

# La predicción del fracaso empresarial de las cooperativas españolas. Aplicación del Algoritmo Extreme Gradient Boosting

Mariano Romero Martínez, Pedro Carmona Ibáñez, José Pozuelo Campillo

**RESUMEN:** El objetivo del estudio realizado es mostrar la utilidad de la metodología *Extreme Gradient Boosting (XGBoost)*, basada en el aprendizaje automático, en la predicción del fracaso empresarial, particularmente en el ámbito de las empresas cooperativas. Para ello, se ha estimado un modelo a partir de una muestra de cooperativas pertenecientes a diversos sectores, que ha permitido identificar que estas empresas presentan una menor propensión al fracaso cuando alcanzan mayores índices de solvencia y rentabilidad y menor endeudamiento. De acuerdo con los resultados, el modelo estimado *XGBoost* obtiene una capacidad predictiva del 100% en la muestra de entrenamiento y de un 86% en una muestra de comprobación independiente. Asimismo, se destaca la facilidad de la interpretación de los resultados alcanzados, siendo posible determinar el signo y el efecto de cada variable sobre la predicción final de todas y cada una de las observaciones. Este trabajo aporta a la literatura del fracaso empresarial del sector cooperativo la constatación de la utilidad de una nueva metodología (*XGBoost*), que proporciona un elevado nivel de aciertos y una sencilla interpretabilidad de los resultados.

**PALABRAS CLAVE:** Fracaso empresarial, cooperativas, modelos de predicción, Extreme Gradient Boosting.

**CLAVES ECONLIT:** C10, M20, M40.

**Cómo citar este artículo/How to cite this article:** ROMERO, M., CARMONA, P. & POZUELO, J. (2021): "La predicción del fracaso empresarial de las cooperativas españolas. Aplicación del Algoritmo Extreme Gradient Boosting", *CIRIEC-España, Revista de Economía Pública, Social y Cooperativa*, 101, 255-288. DOI: 10.7203/CIRIEC-E.101.15572.

**Correspondencia:** Pedro Carmona Ibáñez, Universitat de València, Pedro.Carmona@uv.es, ORCID: 0000-0002-9979-2727; José Pozuelo Campillo, ORCID: 0000-0001-8768-2443; Mariano Romero Martínez. ORCID: 0000-0002-4569-868X.

## Expanded abstract

# The prediction of the business failure of the spanish cooperatives. Application of the Extreme Gradient Boosting Algorithm

Cooperatives are entities with regulatory and operational singularities that make them different from other commercial companies in the business landscape. The relations with the workers, the remuneration of the partner's contribution and the way in which the results are distributed configure them as companies of a deep social nature, even though they also have a commercial nature.

After reviewing the contents of the studies aimed at the identification of business failure prediction models, we have observed that there are still very few which address cooperative societies, despite the importance of this kind of organization in our business landscape. In addition, most of the papers follow traditional statistical methodologies. In fact, the different methodologies used in this field of research have evolved over time; and one of the newest and most innovative is *Extreme Gradient Boosting (XGBoost)*, which to the best of our knowledge has not been applied in this research area.

The objective of this study is to show the usefulness of Extreme Gradient Boosting methodology, which is based on machine learning techniques, in the prediction of business failure, particularly in the field of cooperative companies.

We will determine whether the algorithm is suitable for the identification of the most relevant variables that indicate situations of financial distress. Therefore, it will allow to classify a company as failed or not failed. We will also analyze whether this technique is appropriate to explain the reasons leading to this classification.

We also intend to contribute to overcoming the general absence in the literature of this type of studies in this business environment and to incorporate modern techniques based on artificial intelligence.

To achieve the aims of the study, a model has been estimated from a sample of non-financial cooperatives that failed in 2015, belonging to various productive sectors. To consider a company as failed, a broad definition of failure has been chosen, because it allows us to clearly distinguish between failed and non-failed companies. Thus, the failure has been identified to the legal definition of insolvency, as contemplated by the present Spanish insolvency regulation. The selection made, taking into account this condition, consisted in 56 companies that had filed insolvency proceedings in the year considered. Thereafter, they have been paired with as many non-failed ones with similar characteristics, so that the final sample has consisted of 112 companies.

Since our main goal is to formulate business failure prediction models, we have focused on those variables or financial ratios that provide information on aspects of the firm's solvency and profitability, but without forgetting the influence of the company's debts. To these main cate-

gories, we have added rotation, activity and asset structure. Given the uniqueness of this kind of companies we are considering, some clarifications have been made for some of the selected financial ratios, specifically for the debt and profitability ratios.

Results from the statistical study show the importance of the variables related to solvency, liquidity, debt and profitability. These variables exhibited the most contribution to detecting a situation of business failure.

The model combines in an effective way the variables of long-term solvency, liquidity, activity, and debt, and achieves a 100% accuracy rate in the training sample and 86% in an independent testing sample. The results have confirmed the higher the solvency and immediate liquidity of a cooperative company the lower the probability of failure. Additionally, the higher the variable that measures the ratio value added to sales the higher the probability of failure, as financial and personnel expenses are involved in this relationship. Finally, we have observed that profitability increases when cash flow grows in proportion to the company's assets, therefore being consistent in the sense that as a company performs better the propensity for failure is reduced by being more profitable.

These results are consistent with the empirical evidence: the higher the solvency and profitability rates the lower the failure probability, and the lower the debts the lower the exposure to failure. Consequently, we understand that the applied technique is quite effective for the objectives pursued, that is, the early detection of business failure for cooperative companies.

Probably the variable less easy to compare is profitability since in a cooperative the book result is not comparable with the rest of commercial companies because it does not take into account the potential benefits of their partners through prices of cooperative products and services.

The resulting model, in addition to classifying cooperative as failed and non-failed, has shown its potential to identify the underlying reasons for this classification. That is, the *XGBoost* model algorithm provides explanations about the variable's weights in terms of failure probability for each observation of the test sample, splitting the individual weight and the influence sign of each of the predictors. Therefore, the final model is not only about high accuracy predictions, but also leads to results that are fully interpretable—it is possible to determine the sign and the effect of each variable or financial ratio on the final prediction for each and every one of the observations.

The study is intended to be of a continuity nature because we consider it would be very interesting to carry out this type of analysis again but extending it to a series of several years in order to propose a model capable of detecting failure with a few years in advance.

In short, we have demonstrated that this study contributes to the literature on business failure in the cooperative sector by means of showing the utility of a new methodology (*XGBoost*) capable of identifying cooperatives with a high probability of failure. The results are very accurate, and their interpretation is very easy.

**Keywords:** Business failure, cooperatives, prediction models, Extreme Gradient Boosting.

# 1. Introducción<sup>1</sup>

La dilatada crisis económica y financiera que se inició en 2008 supuso la desaparición de numerosas empresas, lo que ha revitalizado las investigaciones sobre los motivos que llevan a las empresas a situaciones de dificultad financiera que, en ocasiones, pueden provocar su desaparición definitiva.

Revisando los contenidos de los estudios orientados a la elaboración de modelos que tratan de anticipar situaciones de fracaso empresarial, se puede constatar que los centrados en las pymes, fundamentales en nuestro tejido empresarial, han ido ganando terreno, pese a que siguen siendo escasos los que se ocupan de las sociedades cooperativas<sup>2</sup>. Esta circunstancia se atribuye básicamente a la tradicional ausencia de bases de datos fiables de cooperativas, lo que ha dificultado construir muestras de un tamaño adecuado que garanticen la significatividad estadística de los resultados, así como la comparación de las peculiaridades de las cooperativas con el resto de empresas.

Las cooperativas, con una relevante tradición e implantación en diversos sectores económicos de nuestro país, son sociedades de carácter social que pretenden facilitar a sus socios bienes o servicios a menor precio o retribuirles mejor sus prestaciones. Para ello, cuentan con una regulación peculiar diferente de otros tipos de sociedades<sup>3</sup>. Este hecho nos obligará a matizar algunas ratios utilizadas habitualmente en el análisis financiero y estadístico, principalmente en las categorías de endeudamiento y de rentabilidad (Pozuelo, Martínez y Carmona, 2012). Estas matizaciones nos ocuparán unas líneas cuando se proponga en el oportuno apartado las variables explicativas de los posibles modelos.

Entre las diversas características<sup>4</sup> que diferencian a las cooperativas de otras sociedades destacan su capital variable y su remuneración, el reparto del resultado de acuerdo con las operaciones realizadas por cada socio y la asignación obligatoria de parte del resultado a la educación, formación y promoción de sus socios y trabajadores (Juste et al. 2011 y Pozuelo, Martínez y Carmona, 2012). Por su parte, Iturrioz (2010) y Masa et al. (2016) señalan las peculiaridades de quienes toman las decisiones en las cooperativas, ya que en las empresas convencionales los clientes, proveedores y trabajadores no intervienen en la toma de decisiones, mientras sí que lo hacen en las cooperativas, dado el doble rol que éstos suelen tener al ser, simultáneamente, socios de las mismas.

Además, las cooperativas se rigen por una normativa fiscal específica y una adaptación contable (Orden EHA/3360/2010) y dado, que las comunidades autónomas asumen en sus

1. Se agradece el apoyo financiero del Ministerio de Ciencia e Innovación de España y FEDER (Proyecto PGC2018-093645-B-100).

2. Destaca el trabajo pionero de Dietrich et al. (2005) y la revisión de la literatura de Mateos et al. (2011), que además incluye un estudio sobre los modelos de predicción del fracaso empresarial y su aplicabilidad en cooperativas agrarias.

3. Carreras, G. (2011) analiza el carácter mercantil de las cooperativas.

4. Pozuelo, Carmona y Martínez (2012) analizan con profundidad las diferencias de las cooperativas con las sociedades capitalistas.

estatutos algunas competencias en materia cooperativa, dichas comunidades han ido desarrollando sus respectivas normativas en esta materia.

Por otro lado, las distintas metodologías utilizadas en el análisis del fracaso empresarial han evolucionado a través del tiempo, siendo una de las más novedosas, y apenas aplicada en este ámbito, la basada en *Extreme Gradient Boosting*. Este algoritmo de aprendizaje automático se utilizará con el fin de comprobar si permite identificar las variables más relevantes que detectan las situaciones de dificultad financiera, de manera que se pueda clasificar a una empresa como fracasada o no fracasada. Asimismo, comprobaremos si esta técnica puede revelar los motivos que conducen a dicha clasificación.

Así pues, la importancia para los agentes económicos de poder anticipar situaciones de riesgo en las empresas que puedan conducir a su fracaso, la escasez de trabajos en este sentido, centrados en las empresas cooperativas, y las ventajas que puede aportar el uso de la novedosa metodología *Extreme Gradient Boosting*, nos ha llevado a explorar en este trabajo un modelo de predicción de fracaso empresarial de cooperativas españolas utilizando esta metodología.

La delimitación del concepto de fracaso empresarial será fundamental para la selección de las empresas que formarán las diferentes muestras sobre las que se realizará el estudio. Se ha optado por una definición de fracaso que permita distinguir con claridad las empresas sanas de las que no lo son. De esta manera se ha equiparado el fracaso a la calificación jurídica de situación concursal, según lo contemplado en la legislación concursal española (RDL 1/2020 por el que se aprueba el texto refundido de la Ley Concursal)<sup>5</sup>, de modo que una empresa se considerará fracasada si ha presentado o iniciado un expediente concursal y sana o no fracasada en el caso contrario. Aunque este criterio<sup>6</sup>, de marcado carácter jurídico, presenta algunos inconvenientes, como la reducción considerable de los tamaños muestrales, se entiende que son superados por las ventajas de objetividad y fijación de la fecha del fracaso que aporta en el proceso de selección de las empresas que integrarán las distintas muestras.

La estructura de este estudio la detallamos a continuación. En primer lugar, se revisan brevemente las aportaciones de los últimos años con relación al objeto del trabajo. Posteriormente, se pasa a describir tanto los datos con los que trabajaremos -recopilados en las muestras elaboradas a tal fin- como las variables que usaremos. A continuación, se describen los fundamentos de la metodología *Extreme Gradient Boosting* empleada y los resultados obtenidos. Por último, se exponen las conclusiones del trabajo y las posibles líneas futuras de investigación en este campo de estudio.

5. El 7 de mayo de 2020 aparece publicado en el BOE el Real Decreto Legislativo 1/2020 por el que se aprueba el texto refundido de la Ley Concursal. El 1 de septiembre de 2020 entró en vigor y derogó la Ley 22/2003 Concursal. El texto tiene como objetivo regularizar, dilucidar y armonizar las diferentes reformas que ha sufrido la Ley Concursal de 2003, (28 cambios en 5 años).

6. Otras definiciones que aparecen en la literatura sobre el fracaso empresarial son el incumplimiento en las obligaciones de pago, los descubiertos en cuentas bancarias, las pérdidas en ejercicios consecutivos, un patrimonio neto negativo, el cese en el reparto de dividendos, figurar en listados de morosos, cese de actividad, etc. En el sector cooperativo muchas veces el fracaso empresarial se manifiesta como cese de la actividad, significativas bajas de socios o el comienzo de procesos de reestructuración. Sin embargo, la consideración exclusiva de estas circunstancias nos hubiera impedido la elaboración de una base de datos de estudio que ofreciese resultados estadísticamente significativos.

## 2. Antecedentes

Si bien existen revisiones muy completas de la literatura sobre fracaso empresarial (Tseng y Hu, 2010; Tascón y Castaño, 2012; Romero, 2013; Sun, Li, Huang, y He, 2014), nos ocuparemos fundamentalmente de las que hacen referencia a las últimas metodologías incorporadas, entre las que destacan las que forman parte del *Machine Learning*, como las basadas en redes neuronales. Éstas 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 en este campo datan de los años 90 (Bell, Ribar y Verchio, 1990; Odom y Sharda 1992; Wilson y Sharda, 1994; Boritz y Kennedy, 1995), prolongándose hasta la actualidad (López y Pastor, 2015; Popescu, Andreica y Popescu, 2017; Jayanthi, Kaur y Suresh, 2017). Diversos estudios señalan que las metodologías apoyadas en redes neuronales presentan ventajas de robustez respecto a otras anteriores (Tam, 1991; Tam y Kiang, 1992; Wilson y Sharda, 1994; Zhang et al. 1999). Además, muestran una gran capacidad de adaptación a la nueva información (Tam, 1991; Tam y Kiang, 1992) y mejoran significativamente los resultados respecto a otras técnicas estadísticas basadas en los análisis discriminante, logit o probit (Fletcher y Gross, 1999; Ravi, P. y Ravi, V. 2007; Zhang et al., 1999).

Un paso más evolucionado dentro del *Machine Learning* es la metodología denominada *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 metodología permite realizar operaciones más complejas obteniendo resultados muy fiables cuando se dispone de una gran cantidad de datos. Su uso se ha expandido en los últimos años, habiéndose aplicado con éxito en diversas áreas de conocimiento (medicina, multimedia, juegos, procesamiento de datos y del lenguaje, áreas de interés comercial, etc.). Aunque apenas se ha utilizado en el campo del fracaso empresarial, destaca el trabajo de Chaudhuri y Ghosh (2017) donde se consiguen modelos con una alta capacidad predictiva que supera a modelos anteriores basados en otros soportes metodológicos.

Otro conjunto de metodologías actuales de *Machine Learning*, que han evolucionado en paralelo a las anteriormente comentadas, son las denominadas *Boosting* que usan principalmente combinaciones de árboles de decisión individuales (*ensemble trees*) como clasificadores. Su utilidad y capacidad predictiva queda de manifiesto en trabajos sobre fracaso empresarial tanto de ámbito internacional (Kim y Kang, 2010; Kim y Upneja, 2014; Wang, Ma y Yang, 2014; Kim, Kang y Kim, 2015; Zieba et al. 2016) y nacional (Díaz et al. 2004; Alfaro, García et al., 2008; Momparler et al. 2016; Pozuelo et al. 2018; Climent et al. 2019; Carmona et al. 2019).

Alfaro, García et al. (2008) compararon la precisión de predicción del fracaso empresarial de modelos basados en redes neuronales y *Boosting* (en su versión *AdaBoost*) en un conjunto de empresas europeas, teniendo en cuenta las variables habituales de predicción (ratios financieros), así como las variables cualitativas, como el tamaño de la empresa, actividad y estructura jurídica, concluyendo que el enfoque de *AdaBoost* disminuía el error de generalización en aproximadamente un 30% respecto al producido con modelos basados en redes neuronales. Alfaro, Gámez et al. (2008) concluyeron que la metodología *Boosting*, en versión *Adaboosting*,

era adecuada para su uso en la predicción del fracaso empresarial. Para ello la aplicaron a una muestra de 1.180 empresas españolas y demostraron que era más precisa que el análisis discriminante.

Dentro de la metodología *Boosting* han seguido apareciendo nuevas técnicas como la *FS-Boosting* que han mejorado la capacidad predictiva. Destaca el trabajo de Wang, Ma y Yang (2014). Por su parte, Kim et al. (2015), modifican el algoritmo *AdaBoost*, de manera que el cálculo del error geométrico reemplaza al aritmético. Esta técnica, a la que denominan *GMBBoost* logra un poder de predicción y capacidad de aprendizaje muy elevados. Zieba et al. (2016) obtuvo también con esta técnica unos resultados significativamente mejores respecto a otros métodos anteriores.

Por su parte, como ya se ha indicado, Momparler et al. (2016) y Pozuelo et al. (2018) usaron la técnica GBM (*Gradient Boosting Machine*), consiguiendo unas elevadas tasas de acierto en la predicción del fracaso bancario y empresarial, respectivamente.

Finalmente, la variante metodológica del *Extreme Gradient Boosting (XGBoost)* ha demostrado ser mejor cuando se trabaja con datos estructurados y pocas variables. Ésta asigna menor peso a las variables que considera menos significativas, pues incorpora lo que se conoce con el nombre de regularización del peso de las variables. En el ámbito del fracaso empresarial Zieba et al. (2016) utilizaron y compararon diferentes métodos, tanto de base estadística como de *Machine Learning*, concluyendo que el *XGBoost* obtenía mejores resultados. Recientemente, los trabajos llevados a cabo por Climent et al. (2019) y Carmona et al. (2018) centrados en el análisis del fracaso bancario en la eurozona y en los Estados Unidos, respectivamente, han aplicado esta metodología constatando su alta capacidad predictiva.

Por otro lado, aunque el estudio del fracaso empresarial de las sociedades cooperativas no ha sido abordado suficientemente en la literatura científica pese a su importancia en nuestro tejido económico, destacamos los trabajos de Vargas (2010), Iturrioz (2010) e Iturrioz y Martín (2013) donde se analiza el concurso de acreedores de cooperativas, pero sin obtener un modelo predictivo de su fracaso.

Pozuelo, Martínez y Carmona, (2012) llevaron a cabo un estudio centrado en la estimación, a partir de técnicas estadísticas, de modelos aptos para la predicción del fracaso empresarial en las empresas cooperativas. Marí et al. (2014) analizaron cualitativamente los determinantes en la predicción del fracaso de las cooperativas agroalimentarias utilizando el método Delphi, concluyendo en el alto poder predictivo del endeudamiento, la autofinanciación y el *cash flow*.

Masa et al. (2016) analizaron la capacidad predictiva del fracaso empresarial con un modelo de ecuaciones estructurales basado en mínimos cuadrados parciales, que aplicaron a dos muestras, una de cooperativas y otra de empresas capitalistas, constatando que solvencia, liquidez y rentabilidad son las principales variables que pueden anticipar el fracaso. A pesar de encontrar algunas diferencias en los resultados entre las dos submuestras, únicamente resultaron significativas la localización, que influye positiva y significativamente en las cooperativas y no en las empresas capitalistas. También, la antigüedad que influye positiva y significativamente únicamente en las cooperativas, debido a que la dotación anual obligatoria de reservas incide en su solvencia.

Nos encontramos pues con una escasez de trabajos sobre la predicción del fracaso de cooperativas basados en la metodología *Extreme Gradient Boosting*, que constituye un motivo relevante para la realización de este estudio. Se pretende constatar la utilidad de este algoritmo como herramienta en la predicción del fracaso empresarial en el ámbito de las cooperativas.

### 3. Selección y fuentes de la muestra de empresas

En el proceso de selección y obtención de las diferentes muestras de empresas se ha recurrido a la base de datos financieros *SABI* (Sistema de Análisis de Balances Ibéricos).

Para seleccionar las empresas que formarán parte de la muestra de estimación el estudio se limitará a aquellas cooperativas de naturaleza no financiera fracasadas en el año 2015, de acuerdo con la definición adoptada. La selección efectuada, atendiendo a esta condición, redujo la cifra a 74 empresas que habían presentado expedientes concursales en el ejercicio considerado.

En las 74 empresas resultantes se realizaron dos nuevos filtrados. Uno para que la muestra guardase un número proporcional al índice de fracaso del sector en que operaban y otro, por el que se descartaron aquellas empresas de reciente creación (hasta tres años) y las que no contenían datos contables completos de al menos tres ejercicios anteriores a la fecha del fracaso.

Tras estos procesos de selección y filtrado, el número de firmas se redujo a 56 empresas cooperativas fracasadas que son las que definitivamente se integrarán en la muestra de estimación.

Para completar la muestra de estimación y poder aplicar ciertas herramientas estadísticas de clasificación, se recurrió a la técnica del emparejamiento consistente en asociar cada una de las empresas quebradas con una no quebrada de las mismas características elegida aleatoriamente entre aquellas de un tamaño similar, medido en función del volumen de activo, y que operase en el mismo sector económico establecido por el N.A.C.E (rev). (Clasificación Nacional de Actividades Económicas) a nivel de cuatro dígitos y cuando no ha sido posible se ha descendido a tres dígitos. Con este proceso se incorporaron otras 56 empresas quedando la muestra de estimación finalmente compuesta por 112 cooperativas, la mitad sanas y la otra mitad fracasadas.

La agrupación sectorial se ha realizado por categorías atendiendo a las secciones del C.N.A.E. (Clasificación Nacional de Actividades Económicas) de 2009 procurando homogeneizar las actividades previstas en cada una de ellas, lo que nos ha permitido distinguir 9 grupos básicos. El resumen de la distribución sectorial resultante de la muestra de estimación lo mostramos en la tabla 1, donde se puede apreciar el peso específico de cada categoría sectorial.

**Tabla 1.** Distribución de la muestra de estimación por agrupaciones sectoriales

CATEGORÍAS (Grupos básicos)	SECCIONES CNAE 2009	ACTIVIDADES DEL GRUPO	EMPRESAS	%
0	A	Agricultura, ganadería, silvicultura y pesca	12 + 12	21,42
1	B	Industrias extractivas	4 + 4	7,14
2	C	Industria manufacturera	13 + 13	23,21
4	G	Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas	9 + 9	16,07
5	H I J	Transporte y almacenamiento Hostelería Información y comunicaciones	5 + 5	8,92
6	L M	Actividades inmobiliarias Actividades profesionales, científicas y técnicas	3 + 3	5,35
7	N O	Actividades administrativas y servicios auxiliares Administración pública y defensa; seguridad social obligatoria	2 + 2	3,57
8	P Q R	Educación Actividades sanitarias y de servicios sociales Actividades artísticas, recreativas y de entretenimiento	6 + 6	10,71
9	S T U	Otros servicios empleadores de personal doméstico; productores de bienes y servicios para uso propio Actividades de organizaciones y organismos extraterritoriales	2 + 2	3,57
TOTAL EMPRESAS (No fracasadas + fracasadas)			56 + 56	100,00

Para completar las características de la muestra, exponemos en la tabla 2 los grupos de empresas que la constituyen clasificados según el tamaño. En este caso, para diferenciar los distintos tamaños de empresa, nos hemos basado en la clasificación que ofrece la Unión Euro-

pea en el Anexo I del Reglamento (UE) n° 651/2014 de la Comisión tomando como referencia el volumen de activo.

**Tabla 2.** Tamaño de las empresas consideradas

	Número	%	Volumen de Activo (En miles de €) (*)
Microempresas	24 + 24	42,85	Hasta 2.000
Pequeñas	23 + 23	41,07	Entre 2.000 y 10.000
Medianas	5 + 5	8,92	Entre 10.000 y 43.000
Grandes	0	0	Más de 43.000
TOTAL (No fracasadas + fracasadas)	56 + 56	100,00	

(\*) Criterio extraído del Anexo 1 del Reglamento (UE) n° 651/2014 de la Comisión.

## 4. 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 mayoritariamente ratios económico-financieras. La primera dificultad para acometer esta etapa es la ausencia de una teoría general que guíe el proceso de selección, lo que constituye una fuerte limitación a la hora de modelizar el fracaso empresarial<sup>7</sup>. En este trabajo se procurará conciliar la experiencia aportada por otros autores con los objetivos propuestos. Por ello, para seleccionar las ratios con los que comenzaremos a trabajar nos basaremos en:

1. Ratios tradicionales en la literatura sobre análisis contable.
2. Frecuencia de aparición en estudios anteriores.
3. Facilidad para ser calculados y definidos en función de la información contable disponible.

Además, ninguno de los ratios considerados incorpora información de mercados bursátiles ya que ninguna de las cooperativas consideradas en las diferentes muestras cotizaba en el mercado de valores.

7. *Un estudio interesante que aporta una guía para la incorporación de ratios basados en un modelo económico-financiero de solvencia empresarial lo encontramos en Diéguez et al. (2006). También puede consultarse Pozuelo et al. (2010) y Labatut et al. (2009).*

Dado que nuestro objetivo es la formulación de modelos de predicción de fracaso empresarial incidiremos en aquellas variables que, en principio, informen sobre los aspectos de la solvencia y rentabilidad de la firma, sin olvidar la influencia del endeudamiento. A estas categorías añadiremos las de rotación, actividad y estructura de activo:

Atendiendo a la singularidad de las empresas con las que estamos trabajando realizaremos unas matizaciones en algunos de los ratios seleccionados, concretamente en los de endeudamiento y en los de rentabilidad por las razones que vamos a exponer a continuación.

Una característica que diferencia las sociedades cooperativas de las capitalistas es la libre adhesión y la consecuente variabilidad del capital social respetando el capital mínimo estatutario al alza, por la entrada de socios o a la baja por su salida.

Aunque tradicionalmente se ha establecido que el capital<sup>8</sup> de las cooperativas forma parte de los fondos propios y únicamente se configuraba como deuda ante la baja del socio cooperativo, la adaptación de la normativa contable al PGC 2007 y a las Normas Internacionales de Información Financiera (NIIFs) mediante la Orden EHA/3360/2010 establece que, en general, las aportaciones de los socios y otros partícipes a las cooperativas, en la medida que no otorguen a la sociedad el derecho incondicional a rehusar su reembolso, deberán calificarse como pasivo. De esta forma, podemos establecer que en una cooperativa hay dos tipos de socios, los socios-propietarios y los socios-acreedores. A tenor de esta razón y dado el periodo de tiempo que abarca nuestro estudio debemos tener en cuenta que en estos ejercicios el capital de las cooperativas puede haberse considerado como fondos propios en algunos casos y en otros, pensamos que los menos, haberse adaptado antes al cambio impulsado por las normas internacionales y haberse reclasificado en parte como fondos ajenos. Por ello, hemos considerado oportuno mantener la cifra de fondos propios que figura en la base de datos que nos ha servido para elaborar la muestra de estudio.

Referente a la rentabilidad y dada su relación directa con el resultado debemos realizar las siguientes matizaciones con vistas a su determinación en el tipo de empresas que nos ocupa:

Hay dos tipos de resultado en la empresa cooperativa:

1. El que se origina en las operaciones realizadas con los socios (excedente).
2. El que deriva de las operaciones realizadas con terceros no socios y de operaciones extraordinarias.

Desde el ejercicio 2011 la obtención del resultado contable de las cooperativas sigue prácticamente las mismas pautas que el resto de empresas y no es diferente del excedente cooperativo. Este hecho es importante cuando se analizan y estudian las cooperativas por medio de bases de datos, como es nuestro caso, ya que en la mayoría se presentan los datos en formatos similares al resto de empresas, sin tener en cuenta que en las cooperativas era diferente y que el resultado contable de la cuenta de pérdidas y ganancias no es lo mismo que el excedente de la cooperativa que aparece en los fondos propios del balance.

---

8. *La consideración particular de los recursos propios y ajenos en las empresas cooperativas puede ampliarse en Martin et al. (2007) y Bretos et al. (2018).*

Respecto a las ratios de rentabilidad no debemos olvidar que en las cooperativas la mejor rentabilidad de sus socios vendrá probablemente por la obtención de unos buenos precios de los productos y servicios cooperativizados. Dentro de la categoría de rentabilidad hemos considerado unas ratios de naturaleza mixta integrados por las magnitudes de beneficio en sus distintos niveles y de endeudamiento. También, en la categoría de rentabilidad se han integrado algunas ratios planteados con el *cash flow* en su acepción tradicional de recursos generados, es decir, beneficio + amortizaciones + provisiones.

En nuestro estudio se ha partido de la información contenida en los estados financieros de las sociedades cooperativas transformándola en ratios, lo que permitirá establecer con rigor su situación económico-financiera y realizar comparaciones entre las distintas empresas y diferentes periodos económicos. Además de las ratios clásicas en este tipo de estudios, se ha incorporado la variable tamaño, como el montante del activo empresarial. El uso de la variable tamaño expresado en formas diferentes es muy frecuente en los estudios de fracaso empresarial. En algunos trabajos, como Back (2005) y Turetsky y McEwen (2001) aparece expresado como logaritmo natural de la cifra de activo total. Otros autores, como Calvo y García (2006) se inclinan por considerar el criterio establecido por la Comisión Europea (1996) y, también hay autores, como Honjo (2000) que se han decantado por referenciar el tamaño al número de trabajadores de la firma

Por razones de tipo operativo no se ha recurrido a información de distinto origen, como la de naturaleza cualitativa.

La lista de las ratios considerados inicialmente y separados por categorías se muestra en la tabla 3:

**Tabla 3.** Ratios utilizados en el estudio empírico

CLAVE	RENTABILIDAD
REN 1	RESULTADO DE EXPLOTACION/ACTIVO
REN 2	RESULTADO DEL EJERCICIO/FONDOS PROPIOS
REN 3	CASH FLOW RECURSOS GENERADOS/NETO
REN 4	CASH FLOW RECURSOS GENERADOS/FONDOS PROPIOS
REN 5	CASH FLOW RECURSOS GENERADOS/ACTIVO

(\*) El resultado de actividades ordinarias es considerado antes de impuestos.

CLAVE	ESTRUCTURA FINANCIERA
END 1	PASIVO/FONDOS PROPIOS
END 2	GASTOS FINANCIEROS/PASIVO
END 3	GASTOS FINANCIEROS /VENTAS
END 4	PASIVO NO CORRIENTE/FONDOS PROPIOS
END 5	PASIVO CORRIENTE/PFONDOS PROPIOS

CLAVE	ACTIVIDAD
ACT 1	VALOR AÑADIDO/VENTAS
ACT 2	RESULTADO DEL EJERCICIO/VALOR AÑADIDO

CLAVE	ROTACION
ROT 1	VENTAS/ACTIVO

CLAVE	SOLVENCIA (Liquidez)
LIQ 1	ACTIVO CORRIENTE/PASIVO CORRIENTE
LIQ 2	ACTIVO CORRIENTE - EXISTENCIAS/PASIVO CORRIENTE
LIQ. 3	DISPONIBLE/PASIVO CORRIENTE

CLAVE	SOLVENCIA A LARGO PLAZO
SOLV LP 1	ACTIVO /PASIVO

CLAVE	ESTRUCTURA ECONOMICA
EE 1	ACTIVO CORRIENTE/ACTIVO

CLAVE	OTRAS VARIABLES
TAMAÑO	ACTIVO

**Fuente:** Elaboración propia.

Todas las partidas integrantes de las ratios han sido derivadas del Balance de Situación y cuenta de Pérdidas y Ganancias de las empresas que componen las diferentes muestras.

## 5. Metodología

Con la finalidad de clasificar correctamente la muestra disponible como empresas cooperativas en situación de activo o de concurso, se ha aplicado el algoritmo de Aprendizaje Automático *Extreme Gradient Boosting* (XGBoost). Las técnicas de Aprendizaje Automático proporcionan muy buenos resultados en la resolución de problemas, tales como la clasificación de spam en el correo, la detección del fraude, la identificación de eventos anómalos o la asignación de recursos publicitarios (Chen y Guestrin, 2016).

XGBoost es un algoritmo que está inspirado en el modelo del *Gradient Boosting* de Friedman (2001 y 2002), pero permite una implementación más eficiente y mejora la velocidad y resultados del algoritmo original. También incorpora un esquema de regularización que controla la posibilidad de obtener resultados sesgados que no puedan después generalizarse sobre muestras independientes. La regularización añade un factor de penalización que hace que la importancia de las diferentes variables tienda a cero, y así evitar la falta de generalización del modelo resultante. La penalización puede de hecho reducir considerablemente el problema del sobre-ajuste (*overfitting*, en inglés) de los modelos de Aprendizaje Automático, sin que ello conlleve una pérdida en su capacidad predictiva; también permite desarrollar modelos de predicción más simples, con menos variables (James et al., 2017).

Como recogen Kuhn y Johnson (2013), los algoritmos basados en la metodología *boosting* se empezaron a desarrollar a principios de la década de 1990, mediante la combinación de un número de clasificadores débiles (árboles de decisión) que producían un modelo de clasificación agregado o de conjunto con una ratio inferior de errores a los árboles de decisión individuales. Por tanto, y en palabras de Chambers y Dinsmore (2015), *boosting* combina o agrupa varios modelos individuales (en forma de árboles de decisión) en un modelo de conjunto, que se caracteriza porque la capacidad predictiva resultante es muy superior a la de los modelos individuales. La construcción de los modelos se realiza de modo secuencial, en base a los errores del modelo precedente; por tanto, cada árbol de decisión depende de los árboles anteriores y se ajusta a los residuos o errores que van quedando.

Elith et al. (2008), señalan que el modelo resultante agregado es una combinación de cientos o miles de árboles de decisión individuales, que se puede interpretar como un modelo de regresión donde cada término se corresponde con cada uno de los árboles de decisión individuales. La construcción de modelos de aprendizaje formados por múltiples árboles -propio de un modelo *boosting* agregado-, mejora considerablemente la capacidad predictiva de un árbol individual, permitiendo resolver el problema de la baja precisión de los árboles de decisión cuando se elaboran de forma individual.

Por otro lado, el algoritmo XGBoost hace uso de la estocasticidad (*stochastic gradient boosting*), que conlleva un esquema de muestreo aleatorio tanto de las variables como de las observaciones durante el proceso de construcción de los árboles de decisión individuales. La principal ventaja de la estocasticidad reside en el hecho de que incrementa la capacidad predictiva del algoritmo, al tiempo que reduce la varianza del modelo final -pues se selecciona un subconjunto aleatorio de datos para ajustar cada nuevo árbol-.

Otra característica interesante de este tipo de modelos basados en el algoritmo *stochastic gradient boosting*, es que son robustos ante la presencia de valores extremos y valores perdidos, permiten ajustar datos que no son lineales y pueden utilizar variables numéricas y categóricas, sin necesidad de realizar transformaciones previas. Además, las situaciones de multicolinealidad de las variables independientes no interfiere en el ajuste de los modelos (James et al., 2017).

El ajuste de un modelo con el algoritmo de aprendizaje automático *XGBoost* requiere de la especificación de un conjunto de parámetros reguladores de la complejidad de este, que se denominan hiper-parámetros. Sus valores óptimos dependerán de las características de los datos. Por consiguiente, resulta decisivo identificar los valores más adecuados de estos hiper-parámetros para construir un modelo *XGBoost* con la máxima capacidad predictiva. En nuestro estudio, se va a seguir la técnica de la validación cruzada para identificar la combinación óptima de éstos, pues permite estimar la precisión global que tendría el modelo en una muestra independiente o de comprobación. En particular, se hará uso de una validación cruzada de 10 submuestras, de modo que los datos de la muestra de entrenamiento de las sociedades cooperativas se segmentarán aleatoriamente en 10 grupos de tamaño idéntico. De este modo, se calculará 10 veces un error de precisión en cada una de los 10 submuestras de validación, que se promediará para estimar la precisión global del modelo en una muestra de comprobación independiente.

Los creadores de *XGBoost* explican de forma muy extensa y detallada el conjunto de hiper-parámetros propios de este algoritmo, necesarios para obtener un buen ajuste (Xgboost.readthedocs.io, 2019). Los parámetros más significativos son los que se detallan a continuación, cuya combinación óptima de los mismos intentaremos obtener, mediante la técnica de validación cruzada aludida en los párrafos anteriores, a partir de las características intrínsecas de los datos disponibles sobre empresas cooperativas:

- El número máximo de iteraciones del algoritmo.
- La profundidad máxima o tamaño de cada árbol de decisión individual.
- La ratio de aprendizaje o contribución de cada nuevo árbol construido en cada nueva iteración en el modelo.
- La cantidad mínima de disminución de la función de pérdida que se requiere para hacer una nueva partición en el árbol de decisión individual.
- El porcentaje aleatorio de variables que se consideran para la construcción de cada árbol de decisión.
- El porcentaje aleatorio de observaciones que se consideran para la construcción de cada árbol de decisión.
- El número mínimo de observaciones que contiene cada nodo del árbol para que pueda haber una nueva partición.

Para la construcción de los diferentes modelos de aprendizaje automático *XGBoost* y para la identificación de la combinación óptima de los hiper-parámetros que configuran este tipo de algoritmos, se ha utilizado el paquete de programación *R* (R Core Team, 2019) versión 3.5.3, junto con dos de sus librerías externas: *XGBoost* versión 0.82.1 (Chen et al., 2019) y *caret* versión 6.0-82 (Kuhn, 2019).

## 6. Resultados

En la tabla 4 se recogen los principales estadísticos descriptivos de las variables consideradas, diferenciando entre empresas cooperativas sanas y en situación de concurso. Como se ha indicado en el apartado anterior el algoritmo *XGBoost* es robusto ante posibles situaciones de multicolinealidad de las variables independiente, por tanto, se ha incluido en el análisis la totalidad de las variables. Asimismo, se aprecia que no se ha realizado ningún tipo de ajuste en las variables, como tipificar o normalizarlas para evitar valores perdidos o extremos, pues ya se ha señalado en líneas anteriores que este algoritmo no lo precisa.

Las diferencias más significativas las encontramos en las variables relativas a la solvencia y liquidez de las empresas, donde se observa que medias y medianas son mayores siempre en las empresas no fracasadas. Las mayores diferencias se dan en las variables SOLVLP1 y LIQ3, indicativas de la solvencia global y de la liquidez inmediata de la empresa, respectivamente. SOLVLP1 presenta unas diferencias en media de 2,27 vs 1,52 y mediana de 1,58 vs 1,32, mientras que LIQ3 0,51 vs 0,21 en media y 0,21 vs 0,09 en mediana.

También, encontramos diferencias que podrían parecer relevantes en algunas ratios de endeudamiento, donde se obtiene valores superiores en las medianas de las empresas fracasadas, mientras que en algunos casos es a la inversa en el caso de la media, si bien habría que recalcar que la mediana es mejor indicativo ya que es más robusta frente a valores extremos.

Finalmente, entre las variables de rentabilidad y generación de valor, destacan las diferencias que presenta REN4, indicativa del *cash flow* generado en relación con el tamaño de la empresa, que son mayores en las empresas no fracasadas (media 0,05 vs 0,01, y mediana 0,04 vs 0,03) y ACT1, referente al valor añadido en relación con las ventas, con mitad de mediana en empresas no fracasadas (0,28 vs 0,14).

Como se apreciará posteriormente el modelo resultante considerará estas variables entre las más relevantes para estimar el fracaso de la empresa.

**Tabla 4.** Resumen de los estadísticos descriptivos de las variables independientes (un año antes del fracaso)

Variable	media		d. e.		mediana		mínimo		máximo	
	NF	F	NF	F	NF	F	NF	F	NF	F
REN1	0,02	0,01	0,06	0,09	0,02	0,02	-0,19	-0,32	0,20	0,21
REN2	0,12	0,05	0,55	0,38	0,02	0,03	-0,22	-0,94	1,44	4,04
REN3	0,26	0,15	1,19	0,42	0,09	0,12	-0,45	-0,73	1,62	8,91
REN4	0,05	0,01	0,05	0,09	0,04	0,03	-0,04	-0,27	0,23	0,22
END1	7,60	4,79	38,21	8,84	1,41	2,45	-7,32	-6,54	49,63	257,40
END2	0,02	0,04	0,01	0,03	0,02	0,03	0,00	0,00	0,22	0,04
END3	0,06	0,02	0,32	0,02	0,00	0,02	0,00	0,00	0,13	2,14
END4	0,73	1,26	1,81	3,06	0,23	0,60	-2,77	-2,51	19,53	8,96
END5	5,60	3,35	33,10	7,35	0,76	1,67	-5,49	-5,13	49,55	248,45
ACT1	0,56	0,33	2,58	0,23	0,14	0,28	0,01	-0,07	0,86	19,47
ACT2	0,07	0,11	0,28	1,01	0,04	0,02	-0,85	-1,61	6,93	0,80
ROT1	2,07	1,40	2,17	0,84	1,51	1,33	0,01	0,00	4,22	11,82
LIQ1	2,05	2,04	1,98	4,04	1,46	1,21	0,18	0,31	29,50	10,45
LIQ2	1,44	1,01	1,09	0,88	1,19	0,84	0,11	0,07	5,21	5,49
LIQ3	0,51	0,21	0,81	0,40	0,21	0,09	0,00	0,00	2,39	4,69
SOLVLP	2,27	1,52	1,91	0,84	1,58	1,32	0,75	0,10	6,11	10,64
EE	0,56	0,64	0,27	0,24	0,59	0,64	0,06	0,15	1,01	0,98
TAMAÑO	362189	3989,24	475313	5329,98	1952,77	1816,01	56,10	52,38	2279782	20895,31

**Notas:** NF: Empresa no fracasada (56 observaciones) - F: Empresa fracasada (56 observaciones) - d. e.: desviación estándar  
La descripción de las variables se encuentra en la tabla 3. TAMAÑO (cifra de activos) se ha tomado en miles de euros.

## Ajuste del modelo XGBoost

Una de las características del Aprendizaje Automático es que para obtener un modelo que pueda generalizarse resulta necesario identificar la mejor combinación de los hiper-parámetros propios del modelo. En el apartado de este trabajo dedicado a la metodología se ha hecho referencia a estos parámetros del algoritmo *XGBoost*, indicando que para su identificación vamos a emplear la técnica de la validación cruzada. De este modo, esperamos desarrollar un modelo *XGBoost* generalizable, capaz de proporcionar buenos resultados en una muestra independiente diferente de la de entrenamiento. Hay que tener cuidado del riesgo de memorización o sobre-ajuste de los algoritmos de Aprendizaje Automático, que se presenta cuando en lugar de aprender de los datos memorizan su contenido, lo cual se puede detectar cuando los resultados en la muestra de entrenamiento son muy buenos, pero en cambio en la muestra de comprobación dejan bastante que desear.

Se va a tomar el indicador AUC (*Area Under the Curve*) o curva ROC (*Receiver Operating Characteristics*) para medir la capacidad predictiva del algoritmo *XGBoost*. En el Aprendizaje Automático, AUC se emplea habitualmente para valorar la bondad de un modelo de clasificación binario, teniendo en cuenta diferentes puntos de corte de probabilidad para considerar la pertenencia de una observación a un grupo o a otro. Informa pues del potencial que tiene el modelo ajustado para diferenciar entre las dos categorías de observaciones. Este indicador toma valores que van del cero al uno, y cuanto más próximo se encuentra de este último valor mejor. En nuestro caso, cuanto mayor sea el AUC obtenido entonces el modelo *XGBoost* ajustado será capaz de clasificar mejor las entidades cooperativas como en activo o en situación de concurso.

Por consiguiente, mediante la técnica descrita de validación cruzada, ajustaremos diferentes modelos *XGBoost* con distintas combinaciones de hiper-parámetros, y de todos estos modelos seleccionaremos aquél con una combinación específica de hiper-parámetros, que tenga la valoración más alta en el referido indicador AUC. De este modo, estaremos empleando el AUC como función de pérdida o de optimización del algoritmo.

La totalidad de los datos disponibles sobre entidades cooperativas se ha dividido en dos grupos. El primer grupo está formado por el 80% y constituye la muestra de entrenamiento, que alimentará el algoritmo *XGBoost* con el fin de identificar, con una validación cruzada de 10 submuestras, la mejor de las combinaciones de los hiper-parámetros. El 20% restante constituye la muestra independiente de comprobación, que se empleará para medir la capacidad predictiva del modelo resultante. Es importante destacar que esta última muestra permanece oculta durante todo el proceso de entrenamiento del algoritmo y que sólo se emplea al final para evaluar una sola vez si el modelo es generalizable.

La tabla 5 muestra los valores óptimos de los hiper-parámetros que se han obtenido con la validación cruzada de 10 subgrupos ( $k = 10$ ), después de maximizar el valor del indicador de bondad del modelo AUC, que alcanza un valor de 0,748. Este valor recoge la estimación de la capacidad predictiva del modelo *XGBoost* ajustado sobre una hipotética muestra independiente. Por otro lado, la precisión global del modelo estimada mediante la validación cruzada asciende a un 70%.

**Tabla 5.** Valores óptimos de los hiper-parámetros del modelo XGBoost

PARÁMETRO	VALOR
Número máximo de iteraciones del algoritmo	686
Profundidad máxima o tamaño de cada árbol	3
Ratio de aprendizaje o contribución de cada nuevo árbol	0,471
Cantidad mínima de disminución de la función de pérdida para realizar una nueva división en el árbol	7,36
Porcentaje aleatorio de variables en cada árbol	40,46%
Porcentaje aleatorio de observaciones en cada árbol	35,00%
Número mínimo de observaciones que contiene cada nodo del árbol	1

Nota: Los valores óptimos se han identificado por medio de la validación cruzada con un k=10.

## Variables más importantes del modelo XGBoost

La figura 1 muestra la ganancia de las variables más relevantes del modelo, cuanto más alto es el valor de la misma la contribución individual también es mayor. Las 4 primeras variables representan una contribución al modelo de aproximadamente el 50% (15 + 13 + 12 + 11). Se aprecia que las variables más importantes son SOLVLP.1, LIQ.3, ACT.1 y REN.4.

Las variables SOLVLP.1 y LIQ.3 son indicativas de la capacidad de la empresa de hacer frente a los pagos derivados de sus deudas. SOLVLP.1, definida como la ratio

$$\frac{\text{Activo}}{\text{Pasivo}}$$

nos valora la solvencia global de la empresa, mientras que LIQ.3, definida como la ratio

$$\frac{\text{Disponible}}{\text{Pasivo Corriente}}$$

es indicativa de la liquidez inmediata de la empresa. Por su parte, la variable ACT.1 definida como la ratio

$$\frac{\text{Valor añadido}^9}{\text{Ventas}}$$

nos mide el valor añadido con relación a las ventas y, finalmente, REN.4, que representa la relación

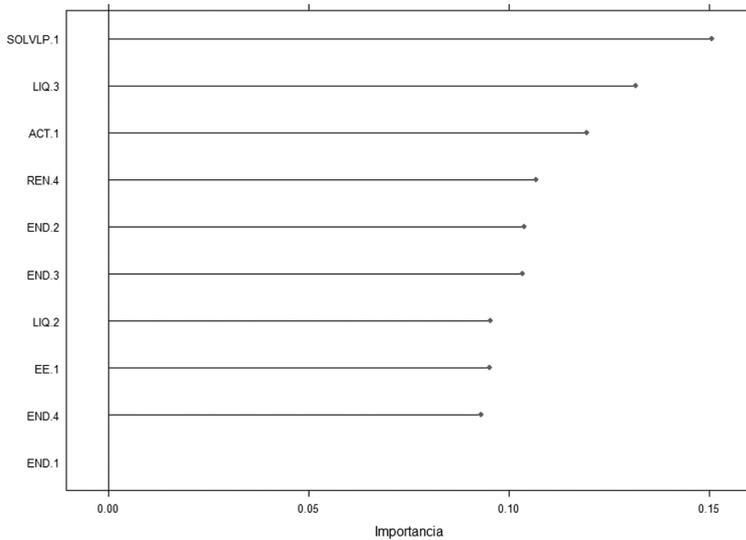
$$\frac{\text{Cash flow\_Recursos generados}^{10}}{\text{Activo}}$$

nos informa del *cash flow* generado por la empresa en relación con el tamaño de la empresa.

9. *Valor añadido* = Resultado antes de impuestos + Amortizaciones + Gastos financieros + Gastos de personal.

10. *Cash flow\\_Recursos generados* = Resultado del ejercicio + Amortizaciones + Fondos de deterioro + Provisiones.

**Figura 1.** Variables más importantes del modelo XGBoost



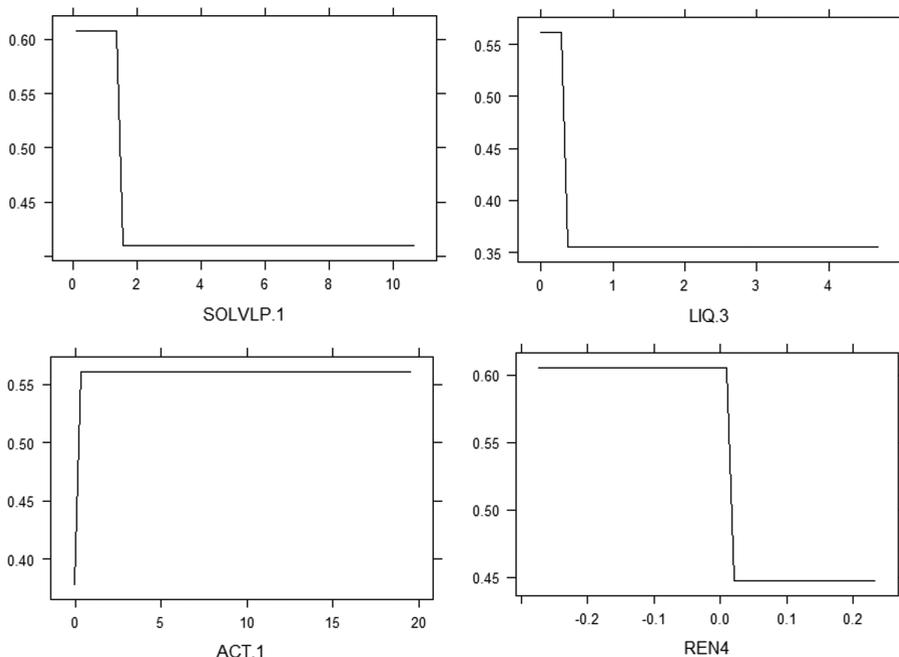
En la figura 2 recogemos los gráficos de dependencia parcial de las cuatro variables más importantes, que ilustran el efecto marginal de cada variable por separado en la probabilidad de que una entidad cooperativa pueda ser considerada como en situación de concurso. El eje de las abscisas muestra los valores que pueden tomar las variables y el eje de ordenadas la probabilidad de fracaso de la entidad cooperativa.

Las curvas descendentes de los gráficos correspondientes a las variables SOLVLP.1 y LIQ.3 indican, respectivamente, que conforme aumenta la solvencia y la liquidez inmediata de la empresa cooperativa disminuye la propensión a su fracaso. Estos dos resultados son plenamente coherentes, ya que si la cooperativa puede hacer frente a sus deudas es razonable que disminuyan las posibilidades de entrar en fases de dificultad financiera.

Atendiendo al gráfico ascendente que presenta la variable ACT.1 de la figura 2, que relaciona el valor añadido con la cifra de ventas, se observa que a medida incrementa su valor también aumentan las posibilidades de fracaso. La ratio ACT.1 puede incrementarse si lo hacen los gastos financieros o los de personal, al ser estas partidas dos componentes importantes del valor añadido, por lo que unos elevados importes de dichos gastos que no vayan acompañados de un paralelo incremento de las ventas, pueden favorecer la aparición de problemas en la empresa cooperativa.

Finalmente, el gráfico descendente de la variable REN.4 de la figura 2, indica que cuanto mayor es su valor, menor es la propensión al fracaso. Esta variable aumenta cuando crece el *cash flow* en proporción al activo de la empresa, por lo que resulta coherente que conforme una empresa tenga mejores resultados se reduzca la propensión al fracaso al ser más rentable.

**Figura 2.** Gráficos de dependencia parcial de las 4 variables más importantes



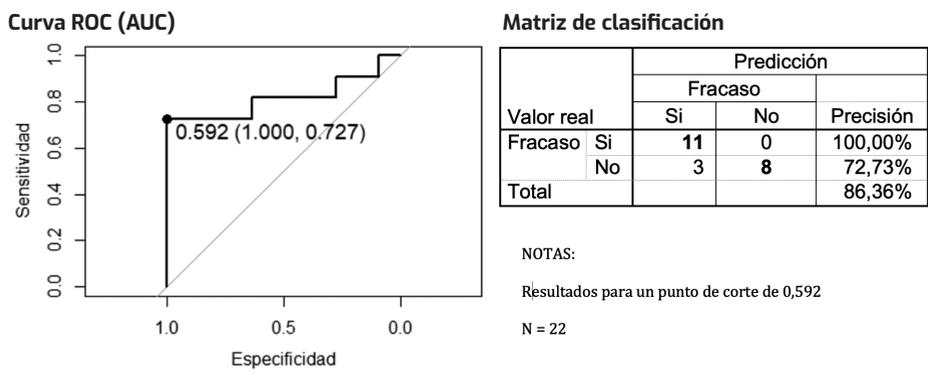
## Precisión del modelo en la muestra de validación

Ya se ha indicado que el modelo *XGBoost* se ha obtenido a partir del 80% de los datos (muestra de entrenamiento), estimándose una precisión global sobre una supuesta muestra independiente de un 70% -mediante validación cruzada con  $k=10$ . Ahora vamos a obtener el valor de la precisión global sobre el 20% de los datos restantes (muestra de comprobación) para comprobar su valor efectivo sobre una muestra real independiente. El panel derecho de la figura 3, la matriz de clasificación recoge que la precisión global del modelo en la muestra de comprobación es de un 86,36%, cifra superior al 70% que se había estimado durante el proceso de ajuste del algoritmo *XGBoost* con la validación cruzada. También se aprecia -en la muestra de comprobación- que la ratio de positivos verdaderos (sensitividad del modelo) es del 100%, por tanto, el modelo es capaz de identificar el 100% de las entidades cooperativas en situación de concurso; y que la ratio de negativos verdaderos (especificidad del modelo) es del 72,73%, de modo que el modelo identifica un porcentaje muy elevado de entidades cooperativas sanas.

En el panel izquierdo de la figura 3 se representa la curva ROC (AUC), apreciándose que el indicador de bondad del modelo, en la muestra de comprobación, asciende a 0,817. Un valor

tan elevado y próximo a uno, es indicativo de la alta capacidad de predicción que tiene el modelo *XGBoost* sobre una muestra independiente. Este valor es superior al que se había estimado anteriormente con la muestra de entrenamiento y haciendo uso de la validación cruzada, que era de 0,748. Todos estos resultados obtenidos sobre la muestra de comprobación, que no son peores que los alcanzados por medio de la estimación cruzada, aseguran que el modelo *XGBoost* resultante es generalizable y no presenta un problema de sobre-ajuste.

**Figura 3.** Precisión del modelo *XGBoost* en la muestra de comprobación



NOTAS: AUC = 0.8170. Precisión = 86,36%

## 7. Interpretabilidad del modelo *XGBoost*

En este apartado se va a mostrar cuáles son los criterios que sigue el modelo *XGBoost* ajustado para asignar a las diferentes observaciones disponibles de entidades cooperativas una determinada probabilidad de propensión al fracaso. En este sentido, se va a abrir la mal denominada caja negra de este tipo de algoritmos mostrando que, en realidad, de una forma similar a una regresión logística, estamos ante lo que podríamos denominar una caja blanca. Así pues, quedará perfectamente explicado cómo clasifica el modelo *XGBoost* cada una de las observaciones de entidades cooperativas como en activo o en situación de concurso, que componen la muestra de comprobación.

Como indica Foster (2017b), un modelo ajustado con el algoritmo *XGBoost* está formado por muchos árboles de decisión individuales, cada uno con un peso específico para conformar el valor de la probabilidad final de predicción de cada observación. Este conjunto tan numeroso de árboles es lo que proporciona la alta capacidad predictiva de este tipo de modelos, pero a costa de reducir la interpretabilidad de los mismos; pues resulta muy difícil determinar la influencia de cada variable en la obtención de las predicciones finales. La librería de *R XGBoost*

*Explainer* (Foster, 2017a), identifica el impacto que tiene cada variable en la predicción de cada una de las observaciones, haciendo que un modelo *XGBoost* sea tan interpretable como una regresión o un árbol de decisión individual.

En particular, para cada una de las entidades cooperativas que forman la muestra de comprobación, la probabilidad de concurso la descompondremos para mostrar la importancia por separado de cada una de las variables o ratios considerados. En realidad, como señala Foster (2017a), lo que se descompone es la proporción de probabilidades o *log-odds*, ya que las probabilidades que proporciona un modelo *XGBoost* son las de una función logística aplicada a los *log-odds*.

La figura 4 muestra la descomposición de los *log-odds* de dos entidades cooperativas fracasadas de la muestra de comprobación, clasificadas correctamente por el algoritmo *XGBoost* (Observaciones 1 y 3). Por su parte, la figura 5 recoge la descomposición de los *log-odds* de dos entidades cooperativas clasificadas correctamente por el algoritmo *XGBoost* como no fracasadas, también de la muestra de comprobación (Observaciones 20 y 21). La probabilidad de fracaso que proporciona el modelo aparece en el eje de las ordenadas y las ratios en el eje de las abscisas, mientras que los *log-odds* se encuentran dentro de las barras verticales.

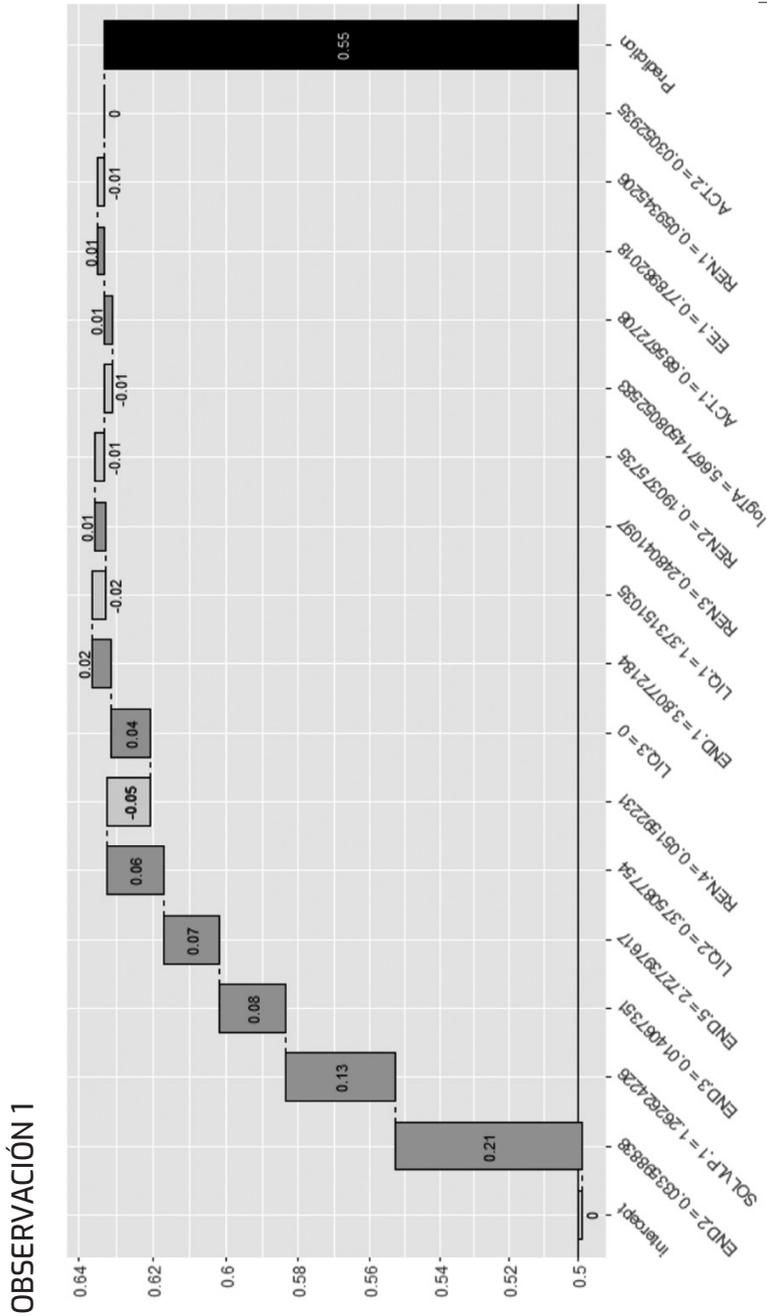
Por ejemplo, para la observación 1 de la figura 4 (entidad cooperativa fracasada), la predicción de fracaso con una probabilidad del 63% (0,633) se descompone para recoger la influencia individual de cada variable: tanto el signo (el color verde recoge una influencia positiva hacia el fracaso y el rojo lo contrario) como la importancia (tamaño de cada una de las barras). Se aprecia que la predicción de los *log-odds* de esta cooperativa que ha entrado en concurso de acreedores, que asciende a 0,546, va variando en función del efecto de las diferentes variables como sigue:

- + 0: Modelo base
- + 0,21: END 2 [la predicción ahora es de 0,21]
- + 0,13: SOLVLP 1 [la predicción es ahora de 0,34]
- + 0,08: END 3 [la predicción es ahora de 0,42]
- + 0,07: END 5 [la predicción ahora es de 0,49]
- + 0,06: LIQ 2 [la predicción ahora es de 0,55]
- 0,05: REN 4 [la predicción ahora es de 0,50]
- + 0,04: LIQ 3 [la predicción ahora es de 0,54]
- + 0,02: END 1 [la predicción ahora es de 0,56]

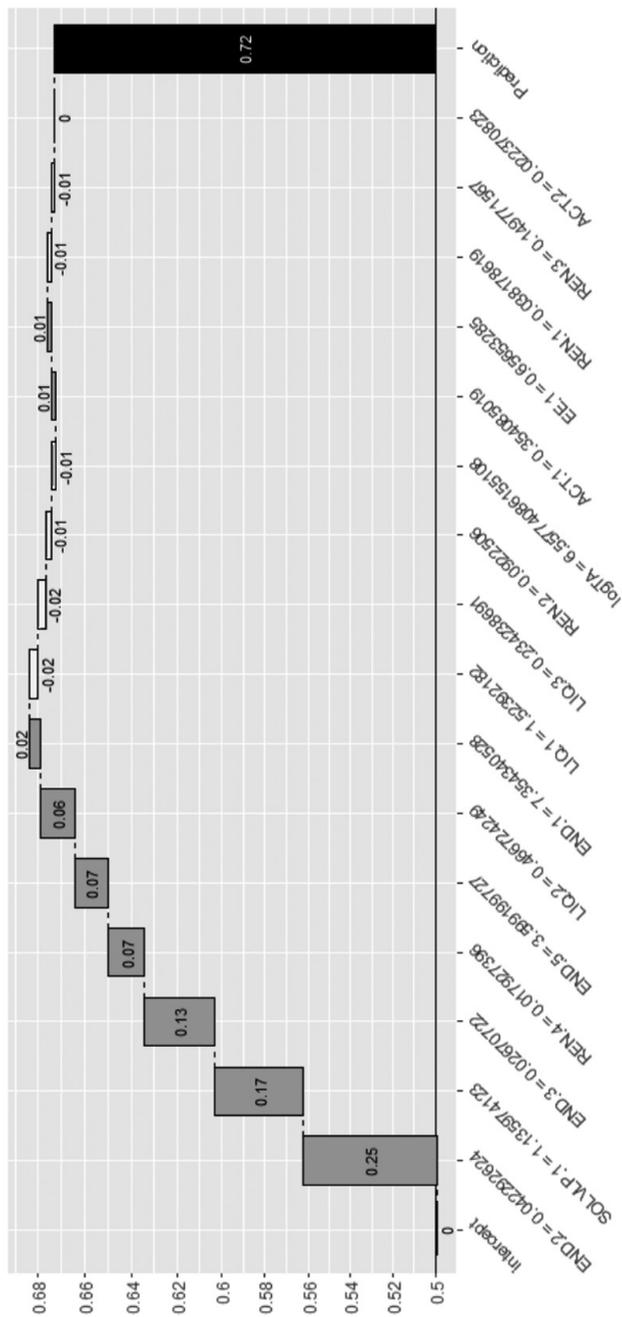
Y, a partir de ahí, faltaría añadir y restar el resto de los *log-odds* de las variables con menor influencia hasta llegar a un *log-odds* final de 0,546, que se corresponde con una probabilidad de fracaso del 63% (0,633).

También se aprecia que, para todas las observaciones, el modelo de descomposición de la probabilidad de fracaso parte siempre de una cifra muy próxima a 0,5 pues la muestra está equilibrada, en el sentido de que tanto las empresas cooperativas sanas como fracasadas representan un 50% de la muestra. Por consiguiente, la probabilidad de fracaso de un modelo base es prácticamente del 50% (0,5).

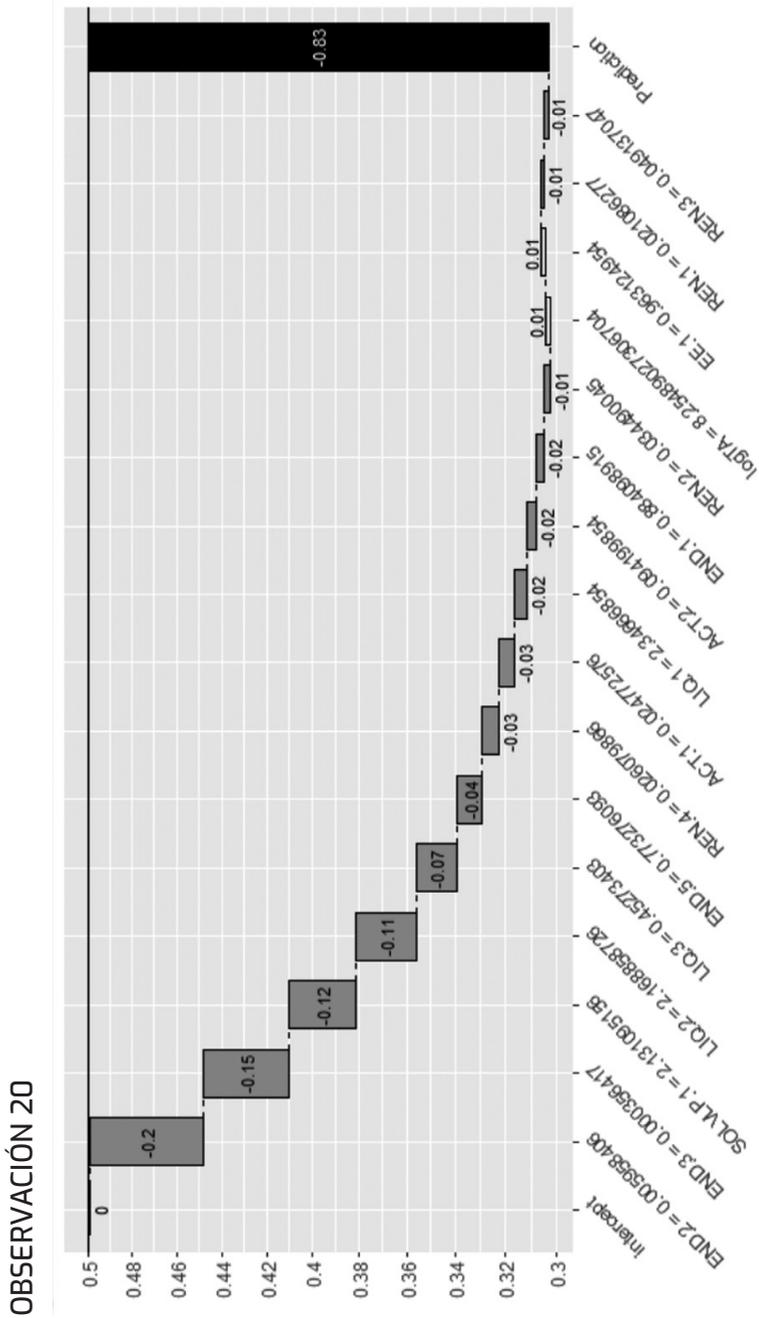
**Figura 4.** Descomposición de la predicción del modelo XGBoost para dos entidades clasificadas correctamente como fracasadas

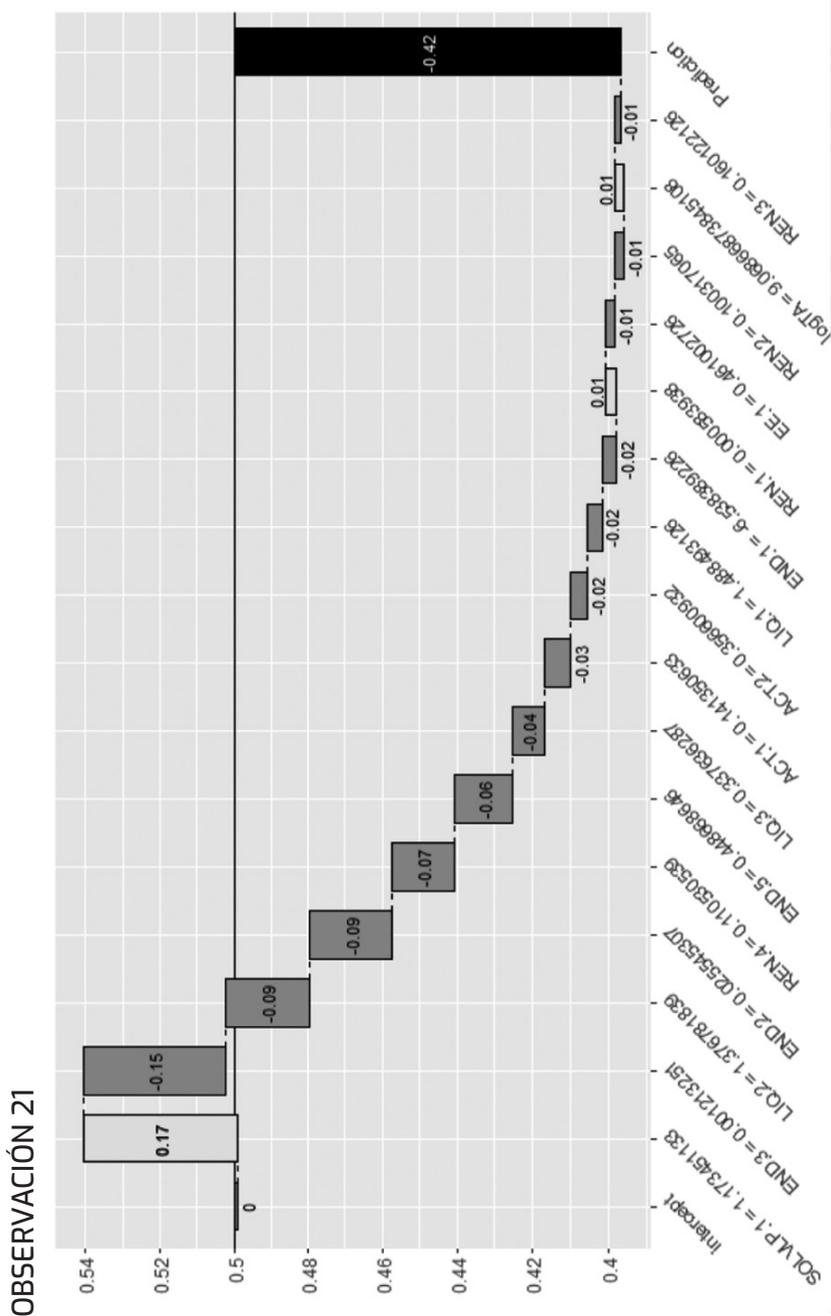


OBSERVACIÓN 3



**Figura 5.** Descomposición de la predicción del modelo XGBoost para dos entidades clasificadas correctamente como no fracasadas





## 8. Conclusiones

Las cooperativas son sociedades de marcado carácter social con unas singularidades tanto normativas como de funcionamiento que las diferencian del resto de sociedades mercantiles del tejido empresarial. Las relaciones con los trabajadores, la remuneración del capital aportado y la forma como se reparten los resultados las configuran como sociedades de un profundo carácter social aun admitiendo también su vertiente mercantil.

La muestra de trabajo ha estado integrada en su totalidad por empresas cooperativas de diversos sectores, excepto del financiero y seguros.

Para el desarrollo del estudio empírico nos hemos servido de la metodología basada en el algoritmo de aprendizaje automático *Extreme Gradient Boosting (XGBoost)* que ha permitido identificar las variables más relevantes que anticipan las situaciones de dificultad financiera, así como la interpretación de los resultados de una forma totalmente transparente, como si de una regresión clásica se tratara.

Los resultados derivados del estudio estadístico indican la relevancia conjunta de las variables indicativas de las categorías de solvencia, liquidez, endeudamiento y rentabilidad. Estas se muestran como las más decisivas a la hora de detectar la situación de fracaso empresarial que alcanza un porcentaje de acierto del 100 % en la muestra de entrenamiento y de un 86% en una muestra de comprobación independiente.

El modelo combina eficazmente de las variables SOLVLP.1, LIQ.3, ACT.1 y REN.4, las dos primeras indicativas de la capacidad de la empresa de hacer frente a los pagos derivados de sus deudas. El estudio confirma que a medida que aumenta la solvencia y la liquidez inmediata de la empresa cooperativa disminuye la propensión a su fracaso. Por su parte, el movimiento de la variable ACT.1, que mide la relación el valor añadido con las ventas, indica que si su valor aumenta las probabilidades de fracaso lo hacen también, al verse involucrados en estos movimientos los gastos financieros y los de personal. Finalmente, REN.4, que informa del *cash flow* generado por la empresa en relación con el tamaño de la empresa, aumenta cuando crece el *cash flow* en proporción al activo de la empresa, por lo que resulta coherente que conforme una empresa tenga mejores resultados se reduzca la propensión al fracaso al ser más rentable.

Estos resultados son coherentes con la evidencia empírica, por lo que entendemos que la técnica aplicada es perfectamente válida para los objetivos perseguidos, que no son otros que la detección del fracaso empresarial en empresas del sector cooperativo.

Tal vez la variable menos homogeneizable sea la rentabilidad, pues en una cooperativa el resultado contable no es comparable con el resto de empresas mercantiles al no tener en cuenta los posibles beneficios de sus socios vía precios de los productos y servicios cooperativizados.

El modelo resultante, además de clasificar a las sociedades cooperativas en fracasadas o no fracasadas, se ha mostrado valido para identificar los motivos de dicha clasificación de forma local. Es decir, para explicar la composición de la probabilidad final de fracaso que proporciona el algoritmo del modelo *XGBoost* para cada observación de la muestra de comprobación, separando el peso individual de cada uno de los predictores, al tiempo que recoge su signo. De este modo, se constata que los resultados demuestran una alta capacidad predictiva y son

totalmente interpretables (es posible determinar el signo y el efecto de cada variable sobre la predicción final de todas y cada una de las observaciones).

Finalmente, nos gustaría conferirle al estudio un carácter de continuidad pues consideramos muy interesante volver a realizar este tipo de análisis extendiéndolo a una serie de varios años con objeto de plantear un modelo capaz de detectar el fracaso con varios años de antelación.

## Bibliografía

ALFARO, E., GARCÍA, N., GÁMEZ, M. & ELIZONDO, D. (2008): "Bankruptcy Forecasting: An Empirical Comparison of AdaBoost and Neural Networks", *Decision Support Systems*, 45(1), 110-122. DOI: 10.1016/j.dss.2007.12.002.

ALFARO CORTÉS, E., GÁMEZ MARTÍNEZ, M. & GARCÍA RUBIO, N. (2008): "Linear discriminant analysis versus AdaBoost for failure forecasting", *Revista Española de Financiación y Contabilidad*, 37(137), 13-32. DOI: 10.1080/02102412.2008.10779637.

BACK, P. (2005): "Explaining financial difficulties based on previous payment behavior, management background variables and financial ratios", *European Accounting Review*, 14(4), 839-868. DOI: 10.1080/09638180500141339.

BELL, T.B., RIBAR, G.S. & VERCHIO, J. (17-18 de mayo de 1990): "Neural Nets Versus Logistic Regression: A Comparison of Each Model's Ability to Predict Commercial Bank Failures" [Comunicación en congreso], En: Srivastava, R.P., *Proceedings of the 1990 Deloitte & Touche/University of Kansas Symposium in Auditing Problems*, 29-53. Auditing Symposium X. Universidad de Kansas, Lawrence.

BORITZ, J.E. & KENNEDY, D.B. (1995): "Effectiveness of Neuronal Network Types for prediction of Business Failure", *Experts Systems with Applications*, 9(4), 503-512. DOI: 10.1016/0957-4174(95)00020-8.

BRETOS, I., DÍAZ-FONCEA, M., MARCUELLO, C. & MARCUELLO, C. (2018): "Cooperativas, capital social y emprendimiento: Una perspectiva teórica", *REVESCO, Revista de Estudios Cooperativos*, 128, 76-98. DOI: 10.5209/REVE.59775.

CALVO-FLORES, A., GARCÍA, D. & MADRID, A. (18-20 de octubre de 2006). *Tamaño, Antigüedad y Fracaso Empresarial* [Comunicación en congreso]. V Workshop de Investigación Empírica en Contabilidad Financiera, Universidad Autónoma de Madrid, Madrid.

CARMONA, P., CLIMENT F. & MOMPALER A. (2019): "Predicting bank failure in the U.S. banking sector: An extreme gradient boosting approach", *Internacional Review of Economic and Finance*, 61, 304-323. DOI: 10.1016/j.iref.2018.03.008.

CARRERAS ROIG, L. (2011): "Consideraciones en torno al posible carácter mercantil de las sociedades cooperativas y acerca de las limitaciones a la realización de operaciones con terceros no socios de estas entidades", *REVESCO, Revista de Estudios Cooperativos*, 106, 55-73. DOI: 10.5209/rev\_REVE.2011.v106.37378.

CHAMBERS, M. & DINSMORE, T.W. (2015): *Advanced Analytics Methodologies: Driving Business Value with Analytics*, Pearson Education, Londres.

CHAUDHURI, A. & GHOSH, S.K. (2017): *Bankruptcy Prediction through Soft Computing based Deep Learning Technique*, Springer Singapore, Singapur.  
<https://doi.org/10.1007/978-981-10-6683-2>.

CHEN, T. & GUESTRIN, C. (2016): "XGBoost. A Scalable Tree Boosting System". En: ACM SIGKDD & ACM SIGMOD (2016), *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. Association for Computing Machinery, Nueva York. DOI: 10.1145/2939672.2939785.

CHEN, T., HE, T., BENESTY, M., KHOTILOVICH, V., TANG, Y., CHO, H., CHEN, K. et al. (2019): *XG-Boost: eXtreme Gradient Boosting. R package version 0.82.1*, 785-794. Recuperado el 30 de junio de 2019. <https://CRAN.R-project.org/package=xgboost>.

CLIMENT F., MOMPARDER A. & CARMONA P. (2019): "Anticipating bank distress in the Eurozone: An extreme gradient boosting approach", *Journal of Business Research*, 101, 885-896. DOI: 10.1016/j.jbusres.2018.11.015.

DÍAZ MARTÍNEZ, Z., FERNÁNDEZ MENÉNDEZ, J. & SEGOVIA VARGAS, 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, Madrid. <http://eprints.ucm.es/6833/>.

DIÉGUEZ SOTO, J., TRUJILLO ARANDA, F. & CISNEROS, A.J. (9 y 10 de noviembre de 2006): "Modelos de predicción de la insolvencia empresarial: la incorporación de ratios a partir de un marco teórico". En: Carrasco, A., *Riesgos, regulación bancaria y derecho concursal. VI Jornadas sobre Predicción del Insolvencia Empresarial*. Carmona, España.

DIETRICH, J., ARCELUS, F.J. & SRINIVASAN, G. (2005): "Predicting financial failure: some empirical evidence from New Brunswick agricultural co-ops", *Annals of Public and Cooperative Economics*, 76, 179-194. DOI: 10.1111/j.1370-4788.2005.00275.x.

ELITH, J., LEATHWICK, J.R. & HASTIE, T. (2008): "A working guide to boosted regression trees", *Journal of Animal Ecology*, 77, 802-813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>.

FLETCHER, D. & GOSS, E. (1993): "Application Forecasting with Neural Networks: An Application Using Bankruptcy Data", *Information and Management*, 24, 159-167. DOI: 10.1016/0378-7206(93)90064-Z.

FOSTER, D. (2017a): *NEW R package that makes XGBoost interpretable*. Recuperado el 30 de junio de 2019, <https://medium.com/applied-data-science/new-r-package-the-xgboost-explainer-51dd7d1aa211>.

FOSTER, D. (2017b): *xgboostExplainer: XGBoost Model Explainer. R package version 0.1*. Recuperado el 30 de junio de 2019. <https://github.com/AppliedDataSciencePartners/xgboostExplainer>.

FRIEDMAN, J.H. (2001): "Greedy function approximation: A gradient boosting machine", *The Annals of Statistics*, 29(5), 1189-1232, <http://www.jstor.org/stable/2699986>.

FRIEDMAN, J.H. (2002): "Stochastic gradient boosting", *Computational Statistics and Data Analysis*, 38(4), 367-378. DOI: 10.1016/S0167-9473(01)00065-2.

HONJO, Y. (2000): "Business failure of new software firms", *Applied Economic Letters*, 7(9), 575-579. DOI: 10.1080/13504850050059032.

ITURRIOZ, J. (2010): "Los procesos concursales en situación de crisis: características de su aplicación a las sociedades cooperativas", *REVESCO, Revista de Estudios Cooperativos*, 100, 134-159, <https://www.redalyc.org/articulo.oa?id=36712366005>.

ITURRIOZ J. & MARTÍN, S. (2013): "El tamaño como elemento determinante de la insolvencia en las sociedades cooperativas: estudio a partir de los procesos concursales", *REVESCO, Revista de Estudios Cooperativos*, 111, 90-107, <http://hdl.handle.net/10637/8311>.

JAMES, G., WITTEN, D., HASTIE, T. & TIBSHIRANI, R. (2017): *An Introduction to Statistical Learning*, Springer, Cham.

JAYANTHI, J., KAUR, G. & SURESH, K. (2017): "Financial forecasting using decision tree (reptree y c4.5) and neural networks (k\*) for handling the missing values", *ICTAC Journal on soft computing*, 7(3), 1473-1477. DOI: 10.21917/ijsc.2017.0204.

JUSTE, J.J., GÓMEZ, J.M. & FERNÁNDEZ, J.E. (2011): "Economía social y desarrollo local/rural. Un análisis de sus sinergias", *Estudios de Economía Aplicada*, 29(1), 189-222, <https://www.redalyc.org/articulo.oa?id=301/30120835007>.

KIM, M.J. & KANG, D.K. (2010): "Ensemble with Neural Networks for Bankruptcy Prediction", *Expert System with Applications*, 37(4), 3373-3379. DOI: 10.1016/j.eswa.2009.10.012.

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. DOI: 10.1016/j.eswa.2014.08.025.

KIM, S.Y. & UPNEJA, A. (2014): "Predicting Restaurant Financial Distrees Using Decision Tree and AdaBoosted Decision Tree Models", *Economic Modelling*, 36, 354-362. DOI: 10.1016/j.econmod.2013.10.005.

KUHN, M. & JOHNSON, K. (2013): *Applied predictive modeling*, Springer, Cham.

KUHN, M. et al. (2019): *Caret: Classification and Regression Training. R package version 6.0-82*. Recuperado el 30 de junio de 2019. <https://cran.rproject.org/web/packages/caret/index.html>.

LABATUT, G., POZUELO, J. & VERES, E. (2009): "Modelización temporal de los ratios contables en la detección del fracaso empresarial de la PYME española", *Revista Española de Financiación y Contabilidad*, 38(143), 423-447. <https://doi.org/10.1080/02102412.2009.10779672>.

LÓPEZ, F.J. & PASTOR, I. (2015): "Bankruptcy visualization and prediction using neural networks: A study of U.S. comercial banks", *Experts Systems with Aplicacions*, 42(6), 2857-2869. DOI: 10.1016/j.eswa.2014.11.025.

MARÍ, S., MARÍN, M. SEGUÍ, E. & MICHAEL-ZAMORANO, M. (2014): "Análisis cualitativo de los determinantes en la predicción del fracaso empresarial en cooperativas agroalimentarias", *ITEA-Información Técnica Económica Agraria*, 110, 300-320. DOI: 10.12706/itea.2014.019.

MARTÍN, S., LEJARRIAGA, G. & ITURRIOZ, J. (2007): "La naturaleza del capital social como aspecto diferenciador entre las sociedades cooperativas y las sociedades laborales", *CIRIEC-España, Revista de economía pública, social y cooperativa*, 58, 59-82. <https://www.redalyc.org/articulo.oa?id=17405803>.

MASA, C.I, ITURRIOZ, J. & MARTÍN, S. (2016): "Aspectos determinantes del fracaso empresarial: efecto de la proyección social de las cooperativas frente a otras formas jurídicas", *CIRIEC-España, Revista Española de Economía Pública, Social y Cooperativa*, 88, 93-125. <https://www.redalyc.org/articulo.oa?id=17449696004>.

MATEOS, A., MARÍN, M.M., MARÍ, S. & SEGUÍ, E. (2011): "Los modelos de predicción del fracaso empresarial y su aplicabilidad en cooperativas agrarias", *CIRIEC-España, Revista de Economía Pública, Social y Cooperativa*, 70, 179-208. <https://www.redalyc.org/articulo.oa?id=17418829007>.

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. DOI: 10.1080/02102412.2015.1118903.

ODOM, M.D. & SHARDA, R. (17-21 de junio de 1990): "A neural network model for bankruptcy prediction" [Comunicación en congreso], *1990 IJCNN International Joint Conference on Neural Networks 2*, San Diego, EE.UU, 163-168. DOI: 10.1109/IJCNN.1990.137710.

POPESCU, M.E., ANDREICA, M. & POPESCU, I.P. (2-4 de noviembre de 2017): "Decision support solution to business failure prediction", *Proceedings of the International Management Conference*, 11(1), 99-106. 11<sup>th</sup> International Management Conference. Faculty of Management, Academy of Economic Studies, Bucarest, Rumania.

POZUELO, J., CARMONA, P. & MARTÍNEZ, J. (2012): "Las sociedades cooperativas y las empresas capitalistas en la Comunidad Valenciana: Análisis comparado de su estructura económica y financiera", *CIRIEC-España, Revista Española de Economía Pública, Social y Cooperativa*, 74, 117-147. <https://www.redalyc.org/articulo.oa?id=17423124005>.

POZUELO, J., LABATUT, G. & VERES, E. (2010): "Análisis descriptivo de los procesos de fracaso empresarial en microempresas mediante técnicas multivariantes", *Revista Europea de Dirección y Economía de la Empresa*, 19(3), 47-66.

POZUELO, J., MARTÍNEZ, J. & CARMONA, P. (2012): "Estudio de la insolvencia empresarial en las cooperativas mediante técnicas multivariantes", *Estudios de Economía Aplicada*, 3, 1-23. <https://www.redalyc.org/articulo.oa?id=30125274015>.

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. DOI: 10.1080/02102412.2018.1442039.

R CORE TEAM (2019): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Recuperado el 30 de junio de 2019. <http://www.R-project.org/>.

RAVI KUMAR, P. & RAVI, V. (2007): "Bankruptcy Prediction in Banks and Firms Via Statistical and Intelligent Techniques - A Review", *European Journal of Operational Research*, 180(1), 1-28. DOI: 10.1016/j.ejor.2006.08.043.

Real Decreto Legislativo 1/2020 de 2020, por el que se aprueba el texto refundido de la Ley Concursal. 5 de mayo de 2020. BOE-A-2020-4859.

Reglamento (UE) 651/2014 de la Comisión, por el que se declaran determinadas categorías de ayudas compatibles con el mercado interior en aplicación de los artículos 107 y 108 del Tratado. 17 de junio de 2014. Diario oficial de la Unión Europea, de 26 de junio de 2014.

ROMERO, F. (2013): "Alcances y limitaciones de los modelos de capacidad predictiva en el análisis del fracaso empresarial", *AD-minister*, 23, 45-70. <https://www.redalyc.org/articulo.oa?id=322329207004>.

SUN, J., LI, H., HUANG, Q.-H. & HE, K.Y. (2014): Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches. *Knowledge-Based Systems*, 57, 41-56. DOI: 10.1016/j.knsys.2013.12.006.

TAM, K.Y. (1991): "Neural Network Models and the Prediction of Bank Bankruptcy", *Omega*, 19(5), 429-445. DOI: 10.1016/0305-0483(91)90060-7.

TAM, K.Y. & KIANG, M.Y. (1992): "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions", *Management Science*, 38(7), 926-947. DOI: 10.1287/mnsc.38.7.926.

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. DOI: 10.1016/S1138-4891(12)70037-7.

TSENG, F.M. & HU, Y.C. (2010): "Comparing for bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks", *Experts Systems with Applications*, 37(3), 1846-1853. DOI: 10.1016/j.eswa.2009.07.081.

TURETSKY, H.F. & McEWEN, R.A. (2001): "An Empirical Investigation of Firm Longevity: A Model of the Ex Ante Predictors of Financial Distress", *Review of Quantitative Finance and Accounting*, 16(4), 323-343. DOI: 10.1023/A:1011291425075.

VARGAS VASSEROT, C. (2010): "El derecho de reembolso del socio en caso de baja y el concurso de las sociedades cooperativas", *CIRIEC-España, Revista de Economía Pública, Social y Cooperativa*, 21, 37-58.

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. DOI: 10.1016/j.eswa.2013.09.033.

WILSON, G.I. & SHARDA, R. (1994): "Bankruptcy prediction using neural network", *Decision support systems*, 11(5), 545-557. DOI: 10.1016/0167-9236(94)90024-8.

XGBOOST.READTHEDOCS.IO (2019): *XGBoost Parameters - xgboost 1.0.0*. Recuperado el 30 de junio de 2019. <https://xgboost.readthedocs.io/en/latest/parameter.html>.

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. DOI: 10.1016/S0377-2217(98)00051-4.

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, 93-101. DOI: 10.1016/j.eswa.2016.04.001.