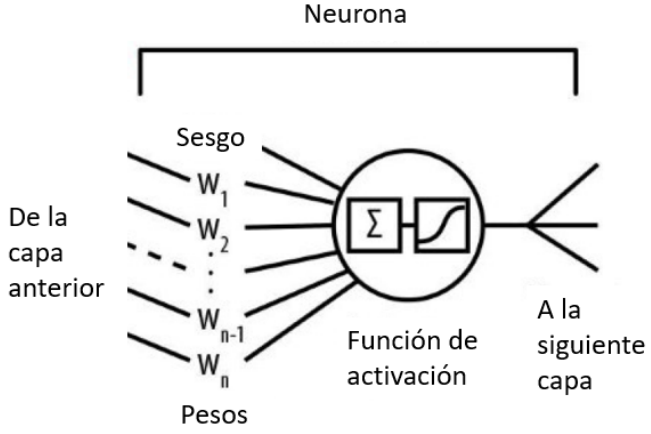
$$output = g(w_0 + \sum_{i=1}^{n} w_i x_i)$$

Si por ejemplo la función de activación que se está empleando fuera la Rectifier, el valor *y* (resultado de la agregación) se transformaría como sigue con *g(y)*, dando lugar al correspondiente output:

$$g(y)\begin{cases} 0, & y < 0 \\ y, & y \geq 0 \end{cases}$$

Como se aprecia en la relación lineal de la formulación de *y*, el sesgo (w_o) es similar a la pendiente de una regresión lineal. Su utilización en las redes neuronales permite obtener predicciones más precisas.

Figura 2. Neurona de una capa oculta.



Fuente: Basado en Ledell (2018).

Una medida muy utilizada (función de coste) para evaluar la precisión del modelo es la suma de los errores al cuadrado. Un error cuadrado denota la proximidad de una predicción con su valor real. El modelo neuronal de AP intenta minimizar estos errores cambiando las reglas que producen la activación de las neuronas (los pesos y sesgo de cada neurona), a través del algoritmo del descenso del gradiente (*gradient descent*). De este modo, los pesos se van ajustando gradualmente mediante la minimización de una función de coste y los valores óptimos de todos los pesos se corresponderán con aquéllos que minimicen la función de coste. La actualización simultánea de todos los pesos de la red neuronal es posible gracias al algoritmo de *backpropagation*. Como explica Ledell (2018), inicialmente los pesos

se asignan aleatoriamente, seguidamente se alimenta el algoritmo con la primera observación de la muestra de entrenamiento y la respuesta correcta, se calcula el error y se ajusta cada uno de los pesos para obtener así un error menor. A continuación, se toma la segunda observación y se repite el proceso hasta completar todas las observaciones de la muestra de entrenamiento, lo que conlleva un *epoch*. De nuevo, se repite esta mecánica con el número de *epochs* especificados; y así el algoritmo encuentra por medio de la iteración la mejor combinación de pesos y sesgo que minimiza la función de coste. De este modo, es posible construir modelos de predicción capaces de detectar la existencia de relaciones no lineales entre las diferentes variables que componen los datos.

Todos estos pesos y sesgos constituyen los parámetros del modelo de la red neuronal; se trata de variables de configuración propias del modelo y cuyos valores deben estimarse a partir de los datos durante el proceso de aprendizaje de los algoritmos. En cambio, la configuración de los hiperparámetros es externa al modelo y sus valores óptimos no es posible determinarla a partir de los datos durante el proceso de entrenamiento del algoritmo; corresponde al investigador su especificación, quien generalmente probará diferentes combinaciones de los mismos hasta encontrar la que produzca mejores resultados (Brownlee, 2017).

El número de parámetros (suma de pesos y sesgos) que tiene que calcular el algoritmo de aprendizaje profundo crece de forma exponencial a medida que se incrementa el número de capas y las neuronas de cada una de ellas. Por ejemplo, para una red neuronal, que resuelva un problema de clasificación binario, compuesta por 15 variables y dos capas de 5 neuronas, el número de parámetros es de:

$$[(15 + 1) \times 5] + [(5 + 1) \times 5] + [(5 + 1) \times 1] = 116 \text{ parámetros}$$

Si aumentamos el diseño de la red a 3 capas de 10 neuronas cada una quedaría así:

$$[(15 + 1) \times 10] + [(10 + 1) \times 10] + [(10 + 1) \times 10] + [(10 + 1) \times 1] = 391 \text{ parámetros}$$

En resumen, en este trabajo vamos a intentar construir un modelo de red neuronal de aprendizaje profundo para diferenciar las empresas sanas de las que se encuentran en una situación concursal. De este modo, el algoritmo debe ser capaz de determinar los valores de los pesos y sesgos que interconectan a las diferentes neuronas de las distintas capas, de acuerdo con el comportamiento especificado en líneas anteriores, para obtener así un modelo de aprendizaje profundo con una capacidad predictiva lo más alta posible en un conjunto de datos diferentes (muestra de comprobación) al proporcionado o utilizado por el algoritmo (muestra de entrenamiento y de validación). Para que los resultados sean lo más satisfactorio posibles es muy importante identificar, para las características intrínsecas de la muestra de datos sobre el fracaso empresarial, la mejor combinación posible de los denominados hiperparámetros

Por último, señalamos que todos los modelos se han construido utilizando la aplicación informática *R* (R Core Team, 2019) versión 3.6.1, así como el paquete *h2o* versión 3.28.0.2 (The H2O.ai team, 2019).

5. Resultados.

En primer lugar, mostramos los principales estadísticos descriptivos de todas las variables utilizadas como predictores un año antes del fracaso empresarial (Tabla 2). Se ha prescindido de la variable país por tratarse de una variable tipo factor. Vamos a considerar la mediana como el estadístico de referencia para comparar los dos tipos de empresas ya que su cuantía no resulta influida por los valores extremos y, por tanto, es una de las medidas de centralidad de la distribución más aceptadas. Para resolver el problema de los valores extremos y con el fin de mejorar los resultados del algoritmo de clasificación empleado, todos los valores que se encuentran fuera del rango inter-cuartil (determinado por la diferencia entre el tercer y primer cuartil) se han sustituido por los valores del primer y tercer cuartil.

En general, no se aprecian grandes diferencias entre los dos grupos de empresas para el conjunto de variables consideradas. Una de las diferencias más significativas la encontramos en la variable V1, indicativa de los ingresos de explotación, que obtienen valores muy superiores en empresas fracasadas.

También encontramos diferencias en la ratio V8 que mide el periodo de cobro, que es considerablemente mayor para las empresas fracasadas que para las sanas (4,03 vs 0,41) y, finalmente destacamos la diferencia en la ratio de solvencia V13, que es mayor en las empresas no fracasadas (42,72) que en las fracasadas (10,70). Como se apreciará posteriormente, el modelo resultante considerará estas variables entre las más relevantes para estimar el fracaso de la empresa.

Como resultado de este primer análisis se observa que, en general, las empresas de mayor tamaño, las que tienen problemas de solvencia y las que presentan mayores plazos de cobro son las que tienen una mayor propensión hacia el fracaso.

Tabla 2. Resumen de los estadísticos descriptivos de las variables independientes (un año antes del fracaso).

	Media		d. e.		Mediana		mínimo		máximo	
Variable	NF	F	NF	F	NF	F	NF	F	NF	F
V1	6,84	100,17	4,94	183,05	7,00	10,00	-1,45	0,00	14,00	647,00
V2	3,43	-14,69	50,14	89,27	1,55	0,98	-123,14	-263,82	103,56	122,43
V3	2,91	-4,01	18,64	18,85	0,44	-0,18	-35,86	-47,95	47,35	31,08
V4	-0,12	-18,34	45,27	77,86	0,77	0,31	-118,61	-236,62	85,43	93,07
V5	1,59	-4,28	16,21	17,77	0,29	0,03	-34,64	-47,95	38,62	26,84
V6	7,10	-3,84	41,06	18,89	4,48	-0,09	-73,42	-49,99	83,70	35,58
V7	10,98	-0,79	41,09	19,25	6,90	0,76	-68,73	-43,51	87,22	45,46
V8	1,41	8,46	2,35	11,16	0,41	4,03	0,01	0,15	9,14	43,08
V9	148,47	82,39	186,97	101,77	66,00	46,00	0,00	0,00	630,00	373,75
V10	62,59	85,07	115,42	99,21	3,00	48,00	0,00	0,00	421,00	376,00
V11	6,08	1,20	10,02	0,77	1,82	1,05	0,07	0,22	38,92	3,37
V12	5,01	0,89	8,49	0,67	1,47	0,76	0,04	0,10	33,21	2,73
V13	43,98	11,06	37,65	29,75	42,72	10,70	-24,01	-54,68	99,31	65,91
V14	822,14	907,93	350,91	1.383,47	265,00	329,00	5,00	29,00	4.365,00	5.417,50

Notas: NF: Empresa no fracasada (49.498 observaciones)
F: Empresa fracasada (12.126 observaciones)
d.e.: desviación estándar
La descripción de las variables se encuentra en la Tabla 1.
V14 o ACTIVO se ha tomado en miles de euros.
Fuente: Elaboración propia.

5.1. Selección de los hiperparámetros óptimos del modelo.

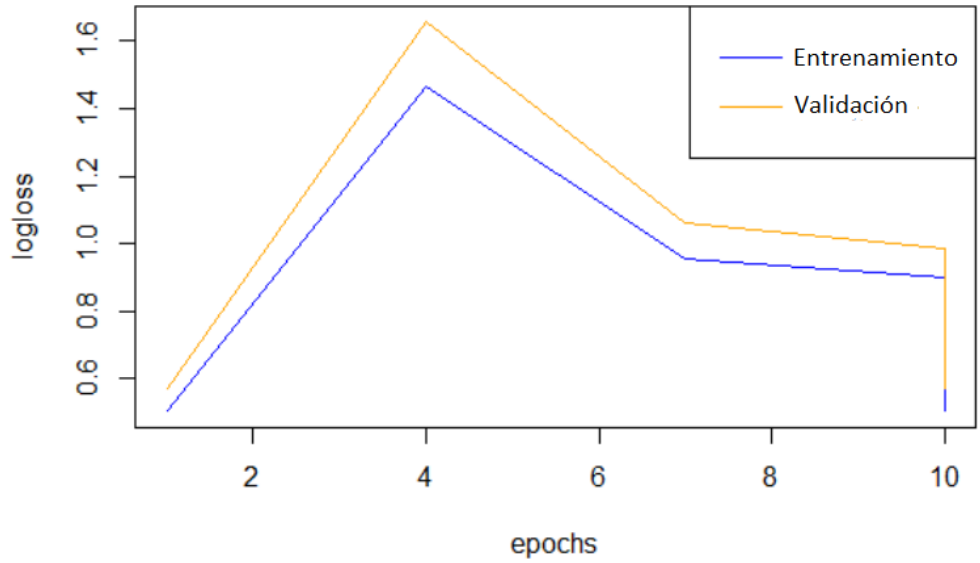
Anteriormente ya se ha indicado que las redes neuronales requieren de la especificación de los valores óptimos de una serie de hiperparámetros que determinan su configuración, para que de este modo sean capaces de producir modelos con una alta capacidad predictiva. Dado que el tamaño de los datos disponible es muy elevado (61.624 observaciones) se ha dividido en tres bloques. El primero, formado por el 60% constituye la muestra de entrenamiento con la que se alimenta el algoritmo para determinar los valores de los pesos y sesgos de todas las neuronas de la red. El segundo, con un 20% de todos los datos, constituye la muestra de validación; se ha utilizado para identificar los valores óptimos de los hiperparámetros. Y el tercero, está formado por la muestra de comprobación; un conjunto de observaciones que ha permanecido oculta durante todo el proceso de entrenamiento y validación del

modelo, y que tiene como finalidad valorar la capacidad predictiva del modelo neuronal ajustado frente a nuevas observaciones, es decir, determinar la generalización del modelo.

El entrenamiento de una red neuronal requiere identificar la combinación de hiperparámetros que conduzca a la obtención de un modelo con una capacidad predictiva muy elevada. En este estudio, como es habitual en los trabajos de aprendizaje automático, emplearemos el estadístico *AUC* (*area under the curve*) que mide el área debajo de la curva *ROC* (*Receiver Operating Characteristics*). *AUC* toma valores entre 0 y 1, pero es por encima de 0,5 cuando indica que un modelo de clasificación binario es mejor que una mera asignación aleatoria. Un valor de 1 significaría que el modelo es perfecto y clasificaría el 100% de los casos correctamente. Entendemos que este indicador de la capacidad predictiva de un modelo es superior al de la precisión global del modelo, tal y como defienden Cortes y Mohri (2004), especialmente cuando la muestra no es balanceada. Es uno de los indicadores más utilizados para valorar el rendimiento de un modelo y compararlo con otros cuando se hace uso de las técnicas del aprendizaje automático. El *AUC* se considera como una medida pura de la eficacia o capacidad predictiva de un modelo, pues es independiente del punto de corte que se utilice para clasificar las predicciones como positivas o negativas. La medida de la precisión global del modelo o porcentaje de aciertos varía según el punto de corte que se emplee para clasificar las probabilidades de ocurrencia que predice el modelo para el evento de interés o el evento sin interés.

Como función de pérdida para la optimización del algoritmo se ha seguido la función *Log-Loss* (*logistic loss*), que penaliza las desviaciones que hay entre las predicciones y los valores reales. Es una medida de precisión que proporciona, además de la pertenencia de las observaciones a una categoría u otra (fracaso frente a no fracaso, en nuestro estudio), la probabilidad de la pertenencia (por ejemplo, la probabilidad de fracaso de una empresa es del 80%). La minimización de la función *Log-Loss* para evaluar la capacidad predictiva de un algoritmo es equivalente a la maximización de la exactitud del algoritmo.

Figura 3. Valores de la función de pérdida Log-Loss para el Modelo 1 de Aprendizaje Profundo con 10 *epochs*.

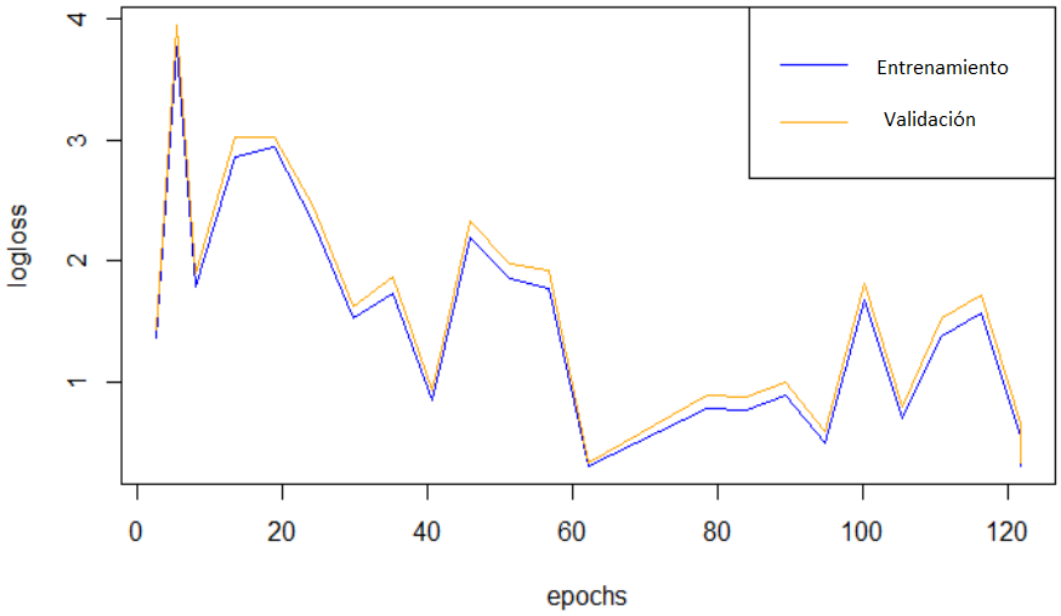


Notas: Entrenamiento = muestra de entrenamiento (60% de las observaciones)
Validación = muestra de validación (20% de las observaciones)
Log-Loss para la muestra de validación: 0,670
AUC para la muestra de validación: 0,8585
Fuente: Elaboración propia.

En primer lugar, vamos a ajustar un modelo neuronal de Aprendizaje Profundo (AP) utilizando la función *h2o.deeplearning()* del paquete de h2o R (The H2O.ai team, 2017) con los parámetros que contiene por defecto (un detalle de los mismos se encuentra en la página web http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/deep-learning.html). La Figura 3 muestra los valores que toma la función de pérdida Log-Loss a medida que se incrementan los *epochs* (número de veces que la totalidad de los datos pasan por el algoritmo) cuando se utiliza *h2o.deeplearning()* sin modificar ninguno de los parámetros que tiene por defecto (Modelo 1). Se observa que el valor óptimo de los *epochs* es de 1, por ello la representación de la pérdida Log-Loss desciende en el último de los *epochs* hasta su valor mínimo de 0,671. El número de *epochs* que utiliza esta función por defecto es de 10. El valor que se obtiene del estadístico *AUC* en la muestra de validación es de 0,8585.

La función *h2o.deeplearning()* está configurada para que después del pase de los *epochs* especificados devuelva el número de *epochs* con un valor mínimo en la función de pérdida. Es por ello que en las figuras que ponen en relación el número de *epochs* con la función de pérdida, se aprecia que al final los valores de pérdida descienden hasta el valor mínimo alcanzado en el mejor de los *epochs*.

Figura 4. Valores de la función de pérdida *Log-Loss* para el Modelo 2 de Aprendizaje Profundo con 200 *epochs.*



Notas: *Log-Loss* para la muestra de validación: 0,670
AUC para la muestra de validación: 0,8585
Fuente: Elaboración propia.

A continuación, de nuevo partiremos de la función *h2o.deeplearning()* con los parámetros que tiene por defecto, pero modificaremos e incrementaremos significativamente el número de *epochs* de 10 a 200, y obtendremos el Modelo 2. Una manera sencilla de mejorar la capacidad predictiva de una red neuronal de aprendizaje profundo es incrementar el número de *epochs*. Tal y como se observa en la Figura 4, no se han llegado a completar los 200 *epochs* porque si no se modifican las especificaciones de la función *h2o.deeplearning()*, el número de *epochs* se interrumpe si después de 5 evaluaciones consecutivas de la función de pérdida el algoritmo observa que no se ha reducido para nada la pérdida. Con el incremento del número de *epochs* la pérdida *Log-Loss* en la muestra de validación se reduce y se sitúa en 0,3388, casi la mitad de la pérdida que se obtuvo con tan solo 10 *epochs*. Asimismo, el estadístico *AUC* mejora y alcanza un valor de 0,9438 en la muestra de validación.

Tabla 3. Detalle de los valores de los hiperparámetros considerados para ajustar el mejor modelo de red neuronal de aprendizaje profundo.

HIPERPARÁMETRO	VALORES CONSIDERADOS
Función de activación de las capas ocultas	Maxout, Rectifier y Tahn
Función de activación de la capa de salida	Sigmoid
Número de capas ocultas	3 y 4
Número de neuronas en cada capa	5, 10, 50, 100 y 200
Regularización L1	0, 0,00001 y 0,0001
Regularización L2	0, 0,00001 y 0,0001
Ratio de exclusión de la capa de entrada	0, 10% y 20%
Ratio de exclusión de las capas ocultas	0, 10%, 20% y 50%
Ratio de aprendizaje	0,005
Número de *epochs*	200

Notas: Las combinaciones de todos los valores de los hiperparámetros asciende a 3.240 modelos diferentes.
Fuente: Elaboración propia.

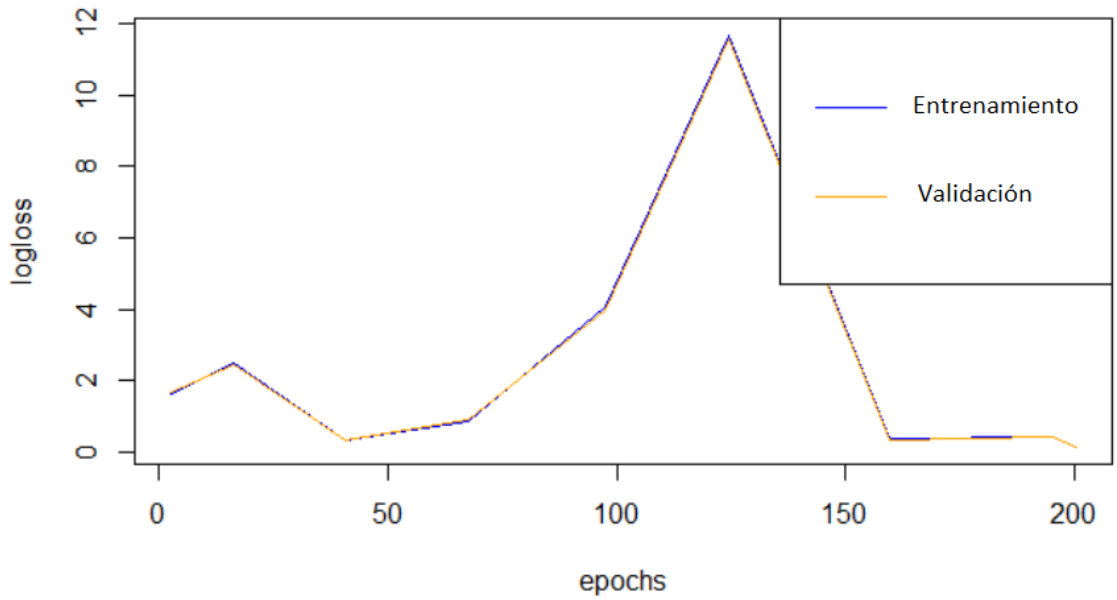
A continuación, para mejorar la red neuronal vamos a intentar identificar una combinación de hiperparámetros que produzca un modelo que sea mejor que el obtenido hasta ahora simplemente incrementando el número de *epochs*. En la Tabla 3 recogemos todos los valores de los hiperparámetros que se han considerado con el fin de identificar la mejor combinación de los mismos. Uno de los problemas de las redes neuronales es la gran cantidad de hiperparámetros que hay que intentar ajustar a sus valores óptimos. No existe ninguna metodología que oriente al investigador sobre cuáles son estos valores óptimos, aunque es posible obtener unos valores razonables de los mismos (esto es, que permitan la construcción de un modelo con unos indicadores de precisión elevados) probando diferentes combinaciones de los mismos. El conjunto de hiperparámetros considerados se ha obtenido a partir de las recomendaciones de Ledell (2018). El número de modelos de aprendizaje profundo que se habrían de ajustar para contemplar la totalidad de las combinaciones posibles de los hiperparámetros contemplados ascendería a 3.240, lo que requeriría de una cantidad de tiempo de trabajo para un ordenador de varios días. Por tanto, en la estrategia de búsqueda de la mejor combinación se ha limitado el tiempo de la CPU a 240 minutos y el número máximo de modelos a construir a 400, combinado la selección de los hiperparámetros de modo aleatorio. Esto significa, que no obtendremos el mejor de todos los posibles modelos para los parámetros considerados; lo que se obtendrá, sin embargo, será uno de los mejores modelos después de todo este proceso de iteración y búsqueda entre todas las combinaciones posibles. De acuerdo con la estrategia de búsqueda señalada, el mejor modelo de aprendizaje profundo que se ha identificado (Modelo 3) reúne las siguientes características:

- función de activación de las capas ocultas: Rectifier;
- número de capas ocultas: 3;
- número de neuronas en cada capa: 50;
- regularización L1: 0,00001;
- regularización L2: 0,0;
- ratio de exclusión de la capa de entrada: 0%;
- ratio de exclusión de las capas ocultas: 10%.

Entre todas las funciones de activación, Rectifier resulta la más apropiada. De este modo, en las capas ocultas de la red neuronal los valores negativos se convierten en 0 y los positivos se mantienen.

La Figura 5 muestra las iteraciones del algoritmo con el número de *epochs* y la función de pérdida. En el mejor modelo identificado (Modelo 3) la función de pérdida *Log-Loss*, para la muestra de validación, asciende a 0,1248 y el estadístico *AUC* alcanza un valor de 0,9834, ambos valores son bastante mejores que los obtenidos con los modelos anteriores. El estadístico *AUC* está muy próximo al valor más óptimo de 1.

Figura 5. Valores de la función de pérdida *Log-Loss* para el Modelo 3 o Modelo Final de Aprendizaje Profundo con la mejor combinación de hiperparámetros.



Notas: función de activación de las capas ocultas: Rectifier; número de capas ocultas: 3; número de neuronas en cada capa: 50; regularización L1: 0,00001; regularización L2: 0,0; y ratio de exclusión de la capa de entrada: 0%; ratio de exclusión de las capas ocultas: 10%.

Fuente: Elaboración propia.

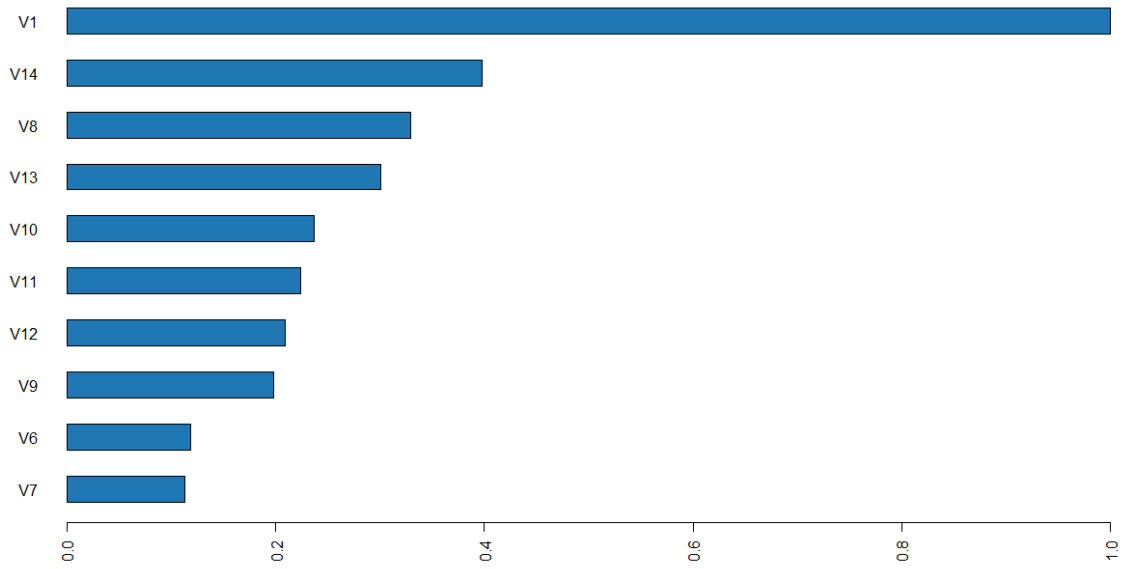
Por último, vamos a tomar los valores óptimos de los hiperparámetros identificados en la construcción del Modelo 3, pero reduciendo la ratio de aprendizaje a un valor muy pequeño (0,0001). El algoritmo del descenso del gradiente para encontrar la solución que minimice la función de coste realiza un recorrido de múltiples iteraciones cuyo movimiento o tamaño de los pasos viene determinado por la ratio de aprendizaje. Muchas veces una reducción de esta ratio permite encontrar una solución mejor, aunque a costa de incrementar el tiempo de optimización, pues para un recorrido similar si los pasos son más pequeños el tiempo de la CPU de un ordenador aumenta considerablemente. De este modo, se ha obtenido un Modelo 4, con unos resultados peores (*AUC* de 0,9541 y pérdida *Log-Loss* de 0,1932, en la muestra de validación) que los obtenidos en el modelo anterior. Por tanto, es el Modelo 3 el que constituirá nuestro Modelo Final.

5.2. Importancia de las variables del modelo de aprendizaje profundo.

Aunque el modelo de aprendizaje profundo es más difícil de interpretar que un modelo sencillo como una regresión o árbol de decisión individual, se puede obtener una valoración global de la importancia de cada variable midiendo la mejora que aporta cada uno de los predictores en la búsqueda de la solución óptima del algoritmo. En este sentido, los gráficos de dependencia parcial ofrecen una representación gráfica del efecto marginal de una variable en la respuesta del modelo. El efecto de una variable se mide en el cambio que produce sobre la media de la variable de respuesta.

La Figura 6 muestra la importancia de las variables del modelo de aprendizaje profundo. Se aprecia que las variables de mayor peso son, por una parte, V1, que expresa el valor de los ingresos de explotación y V14, indicativa de la cifra de activos totales. Ambas variables están directamente relacionadas con el tamaño de la empresa. Por otra parte, son especialmente significativas también V8, que muestra la relación entre ventas sobre recursos permanentes y V13 que indica el cociente entre el patrimonio neto y el activo.

Figura 6. Importancia de las variables del modelo de aprendizaje profundo.



Fuente: Elaboración propia.

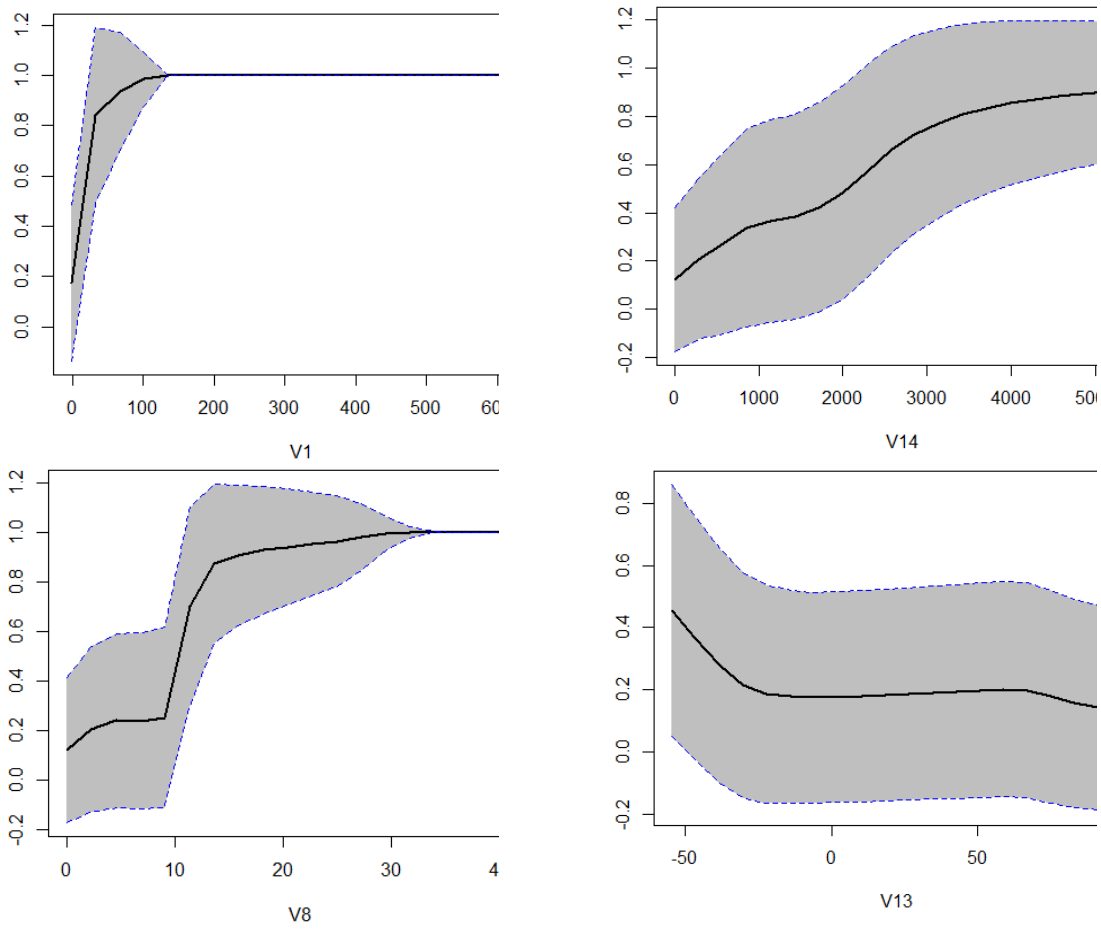
La Figura 7 muestra los gráficos de dependencia parcial para las cuatro variables más relevantes en la respuesta al fracaso empresarial identificados. Se presenta el efecto marginal de cada variable en la probabilidad de que una empresa entre en una situación de concurso. Esto es, cada gráfico ilustra la influencia parcial de una variable independiente en el modelo resultante tras considerar el efecto en conjunto y promediado de las otras variables independientes. Aunque la forma de estos gráficos no proporciona una explicación muy clara, sí que recogen la tendencia general y pueden utilizarse para interpretar el efecto de los predictores en la variable de respuesta (Friedman, 2002; Natekin & Knoll, 2013). El eje de las abscisas muestra los valores de la variable y el de ordenadas la probabilidad de fracaso empresarial. Las franjas de color gris reflejan la desviación estándar de las curvas de dependencia parciales.

Las curvas ascendentes de los gráficos correspondientes a las variables V1 y V14 indican que conforme aumenta el valor de los ingresos de explotación aumenta la propensión al fracaso empresarial. Atendiendo al gráfico representativo de la variable V14 se observa que a medida que aumenta el tamaño de la empresa aumentan las posibilidades de entrar en fase de concurso, es decir, aumentan las posibilidades de entrar en fases de dificultad financiera. Los resultados obtenidos en estas dos variables, relacionadas directamente con el tamaño de la empresa, apuntan a que cuanto mayor es el tamaño de la firma, mayor es la propensión a fracasar. Este hecho se justifica en que las empresas más grandes están sujetas a rigideces en sus estructuras económicas y financieras que las exponen a mayores riesgos por su menor flexibilidad en la adaptación a los cambios de coyuntura. Este resultado es consecuente con la situación de numerosas empresas y similar al obtenido en recientes estudios, como Momparler, Carmona y Climent (2016).

Por su parte, el gráfico ascendente de la variable V8, cociente entre ventas y recursos permanentes, nos indica que cuanto mayor es su valor, mayor es la propensión al fracaso. Esta variable aumenta cuando lo hacen los ingresos (lo que es coherente con lo indicado anteriormente con la variable V1) o cuando disminuyen los recursos permanentes, por una situación de dependencia del endeudamiento.

Por último, la variable de solvencia V13 resulta también significativa, de modo que cuando aumenta se reduce la propensión al fracaso al ser la empresa cada vez más solvente.

Figura 7. Gráficos de dependencia parcial.



Notas: Eje de las ordenadas recoge la probabilidad de fracaso empresarial.
Fuente: Elaboración propia.

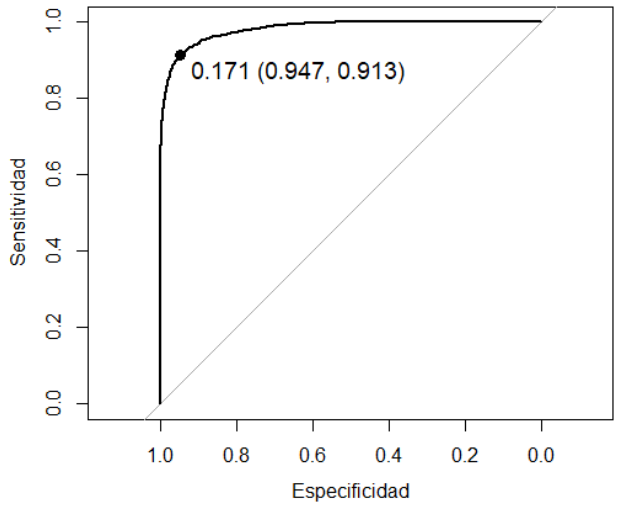
5.3. Precisión del modelo y capacidad predictiva sobre una muestra independiente.

La Figura 8 recoge los resultados la aplicación del modelo de aprendizaje profundo sobre la muestra de comprobación. Como se ha especificado anteriormente, la muestra de datos disponibles se ha dividido en tres bloques: muestra de entrenamiento (60%), muestra de validación (20%) y muestra de comprobación (20%). Las dos primeras muestras se han empleado para ajustar y extraer el mejor modelo, de modo que la muestra de comprobación ha permanecido siempre oculta para evitar incurrir en un problema de sobreajuste (*overfitting*, en inglés). Ahora, es el momento de utilizar esta muestra de comprobación para determinar la validez del modelo con unas observaciones totalmente independientes de las utilizadas durante la construcción del modelo. El valor del estadístico *AUC* en esta muestra de comprobación es 0,9816, por lo que el modelo es capaz de clasificar correctamente la mayoría de las empresas, diferenciando entre activas y en concurso. Una curva ideal *ROC* se extendería y alcanzaría el vértice superior izquierdo, de su representación gráfica; por tanto, cuanto mayor sea el área de debajo de la curva mejor es el algoritmo del clasificador y el *AUC* toma un valor más próximo a 1 (Figura 8). La gráfica muestra también la sensibilidad (el porcentaje de aciertos del evento de interés o fracaso, en nuestro caso) y la especificidad (el porcentaje de aciertos del evento que no es de interés o no fracaso, en nuestro caso) del modelo para el mejor punto de corte (0,171), que indica la probabilidad de corte para que el modelo realice su labor de clasificación entre los dos tipos de empresas. En la parte de la derecha, la matriz de confusión refleja que el porcentaje global de aciertos es de un 94,01%. También se extrae que se clasifican correctamente un 94,67% de las empresas en activo y un 91,30% de la que están en situación de concurso (especificidad y sensibilidad del modelo, respectivamente).

En suma, estos resultados, obtenidos con la muestra de comprobación, vienen a demostrar que el modelo de aprendizaje profundo construido con la muestra de entrenamiento y de validación tiene un nivel de precisión y una capacidad predictiva también muy altos sobre unos datos independientes no empleados para su elaboración.

Figura 8. Precisión del modelo de aprendizaje profundo en la muestra de comprobación.

Curva ROC



NOTAS: AUC = 0,98. Exactitud = 94,01%

Matriz de confusión

NOTAS:

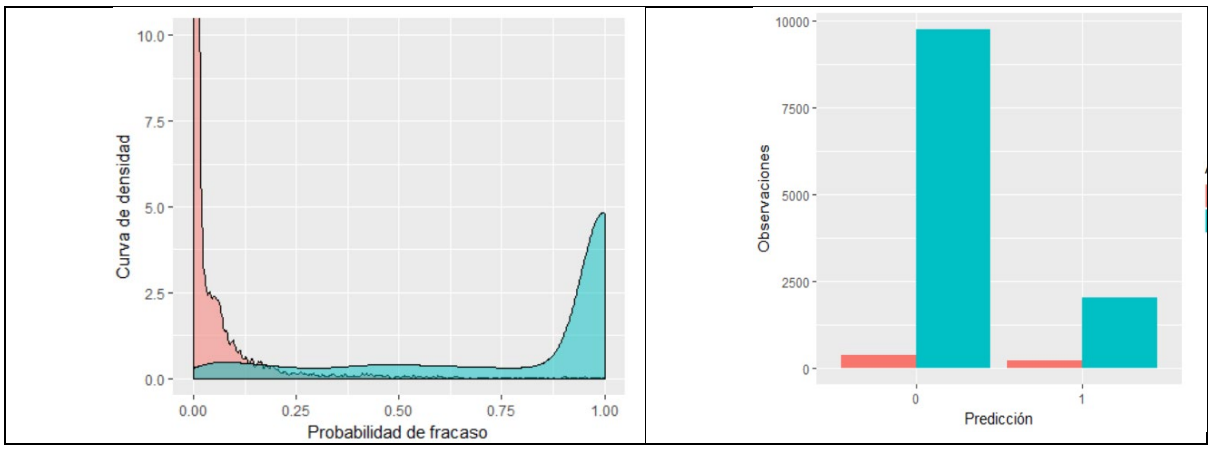
Real		Predicción: Fracaso: No	Predicción: Fracaso: Sí	Exactitud
Fracaso	No	**9.403**	529	94,67%
	Sí	209	**2.173**	91,30%
Total				94,01%

Resultados para un punto de corte de 0,171.
N = 12.314

Fuente: Elaboración propia.

A continuación, para constatar de modo visual que el modelo de aprendizaje profundo construido es capaz de diferenciar entre las empresas fracasadas y no fracasadas se ha dibujado la curva de densidad de las probabilidades que produce el modelo para ambas categorías de empresa (Figura 9, panel de la izquierda).

Figura 9. Curva de densidad de las probabilidades de las predicciones y distribución de los aciertos del modelo de aprendizaje profundo.



Fuente: Elaboración propia.

Se observa que efectivamente las dos distribuciones de probabilidad son muy diferentes, aunque la existencia de un área de intersección corrobora que el modelo no es perfecto y comete algunos errores como ya se ha comentado. El panel de la derecha de esta figura muestra gráficamente la existencia de estos errores, que son mucho menores que los aciertos (empresas clasificadas correctamente).

6. Conclusiones.

El objetivo de este estudio ha sido analizar la idoneidad de las redes neuronales del *Deep Learning,* en particular, de las redes de alimentación hacia adelante (feedforward neuronal networks) en la predicción del fracaso empresarial mediante la formulación de un modelo que permita predecir el fracaso empresarial en empresas europeas, intentando cubrir así la práctica inexistencia de investigación detectada, ya que apenas había literatura en este campo que utilizara esta novedosa metodología.

Se ha obtenido un modelo que permite clasificar, de una manera general, entre empresas sanas y en peligro de entrar en situación concursal. Para ello se ha realizado un análisis empírico aplicando la metodología *Deep Learning*, novedosa en este campo de investigación. La capacidad predictiva que muestra el modelo estimado alcanza un 94% sobre una muestra de comprobación independiente formada por el 20% de las observaciones. Atendiendo a estos resultados, podemos constatar la adecuación y utilidad de esta técnica en el campo de investigación de la predicción del fracaso empresarial.

Se ha podido comprobar que las variables más relevantes son las indicativas del tamaño empresarial, representadas por el volumen de ingresos y el tamaño de activo, lo que demuestra que, en general, el aumento del tamaño empresarial influiría decisivamente en el riesgo de fracasar, convirtiendo a las empresas de mayor tamaño en más vulnerables. Este hecho se justifica en que las empresas más grandes están sujetas a rigideces en sus estructuras económicas y financieras que las exponen a mayores riesgos por su menor flexibilidad para adaptarse a los cambios de coyuntura.

Además, se ha comprobado que la probabilidad de una situación de quiebra aumentaría cuando no se dispone de recursos permanentes suficientes y se recurre excesivamente al endeudamiento. En definitiva, mayor endeudamiento supone mayor riesgo de fracaso.

Finalmente, se ha observado que la posibilidad de entrar en concurso está íntimamente relacionada con la incapacidad de atender puntualmente las obligaciones financieras. Este extremo queda demostrado con la evolución de la variable solvencia que permite afirmar que la probabilidad al fracaso se reduce en aquellas entidades que muestran valores elevados en esta variable.

Entendemos que además de los indicadores clásicos de la viabilidad financiera constatados en este estudio, endeudamiento y solvencia, el modelo estimado incluye como novedad la variable tamaño, clave en el futuro de la empresa. La combinación de estas variables puede aplicarse por los usuarios para diagnosticar si una empresa presenta síntomas de dificultad financiera que podrían derivar en una situación de fracaso.

Una línea de investigación complementaria consistiría en añadir al estudio variables de carácter cualitativo o no financieras, en definitiva, información de mercado que refleje también la influencia de la coyuntura económica en la empresa analizada. En este sentido, pensamos en la conveniencia de ampliar la definición de fracaso empresarial de modo que además de criterios jurídicos se incorporen también criterios económicos, lo que sin duda incrementará la capacidad de generalización de los modelos obtenidos. También consideramos abierta la posibilidad de centrar este tipo de trabajos en sectores productivos particulares.

Referencias

Abbott, A., Deshowitz, A., Murray, D., & Larson, E. (2018). WalkNet: A Deep Learning Approach to Improving Sidewalk Quality and Accessibility. *SMU Data Science Review, 1*(1-7), 1-12.

Ahn, B.S., Cho, S.S., & Kim, C.Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Experts Systems with Aplications, 18*, 65-74.

Alfaro, E., Gámez, M., & García, N. (2007a). A Boosting Approach for Corporate Failure Prediction. *Applied Intelligence, 27*(1), 29-37.

Alfaro, E., Gámez, M., & García, N. (2007b). Multiclass Corporate Failure Prediction by Adaboost.M1. *International Advances in Economic Research, 13*(3), 301-312.

Alfaro, E., Gámez, M., García, N., (2008). Linear Discriminant Analysis Versus Adaboost for Failure Forecasting. *Revista Española de Financiación & Contabilidad, 37*(137), 13-32.

Alfaro, E., García, N., Gámez, M., & Elizondo, D. (2008). Bankruptcy Forecasting: An Empirical Comparison of AdaBoost and Neural Networks. *Decisión Support Systems, 45*(1), 110-122.

Alfaro, E., Gámez, M., & García, N. (2013). Adabag: An R Package for Classification with Boosting and Bagging. *Journal of Statistical Software, 54*(2), 1-35.

Bell, T.B., Ribar, G.S., & Verchio, J. (1990). Neural Nets Versus Logistic Regression: A Comparison of Each Model's Ability to Predict Commercial Bank Failures. En Srivastava, R.P. (ed). *Proceedings of the 1990 Deloitte & Touche/University of Kansas Symposium in Auditing Problems,* 29-53. Auditing Symposium X. Universidad de Kansas, Lawrence.

Boritz, J., & Kennedy, D.B. (1995). Effectiveness of Neuronal Network Types for prediction of Business Failure. *Experts Systems with Aplications, 9*(4), 503-512.

Brownlee, J. (2017). *What is the Difference Between a Parameter and a Hyperparameter?* https://machinelearningmastery.com/difference-between-a-parameter-and-a-hyperparameter/ (visitado el 28 de marzo de 2018)

Chaudhuri, A., & Ghosh, S. K. (2017). *Bankruptcy Prediction through Soft Computing based Deep Learning Technique*. Germany: Springer.

Cook, D. (2017). *Practical Machine Learning with H2O*. Sebastopol, CA (USA): O'Reilly Media.

Cortes, C., & Mohri, M. (2004). AUC optimization vs, error rate minimization, *Advances in Neural Information Processing Systems*, *16*(16), 313-320.

Crawford, C. (2016). *An Introduction to Deep Learning*. https://blog.algorithmia.com/introduction-to-deep-learning-2016/ (visitado el 30 de marzo de 2018)

Deng, L., & Yu, D. (2013). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing, 7*(3-4), 197-387. http://dx.doi.org/10.1561/2000000039

Díaz, Z., Fernández, J., & Segovia, M.J. (2004). Sistemas de inducción de reglas y árboles de decisión aplicados a la predicción de insolvencias en empresas aseguradoras. *Documentos de Trabajo de la Facultad de Ciencias Económicas y Empresariales, 9.* Universidad Complutense de Madrid.

Fletcher, D., & Goss, E. (1993). Application Forecasting with Neural Networks: An Application Using Bankruptcy Data. *Information and Management, 24*, 159-167.

Friedman, J.H. (2002). Stochastic Gradient Boosting. *Journal of Computational Statistics & Data Analysis*, *38*(4), 367-378.

Goodfellow, I., Bengio, Y., & Aaron, C., (2016). *Deep Learning*. Cambridge, MA, USA: The MIT Press.

Gulli, A., & Pal, S. (2017). *Deep Learning with Keras. Implement neural networks with Keras on Theano and Tensorflow*. Birmingham, Reino Unido: Packt Publishing Ltd.

Jayanthi, J., Kaur, G., & Suresh, K. (2017). Financial forecasting using decision tree (reptree & c4.5) and neural networks (k*) for handling the missing values. ICTAC *Journal on Soft Computing, 7*(3), 1473-1477.

Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. *Expert Systems with Applications*, *37*(9), 6233-6239.

Kim, M.J., & Kang, D.K. (2010). Ensemble with Neural Networks for Bankruptcy Prediction. *Expert System with Applications*, *37*(4), 3373-3379.

Kim, M.J., Kang, D.K., & Kim, H.B. (2015). Geometric Mean Based Boosting Algorithm with Over-Sampling to Resolve Data Imbalance Problem for Bankruptcy Prediction. *Expert Systems with Applications*, *42*(3), 1074-1082.

Kim, S.Y., & Upneja, A. (2014). Predicting Restaurant Financial Distrees Using Decisión Tree and AdaBoosted Decision Tree Models. *Economic Modelling, 35*, 354-362.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436-444. http://dx.doi.org/10.1038/nature14539

Ledell, E. (2018). *Deep Learning with H2O*. https://github.com/ledell (visitado el 20 de marzo de 2018)

López, F.J., & Pastor, I. (2015). Bankruptcy visualization and prediction using neural netwoks: A study of U.S. comercial banks. *Experts Systems with Aplications, 42*, 2857-2869.

Momparler, A., Carmona, P., & Climent, F.J. (2016). La predicción del fracaso bancario con la metodología Boosting Classification Tree. *Revista Española de Financiación y Contabilidad, 45*(1), 63-91.

Natekin A., & Knoll, A. (2013). Gradient Boosting Machines. A Tutorial. *Frontiers in Neurorobotics. 7*(21), 1-21.

Odom, M.D., & Sharda, R. (1992). A Neural Network Model for Bankruptcy Prediction. En R.R. Trippi and E. Turban (Eds.). *Neural networks in Finance and Investing*. Chicago: Probus Publishing.

Patel, V., Armstrong, D., Ganguli, M.P., Roopa, S., Kantipudi, N., Aalbashir, S., & Kamarth, M. (2016). Deep Learning in Gastrointestinal Endoscopy. *Critical Reviews™ in Biomedical Engineering, 44*(6), 493-504. http://dx.doi.org/10.1615/CritRevBiomedEng.2017025035

Popescu, M.E., Andreica, M., & Popescu, I-P. (2017). Decision support solution to business failure prediction. *Proceedings of the International Management Conference, Faculty of Management,* Academy of Economic Studies, Bucharest, Romania, *11*(1), 99-106.

Pozuelo, J., Martínez, J., & Carmona, P. (2018). Análisis de la utilidad del algoritmo Gradient Boosting Machine (GBM) en la predicción del fracaso empresarial. *Spanish Journal of Finance and*

Accounting / Revista Española de Financiación y Contabilidad, *47*(4) 507-532. http://dx.doi.org/10.1080/02102412.2018.1442039.

R Core Team (2019). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, Austria. URL http://www.R-project.org/.

Ravi, P., & Ravi, V. (2007). Bankruptcy Prediction in Banks and Firms ViaStatistical and Intelligent Techniques - A Review. *European Journal of Operational Research, 180*(1), 1-28.

Romero, F. (2013). Alcances y limitaciones de los modelos de capacidad predictiva en el análisis del fracaso empresarial. *AD-minister, 23*, 45-70.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85-117.

Tam, K.Y. (1991). Neural Network Models and the Prediction of Bank Bankruptcy, *Omega, 19*(5), 429-445.

Tam, K.Y., & Kiang, M.Y. (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. *Management Science, 38*(7), 926-947.

Tascón, M.T., & Castaño, F.J. (2012). Variables y modelos para la identificación y predicción del fracaso empresarial: revisión de la investigación empírica reciente. *Revista de Contabilidad, 15*(1), 7-58.

The H2O.ai team (2019). h2o: R Interface for H2O. R package version 3.26.0.2. https://CRAN.R-project.org/package=h2o.

Tseng, F-M., & Hu, Y-Ch. (2010). Comparing fou nakrupcty prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks. *Experts Systems with Aplications, 37*, 1846-1853.

Wang, G., Ma, J., & Yang, S. (2014). An Improved Boosting Based on Feature Selection for Corporate Bankruptcy Prediction. *Expert Systems with Applications, 41*(5), 2353-2361.

Wang, L., Zeng, Y., & Chen, T. (2015). Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Systems with Applications, 42*, 855-863.

West, D., Dellana, S., & Qian, J. (2005). Neural Network Ensemble Strategies for Financial Decision Applications. *Computers & Operations Research, 32*(10), 2543-2559.

Wilson, G.I., & Sharda, R. (1994). Bankruptcy Prediction Using Neural Network. *Decision Support Systems, 11*, 545-557.

Zhang, G.P., Hu, M.Y., Patuwo, B.E., & Indro, D.C. (1999). Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis. *European Journal of Operational Research, 116*(1), 16-32.

Zieba, M., Tomczak, S.K., & Tomczak, J.M. (2016). Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction. *Expert Systems with Applications, 58*(1), 93-101.