



UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Programa de Doctorado en Desarrollo Integral e
Innovación de Destinos Turísticos.

Tesis Doctoral

**Aplicación de Inteligencia
Artificial en la detección
de problemas financieros**

JESÚS MARÍA GONZÁLEZ MARTÍN

Las Palmas de Gran Canaria, junio 2017

Director: **Dr. Agustín J. Sánchez Medina**

D. JERÓNIMO PÉREZ ALEMÁN, VICEDECANO DE POSGRADO Y FORMACIÓN CONTINUA DE LA FACULTAD DE ECONOMÍA, EMPRESA Y TURISMO DE LA UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA, EN CALIDAD DE PRESIDENTE DE LA COMISIÓN DE DOCTORADOS EN EXTINCIÓN

CERTIFICA,

Que la citada Comisión en su sesión de fecha 9 de junio de 2017 tomó el acuerdo de dar el consentimiento para su tramitación a la tesis doctoral titulada “**APLICACIÓN DE INTELIGENCIA ARTIFICIAL EN LA DETECCIÓN DE PROBLEMAS FINANCIEROS**” presentada por el doctorando **D. Jesús María González Martín** y dirigida por el Doctor **Agustín Jesús Sánchez Medina**.

Y para que así conste, y a efectos de lo previsto en el Artº 6 del Reglamento para la elaboración, defensa, tribunal y evaluación de tesis doctorales de la Universidad de Las Palmas de Gran Canaria, firmo la presente en Las Palmas de Gran Canaria, a 9 de junio de dos mil diecisiete.



Jerónimo Pérez Alemán



UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Programa de Doctorado en Desarrollo Integral e
Innovación de Destinos Turísticos.

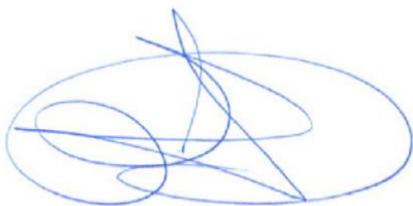
Facultad de Economía, Empresa y Turismo

Aplicación de Inteligencia Artificial en la
detección de problemas financieros

Tesis doctoral presentada por: D. Jesús María González Martín

Dirigida por el Dr. D. Agustín J. Sánchez Medina

El Director



El Doctorando



Las Palmas de Gran Canaria, junio del 2017

a Fátima, Marta, Inés, Asier
y mis padres

Agradecimientos

En primer lugar, me gustaría darle las gracias a Fátima por los esfuerzos que ha realizado para que yo pudiera sacar adelante este proyecto personal, sin su ayuda, no hubiese sido posible.

A mi hijo Asier, me gustaría darle las gracias por ser como es y decirle que el próximo proyecto va a ser que él aprenda a andar en bicicleta.

A Marta, Inés y Asier, me gustaría agradecerles su paciencia y decirles que tienen que aprender a esforzarse para que sus sueños se conviertan en realidad.

A mis padres, por lo todo lo que han hecho por mis hermanos y por mí durante todos estos años. Y a mis hermanos, por esos buenos momentos vividos y por todo lo aprendido junto a ellos.

Y como no podía faltar, a los “mortas”, a pesar de la distancia, saber que siempre están presentes.

Por último, pero no por ello menos importante, a mi director de tesis y las personas que han colaborado. A Agustín Sánchez, por confiar en mí y hacer este proyecto realidad y a Jesús B. Alonso y Félix Blázquez por realizar aportaciones decisivas y de esta forma, poner la guinda del proyecto.

ÍNDICE GENERAL

INTRODUCCIÓN Y OBJETIVO DE INVESTIGACIÓN	17
CAPÍTULO 1: APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA LA PREDICCIÓN DEL ESTRÉS FINANCIERO EN EMPRESAS DEL SECTOR DE RECOGIDA Y TRATAMIENTO DE AGUAS RESIDUALES.....	27
1.1 RESUMEN	29
1.2 ABSTRACT	30
1.3 INTRODUCCIÓN	31
1.4 REVISIÓN DE LA LITERATURA	35
1.5 METODOLOGÍA.....	41
1.5.1 Datos	41
1.5.2 Técnicas de Inteligencia Artificial utilizadas	42
1.5.2.1 Support Vector Machine	42
1.5.2.2 Fuzzy Logic	43
1.5.2.3 Self-organizing maps (SOM)	43
1.5.2.4 Classifier ensembles	44
1.5.2.5 Técnica de los Algoritmos Genéticos	45
1.5.3 Análisis estadístico	46
1.5.4 Medidas de Rendimiento.....	49
1.6 RESULTADOS.....	49
1.7 DISCUSIÓN	55
1.8 REFERENCIAS.....	56
CAPÍTULO 2: OPTIMIZACIÓN DE LA PREDICCIÓN DE PROBLEMAS FINANCIEROS EN EMPRESAS SANITARIAS PRIVADAS ESPAÑOLAS APLICANDO ALGORITMOS GENÉTICOS.....	65
2.1 RESUMEN	67
2.2 ABSTRACT	68
2.3 INTRODUCCIÓN Y REVISIÓN DE LA LITERATURA	69
2.4 METODOLOGÍA.....	74
2.4.1 Datos	74
2.4.2 Técnica de los Algoritmos Genéticos.....	75
2.4.3 Datos no balanceados.....	76
2.4.4 Análisis estadístico	76
2.4.5 Medidas de rendimiento.....	78
2.5 RESULTADOS.....	80
2.6 DISCUSIÓN	83
2.7 REFERENCIAS.....	84
CAPÍTULO 3: PREDICCIÓN DEL ESTRÉS FINANCIERO EN EMPRESAS DEL SECTOR PRIVADO DEL AGUA, UNA APROXIMACIÓN DESDE EL PUNTO DE VISTA DE LA INTELIGENCIA ARTIFICIAL	93
3.1 RESUMEN	95
3.2 ABSTRACT	96
3.3 INTRODUCCIÓN	97
3.4 REVISIÓN DE LA LITERATURA	100
3.5 METODOLOGÍA.....	105

3.5.1	<i>Datos</i>	105
3.5.2	<i>Técnicas de Inteligencia Artificial utilizadas</i>	106
3.5.2.1	Redes Neuronales (NN)	106
3.5.2.2	Support Vector Machine	107
3.5.2.3	Decision Tree J48	108
3.5.2.4	Random Forest	108
3.5.2.5	Técnica de los Algoritmos Genéticos	109
3.5.3	<i>Análisis estadístico</i>	110
3.5.4	<i>Medidas de Rendimiento</i>	113
3.6	RESULTADOS	113
3.7	DISCUSIÓN	115
3.8	REFERENCIAS	116
CONCLUSIONES		125
4.1	CONCLUSIONES	127
4.2	LIMITACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	128
4.3	REFERENCIAS	129

ÍNDICE DE TABLAS

TABLA 1.1 EJEMPLO DE RATIOS FINANCIEROS, SUS CATEGORÍAS Y REFERENCIAS	36
TABLA 1.2 ESTADÍSTICA DESCRIPTIVA DE LAS RATIOS FINANCIEROS Y SELECCIÓN DE VARIABLES	47
TABLA 1.3 MATRIZ DE CONFUSIÓN	49
TABLA 1.4 MEDIDAS DE RENDIMIENTO DE LOS DATOS DE VALIDACIÓN	51
TABLA 1.5 COMPARACIONES PAREADAS DE LA EXACTITUD CON LA CORRECCIÓN DE BONFERRONI	51
TABLA 2.1 EJEMPLOS DE RATIOS FINANCIEROS, SUS CATEGORÍAS Y REFERENCIAS.....	70
TABLA 2.2 GASTO SANITARIO RESPECTO DEL PIB EN PAÍSES DE LA UE	74
TABLA 2.3 ESTADÍSTICA DESCRIPTIVA DE LA MUESTRA UTILIZADA.....	74
TABLA 2.4 PARÁMETROS UTILIZADOS EN LOS AG	76
TABLA 2.5. MATRIZ DE CONFUSIÓN	78
TABLA 2.6. ESTADÍSTICA DESCRIPTIVA DE LOS REGISTROS DE LA MUESTRA	80
TABLA 2.7. COEFICIENTES Z'-SCORE DE ALTMAN Y ESTADÍSTICA DESCRIPTIVA DE LAS SOLUCIONES OBTENIDAS EN EL PROCESO AG.....	80
TABLA 2.8. MATRIZ DE CONFUSIÓN DE LOS MODELOS CON LOS DATOS TEST Y CON TODOS LOS REGISTROS..	82
TABLA 2.9. COMPARACIÓN DE LAS MEDIDAS DE RENDIMIENTO CON LOS DATOS TEST Y TODOS LOS REGISTROS.....	82
TABLA 3.1 EJEMPLOS DE RATIOS FINANCIEROS, SUS CATEGORÍAS Y REFERENCIAS	101
TABLA 3.2 ESTADÍSTICA DESCRIPTIVA DE LAS RATIOS FINANCIEROS Y SELECCIÓN DE VARIABLES	110
TABLA 3.3 VARIABLES EXPLICATIVAS DEL ESTRÉS FINANCIERO INCLUIDAS EN EL MODELO 1 Y DEL MODELO 2	112
TABLA 3.4 MATRIZ DE CONFUSIÓN	113
TABLA 3.5 MEDIDAS DE RENDIMIENTO DE LOS DATOS DE VALIDACIÓN	114
TABLA 3.6 COMPARACIONES PAREADAS DE LA EXACTITUD CON LA CORRECCIÓN DE BONFERRONI. TEST DATASET A (P-VALOR).....	115

ÍNDICE DE FIGURAS

FIGURA 1.1. COMPARACIÓN DE LA EXACTITUD ENTRE TÉCNICAS Y MODELOS	52
FIGURA 1.2. MODELO 1. SOM CON TODOS LOS REGISTROS DE LA MUESTRA Y DESGLOSADO POR ESTADO (UNSUPERVISED LEARNING).....	53
FIGURA 1.3. MODELO 2. SOM CON TODOS LOS REGISTROS DE LA MUESTRA Y DESGLOSADO POR ESTADO (UNSUPERVISED LEARNING).....	54
FIGURA 1.4. MODELO3. SOM CON TODOS LOS REGISTROS DE LA MUESTRA Y DESGLOSADO POR ESTADO (UNSUPERVISED LEARNING).....	54
FIGURA 2.1. PROCESO DE INFRA-MUESTREO CON LOOCV Y AG.....	77
FIGURA 2.2 COMPARACIÓN CURVAS ROC ENTRE EL MODELO AG Y Z'-SCORE ALTMAN.....	82
FIGURA 3.1. COMPARACIÓN DE LA EXACTITUD ENTRE MODELOS Y TÉCNICAS	115

Introducción y objetivo de investigación

Los métodos para predecir la bancarrota se han convertido en un importante tema de investigación desde la década de 1960, llegándose a considerar que el énfasis realizado sobre este tema podría ser catalogado como un indicador del grado de desarrollo y robustez económica de un país. Los altos costes económicos y sociales inherentes a los fracasos corporativos, han impulsado los esfuerzos para proporcionar una mejor predicción de los eventos de quiebra. Por ello, disponer de herramientas que permitan predecir la bancarrota con suficiente antelación resultaría de gran utilidad a las empresas, sociedad, administraciones e incluso países, para evitar las consecuencias desastrosas asociadas a la bancarrota. Desde el punto de vista estadístico, los modelos predictivos de la bancarrota pueden ser clasificados en dos grandes grupos, los que utilizan la estadística convencional para realizar las predicciones y los que usan técnicas de Inteligencia Artificial. Beaver (1966) fue uno de los pioneros en la predicción de la bancarrota utilizando para ello ratios financieros desde un punto de vista univariante. Posteriormente, Altman (1968) presentó un trabajo seminal, donde desarrolló un modelo multivariante con cinco ratios y en el que utilizaba el análisis discriminante múltiple para la predicción de problemas financieros severos en empresas manufactureras. Este modelo realizaba buenas predicciones hasta con dos años de antelación. Posteriormente, en el año 2000, este autor adaptó su fórmula para empresas privadas modificando uno de los 5 ratios propuestos en su trabajo seminal, concretamente, cambió la variable “*Market Value of Equity*” (capitalización bursátil) por “*Book values of equity*” (patrimonio neto).

A partir de la década de 1990 y con el desarrollo de software más potentes, se empezaron a utilizar en la predicción de la bancarrota, las técnicas de Inteligencia Artificial (IA). Estas pruebas han demostrado ser más robustas frente a las técnicas convencionales, ya que necesitan menos suposiciones, como la linealidad o normalidad de los datos y, además, proporcionan modelos predictivos más exactos. Por ello, han ido desplazando a las técnicas convencionales con el paso del tiempo.

Por ello, en el presente documento, se utilizan diferentes técnicas de IA como redes neuronales, *support vector machine*, árboles de decisión, *random forest*, *fuzzy logic*, mapas auto-organizados, algoritmos genéticos (AG) y combinación de técnicas de clasificación (*classifier ensembles*) para la predicción de la bancarrota o estrés financiero en tres sectores empresariales claves para la sociedad, los cuales se detallan a continuación. Además, se comparó el rendimiento ofrecido por dichas técnicas para la realización de una misma predicción.

Centrando la atención ahora el sector objeto de estudio, cabe mencionar que se utilizaron tres sectores, el de suministro de agua, el de tratamiento de aguas residuales y el sector de la medicina privada. Así, con el objeto de comprobar la robustez de las técnicas de IA se eligieron dos sectores que compartían muchas características y otro totalmente diferente pero en gran auge en España en los últimos años.

Hasta finales del siglo XX, la preocupación por el acceso al agua estuvo muy reducida a determinadas zonas geográficas de gran escasez y no se planteó como un problema de escala global. Sin embargo, en la actualidad se concibe como un recurso básico para la sociedad y que debe ser gestionado como un recurso de gran valor.

Por otra parte, desde la década de 1980 numerosas naciones y países, como Inglaterra, Gales o Chile, han decidido privatizar, quizás por la falta de recursos económicos, la industria del suministro de agua. España, no se ha visto ajena a esta situación y ya la gestión privada (a través de empresas privadas o mixtas) es del 50%.

Sin entrar en el debate de quien es más eficiente en la gestión de los recursos hídricos, se ha creído necesario aportar a este sector herramientas financieras que les permitan prever la bancarrota con suficiente antelación para evitar los problemas derivados de una falta de suministro de agua para la sociedad.

Al hilo de lo anterior, se puede afirmar que el agua es un recurso limitado e imprescindible y, sin embargo, se ve sometido a muchas presiones generadas por el ser humano que pueden afectar potencialmente su calidad. Por ello, el tratamiento y depuración de las aguas residuales son dos de las primeras medidas a implantar para conseguir la salubridad de la población. En esta situación, las empresas de recogida y tratamiento de aguas residuales se han convertido en invitados imprescindibles en cualquier sociedad desde un punto de vista tanto económico, como ambiental y sanitario.

Estas empresas, han estado muy presente en la literatura científica en los últimos años, pero las temáticas que se abordaban era en su mayor parte, desde un punto de vista técnico, debido al desarrollo de nuevas tecnologías, a cambios en procesos o materiales, no existiendo trabajos que estudiaran la problemática de la estabilidad financiera y bancarrota de las empresas en el sector. Por tanto, desde el ámbito privado, es necesario conocer si son financieramente autosuficientes o si presentan un deterioro financiero, que

puede ser provocado debido a las elevadas inversiones que se deben realizar en infraestructuras.

En relación al sector sanitario en España, en el año 2015, el gasto de la sanidad supuso el 9% del Producto Interior Bruto nacional (PIB) (OECD/EU, 2016), el 2,7% si se hace referencia solo a la sanidad privada, siendo este dato superior respecto al de países del entorno. Por otra parte, es evidente la cada vez mayor colaboración entre el sector público y privado. En la misma línea, también se pretende que la industria farmacéutica evolucione hacia la creación de nuevos modelos de colaboración y que aporte un mayor desarrollo de los servicios (Deloitte, 2014).

Tras analizar brevemente la importancia de este sector, se pretende obtener un modelo predictivo ad-hoc, por ello, partiendo de las ratios financieras propuestas por Altman (2000) para empresas privadas (Z' -Score), se pretende encontrar unos coeficientes para dicha fórmula calculados para el mencionado sector.

Por todo lo comentado previamente, el problema de investigación que se plantea en esta tesis surge debido a la falta de modelos predictivos desde un punto de vista financiero en las empresas privadas del sector del suministro del agua, del sector de recogida y tratamiento de aguas residuales y del sector sanitario privado español. Por ello, se considera necesaria la obtención de herramientas financieras que permitan ofrecer información a estos sectores con la mayor antelación posible y que, de esta forma, los *stakeholder* y la propia empresa puedan ganar en estabilidad financiera.

Para crear estos modelos predictivos, se han utilizado técnicas de IA (e.g. redes neuronales, *support vector machine*, *random forest*, etc.) y otras evoluciones de éstas técnicas como son la combinación de técnicas de clasificación (*classifier ensembles*). Para obtener los coeficientes del sector sanitario, se ha utilizado la técnica de los algoritmos genéticos.

Por otra parte, para desarrollar estas predicciones, se ha realizado una exhaustiva revisión de la literatura, donde se han seleccionado las 48 ratios financieras más utilizados por los principales autores de esta materia. Se han utilizado distintas técnicas para la selección de aquellas ratios que presentaban mayor poder predictivo como la selección binaria de variables a través de los AG o las curvas ROC (*Receiving Operating*

Curve) y se han desarrollado distintos modelos predictivos para comparar su rendimiento y de esta forma seleccionar el mejor.

A la luz de los resultados logrados, se puede concluir que se han obtenido varios modelos predictivos con un excelente rendimiento para la predicción del estrés financiero en las empresas de suministro de agua y en las empresas de tratamiento y depuración de aguas. En relación a las empresas del sector sanitario privado, se obtuvieron unos coeficientes que mejoraban las predicciones que realizaba la fórmula Z'-Score de Altman.

Por otra parte, también se ha demostrado que la técnica SOM, así como la combinación de técnicas de clasificación (*classifier ensembles*) realiza predicciones con una gran exactitud.

Por último, los procedimientos empleados a la hora de seleccionar las variables predictivas, han sido eficientes, entre las que cabe destacar, que un modelo que mayor número de variables predictivas, no tiene por qué tener mejor rendimiento.

Objetivos

Dada la importancia de los tres subsectores empresariales objeto de análisis, y ante la poca o nula existencia de herramientas predictivas específicas de estos sectores, el objetivo principal del presente estudio es aportar a cada uno de ellos modelos predictivos que sean capaces de realizar pronósticos fiables con una antelación de dos años. De esta forma, se amplían los conocimientos de estas empresas desde un punto de vista financiero, lo que puede ser clave para una mejor optimización de los recursos que les permitan realizar otros proyectos, como inversiones en infraestructuras o innovación, o que les permita aumentar sus reservas, en caso de un futuro incierto.

Así mismo, este mejor conocimiento financiero, implica aportar una estabilidad a la sociedad, debido a que no tendrá que ver recortado su bienestar debido a problemas derivados de la mala gestión de las empresas relacionadas con la gestión del agua o las o sanitarios.

También es un objetivo de este estudio analizar cómo conformar modelos predictivos a través de ratios financieros, es decir, como seleccionar aquellas variables

que permiten realizar las mejores predicciones posibles. En este punto se han aplicado varios procesos: a. minimizar el error AIC (error “*Akaike Information Criterion*”) a través de una regresión logística aplicando los AG para la clasificación binaria de variables, b. minimizar el error BIC (error “*Bayesian Information Criterion*”) en lugar del AIC o c. utilizar el área bajo las curvas ROC (*AUCROC*).

Otro gran objetivo de este trabajo reside en la comparación del rendimiento de las principales técnicas de IA, así como la combinación de técnicas de clasificación, aplicadas hoy en día en la creación de modelos predictivos. Seguidamente se detallan de forma sucinta las técnicas de IA aplicadas:

- Redes Neuronales (NN): Las NN ofrecen un paradigma computacional inspirado en las redes neuronales biológicas del sistema nervioso humano (Kumar *et al.*, 2007). Contienen unidades de procesamiento de información similares a las neuronas disponibles en el cerebro humano (Tsai *et al.*, 2014).
- SVM: Utiliza un modelo lineal para implementar los límites de clase no lineales mediante la asignación de vectores de entrada no lineales en un espacio de alta dimensional característica.
- Árboles de decisión (DT): Toma la forma de una estructura de árbol inversa, es decir, de arriba a abajo, donde se van desglosando los datos para crear las ramas. Un árbol de decisión se construye con nodos, donde cada nodo denota una prueba en una variable, y cada rama que sale del nodo representa un resultado de la prueba.
- Random Forest (RF): Es un algoritmo de clasificación que utiliza un conjunto de árboles de clasificación (*decisión tree*). Cada uno de los árboles de clasificación se construye utilizando una muestra de *bootstrap* de los datos, y en cada división el conjunto de variables candidatas es un subconjunto aleatorio del total de las variables.
- Fuzzy Logic (FL): Proporciona un modelo matemático para emular el proceso cognitivo humano. Esta teoría busca clasificar el razonamiento subjetivo y asignar grados de posibilidades para llegar a conclusiones. Proporciona un medio para tomar decisiones basadas en datos ambiguos, imprecisos o incompletos.

- Self-organizing Maps (SOM): Es una red neuronal no supervisada (un-supervised) que se compone de dos capas neuronales. La capa de entrada tiene tantas neuronas como variables, y su función es simplemente capturar la información. La capa de salida, tiene forma rectangular con “x” filas e “y” columnas (dos dimensiones). Cada neurona en la capa de entrada está conectada a cada neurona en la capa de salida a través de los pesos sinápticos.
- Classifier Ensembles (Combinación de Técnicas de clasificación): A la combinación de múltiples clasificadores se le llama “*classifier ensembles*”. Tiene como objetivo la obtención de clasificadores de mayor exactitud combinando técnicas menos precisas. Básicamente se proponen para mejorar el rendimiento de clasificación de un único clasificador.
- Algoritmos Genéticos (AG): Simulan la teoría de la evolución propuesta por Darwin y son procesos estocásticos robustos que pueden usarse para selección binaria de variables o para resolver problemas de búsqueda y optimización a través de una función *fitness* o función objetivo.

Para lograr todos estos objetivos planteados y poder implementar la experimentación se utilizaron datos financieros de las empresas que componían el sector. Dichos datos fueron extraídos de la base de datos SABI-Sistema de Análisis de Balances Ibéricos (Bureau Van Dijk, 2017). Finalmente, indicar que el software utilizado para analizar los datos fue R Core Team (2017).

Estructura

El presente trabajo está estructurado, además de la presente introducción y de las conclusiones, a través de tres capítulos, los cuales pasan a presentarse brevemente a continuación.

El primer capítulo comienza con una descripción de la situación de las empresas del sector de recogida y tratamiento de aguas residuales, así como la evolución histórica de los distintos modelos predictivos de la bancarrota. Dada la relevancia de este sector y a la falta de información financiera, se plantea la necesidad de crear un modelo para

predecir el estrés financiero con dos años de antelación, el cual consideramos que es un tiempo sustancial para que las empresas adopten las medidas oportunas para reconducir la situación en caso de un escenario desfavorable.

Tras la revisión de la literatura, se seleccionaron inicialmente 48 ratios financieros, aunque 10 de ellos fueron desestimados al no estar correctamente informados. Se utilizaron cuatro técnicas de IA para realizar las predicciones:

- Support vector machine
- Classifier Ensembles o combinación de técnicas de clasificación
- Fuzzy Logic (lógica difusa)
- Self-Organizing Maps (SOM) o mapas auto-organizados

Tras ejecutar los datos con los tres modelos y las 4 técnicas, se pudo observar que la técnica que presenta mejor rendimiento, por encima del 95%, es la técnica SOM.

El segundo capítulo incluye una descripción de la literatura sobre la predicción de la bancarrota y también se muestra la cada vez mayor importancia que está adquiriendo la sanidad privada en España, la cual, se ha visto incrementada durante la crisis económica por la que atraviesa España desde el año 2007/2008. Por ello, se desea realizar un modelo predictivo ad-hoc para este sector, para ello, se parte de la fórmula Z'-Score de Altman para empresas privadas (2000), que está formada por cinco ratios financieros y sus correspondientes coeficientes. Entonces, conservando las 5 ratios propuestos por Altman, y utilizando la técnica de los AG, se obtienen cinco coeficientes nuevos asociados a las empresas de este sector. Los resultados muestran que los coeficientes para el modelo predictivo obtenido con los AG presentan mayor exactitud, sensibilidad y especificidad que el propuestos originalmente por Altman.

El tercer capítulo se inicia con una descripción de la importancia del agua para la sociedad, por este motivo, se destaca la relevancia que tienen estas empresas para el desarrollo y bienestar social. Por otra parte, también se comenta como estas organizaciones se han ido privatizando desde la década de 1980. En España, la gestión privada o mixta de estas empresas supera el 50%. Teniendo en cuenta la indiscutible importancia de este sector, se propone la creación de un modelo para predecir el estrés financiero con dos años de antelación.

Se utilizaron cinco técnicas de IA para realizar las mencionadas predicciones:

Introducción y objetivo de la investigación

- Redes Neuronales (NN)
- Support vector machine (SVM)
- Support vector machine optimizado (SVM-T)
- Árboles de decisión (DT)
- Random Forest (RF)

Se observó que la técnica que presenta mejor rendimiento, por encima del 85% para es la técnica SVM-T, seguida muy de cerca por la técnica RF.

Capítulo 1: Aplicación de técnicas de Inteligencia Artificial para la predicción del estrés financiero en empresas del sector de recogida y tratamiento de aguas residuales

1.1 Resumen

La gestión de las aguas residuales es una cuestión de gran importancia para la sociedad actual. Dicha importancia, la cual se ha visto incrementada en los últimos años por la escasez de recursos hídricos, se puede observar además desde diferentes puntos de vista, el económico, el ambiental, el sanitario, etc. Actores fundamentales en la gestión de este recurso son las empresas proveedoras del servicio. Así, el que éstas tengan un buen desempeño se vuelve un aspecto clave para garantizar la calidad del servicio y, para que puedan lograr el citado buen desempeño, el que cuenten con salud financiera resulta clave. En este sentido, disponer de una herramienta que permita prever la bancarrota de este tipo de empresas con años de antelación resultaría de gran utilidad para evitar a todos los *stakeholders*, sociedad en general, administración pública, entorno, etc., los problemas derivados de que no se lleve a cabo el tratamiento de este tipo de aguas. Por este motivo, en el presente trabajo se plantea establecer un modelo de predicción basado en inteligencia artificial (artificial intelligence, AI) y, para ello, se ha realizado la comparación del rendimiento para el problema mencionado de cuatro técnicas de inteligencia artificial: a. *Support Vector Machine* (SVM), b. combinación de técnicas de clasificación (*classifier ensembles*, CE), entre las que se encuentran: *random forest*, *bagged neural networks* o *bagged support vector machine*, c. *Fuzzy Logic* (FL) y d. *self-organizing maps* (SOM). Los resultados empíricos demuestran que la técnica que mejores predicciones realiza con los datos de validación es SOM, seguido de CE.

Palabras Claves: estrés financiero, empresas gestoras de servicios de aguas residuales, Z-Altman, inteligencia artificial, combinación de técnicas de clasificación.

1.2 Abstract

Wastewater management is a matter of great importance for today's society. This importance, which has been increased in recent years by the scarcity of water resources, can also be observed from different points of view, economic, environmental, health, etc. Fundamental players in the management of this resource are the companies that provide the service. Thus, having a good performance becomes a key aspect to guarantee the quality of the service and, to achieve the aforementioned good performance, having financial health is essential. In this sense, having a tool to predict the bankruptcy of this type of companies with years in advance would be very useful to avoid all the stakeholders, society in general, public administration, environment, etc., the problems arising from the non-treatment of this type of water. For this reason, the present work proposes to establish a prediction model based on artificial intelligence (AI) and, for this, the performance comparison for the mentioned problem was realized using four techniques of artificial intelligence: a. Support Vector Machine (SVM), b. Combination of classification techniques (classifier ensembles, CE), including random forest, bagged neural networks or bagged support vector machine, c. Fuzzy Logic (FL) and d. Self-organizing maps (SOM). The empirical results show that the technique that best predicts with validation data is SOM, followed by CE.

Keywords: Financial stress, wastewater service management companies, Z-Altman, artificial intelligence, classifier ensembles

1.3 Introducción

El agua es un recurso limitado y fundamental para el desarrollo de la vida. Sin embargo, se ve sometido a muchas presiones generadas por el ser humano que pueden afectar potencialmente su calidad. Por ello, el tratamiento y depuración de las aguas residuales son dos de las primeras medidas a implantar para conseguir la salubridad de la población.

El tratamiento de aguas residuales es un eslabón importante en el ciclo del agua que permite el saneamiento y la reutilización de ésta, facilita la generación de energía y permite la recuperación de los productos de los desechos. En este sentido, la gestión y capacidad de adaptación de los recursos hídricos se ha convertido en una necesidad estratégica de carácter mundial, donde la importancia que han adquirido las aguas residuales ha llegado a tal magnitud que la edición correspondiente al año 2017 del Informe Mundial de las Naciones Unidas sobre el Desarrollo de los Recursos Hídricos (ONU, 2017) se ha centrado exclusivamente en este recurso tan valioso para la sostenibilidad del medio ambiente.

Uno de los aspectos más importante sobre el que advierte dicho informe, está relacionado con el daño que ocasiona las aguas residuales no tratadas, ya que pueden quedar contaminadas por bacterias y virus nocivos y representar, por tanto, un riesgo para la salud pública, las aguas continentales y el medio marino. Por otra parte, se estima que más del 80% de dichas aguas, se vierten al medio ambiente sin tratamiento alguno, cuya consecuencia más directa es la alta contaminación de las aguas en la mayoría de los ríos de África, Asia y América Latina, estimándose en unos 245.000 km² de zonas muertas desoxigenadas en mares y océanos, con nefasta repercusión en la industria pesquera, medios de subsistencia y cadenas alimenticias (ONU, 2017).

Por tanto, el tratamiento de las aguas residuales puede considerarse una necesidad imperiosa a nivel mundial, no sólo porque reduciría de forma importante su impacto en el medio ambiente, sino que, además, permitiría reutilizar las aguas recicladas en el consumo humano, fundamentalmente a través de su utilización en el sector agrícola en el que se emplea más del 70% del volumen de dichas aguas. En este sentido, no es de extrañar que dicho informe concluya que *“en un mundo donde la demanda de agua dulce está en constante aumento y los escasos recursos hídricos se ven cada vez más exigidos*

Capítulo1: Aplicación de técnicas de Inteligencia Artificial para la predicción del estrés financiero en empresas del sector de recogida y tratamiento de aguas residuales

por la captación excesiva, la contaminación y el cambio climático, sería sencillamente impensable no aprovechar las oportunidades que brinda una mejor gestión de las aguas residuales” (ONU, 2017: 7).

Sin embargo, no todos los países tienen capacidad para gestionar correctamente las aguas residuales, de hecho según Sato *et al.* (2013), su tratamiento está relacionado con el nivel de ingresos del país. Dichos autores señalan que, en promedio, los países con ingresos altos tratan cerca del 70% de las aguas residuales que generan, mientras que esa proporción cae al 38% en los países de ingresos medios-altos y al 28% en los países de ingresos medios-bajos. En países de bajos ingresos, solo el 8% de las aguas residuales industriales y municipales se someten a algún tipo de tratamiento.

En la Unión Europea, y consecuentemente en España, la depuración de aguas residuales urbanas es una práctica obligada desde la entrada en vigor de la Directiva 91/271/CE de 21 de mayo, donde se define como principal objetivo proteger al medio ambiente de los efectos negativos de los vertidos de las aguas residuales urbanas y de los sectores industriales. El cumplimiento de este objetivo ha supuesto la construcción y puesta en marcha de multitud de instalaciones, que garanticen la recogida y tratamiento de aguas, en función del tamaño de la aglomeración y características de la zona de vertido.

Esta obligación, junto con el incremento de la conciencia en la protección de los recursos hídricos, ha provocado una fuerte inversión por parte de las Administraciones Públicas a todos los niveles (europeas, nacionales, autonómicas y locales) en el diseño, construcción y mantenimiento de estaciones depuradoras de aguas residuales. En concreto, el gasto en gestión y tratamiento de aguas residuales en la Unión Europea, con 28 Estados miembros, ascendió a alrededor del 0,60% del PIB en 2011 (Guerrini *et al.*, 2016).

Es evidente que la gestión de aguas residuales es costosa y requiere de una importante inversión en infraestructura, por lo que la falta de recursos financieros constituye una limitación importante en el desarrollo de este sector y, por tanto, en la sostenibilidad del medio ambiente. De ahí que en la mayoría de los países industrializados fuesen las empresas públicas las encargadas de llevar a cabo esta actividad por el claro beneficio social que producían. No obstante, poco a poco se han ido

introduciendo empresas privadas o se han ido privatizando empresas públicas con la intención de mejorar la eficiencia de dichas empresas.

La privatización ha sido una actividad frecuente en los gobiernos en muchos países desde finales de 1970, con la intención de mejorar la eficiencia de la prestación de servicios públicos y, por supuesto, mejorar la situación financiera de los gobiernos, sobretodo en épocas de crisis financieras. Esto ha permitido al sector privado participar en la prestación de ciertos servicios públicos a través de mecanismos como las asociaciones público-privadas (Lee y Yu, 2012). En particular, en el sector del agua y aguas residuales, estas asociaciones se han vuelto más frecuentes desde los años noventa, particularmente en algunos países en vías de desarrollo, como China, India o Ghana (Zhong *et al.*, 2008).

Por tanto, desde el ámbito privado, se hace necesario comprobar la viabilidad del sector de aguas residuales, de manera que no sólo sea interesante desde el punto de vista social, sino empresarial. La necesidad de conocer si las empresas de este sector son financieramente autosuficientes o el posible deterioro financiero que se pueda producir por diversas circunstancias, permite por un lado, alentar la inversión privada, mientras que por otro lado supone un desafío para los organismos públicos cuyo objetivo debería centrarse en desarrollar normativas para la promoción del sector.

Por otra parte, la bancarrota es un evento crítico que puede causar grandes pérdidas a los distintos *stakeholders*, por ello, el desarrollo de modelos efectivos de predicción es una tarea importante, a la vez que complicada. Además, si esta circunstancia es crítica en cualquier empresa, en las que realizan depuración y tratamiento de aguas residuales lo es más si cabe. Ello es así, ya que a los problemas propios de una situación como la que se describe en cualquier empresa, se une los problemas de salud y ambientales que puede generar el que estas empresas no funcionen correctamente.

El objetivo de los modelos de predicción de la bancarrota es predecir con cierta antelación si una empresa irá a la bancarrota o no. Si el modelo predictivo no presenta un buen rendimiento, podría provocar la toma incorrecta de decisiones tanto en la propia empresa como en los *stakeholders* y, como consecuencia de ello, una grave crisis financiera (Tsai *et al.*, 2014). Similar al objetivo de la predicción de la bancarrota, ocurre con el estrés financiero, es decir, un buen modelo predictivo para el estrés financiero puede ayudar a tomar decisiones financieras con menor riesgo para los *stakeholders*. Por

ejemplo, si un organismo público pudiese detectar con años de antelación que una empresa concesionaria para el tratamiento de aguas residuales se encontrará en una grave crisis financiera, podría tomar medidas antes de que ésta quiebre y se tenga problemas con la prestación del servicio.

Desde un punto de vista estadístico, hay dos principales categorías en la literatura para tratar este tema. La primera utiliza técnicas estadísticas convencionales como el análisis discriminante (E. I. Altman, 1968), el modelo logit (Ohlson, 1980), el modelo probit (Zmijewski, 1984). La segunda, utiliza técnicas de IA, como redes neuronales, *Support Vector Machine* (SVM), árboles de decisión, lógica difusa (*fuzzy logic*), etc. (Odom y Sharda, 1990; Ravi Kumar y Ravi, 2007; Shin *et al.*, 2005).

Varios estudios han demostrado que las técnicas de IA presentan mejor rendimiento que las técnicas convencionales en términos de exactitud y error (Tsai, 2008; Tsai y Wu, 2008; Tsai y Wu, 2008). Además, recientemente se ha demostrado que la combinación de técnicas de IA, (*classifier ensembles*) presenta mejores resultados que las técnicas individuales (Tsai *et al.*, 2014).

Ante todo ello, el presente trabajo tiene como objetivo mejorar el conocimiento de un sector de actividad económica con unas características especiales ya que, si bien en los últimos años, se ha desarrollado mucha literatura entorno a este tipo de empresas, estos se han centrado principalmente en aspectos técnicos como el empleo de nuevas tecnologías, procesos y materiales (Guerrini *et al.*, 2016), y ninguno en predecir la futura quiebra o estrés financiero. Este tipo de predicciones ya se hace en otras empresas y sectores (e.g. diferentes tipos de empresas del Reino Unido (Almamy *et al.*, 2016), empresas de tamaño medio y pequeñas italianas (Gordini, 2014), empresas de tamaño medio manufactureras de Korea (Shin y Lee, 2002)). Otro de los objetivos que se plantean es realizar una comparativa entre tres modelos con distinto número de variables predictivas para comparar cuál de ellos presenta resultados más exactos (en el apartado Análisis Estadístico y Tabla 1.2, que se comentan con posterioridad, aparecen detallados como se han creado estos tres modelos).

Este artículo se ha organizado de la siguiente forma. En el presente epígrafe 1 se expone la introducción del trabajo. En la sección 2 se expone la revisión de la literatura, en el apartado 3 la metodología (descripción de técnicas empleadas, medidas de

rendimiento y datos utilizados). En la sección 4 se describen los resultados y, por último en la 5, se muestran las conclusiones.

1.4 Revisión de la Literatura

La grave crisis financiera que ha venido afectando a la gran mayoría de países en los últimos años, ha generado una cada vez mayor preocupación por el riesgo financiero y ha relanzado el interés por trabajos que analizan las causas del fracaso empresarial, con el claro objetivo de establecer un método fiable que permita detectar y paliar, con la suficiente anticipación, los graves efectos que suelen generar este tipo de situaciones. Es evidente que detectar el fracaso empresarial resulta necesario para garantizar la estabilidad de los *stakeholders*, los mercados financieros y la prosperidad económica general (Jones, 1987). Aunque resulta evidente la importancia de la identificación temprana de los problemas financieros, en el caso concreto del sector de recogida y tratamiento de aguas residuales adquiere mayor significación al tratarse de un servicio vital para la salubridad de la población, la sostenibilidad ambiental y el poder reutilizar un bien escaso, básico y necesario como es el agua. Sin olvidarnos, del cumplimiento de la normativa europea en este sentido.

La literatura de la predicción de la bancarrota empieza en la década de 1930. Así, FitzPatrick (1932) comparó 13 ratios de empresas en bancarrota y empresas en no-bancarrota (activas) y descubrió que las empresas activas poseían ratios más favorables. Hasta mediados de la década de los años 60, las investigaciones se basaban en análisis univariable. El más reconocido fue el de Beaver (1966), que encontró que un número determinado de ratios podían discriminar entre bancarrota y no-bancarrota con hasta cinco años de antelación. Dicho autor sugirió la posibilidad de que múltiples ratios considerados al mismo tiempo podrían aumentar el poder predictivo y de esta forma, empezó la evolución de los modelos predictivos de bancarrota. En esta línea, Altman, (1968) publicó el primer análisis multivariable. Altman utilizó el análisis discriminante múltiple (ADM) para desarrollar un modelo predictivo usando cinco ratios financieros para predecir la bancarrota de empresas manufactureras.

Por su parte, Martin (1977) utilizó la regresión logística (logit) para predecir la probabilidad de fracaso de los bancos sobre una base de datos obtenida del Sistema de Reserva Federal. Ohlson (1980) utilizó un modelo logit para superar las desventajas

Capítulo 1: Aplicación de técnicas de Inteligencia Artificial para la predicción del estrés financiero en empresas del sector de recogida y tratamiento de aguas residuales

asociadas a los modelos ADM. Este estudio utilizaba 9 variables, de las cuales, 8 eran ratios financieros. Dietrich y Kaplan (1982) desarrollaron un modelo lineal de tres variables para clasificar los riesgos de préstamo. Compararon su modelo con el modelo de Altman y con el modelo de predicción de quiebra de Wilcox (1973) y obtuvieron mejores resultados con respecto a los otros dos modelos. Zmijewski (1984) desarrollo un nueva fórmula utilizando un modelo probit. Este estudio utilizaba 3 ratios financieros y la exactitud de este modelo con los datos de su muestra fue del 99%. Altman (2000) adaptó su fórmula propuesta en 1968 para el sector de las empresas privadas, cambiando la variable “*Market Value of Equity*” (capitalización bursátil) por “*Book values of equity*” (patrimonio neto).

La Tabla 1.1 (Altman, 1968; Altman, 2000; Aziz *et al.*, 1988; Bauer y Agarwal, 2014; Beaver, 1966; Becchetti y Sierra, 2003; Beynon y Peel, 2001; Blanco-Oliver *et al.*, 2015; Charitou *et al.*, 2004; Chi y Tang, 2006; Cho *et al.*, 2009; Chung *et al.*, 2016; Cielen *et al.*, 2004; Cultrera y Brédart, 2016; Dambolena y Khoury, 1980; Divsalar y Roodsaz, 2012; du Jardin, 2017; Dugan y Zavgren, 1989; Gu y Gao, 1999; Jeong *et al.*, 2012; Joo y Jin, 2000; Khan, 1985; Kim *et al.*, 2016; Kim y Gu, 2006; Kim y Kang, 2012; Laitinen y Laitinen, 2000; Lennox, 1999; Li y Miu, 2010; Mousavi *et al.*, 2015; Nouri y Soltani, 2016; Ohlson, 1980; Pompe y Bilderbeek, 2005; Shumway, 2001; Wang *et al.*, 2014; Wilson y Summers, 2002; Xu y Zhang, 2009) muestra las principales ratios financieros utilizados y sus referencias.

Tabla 1.1 Ejemplo de ratios financieros, sus categorías y referencias

Categorías	Variables y definiciones	Estudios Previos
I. CAPITAL STRUCTURE VARIABLE	Long Term Debt to Total Asset	Beaver 1966; Ohlson 1980; Gu y Gao, 2000; Mousavi et al, 2015
	Short and Long Term Debt to Equity	Chi et al 2006; Zmijewski,1984
	Short y Long Term Debt to Total Assets	Beaver 1966; Laitinen et al 2000; Lennox, 1999,
	Cash-Holdings to Total Assets	Dugan et al 1989; Laitinen et al 2000
	Equity Share of Total Assets	Altman,1968; Cielen et al 2004; Charitou et al , 2004; Pantalone y Platt, 1987; Pompe et al 2005; Shumway,2001
	Current Liabilities to Total Assets	Blanco-Oliver et al, 2015; Gu y Gao 2000; Jeong et al, 2012; Lo, 1985; Pantalone y Platt ,1987; Zmijewski,1984
	Total liabilities to total assets	Divsalar et al, 2012, Mousavi et al, 2015; Xu y Zhang, 2009
	Total liabilities to current assets	Blanco-Oliver et al, 2015
	Total liabilities to working capital	Mousavi et al, 2015
	Quick assets to total assets	Divsalar et al, 2012

Capítulo1: Aplicación de técnicas de Inteligencia Artificial para la predicción del estrés financiero en empresas del sector de recogida y tratamiento de aguas residuales

	Quick assets to current assets	Blanco-Oliver et al, 2015
	Quick assets to current liabilities	Du Jardin, 2017
	Equity to total assets	Cultrera y Brédart, 2016
	Current assets to total assets	Kim y Kang, 2012; Wang et al, 2014; Zhao et al, 2017
	Current debt to total debt	Wang et al, 2014
	Retained earnings to total assets	Blanco-Oliver et al, 2015; Li y Miu, 2010; Kim y Kang, 2012
	Current Liabilities to total capital	Cho et al, 2009
	Current ratio	Blanco-Oliver et al, 2015; Cultrera y Brédart, 2016; Chung et al, 2016; Wang et al, 2014; Xu y Zhang, 2009; Zhao et al, 2017
	Debt-total assets	Chung et al, 2016; Kim y Kang, 2012
	Total borrowings to total assets	Kim et al, 2016
II. LIQUIDITY RATIOS	Quick Ratio	Altman,1968; Cielen et al 2004; Charitou et al , 2004; Chung et al, 2016; Nam y Jinn ,2001; Kim y Kam, 2016; Shumways, 2001
	Working capital to total asset	Nouri y Soltari, 2016; Li y Miu, 2010; Mousavi et al, 2015; Xu y Zhang, 2009; Zhao et al, 2017
	Working capital to sales	Zhao et al, 2017
	Cash to current debt	Du Jardin, 2017; Wang et al, 2014; Zhao et al, 2017
	Cash to current assets	Du Jardin, 2017
	Cash to total assets	Blanco-Oliver et al, 2015; Du Jardin, 2017, Zhao et al, 2017
	Cash to Long y Short Term Debt	Dugan y Zavgren 1989; Charitou et al , 2004
III. PROFITABILITY RATIOS	Return on Equity	Altman, 1968; Pompe et al 2005; Shumways, 2001
	Return on Assets	Becchetti et al 2003; Chung et al, 2016; Dambolena y Khoury 1980; Kim et al, 2016; Nouri y Soltari, 2016; Ohlson, 1980; Zmijewski, 1984
	Net Profit Margin	Beaver 1966; Kim et al 2006
	Net Profit to total assets	Zhao et al, 2017
	Net Profit to current assets	Zhao et al, 2017
	Gross Profit Margin	Cielen et al 2004; Kim et al 2006
	Gross Profit to sales	Zhao et al, 2017
	Profit Margin	Becchetti et al 2003; Pompe et al 2005
	Net profit to liabilities	Zhao et al, 2017
	Net incomen to current assets	Mousavi et al, 2015
	Net incomen to total assets	Bauer y Agarwal, 2014; Xu y Zhang, 2009
	Net incomen to capital	Mousavi et al, 2015
	Net incomen to sales	Jeong et al, 2012
	Ordinary income to sales	Jeong et al, 2012; Kim y Kang, 2012
	Ordinary income to total assets	Kim y Kang, 2012
Ordinary income to capital	Kim y Kang, 2012	
Return on capital employment	Nouri y Soltari, 2016	
Profit to sales	Cho et al, 2009	

Capítulo1: Aplicación de técnicas de Inteligencia Artificial para la predicción del estrés financiero en empresas del sector de recogida y tratamiento de aguas residuales

	Operating profit to sales	Cho et al, 2009
IV. CASHFLOW	Cash flow to Sales	Beaver 1966; Wang et al, 2014
	Cash flow to Total Assets	Aziz et al. , 1988; Pompe et al 2005; Young et al, 2005
	Cash flow to total debt	Cultrera y Brédart, 2016
	Operating cash flow to sales	Wang et al, 2014
	Operating cash flow to total debt	Kim y Kang, 2012; Wang et al, 2014
	Operating cash flow to total assets	Kim et al, 2016; Wang et al, 2014
	Cash flow to long y short term debt	Beaver 1966; Chung et al, 2016
	Sales to Total Assets	Altman 1968; ; Li y Miu, 2010
	Sales to current assets	Divsalar et al, 2012
	EBITDA to Total Assets	Altman 1968; Cultrera y Brédart, 2016; Du Jardin, 2017; Gu y Gao, 2000; Young et al 2005
	EBITDA to sales	Kim et al, 2016
	EBIT to total assets	Mousavi et al, 2015
	EBITDA to interest expenses	Kim y Kang, 2012
	Retained Earnings to Total Assets	Altman 1968
	Financial Expenses to Sales	Becchetti et al 2003
V. DEVELOPMENT AND FIRM	Change in Sales from Last Year	Arshad 1985; Pompe et al 2005
	Change in Total Assets from Last Year	Arshad 1985; Becchetti et al 2003
	Firms Age	Beaver 1966; Kim et al 2006; Altman, 2000
	Firm Size (Natural Logarithm of sales)	Cielen et al 2004; Kim et al 2006
	Firm Size (Natural Logarithm of total assets)	Blanco-Oliver et al, 2015; Beynon, y Peel, 2001; Kim y Kang, 2012; Ohlson, 1980; Wilson y Summer, 2002
	Growth rate of tangible assets	Cho et al, 2009
	Growth rate of total assets	Jeong, et a, 2012
	Growth rate of current assets	Cho et al, 2009
VI. TURNOVER	Accounts Receivable to Sales	Beaver 1966
	Inventory to Sales	Beaver 1966; Du Jardin, 2017; Kim y Kang, 2012; Mousavi et al, 2015; Young et al 2005
	Inventory to Total Assets	Cielen et al 2004
	Current assets to sales	Du Jardin, 2017; Wang et al, 2014
	Current liabilities to sales	Du Jadin, 2017
	Receivables to sales	Du Jadin, 2017
	Sales to total assets	Du Jadin, 2017
	Asset turnover ratio	Chung et al, 2016; Jeong et al, 2012; Nouri y Soltari, 2016
	Inventory to turnover ratio	Nouri y Soltari, 2016
	Turnover of managerial assets	Cho et al, 2009
VII. MARKET VARIABLES	Market value to total debt	Nouri y Soltari, 2016
	Market value of total assets	Bauer y Agarwal, 2014
	Market value of equity to total assets	Bauer y Agarwal, 2014; Li y Miu, 2010
	Stock price	Nouri y Soltari, 2016
	Stock return	Nouri y Soltari, 2016

VIII. OTROS	Productivity of capital	Kim et al, 2016
	Interest expenses-gross profit	Chung et al, 2016
	Interest expenses to sales	Cho et al, 2009
	Financial expenses to total debt	Kim y Kang, 2012
	Capital employed to total liabilities	Blanco-Oliver et al, 2015

Por otro lado, hasta 1990, las técnicas estadísticas multivariantes más usadas para la predicción de la bancarrota eran el ADM (e.g. Altman, 1968, 1983), el logit (e.g. Martin, 1977; Ohlson, 1980) y el probit (e.g. Zmijewski, 1984). Sin embargo, estas técnicas deben cumplir unos supuestos como la linealidad, la normalidad e independencia de las observaciones, que no siempre se cumplen en el ámbito financiero (Deakin, 1972). En consecuencia, dichos métodos pueden tener limitaciones de validez y efectividad (Shin y Lee, 2002).

Desde finales de los años 80, las técnicas de AI, que son menos vulnerables a las suposiciones comentadas previamente (Chung y Tam, 1993; Messier y Hansen, 1988), han sido aplicadas con éxito en la predicción de la bancarrota (e.g. Angelini *et al.*, 2008; Atiya, 2001; Paliwal y Kumar, 2009). El primer artículo en el que se utilizaron redes neuronales (NN) en la predicción de la bancarrota fue realizado en 1990 por Odom y Sharda (1990). Dicho modelo clasificaba correctamente (exactitud) al 81.81% de las empresas de validación (test). Shin y Lee (2002) utilizaron los AG para predecir la bancarrota en 528 empresas manufactureras de tamaño medio. Este modelo clasificaba correctamente el 80.8% de las empresas.

Fan y Palaniswami (2000) fueron los primeros en utilizar SVMs en problemas de predicción de bancarrota. La técnica SVM fue comparada con las técnicas “*linear discriminant analysis*” (LDA), NN y “*learning vector quantization*” (LVQ) con un fichero de empresas australianas. SVM obtuvo los mejores resultados con una exactitud del 70%, seguido por las NN con un 68%, seguido de LVQ con un 63% y seguido de LDA con un 63% (Atiya, 2001).

Hachicha *et al.* (2011) utilizaron FL y las técnicas de evolución diferencial para explicar la dinámica del mercado financiero. Michael *et al.* (1999) propuso un método generador de reglas difusas para la predicción de la bancarrota y lo comparó con el análisis discriminante, logit y probit. Se utilizaron dos muestras de datos de empresas griegas. El clasificador basado en reglas difusas obtuvo un error general del 26.16%. El

error general del análisis discriminante, logit y probit fue 29,35%, 27,15% y 30,63%, respectivamente.

Por su parte, Alam *et al.* (2000) propusieron “*fuzzy clustering*” para identificar bancos con riesgo potencial de quiebra y compararon dicha técnica con dos tipos de redes neuronales, “*competitive neural network*” y SOM. Los autores concluyeron que tanto “*fuzzy clustering*” como SOM son buenas técnicas para identificar los bancos potenciales de quiebra (Ravi Kumar *et al.*, 2007).

Teuvo Kohonen (1982) presentó un modelo de red denominado mapas auto-organizados o SOM (Self-Organizing Maps). Dicha técnica, un tipo red neuronal, fue utilizada posteriormente para representar gráficamente la predicción de problemas financieros (e.g. López Iturriaga y Sanz, 2015, N. Chen *et al.*, 2013). Serrano-Cinca (1996) comparó el rendimiento de SOM con análisis discriminante lineal y NN en el diagnóstico financiero. Como variables independientes su utilizaron las variables propuestas por Altman (1968). La arquitectura de SOM utilizada era 4 nodos de entrada y 144 nodos de salida dispuestos en una cuadrícula cuadrada de 12*12. Kaski *et al.* (2001) introdujeron la métrica basada en la matriz de información de Fisher y la implementaron con SOM para representar la predicción de la bancarrota.

Ya se ha comentado previamente que las técnicas de IA, tales como la FL superan a las técnicas estadísticas convencionales, incluyendo la regresión logística, en términos de predicción, exactitud y error (e.g. Tsai, 2008, Tsai y Wu, 2008). Por otra parte, la combinación de múltiples técnicas de clasificación (CE) ofrecen un mejor rendimiento que las técnicas de IA utilizadas de forma individual, ya que la combinación compensa los errores cometidos por las técnicas individuales (e.g. Lin *et al.*, 2012, Tin Kam Ho *et al.*, 1994).

Hansen y Salamon (1990) aportaron algunos de los primeros resultados de CE para demostrar que el error de una NN se podía reducir significativamente mediante la combinación de redes similares, todas ellas entrenadas con el mismo fichero. West *et al.* (2005) compararon un modelo de NN con tres CE y concluyeron que estas últimas técnicas tenían mejor rendimiento que las NN.

1.5 Metodología

1.5.1 Datos

Los datos utilizados para este trabajo se obtuvieron de la base de datos SABI-Sistema de Análisis de Balances Ibéricos (Bureau Van Dijk, 2017). Se seleccionaron aquellas empresas cuyo CNAE (Clasificación Nacional de Actividades Económicas) era el 3700 (Recogida y Tratamiento de Aguas residuales) y, además, se encontraban en situación “activa”. Se incluyeron un total de 230 empresas y se seleccionaron las que tenían todos los ratios informados. Ello hizo que la muestra final quedase en 131 empresas, 39 se encontraban en estrés financiero y 92 que no lo estaban. Para definir si la empresa se encontraba dicha situación, se utilizó la fórmula Z'-Score de Altman para empresas privadas (Altman, 2000). Esta elección se debió a que Altman es uno de los autores más citados en relación a la predicción de la bancarrota (Almamy *et al.*, 2016):

$$Z' = 0.717 * X1 + 0.847 * X2 + 3.107 * X3 + 0.420 * X4 + 0.998 * X5 \quad (\text{fórmula 1.1})$$

donde:

X1: Working Capital/Total Assets

X2: Retained Earnings/Total Assets

X3: EBIT/Total Assets

X4: Book Value Equity/Total liabilities

X5: Sales/Total Assets

Los valores umbrales que estableció el autor fueron los siguientes: puntuaciones del valor Z' superiores a 2,90, indican que la empresa está en una zona segura. Si la puntuación está entre 1,23 y 2,90, la empresa está en una zona gris y si la puntuación está por debajo de 1,23, entonces la empresa está en riesgo de bancarrota. En nuestro caso, se clasificó a la empresa en estrés financiero si su valor Z' era inferior a 1,23.

Teniendo en cuenta que la fórmula Z-Score de Altman presenta buen rendimiento para predecir la bancarrota tanto con un año de antelación, como con dos años de antelación (Altman, 2000), se seleccionaron los datos del último año disponible para

Capítulo 1: Aplicación de técnicas de Inteligencia Artificial para la predicción del estrés financiero en empresas del sector de recogida y tratamiento de aguas residuales

calcular si las empresas presentaban estrés financiero o no y por otra parte, se seleccionaron los datos de dos años antes para utilizarlas como variables predictivas.

1.5.2 Técnicas de Inteligencia Artificial utilizadas

A continuación se detallan sucintamente las técnicas de IA utilizadas en el presente trabajo.

1.5.2.1 Support Vector Machine

Vapnik (1998) fue el primero en utilizar la técnica SVM para realizar clasificaciones binarias. SVMs utilizan un modelo lineal para implementar límites de clases no lineales a través de algunos vectores de entrada de mapeo no lineales en un espacio de entidad de alta dimensionalidad. El modelo lineal construido en el nuevo espacio puede representar un límite de decisión no lineal en el espacio original. En el nuevo espacio, se construye un hiperplano de separación óptimo. Así, SVM se conoce como el algoritmo que encuentra un tipo especial de modelo lineal, el hiperplano de margen máximo. El hiperplano de margen máximo da la máxima separación entre las clases de decisión. Los ejemplos de entrenamiento más cercanos al hiperplano de margen máximo se llaman vectores de soporte. Todos los demás ejemplos de entrenamiento son irrelevantes para definir los límites de la clase binaria.

SVM es lo suficientemente simple como para ser analizado matemáticamente, ya que se puede demostrar que corresponde a un método lineal en un espacio de características de alta dimensional no linealmente relacionado con el espacio de entrada. En este sentido, SVM puede servir como una alternativa prometedora que combina las fortalezas de los métodos estadísticos convencionales que son más teóricos y fáciles de analizar, y más métodos de aprendizaje automático, libres de distribución y robustos (Min y Lee, 2005).

En este estudio, la función kernel utilizada fue la “función radial basis”. Además, hay dos parámetros asociados a la “función radial basis” que juegan un papel muy importante en el desarrollo de la ejecución del SVM: C (cost) y γ (gamma). C es el

Capítulo 1: Aplicación de técnicas de Inteligencia Artificial para la predicción del estrés financiero en empresas del sector de recogida y tratamiento de aguas residuales

parámetro de penalización del término de error. Y γ (gamma) es el coeficiente de la función kernel, que en este caso es la “función radial basis”. La incorrecta selección de estos parámetros pueden causar problemas de “*overfitting*” o “*underfitting*” (Hsu *et al.*, 2016; Tay y Cao, 2001).

Para comprobar la importancia de la correcta selección de los dos parámetros, y teniendo en cuenta que no se sabe de antemano cuales van a ser los valores más óptimos para un problema dado (Hsu *et al.*, 2016), el valor C se seleccionó entre los valores 0,1, 1, 10 ó 100 y el valor de gamma entre 0,5, 1 ó 2.

1.5.2.2 Fuzzy Logic

La FL se ha convertido en un enfoque común desde que se presentó por primera vez, cuando Zadeh (1965) publicó su trabajo seminal "*Fuzzy Sets*". La teoría del conjunto difuso proporciona un modelo matemático para emular el proceso cognitivo humano. En comparación con los modelos probabilísticos o estadísticos, la teoría del conjunto difuso busca clasificar el razonamiento subjetivo y asignar grados de posibilidades para llegar a conclusiones (Alam *et al.*, 2000). Proporciona un medio para tomar decisiones basadas en datos ambiguos, imprecisos o incompletos. FL representa modelos o conocimiento usando reglas "*if-then*" en forma de " si A y B, entonces C " (Chen, 2013). Las cualidades de este método incluyen su simplicidad, rendimiento satisfactorio y robustez.

Sin embargo, la construcción de modelos difusos de grandes y complejos sistemas es una tarea difícil que exige la identificación de muchos parámetros y el éxito va a depender de la elección de estos parámetros (Hachicha *et al.*, 2011) y de la función de membresía o pertenencia (membership function). Una forma de resolver este problema es utilizar un método de optimización, en este caso, se ha utilizado la técnica de los AG.

1.5.2.3 Self-organizing maps (SOM)

Este sistema neuronal fue desarrollado por Teuvo Kohonen (1982) (también se conocen como Kohonen Maps). Ha demostrado su eficacia en dominios reales, incluyendo *clustering*, reconocimiento de patrones, reducción de dimensiones y la

extracción de características. El modelo SOM es una red neuronal no supervisada (*unsupervised*) y se compone de dos capas neuronales. La capa de entrada tiene tantas neuronas como variables, y su función es simplemente capturar la información. La capa de salida, tiene forma rectangular con “x” filas e “y” columnas (dos dimensiones). Cada neurona en la capa de entrada está conectada a cada neurona en la capa de salida a través de los pesos sinápticos. Por lo tanto, cada neurona en la capa de salida tiene n conexiones a la capa de entrada. Cada una de estas conexiones tiene un peso sináptico asociado con ella.

SOM intenta proyectar el espacio de entrada multidimensional, que en nuestro caso podría ser información financiera, en el espacio de salida de tal manera que los patrones de entrada cuyas variables presenten valores similares aparezcan cercanos entre sí en el mapa que se crea. Cada neurona aprende a reconocer un tipo específico de patrón de entrada. Las neuronas que están cerca en el mapa reconocerán patrones de entrada similares cuyas imágenes por lo tanto, aparecerán cerca unas de otras en el mapa creado. De esta manera, la topología esencial del espacio de entrada se conserva en el espacio de salida (Serrano-Cinca, 1996). De este modo, es posible establecer en un mapa bidimensional diferentes zonas tales empresas activas y no activas y crear trayectorias de quiebra (López Iturriaga y Sanz, 2015).

1.5.2.4 Classifier ensembles

A la combinación de múltiples clasificadores se le llama “*classifier ensembles*”. El concepto de “*ensemble*” apareció en la literatura de la clasificación en 1965 (Nilsson, 1965). Sin embargo, es una de las últimas tendencias en las líneas de investigación asociada a las técnicas de clasificación (Lin *et al.*, 2012) (e.g. Oza y Tumer, 2007, Ranawana y Palade, 2006, Rokach, 2009). Tiene como objetivo la obtención de clasificadores de mayor exactitud combinando técnicas menos precisas. Básicamente se proponen para mejorar el rendimiento de clasificación de un único clasificador (Kittler *et al.*, 1998), ya que la combinación compensa los errores cometidos por clasificadores individuales en diferentes partes del espacio de entrada (Lin *et al.*, 2012).

El concepto es muy sencillo, un número determinado de clasificadores comparten los datos de entrada (inputs) y sus respectivos datos de salida (“outputs”) son combinados

Capítulo 1: Aplicación de técnicas de Inteligencia Artificial para la predicción del estrés financiero en empresas del sector de recogida y tratamiento de aguas residuales

(media aritmética, *voting*, *bagging* o *boosting*) para crear un único “output” (Tsai *et al.*, 2014). De esta forma, unas técnicas se complementan con otras y se obtienen mejores soluciones y más robustas (Oza y Tumer, 2007).

1.5.2.5 Técnica de los Algoritmos Genéticos

Los AG simulan la teoría de la evolución propuesta por Darwin (Gordini, 2014). Fueron desarrollados por Holland (1975) y son procesos estocásticos robustos que pueden usarse para selección binaria de variables o para resolver problemas de búsqueda y optimización a través de una función *fitness* o función objetivo. Esta función, es una función de un cromosoma cuyo valor siempre se utiliza para medir el rendimiento de un grupo de parámetros asociados al cromosoma correspondiente y que siempre se debe maximizar. La función de *fitness* se puede elegir en función de las necesidades de los objetivos de negocio (Zhou *et al.*, 2012).

Los AG realizan el proceso de búsqueda en cuatro pasos: inicialización, selección, cruce y mutación (Shin y Lee, 2002).

En el proceso de inicialización, una población de individuos, llamados cromosomas, se distribuyen de forma aleatoria por el rango de búsqueda definido previamente para cada variable y son evaluados a través de la función *fitness* (Gordini, 2014).

El proceso de selección se basa en el principio de supervivencia de los individuos mejor adaptados, es decir, aquellos cromosomas que presenten valores más óptimos en la función *fitness*, serán seleccionados para la crear la siguiente generación (Gordini, 2014). En el proceso de cruce, se seleccionan dos individuos con valores óptimos en la función *fitness* y se unen de manera que el descendiente contiene información parcial de cada uno de sus progenitores, de esta forma, cabe esperar, que el descendiente mejore el resultado de la función *fitness* obtenido por sus progenitores (Shin y Lee, 2002).

La mutación en los AG es un mecanismo donde de forma aleatoria se seleccionan miembros de la población y se cambia alguno de sus genes. De esta forma, se evitan máximos locales, ya que se desplazan valores numéricos hacia zonas del espacio de

búsqueda que no pueden ser alcanzadas por los otros operadores genéticos (Shin y Lee, 2002).

En este análisis, se han utilizado los AG para selección binaria de variables a través de modelos de regresión logística (Scrucca, 2013), las variables que sean seleccionadas en este paso, serán las que se utilicen posteriormente con las técnicas de IA. Siedlecki y Sklansky (1989) fueron los primeros que propusieron la utilización de los AG para la selección de variables, pero no se han utilizado como una técnica predictiva.

1.5.3 Análisis estadístico

A través de una revisión de la bibliografía (Altman, 1968; Ciampi *et al.*, 2009; Ciampi y Gordini, 2013; Fedorova *et al.*, 2013; Yazdanfar y Nilsson, 2008; Zmijewski, 1984), se seleccionaron inicialmente 38 ratios financieros (10 fueron eliminados al no estar informados en su totalidad y de esta forma no reducir el tamaño muestral).

En un primer momento, se comprobó el rendimiento de cada variable basándose en el análisis de “*receiver operating characteristic curve*” (curvas ROC). Se calculó para cada uno de los ratios la curva ROC, donde la variable dicotómica era el estrés financiero. La curva ROC representa de forma gráfica el porcentaje de empresas activas clasificadas incorrectamente (1-Especificidad) en el eje de abscisas y el porcentaje de empresas en situación de estrés clasificadas correctamente (Sensibilidad) (Barandela *et al.*, 2003). Para evaluar el rendimiento de cada variable en las curvas ROC se utilizó el área bajo la curva ROC (AUROC), que equivale a la probabilidad de que si se eligiesen al azar a dos empresas, una en situación de estrés y otra sin estrés, la prueba las clasificase a ambas correctamente. Si el AUROC es 1, significa que es una variable perfecta para predecir. Generalmente, una variable tiene un AUROC entre 0,5 y 1, y cuanto más cerca de 1, mayor poder discriminatorio tendrá (Kim y Kang, 2012).

En la Tabla 1.2 se muestra la estadística descriptiva y el AUCROC de los ratios (ordenados de forma descendente en relación al área bajo la curva ROC).

Capítulo 1: Aplicación de técnicas de Inteligencia Artificial para la predicción del estrés financiero en empresas del sector de recogida y tratamiento de aguas residuales

Tabla 1.2 Estadística descriptiva de las ratios financieras y selección de variables

Variable	N	Media (Dt)	Mediana (P25 – P75)	AUCROC
Ventas/Pasivo ^{b, c}	R35 131	2.74(4.04)	1.64(0.75-2.83)	0.85
Cash Flow/Deuda total ^{b, c}	R16 131	0.91(2.07)	0.18(0.06-0.69)	0.82
Ingresos explotación/ Total activo ^a	R05 131	0.96(0.75)	0.84(0.52-1.22)	0.77
CC/Total activo	R07 131	0.12(0.43)	0.14(-0.04-0.4)	0.77
PN/Pasivo ^a	R08 131	3.02(6.92)	0.56(0.16-1.96)	0.76
PT/AT	R10 131	0.63(0.4)	0.64(0.34-0.86)	0.76
Total deuda/Total deuda + PN	R17 131	0.63(0.4)	0.64(0.34-0.86)	0.76
SOLV	R41 131	4.02(6.92)	1.56(1.16-2.96)	0.76
Reservas/Total activo ^a	R06 131	0.27(0.35)	0.13(0-0.46)	0.75
AC/PC	R11 131	3.26(4.62)	1.67(0.92-3.87)	0.75
Test acido	R18 131	3.14(4.63)	1.46(0.71-3.75)	0.75
Quick Ratio (AC-exis/PC)	R24 131	3.15(4.63)	1.46(0.72-3.75)	0.75
Tesorería	R39 131	1.59(3.04)	0.37(0.1-1.73)	0.75
Cash Flow/AT ^{a, b, c}	R30 131	0.06(0.15)	0.06(0.01-0.11)	0.74
Tesorería/Total activo	R13 131	0.12(0.14)	0.07(0.02-0.17)	0.73
CBF	R45 131	1.57(3.66)	1.09(0.85-1.49)	0.73
Resl/AT	R09 131	0.01(0.15)	0.02(0-0.05)	0.72
CC (AC-PC) ^c	R37 131	205.28(3214.08)	63.36(-8.74-267.5)	0.71
Pasnocial ^c	R43 131	3692.81(25110.67)	95.51(27.74-421.58)	0.71
Pasivo/PN ^c	R01 131	8.93(51.2)	0.99(0.19-2.79)	0.70
RE	R04 131	0.03(0.18)	0.04(0-0.08)	0.70
EBITDA/Total activo	R14 131	0.06(0.18)	0.08(0.01-0.12)	0.68
Rexplo	R28 131	0.06(0.61)	0.04(0-0.09)	0.67
Cash Flow	R29 131	295.62(1306.62)	23.09(3.9-122.48)	0.64
Gastos Personal/Activo total ^b	R47 131	0.35(0.37)	0.27(0.08-0.47)	0.64
PC/ Pasivo	R02 131	0.66(0.33)	0.74(0.34-1)	0.62
Resultado explotación	R27 131	170.15(1316.24)	12.96(-0.03-77.38)	0.62
RAIT	R03 131	215.71(1406.48)	13.62(0.3-83.44)	0.61
EBITDA	R40 131	370.84(1614.25)	28.89(4.87-116.71)	0.6
PASNC/EBITDA	R46 131	1614.84(18786.6)	1.5(0-5)	0.6
Ventas/CC	R36 131	-0.88(52.26)	1.65(-0.26-4.32)	0.58
Deudas CP/PN	R12 131	0.54(4.84)	0.04(0-0.27)	0.57
Activos explotación	R26 131	6074.49(35371.76)	355.47(152.86-1610.38)	0.57
Ventas/AC	R34 131	2.19(2.9)	1.65(1.01-2.8)	0.57
Act. Extrf	R42 131	521.55(2146.75)	5.9(0-105.44)	0.57
RF	R25 131	-1.1(14.2)	0.04(-0.01-0.12)	0.55
Cash Flow/PN	R31 131	-0.88(13.49)	0.14(0.01-0.3)	0.52
NOF	R38 131	314.78(1383.52)	32(-1.14-172.83)	0.51
EBITDA/Gastos financieros	R15 114	160.4(1290.47)	6.31(1.22-15.38)	0
Interest/Turn over	R19 124	0.03(0.06)	0.01(0-0.03)	0
Equity/Long term material asset	R20 114	79.99(828.53)	1.03(0.35-3.03)	0
Net financial position /Turn over	R21 124	0.57(2.94)	0.16(0.06-0.35)	0
Long term assests/n2 employees	R22 114	-14.63(87.32)	-1.24(-3.07--0.53)	0
Interes charges/bank loan	R23 94	0.16(0.38)	0.05(0.03-0.1)	0
Ebitda/Ventas	R32 124	0.07(0.38)	0.1(0.04-0.18)	0
Ventas/ANC	R33 128	25.81(228.47)	1.86(0.73-5.02)	0
Solvextrf	R44 121	28.24(260.6)	0.05(0-0.47)	0
Gastos Personal/Importe neto cif. negoc.	R48 124	0.39(0.35)	0.36(0.19-0.5)	0

^a Ratios del modelo 1.

^b Ratios del modelo 2.

^c Ratios del modelo 3.

Se seleccionaron aquellas ratios cuyo AUROC fuese igual o superior a 0,7 (21 ratios) y posteriormente se realizó una regresión logística donde se utilizó la técnica de los AG (empleando la clasificación binaria, type="binary") para seleccionar aquellas ratios que minimizasen el error "Akaike information criterion" (AIC), por ello, la función fitness utilizada fue "-AIC". En este modelo, fueron seleccionados 13 ratios (R35, R05, R07, R10, R17, R41, R08, R06, R11, R18, R30, R13 y R04). Teniendo en cuenta que algunos estudios han demostrado que las técnicas de IA pueden ver afectado su rendimiento por problemas de multicolinealidad (e.g. Eom *et al.*, 2008, Dong y Han,

2004), se analizó el factor de inflación de la varianza (*variance information factors*, VIF) obtenido sobre las trece variables seleccionadas. Debido a que cinco de ellas presentaban valores superiores a 5 (Field *et al.*, 2012; Kim y Kang, 2012), se fueron excluyendo de forma gradual para garantizar un modelo sin problemas de multicolinealidad. Finalmente fueron seleccionadas cuatro variables (R05, R06, R08 y R30), las cuales, forman el modelo 1.

Como tres de las cuatro ratios seleccionadas en el modelo 1, formaban parte de la fórmula Z' -Score (véase fórmula 1.1), y se podría interpretar que la variable dependiente (estrés financiero) pudiera ser una combinación lineal de estas tres ratios, a pesar de contener información de dos años anteriores a los utilizados en la fórmula Z' -Score, se crearon dos modelos adicionales sin tener en cuenta ninguna de las ratios utilizadas en la fórmula Z' -Score.

Para obtener las variables de este segundo modelo, se volvió a ejecutar el proceso de la regresión logística con la técnica de selección de variables de los AG incluyendo aquellas cuya AUCROC fuese igual o superior a 0,70, pero sin tener en cuenta a las cinco ratios financieras utilizados en la fórmula Z' -Score (R04, R05, R06, R07 y R08) utilizándose como función *fitness* “-BIC”. Se obtuvieron 3 ratios (R35, R16, R30) sin problemas de multicolinealidad, los cuales forman el modelo 2.

Por último, se volvió a repetir el paso anterior, pero en este caso, la función *fitness* utilizada fue “-AIC”. Se obtuvieron 6 ratios (R35, R16, R30, R37, R43 y R01) sin problemas de multicolinealidad (véase Tabla 1.2). De esta forma, se puede comprobar qué modelo genera un error menor y por lo tanto, concluir cual de ellos presenta mejor rendimiento.

Una vez establecidos ambos modelos, se utilizaron las técnicas de IA comentadas previamente: SVM, CE, FL y SOM para establecer cuál de las técnicas y modelos empleados presentaba mejor rendimiento.

Con la finalidad de obtener unos resultados robustos, y no debidos al azar, el proceso de creación de ficheros de entrenamiento y validación, sus correspondientes ejecuciones y predicciones se repitió 1.000 veces tanto para el modelo 1 como para el modelo 2). El resultado final se calculó como el valor promedio obtenido en las citadas 1.000 ejecuciones.

El programa estadístico utilizado ha sido R Core Team, 2017. Los paquetes utilizados han sido: pROC (Robin *et al.*, 2011), GA (Scrucca, 2013), e1071 (Meyer, 2001), hybridEnsemble (Package “hybridEnsemble”, 2015), Kohonen (Wehrens y Buydens, 2007) y ggplot2 (Wickham, 2009).

1.5.4 Medidas de Rendimiento

Para comprobar el rendimiento del proceso se utilizó una matriz de confusión (Kim *et al.*, 2016) como se muestra en la Tabla 1.3. Los Verdaderos Positivos (VP) indican el número de empresas en situación de estrés correctamente clasificadas. Los Falsos Positivos (FP) indican el número de empresas sin estrés financiero incorrectamente clasificadas. Los Falsos Negativos (FN) indican el número de empresas en estrés financiero incorrectamente clasificadas y los Verdaderos Negativos (VN) indican el número de empresas sin estrés correctamente clasificadas. La Sensibilidad se define como $VP/(VP + FN)$, la especificidad como $VN / (VN + FP)$, la exactitud (*accuracy*) es la proporción de empresas correctamente clasificadas, i.e., $(VP + VN) / (VP + FN + FP + VN)$ (Altman y Bland, 1994) y la media geométrica se define como $\sqrt{Sensibilidad * Especificidad}$, dando la misma importancia a la sensibilidad y a la especificidad.

Tabla 1.3 Matriz de confusión

Actual	Predicción	
	Estrés	No-estrés
Estrés	Verdaderos Positivos (VP)	Falsos Negativos (FN)
No-estrés	Falsos Positivos (FP)	Verdaderos Negativos (VN)

1.6 Resultados

Tal y como se indica en la metodología, se utilizaron 4 técnicas de IA para procesar los datos: a. SVM, b. CE, c. FL y d. SOM y tres modelos predictivos con distintas ratios. Se utilizó un porcentaje del 80-20% para crear los ficheros de entrenamiento (training) y validación (test). El objetivo era evaluar el rendimiento de las distintas técnicas predictivas aplicadas en los modelos para comprobar cuál es la mejor

técnica y modelo. Se realizaron 1.000 iteraciones y en cada una de ellas se realizó una selección aleatoria de los registros de la base de datos para crear el fichero de entrenamiento (training dataset) y el fichero de validación (test dataset). Los resultados del fichero de validación de cada una de las 1.000 iteraciones, se iban acumulando en una matriz de confusión (véase Tabla 1.3) para cada una de las cuatro técnicas y cada uno de los tres modelos. Al finalizar el proceso, se calcularon la sensibilidad, especificidad y exactitud con los datos acumulados en cada una de las matrices de confusión. La Tabla 1.4 muestra las medidas de rendimiento obtenidas con el fichero de validación (test). Se puede observar, el buen rendimiento de todas las técnicas (excepto la técnica SVM del modelo 3 (78,21%), el resto de técnicas y modelos presentan una exactitud superior al 80%) y en particular, el excelente rendimiento de la técnica SOM, con una exactitud del 96,00%, 98,15% y 95,67% en los modelos 1, 2 y 3, respectivamente. Teniendo en cuenta que una de las técnicas que forman el CE en este estudio es el SVM, se podría presuponer que ésta arroja mejores resultados. Esto es así en los modelos 2 y 3 donde la técnica CE mejora el resultado de las técnicas individuales (84,17% vs 82,59% para el modelo 2 y 82,09% vs 78,21% para el modelo 3). Sin embargo, este hecho no sucede en el modelo 1, donde la exactitud de la técnica SVM es del 86,11% frente al 84,90% de la técnica CE. Para intentar explicar este resultado, se analizan los valores de la sensibilidad y especificidad, observándose que la primera es 4,71 puntos superior en la técnica SVM respecto de CE (95,78% vs 91,07%). Sin embargo, la sensibilidad es 7,11 puntos superior en la técnica CE (70,17% vs 63,06%), pero dado que la prevalencia del análisis es del 29,54, adquiere más peso la especificidad que la sensibilidad y esto hace que la exactitud sea mayor en el modelo que presenta mayor especificidad. Para confirmar esta situación, se observa que la media geométrica de la sensibilidad y especificidad es superior en la técnica CE que en la de SVM.

Siguiendo con el análisis de sensibilidad y especificidad, se puede observar que en todos los casos en que la especificidad es superior a la sensibilidad, el valor mínimo de la especificidad es 83,84% (modelo 3, técnica CE), frente a un 42,34%, que es el valor mínimo de la sensibilidad (modelo 3, técnica SVM), por lo que se puede afirmar, que independientemente del modelo seleccionado, es un proceso más específico que sensible.

Tabla 1.4 Medidas de Rendimiento de los datos de validación

Medidas Rendimiento	Prevalencia	Sensibilidad	Especificidad	Exactitud	Media Geom.
Mod1-SVM	29,54	63,06	95,78	86,11	77,72
Mod1-CE	29,54	70,17	91,07	84,90	79,94
Mod1-FL	29,54	67,55	88,02	81,97	77,11
Mod1-SOM	29,54	89,54	98,71	96,00	94,01
Mod1-Promedio	29,54	72,58	93,40	87,25	82,33
Mod2-SVM	29,54	66,42	89,30	82,54	77,01
Mod2-CE	29,54	73,65	88,58	84,17	80,77
Mod2-FL	29,54	71,89	86,86	82,44	79,02
Mod2-SOM	29,54	97,77	98,31	98,15	98,04
Mod2-Promedio	29,54	77,43	90,76	86,83	83,83
Mod3-SVM	29,54	42,34	93,24	78,21	62,83
Mod3-CE	29,54	77,92	83,84	82,09	80,83
Mod3-FL	29,54	69,65	85,49	80,81	77,16
Mod3-SOM	29,54	93,01	96,78	95,67	94,88
Mod3-Promedio	29,54	70,73	89,84	84,20	79,71

Si se observa los promedios de las medidas de rendimiento, se puede comprobar que la sensibilidad obtenida con modelo 2 es superior a la de los otros dos modelos (77,43%, 72,58% y 70,73% respectivamente). La especificidad del modelo 1 es superior al modelo 2 y 3 (93,40%, 90,76% y 89,84%). La exactitud del modelo 1 es superior al modelo 2 (87,25% vs 86,83%) y por el contrario, la media geométrica del modelo 2 es superior al modelo 1 (83,83% vs 82,33%). Así pues, el modelo 1 es más específico y el modelo 2 es más sensible. Por otra parte, el modelo que peores medidas de rendimiento presenta es el modelo 3.

La Tabla 1.5 muestra las comparaciones dos a dos de la exactitud entre todas las técnicas con la corrección de Bonferroni (Armstrong, 2014) únicamente entre el modelo 1 y modelo 2 (no se incluye el modelo 3 ya que sin lugar a dudas es el peor modelo predictivo y de esta forma, no se penaliza aun más el p-valor). Se puede observar que las diferencias significativas se aprecian entre las técnicas FL con SOM, en ambos modelos. Además, la técnica SOM del modelo 2, presenta diferencias con las técnicas SVM y CE de ese mismo modelo.

Tabla 1.5 Comparaciones pareadas de la exactitud con la corrección de Bonferroni

	Mod1 SVM	Mod1 CE	Mod1 FL	Mod1 SOM	Mod2 SVM	Mod2 CE	Mod2 FL
Mod1-CE	1,00	-	-	-	-	-	-
Mod1-FL	1,00	1,00	-	-	-	-	-
Mod1-SOM	0,77	0,42	0,09	-	-	-	-
Mod2-SVM	1,00	1,00	1,00	0,12	-	-	-
Mod2-CE	1,00	1,00	1,00	0,29	1,00	-	-
Mod2-FL	1,00	1,00	1,00	0,11	1,00	1,00	-
Mod2-SOM	0,10	0,05	0,01	1,00	0,01	0,03	0,01

^a Los cálculos se han realizado en términos porcentuales, no con los valores brutos obtenidos en la matriz de confusión del proceso ejecutado.

La Figura 1.1 muestra que el modelo 3 es el que realiza peores predicciones. Por otra parte, el modelo 1 presenta mayor exactitud en las técnicas SVM y CE y sin embargo, el modelo 2 es superior en las técnicas FL y SOM.

En resumen, la mejor técnica predictiva es SOM en el modelo 2. Por otra parte, el modelo 1 va a realizar predicciones más exactas y específicas que el resto de modelos y el modelo 2, es más sensible que el resto de modelos. También se puede concluir que el modelo predictivo construido minimizando el error “BIC” (modelo 2) presenta mejor rendimiento que el modelo construido minimizando el error “AIC” (modelo 3).

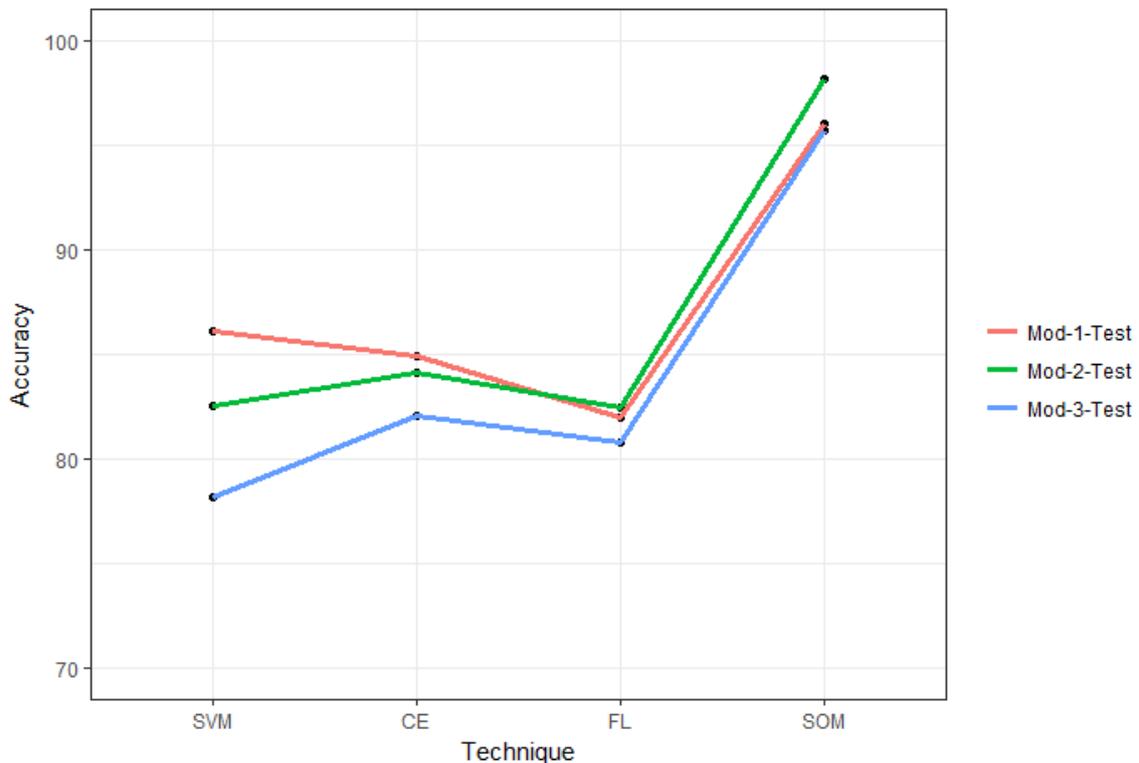


Figura 1.1. Comparación de la exactitud entre técnicas y modelos

En las siguientes figuras, se representan los mapas auto-organizados no supervisados (*unsupervised learning self-organizing map*). Al ser un mapa no supervisado, significa que no se tiene en cuenta a la variable dependiente, en este caso estrés financiero, de tal forma, que la red auto-organizada debe descubrir por sí misma rasgos comunes, correlaciones o categorías en los datos de entrada, e incorporarlos a su estructura interna de conexiones, de tal forma que cuando se presente a la red un patrón de entrada, solo una de las neuronas de salida se active, de tal forma que las

informaciones similares sean clasificadas formando parte de la misma categoría y por tanto, activando la misma neurona de salida. De esta forma, las neuronas de salida tienen asociado un vector de pesos llamado vector de referencia o “*codebook vectors*” el cual constituye el vector prototipo o promedio de la categoría a la que representa. Estos vectores de referencia, son los que se representan en las figuras que se muestran seguidamente.

En la Figura 1.2 se observa el SOM (unsupervised learning) asociado al modelo 1. En las empresas sin estrés financiero se puede observar como la ratio R06 está más presente que en las que sí presentan estrés financiero, esto significa que los “*codebook vectors*” toman valores similares para esta ratio, pero valores diferentes para el resto de ratios. Con el resto de ratios (R05, R08 y R30), sucede lo contrario, están más presentes en las empresas que sí presentan estrés financiero.

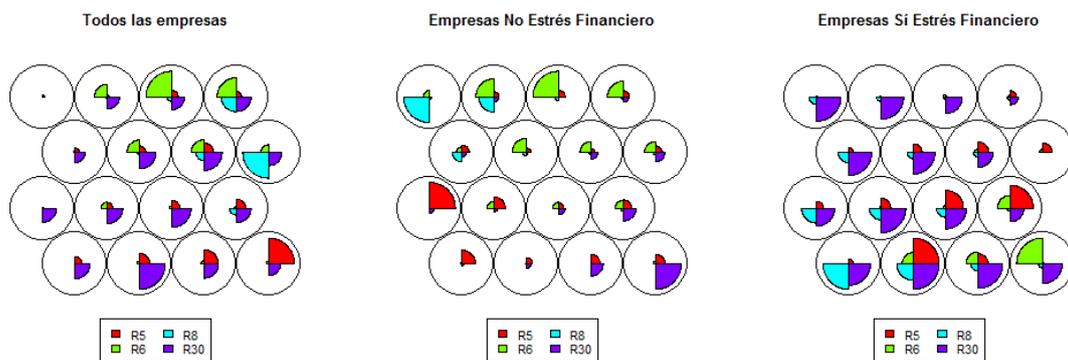


Figura 1.2. Modelo 1. SOM con todos los registros de la muestra y desglosado por estado (unsupervised learning)

En la Figura 1.3 se muestra el SOM (*unsupervised learning*) obtenido con el modelo 2. En las empresas sin estrés financiero se puede observar como la ratio R35 presenta semejanza con la ratio R16, la cual es representada en la neurona situada en el borde superior derecho. En las empresas con estrés financiero, las semejanzas entre las tres ratios se representan en la parte superior derecha.

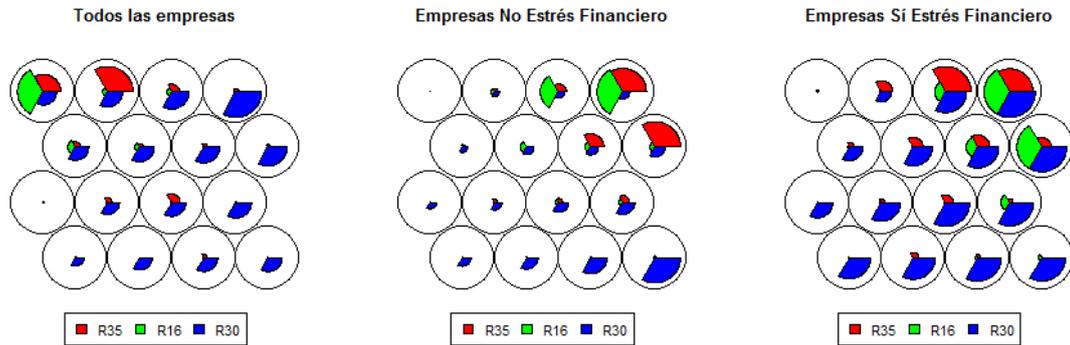


Figura 1.3. Modelo 2. SOM con todos los registros de la muestra y desglosado por estado (unsupervised learning)

En la Figura 1.4 se presenta el SOM (*unsupervised learning*) vinculado al modelo 3. En las empresas sin estrés financiero se puede observar como la ratio R35 se asocia con la ratio R16 y R01 en el borde inferior izquierdo y la ratio R37, R43 y R1 en el borde inferior derecho. Por otro lado, en las empresas con estrés financiero se asocian las ratios R30, R37 y R35 en la parte inferior derecha. También hay que destacar que la ratio R30 está muy presente en el mapa de empresas con estrés financiero y apenas aparece en el mapa de empresas que no presentan estrés financiero.

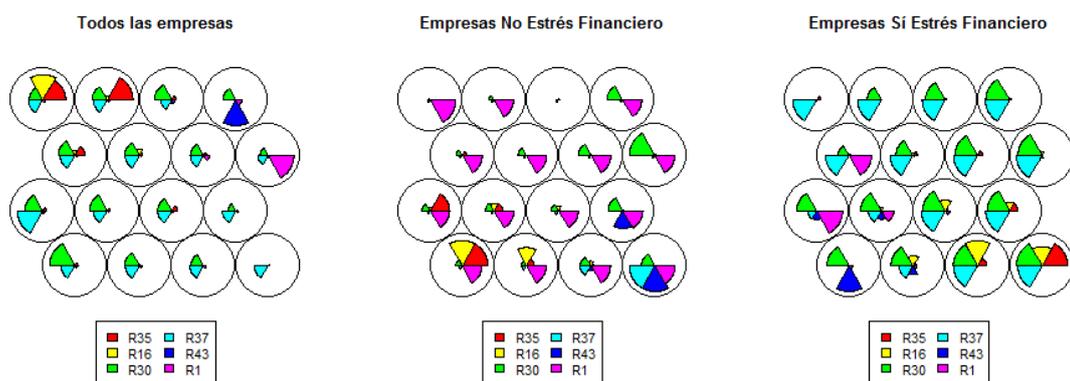


Figura 1.4. Modelo 3. SOM con todos los registros de la muestra y desglosado por estado (unsupervised learning)

1.7 Discusión

Debido a la cada vez más importante presencia de las empresas privadas en el sector de recogida y tratamiento de aguas residuales y teniendo en cuenta la falta en la literatura de modelos predictivos financieros para este sector, se han propuesto la creación de diferentes modelos predictivos que sirvan para dar estabilidad financiera. De esta forma, se pretende ampliar el conocimiento del comportamiento financiero, ya que desde un punto de vista técnico, hay bastante literatura al respecto.

La principal contribución de este estudio es la elaboración de tres modelos predictivos, dos de los cuales, modelo 1 y modelo 2, presentan un gran rendimiento para predecir el estrés financiero con dos años de antelación en estas empresas. También se puede observar que el modelo 2 es más sensible y presenta mayor media geométrica que el modelo 1 y por otra parte, el modelo 1 presenta mayor especificidad y exactitud.

La otra gran aportación está relacionada con las técnicas empleadas. Así, sin lugar a dudas, la técnica que mejores predicciones realiza es la SOM, es un modelo muy sensible y específico al mismo tiempo, seguido de CE. Por último, cabe destacar que el modelo obtenido minimizando el error “BIC” presenta mejor rendimiento que el obtenido minimizando el error “AIC”, por lo que se puede afirmar que utilizando el error BIC para elaborar modelos predictivos, se obtienen mejores resultados que utilizando el error AIC. Como consecuencia de los buenos resultados obtenidos, se puede predecir con suficiente antelación las empresas que se encuentran en un riesgo elevado de entrar en bancarrota. Ello supone que tanto los *stakeholders* como los gestores de la propia empresa, cuentan con tiempo suficiente para tomar las medidas oportunas. Con esto se puede reducir la probabilidad de problemas posteriores de insolvencias con el consecuente abandono de la prestación del servicio, con las correspondientes consecuencias para las personas que reciben el servicio, las empresas que lo contratan (evitando tener que intervenir para garantizar el servicio), la propia empresa y hasta el propio medio ambiente. Por otra parte, aquellas empresas que no presenten riesgo financiero, seguirán teniendo la confianza de los acreedores y podrán obtener sin problemas la financiación para realizar las inversiones necesarias en infraestructuras o innovación o tener la confianza de quien los contrate.

Finalmente, señalar que las limitaciones de este estudio radican en el reducido tamaño muestral de la base de datos. Por otra parte, el uso de los AG no garantiza que se encuentre la solución óptima del problema. Estas soluciones dependerán del rango de búsqueda aplicado para el cálculo de los coeficientes, así como el tamaño de la población, la ratio de cruce y la ratio de mutación aplicados en el modelo, ya que estos parámetros pueden afectar a las soluciones obtenidas. Por otra parte, el principal inconveniente de técnicas como las NN, SVM o RF es que son cajas negras y es muy difícil explicar cómo realizan sus predicciones. En cuanto a futuras líneas de investigación, cabe destacar que podrían utilizarse otras técnicas de “*classifiers ensemble*” como *bagging* o *boosting* junto a técnicas de inteligencia artificial como las RF o SVM, o técnicas de “*hybrid classifiers*” con la intención de mejorar las predicciones y reducir los errores.

1.8 Referencias

- AEAS. (2014). XIII Encuesta de suministro de agua potable y saneamiento en España.
- Alam, P., Booth, D., Lee, K., & Thordarson, T. (2000). The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study. *Expert Systems with Applications*, 18(3), 185-199. [https://doi.org/10.1016/S0957-4174\(99\)00061-5](https://doi.org/10.1016/S0957-4174(99)00061-5)
- Almamy, J., Aston, J., & Ngwa, L. N. (2016). An evaluation of Altman's Z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence from the UK. *Journal of Corporate Finance*, 36, 278-285. <https://doi.org/10.1016/j.jcorpfin.2015.12.009>
- Altman. (2000). Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA ® Models Background. *Stern School of Business, New York University*, 1-54.
- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 1 : sensitivity and specificity. *Bmj*, 308(6943), 1552. <https://doi.org/10.1258/phleb.2012.012J05>
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Altman, E. I. (1983). *Corporate financial distress : a complete guide to predicting, avoiding, and dealing with bankruptcy* (3rd Editio). Wiley.
- Altman, E. I., Sabato, G., & Wilson, N. (2010). The Value of non-financial information in small-sized enterprise risk management. *The Journal of Credit Risk*, 6(2), 1-33.
- Angelini, E., di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance*, 48(4), 733-755. <https://doi.org/10.1016/j.qref.2007.04.001>
- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, 34(5), 502-508. <https://doi.org/10.1111/opo.12131>
- Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks*, 12(4), 929-935.

- <https://doi.org/10.1109/72.935101>
- Aziz, A., Emanuel, D., & Lawson, G. (1988). *Journal of management studies. Journal of Management Studies* (Vol. 25). Blackwell Publishers.
- Barandela, R., Sánchez, J. S., García, V., & Rangel, E. (2003). Strategies for learning in class imbalance problems. *Pattern Recognition*, 36(3), 849-851. [https://doi.org/10.1016/S0031-3203\(02\)00257-1](https://doi.org/10.1016/S0031-3203(02)00257-1)
- Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 40, 432-442. <https://doi.org/10.1016/j.jbankfin.2013.12.013>
- Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Source Journal of Accounting Research Empirical Research in Accounting: Selected Studies*, 4, 71-111.
- Becchetti, L., & Sierra, J. (2003). Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of Banking & Finance*, 27(11), 2099-2120. [https://doi.org/10.1016/S0378-4266\(02\)00319-9](https://doi.org/10.1016/S0378-4266(02)00319-9)
- Beynon, M. J., & Peel, M. J. (2001). Variable precision rough set theory and data discretisation: an application to corporate failure prediction. *Omega*, 29(6), 561-576. [https://doi.org/10.1016/S0305-0483\(01\)00045-7](https://doi.org/10.1016/S0305-0483(01)00045-7)
- Blanco-Oliver, A., Irimia-Dieguez, A., Oliver-Alfonso, M., & Wilson, N. (2015). Improving Bankruptcy Prediction in Micro-Entities by Using Nonlinear Effects and Non-Financial Variables, (2), 144-167.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Bureau Van Dijk. (2017). SABI. Sistemas de Análisis de Balances Ibéricos.
- Charitou, A., Neophytou, E., & Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, 13(3), 465-497. <https://doi.org/10.1080/0963818042000216811>
- Chen, M.-Y. (2011). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers and Mathematics with Applications*, 62, 4514-4524. <https://doi.org/10.1016/j.camwa.2011.10.030>
- Chen, M.-Y. (2013). A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. *Information Sciences*, 220, 180-195. <https://doi.org/10.1016/j.ins.2011.09.013>
- Chen, N., Ribeiro, B., Vieira, A., & Chen, A. (2013). Clustering and visualization of bankruptcy trajectory using self-organizing map. *Expert Systems With Applications*, 40, 385-393. <https://doi.org/10.1016/j.eswa.2012.07.047>
- Chi, L.-C., & Tang, T.-C. (2006). Bankruptcy Prediction: Application of Logit Analysis in Export Credit Risks. *Australian Journal of Management*, 31(1), 17-27. <https://doi.org/10.1177/031289620603100102>
- Cho, S., Kim, J., & Bae, J. K. (2009). An integrative model with subject weight based on neural network learning for bankruptcy prediction. *Expert Systems with Applications*, 36(1), 403-410. <https://doi.org/10.1016/j.eswa.2007.09.060>
- Chung, C.-C., Chen, T.-S., Lin, L.-H., Lin, Y.-C., & Lin, C.-M. (2016). Bankruptcy Prediction Using Cerebellar Model Neural Networks. *International Journal of Fuzzy Systems*, 18(2), 160-167. <https://doi.org/10.1007/s40815-015-0121-5>
- Chung, H.-M. M., & Tam, K. Y. (1993). A Comparative Analysis of Inductive-Learning Algorithms. *Intelligent Systems in Accounting, Finance and Management*, 2(1), 3-18. <https://doi.org/10.1002/j.1099-1174.1993.tb00031.x>
- Ciampi, F., & Gordini, N. N. (2013). Small Enterprise Default Prediction Modeling

- through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises. *Journal of Small Business Management*, 51(1), 23-45. <https://doi.org/10.1111/j.1540-627X.2012.00376.x>
- Ciampi, F., Vallini, C., Gordini, N., & Benvenuti, M. (2009). Are Credit Scoring Models Able to Predict Small Enterprise Default? Statistical Evidence from Italian Small Enterprises.
- Cielen, A., Peeters, L., & Vanhoof, K. (2004). Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research*, 154(2), 526-532. [https://doi.org/10.1016/S0377-2217\(03\)00186-3](https://doi.org/10.1016/S0377-2217(03)00186-3)
- Cultrera, L., & Brédart, X. (2016). Bankruptcy prediction: the case of Belgian SMEs. *Review of Accounting and Finance*, 15(1), 101-119. <https://doi.org/10.1108/RAF-06-2014-0059>
- Dambolena, I. G., & Khoury, S. J. (1980). American Finance Association Ratio Stability and Corporate Failure. *Source: The Journal of Finance*, 35(4), 1017-1026.
- Deakin, E. B. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 10(1), 167. <https://doi.org/10.2307/2490225>
- Deloitte. (2014). El sector sanitario en España, situación actual y perspectivas de futuro.
- Díaz-Uriarte, R., & Alvarez De Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7:3. <https://doi.org/10.1186/1471-2105-7-3>
- Dietrich, J. R., & Kaplan, R. S. (1982). Empirical Analysis of the Commercial Loan Classification Decision. *Source: The Accounting Review*, 57(1), 18-38.
- Divsalar, M., & Roodsaz, H. (2012). A Robust Data-Mining Approach to Bankruptcy Prediction. *Journal of Forecasting*, 523(March 2011), 504-523.
- Dong, Y.-S. D. Y.-S., & Han, K.-S. H. K.-S. (2004). A comparison of several ensemble methods for text categorization. *IEEE International Conference on Services Computing, 2004. (SCC 2004). Proceedings. 2004*, 0-3. <https://doi.org/10.1109/SCC.2004.1358033>
- du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications*, 75, 25-43. <https://doi.org/10.1016/j.eswa.2017.01.016>
- Du Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, 242(1), 286-303. <https://doi.org/10.1016/j.ejor.2014.09.059>
- Dugan, M. T., & Zavgren, C. (1989). How a Bankruptcy Model Could Be Incorporated as an Analytical Procedure. *The CPA journal*, 94-65.
- Eom, J.-H., Kim, S.-C., & Zhang, B.-T. (2008). AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction. *Expert Systems with Applications*, 34(4), 2465-2479. <https://doi.org/10.1016/j.eswa.2007.04.015>
- Etheridge, H. L., Sriram, R. S., & Hsu, H. Y. K. (2000). A Comparison of Selected Artificial Neural Networks that Help Auditors Evaluate Client Financial Viability. *Decision Sciences*, 31(2), 531-550. <https://doi.org/10.1111/j.1540-5915.2000.tb01633.x>
- Fan, A., & Palaniswami, M. (2000). Selecting bankruptcy predictors using a support vector machine approach. En *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium* (pp. 354-359 vol.6). IEEE. <https://doi.org/10.1109/IJCNN.2000.859421>
- Fedorova, E., Gilenko, E., & Dovzhenko, S. (2013). Bankruptcy prediction for Russian

- companies: Application of combined classifiers. *Expert Systems with Applications*, 40(18), 7285-7293. <https://doi.org/10.1016/j.eswa.2013.07.032>
- Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.
- FitzPatrick, P. (1932). *A comparison of the ratios of successful industrial enterprises with those of failed companies*. [Washington].
- Fritsch, S., & Guenther, F. (2010). neuralnet: Training of neural networks. *The R Journal*, 2(1), 30-38. <https://doi.org/10.1109/SP.2010.25>
- González-Gómez, F., García-Rubio, M. A., & Gonz Alez-Martínez, J. (2014). Beyond the publicprivate controversy in urban water management in Spain. *Utilities Policy*, 31, 1-9. <https://doi.org/10.1016/j.jup.2014.07.004>
- Gordini, N. A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy, 41 *Expert Systems with Applications* § (2014). Elsevier Ltd. <https://doi.org/10.1016/j.eswa.2014.04.026>
- Gu, Z., & Gao, L. (1999). A multivariate model for predicting business failures of hospitality firms. *Tourism and Hospitality Research*, 2(1).
- Guerrini, A., Romano, G., Ferretti, S., Fibbi, D., & Daddi, D. (2016). A Performance Measurement Tool Leading Wastewater Treatment Plants toward Economic Efficiency and Sustainability. *Sustainability*, 8(12), 1250. <https://doi.org/10.3390/su8121250>
- Hachicha, N., Jarboui, B., & Siarry, P. (2011). A fuzzy logic control using a differential evolution algorithm aimed at modelling the financial market dynamics. *Information Sciences*, 181(1), 79-91. <https://doi.org/10.1016/j.ins.2010.09.010>
- Hadipuro, W. (2010). Indonesia ' s Water Supply Regulatory Framework : Between Commercialisation and Public Service ? *Water Alternatives*, 3(3), 475-491.
- Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10), 993-1001. <https://doi.org/10.1109/34.58871>
- Hennawy, R. H. A., & Morris, R. C. (1983). The significance of base year in developing failure prediction models. *Journal of Business Finance & Accounting*, 10(2), 209-223. <https://doi.org/10.1111/j.1468-5957.1983.tb00424.x>
- Holland, J. H. (1975). *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press.
- Hornik, K., Buchta, C., & Zeileis Wirtschaftsuniversität Wien, A. W. (2009). Open-Source Machine Learning: R Meets Weka. <https://doi.org/10.1007/s00180-008-0119-7>
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2016). *A Practical Guide to Support Vector Classification*.
- Hua, Z., Wang, Y., Xu, X., Zhang, B., & Liang, L. (2007). Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications*, 33(2), 434-440. <https://doi.org/10.1016/j.eswa.2006.05.006>
- Jeong, C., Min, J. H., & Kim, M. S. (2012). A tuning method for the architecture of neural network models incorporating GAM and GA as applied to bankruptcy prediction. *Expert Systems with Applications*, 39(3), 3650-3658. <https://doi.org/10.1016/j.eswa.2011.09.056>
- Jones, F. L. (1987). Current Techniques in Bankruptcy Prediction. *Journal of Accounting Literature*, 6, 131-164.
- Joo, H. N., & Jin, T. (2000). Bankruptcy Prediction: Evidence from Korea Listed Companies During the IMF Crisis. *Journal of International Financial Management*

- and Accounting*, 11(3), 195-217.
- Kaski, S., Sinkkonen, J., & Peltonen, J. (2001). Bankruptcy analysis with self-organizing maps in learning metrics. *IEEE Transactions on Neural Networks*, 12(4), 936-947. <https://doi.org/10.1109/72.935102>
- Khan, A. M. (1985). Analyzing financial statements for managerial performance measurement and bankruptcy prediction. *Engineering Management International*, 3(3), 165-174. [https://doi.org/10.1016/0167-5419\(85\)90003-1](https://doi.org/10.1016/0167-5419(85)90003-1)
- Kim, H.-J. J., Jo, N.-O. O., & Shin, K.-S. S. (2016). Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. *Expert Systems with Applications*, 59, 226-234. <https://doi.org/10.1016/j.eswa.2016.04.027>
- Kim, H., & Gu, Z. (2006). Predicting Restaurant Bankruptcy: A Logit Model in Comparison with a Discriminant Model. *Journal of Hospitality & Tourism Research*, 30(4), 474-493. <https://doi.org/10.1177/1096348006290114>
- Kim, M.-J. J., & Kang, D.-K. K. (2012). Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction. *Expert Systems with Applications*, 39(10), 9308-9314. <https://doi.org/10.1016/j.eswa.2012.02.072>
- Kittler, J., Hatef, M., Duin, R. P. W., & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 226-239. <https://doi.org/10.1109/34.667881>
- Ko, P.-C., & Lin, P.-C. (2006). An evolution-based approach with modularized evaluations to forecast financial distress. *Knowledge-Based Systems*, 19(1), 84-91. <https://doi.org/10.1016/j.knosys.2005.11.006>
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence*, 2, 1137-1143. <https://doi.org/10.1067/mod.2000.109031>
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59-69. <https://doi.org/10.1007/BF00337288>
- Laitinen, E. K., & Laitinen, T. (2000). Bankruptcy prediction Application of the Taylor's expansion in logistic regression. *International Review of Financial Analysis*, 9, 327-349.
- Lee, C.-H., & Yu, Y.-H. (2012). Characteristics of public-private partnerships for municipal wastewater treatment in Taiwan. *Journal of the Chinese Institute of Engineers*, 35(2), 245-258. <https://doi.org/10.1080/02533839.2012.638535>
- Lennox, C. (1999). Identifying failing companies: a re-evaluation of the logit, probit and DA approaches. *Journal of Economics and Business*, 51(4), 347-364. [https://doi.org/10.1016/S0148-6195\(99\)00009-0](https://doi.org/10.1016/S0148-6195(99)00009-0)
- Li, M. Y. L., & Miu, P. (2010). A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach. *Journal of Empirical Finance*, 17(4), 818-833. <https://doi.org/10.1016/j.jempfin.2010.04.004>
- Liaw, A., Wiener, M., & Andy Liaw, M. (2015). Breiman and Cutler's Random Forests for Classification and Regression Description Classification and regression based on a forest of trees using random inputs.
- Lin, W.-Y., Hu, Y.-H., & Tsai, C.-F. (2012). Machine Learning in Financial Crisis Prediction: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4), 421-436. <https://doi.org/10.1109/TSMCC.2011.2170420>
- López Iturriaga, F. J., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems with*

- Applications*, 42(6), 2857-2868. <https://doi.org/10.1016/j.eswa.2014.11.025>
- Martin, D. (1977). Early warning of bank failure. *Journal of Banking & Finance*, 1(3), 249-276. [https://doi.org/10.1016/0378-4266\(77\)90022-X](https://doi.org/10.1016/0378-4266(77)90022-X)
- Mekonnen, M. M., & Hoekstra, A. Y. (2016). Four billion people facing severe water scarcity. *Science Advances*, 2(2), e1500323-e1500323. <https://doi.org/10.1126/sciadv.1500323>
- Messier, W. F., & Hansen, J. V. (1988). Inducing Rules for Expert System Development: An Example Using Default and Bankruptcy Data. *Management Science*, 34(12), 1403-1415. <https://doi.org/10.1287/mnsc.34.12.1403>
- Meyer, D. (2001). Support Vector Machines * The Interface to libsvm in package e1071, 13(9).
- Michael, S., Georgios, D., Nikolaos, M., & Constantin, Z. (1999). A Fuzzy Knowledge-based Decision Aiding Method for the Assessment of Financial Risk: The Case of Corporate Bankruptcy Prediction. *European Symposium on intelligent techniques*, (February).
- Min, J. H., & Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4), 603-614. <https://doi.org/10.1016/j.eswa.2004.12.008>
- Mousavi, M. M., Ouenniche, J., & Xu, B. (2015). Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. *International Review of Financial Analysis*, 42, 64-75. <https://doi.org/10.1016/j.irfa.2015.01.006>
- Nakazawa, M. (2015). fmsb: Functions for medical statistics book with some demographic data. *R package version 0.5.2*, 1-40.
- Nilsson, N. J. (1965). *Learning machines : foundations of trainable pattern-classifying systems*. McGraw-Hill.
- Nouri, B. A., & Soltani, M. (2016). Designing a Bankruptcy Prediction Model Based on Account, Market and Macroeconomic Variables, 9(1), 125-147.
- Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *International Joint Conference on Neural Networks*, 2, 163-168. <https://doi.org/10.1109/IJCNN.1990.137710>
- OECD/EU. (2016). Health at a Glance: Europe 2016 State of Health in the EU Cycle. <https://doi.org/10.1787/9789264265592-en>
- Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), 109-131. <https://doi.org/10.2307/2490395>
- ONU. (2015). *World Population Prospects*. New York.
- ONU. (2017). Informe Mundial sobre el Desarrollo de los Recursos Hídricos de las Naciones Unidas 2017: Las aguas residuales: el recurso desaprovechado, resumen ejecutivo; 2017.
- Oza, N. C., & Tumer, K. (2007). Classifier ensembles: Select real-world applications. <https://doi.org/10.1016/j.inffus.2007.07.002>
- Package «hybridEnsemble». (2015).
- Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems With Applications*, 36, 2-17. <https://doi.org/10.1016/j.eswa.2007.10.005>
- Pompe, P. P. M., & Bilderbeek, J. (2005). The prediction of bankruptcy of small- and medium-sized industrial firms. *Journal of Business Venturing*, 20(6), 847-868. <https://doi.org/10.1016/j.jbusvent.2004.07.003>
- PWC. (2014). *La gestión del agua en España, análisis de la situación actual del sector y retos futuros*.

- R Core, T. (2017). R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ranawana, R., & Palade, V. (2006). Multi-Classifer Systems: Review and a roadmap for developers. *International Journal of Hybrid Intelligent Systems*, 3(1), 35-61. <https://doi.org/10.3233/HIS-2006-3104>
- 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. <https://doi.org/10.1016/j.ejor.2006.08.043>
- Ravi Kumar, P., Ravi, V., Kumar, P. R., & 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. <https://doi.org/10.1016/j.ejor.2006.08.043>
- Rijsberman, F. R. (2006). Water scarcity: Fact or fiction? *Agricultural Water Management*, 80(1), 5-22. <https://doi.org/10.1016/j.agwat.2005.07.001>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., ... Müller Markus. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12(4116), 1-77. <https://doi.org/10.1126/science.182.4116.990>
- Rokach, L. (2009). Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography. *Computational Statistics & Data Analysis*, 53(12), 4046-4072. <https://doi.org/10.1016/j.csda.2009.07.017>
- Sato, T., Qadir, M., Yamamoto, S., Endo, T., & Zahoor, A. (2013). Global, regional, and country level need for data on wastewater generation, treatment, and use.
- Scrucca, L. (2013). GA : A Package for Genetic Algorithms in R. *Journal of Statistical Software*, 53(4), 1-37. <https://doi.org/10.18637/jss.v053.i04>
- Serrano-Cinca, C. (1996). Self organizing neural networks for financial diagnosis. *Decision Support Systems*, 17, 227-238.
- Shin, K.-S., Lee, T. S., & Kim, H.-J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28(1), 127-135. <https://doi.org/10.1016/j.eswa.2004.08.009>
- Shin, K.-S., & Lee, Y.-J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications*, 23(3), 321-328. [https://doi.org/10.1016/S0957-4174\(02\)00051-9](https://doi.org/10.1016/S0957-4174(02)00051-9)
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*, 74(1), 101-124. <https://doi.org/10.1086/209665>
- Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, 10(5), 335-347. [https://doi.org/10.1016/0167-8655\(89\)90037-8](https://doi.org/10.1016/0167-8655(89)90037-8)
- Tay, F. E. H., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29, 309-317.
- Tin Kam Ho, Hull, J. J., & Srihari, S. N. (1994). Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1), 66-75. <https://doi.org/10.1109/34.273716>
- Tsai. (2008). Financial decision support using neural networks and support vector machines. *Expert Systems*, 25(4), 380-393. <https://doi.org/10.1111/j.1468-0394.2008.00449.x>
- Tsai, Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing Journal*, 24, 977-984. <https://doi.org/10.1016/j.asoc.2014.08.047>

- Tsai, & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34(4), 2639-2649.
<https://doi.org/10.1016/j.eswa.2007.05.019>
- Vapnik, V. N. (1998). *Statistical learning theory*. Wiley.
- 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. <https://doi.org/10.1016/j.eswa.2013.09.033>
- Wehrens, R., & Buydens, L. M. C. (2007). Self- and super-organizing maps in R: The kohonen package. *Journal of Statistical Software*, 21(5), 1-19.
<https://doi.org/10.18637/jss.v021.i05>
- West, D., Dellana, S., & Qian, J. (2005). Neural network ensemble strategies for financial decision applications. *Computers & Operations Research*, 32(10), 2543-2559.
<https://doi.org/10.1016/j.cor.2004.03.017>
- Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, 18(1), 30-55. <https://doi.org/10.1007/s10618-008-0116-z>
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Comprehensive R Archive Network (CRAN).
- Wilcox, J. W. (1973). A Prediction of Business Failure Using Accounting Data. *Journal of Accounting Research*, 11, 163. <https://doi.org/10.2307/2490035>
- Wilson, N., & Summers, B. (2002). Trade credit terms offered by small firms: Survey evidence and empirical analysis. *Journal of Business Finance and Accounting*, 29(3-4), 317-351. <https://doi.org/10.1111/1468-5957.00434>
- World Economic Forum. (2015). *Global Risks 2015. 10th Edition*. Geneva.
- Xie, Y., Li, X., Ngai, E. W. T., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36(3), 5445-5449. <https://doi.org/10.1016/j.eswa.2008.06.121>
- Xu, M., & Zhang, C. (2009). Bankruptcy prediction: The case of Japanese listed companies. *Review of Accounting Studies*, 14(4), 534-558.
<https://doi.org/10.1007/s11142-008-9080-5>
- Yazdanfar, D., & Nilsson, M. (2008). The bankruptcy determinants of Swedish SMEs. *Institute for Small Business & Entrepreneurship*, (November), 1-14.
- Zadeh, L. a. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.
[https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- Zhang, Z. (2016). Variable selection with stepwise and best subset approaches. *Annals of translational medicine*, 4(7), 136. <https://doi.org/10.21037/atm.2016.03.35>
- Zhong, L., Mol, A. P. J., & Fu, T. (2008). Public-private partnerships in China's urban water sector. *Environmental Management*, 41(6), 863-877.
<https://doi.org/10.1007/s00267-008-9070-1>
- Zhou, L., Lai, K. K., & Yen, J. (2012). Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation. *International Journal of Systems Science*, 45(July 2015), 1-13.
<https://doi.org/10.1080/00207721.2012.720293>
- Zibanezhad, E., Foroghi, D., & Monadjemi, A. (2011). Applying decision tree to predict bankruptcy. *Proceedings - 2011 IEEE International Conference on Computer Science and Automation Engineering, CSAE 2011*, 4, 165-169.
<https://doi.org/10.1109/CSAE.2011.5952826>
- Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22(1), 59-82.
<https://doi.org/10.2307/2490859>

Capítulo 2: Optimización de la predicción de problemas financieros en empresas sanitarias privadas españolas aplicando Algoritmos Genéticos

(este artículo se encuentra en segunda revisión en la revista “Gaceta Sanitaria”, factor de impacto JCR 1,51 en 2015, 0,489 SCImago Journal Rank)

2.1 Resumen

El presente trabajo presenta una metodología para optimizar, a través de la Z'-Score de Altman para empresas privadas, la predicción de entrada en situación de concurso de acreedores (bancarrota) en empresas privadas del sector sanitario español. El método propuesto consiste en la aplicación de los algoritmos genéticos (AG) para encontrar los coeficientes de la fórmula de la cadena de ratios propuesta por Altman en su versión para empresas privadas que optimicen la predicción, maximizando la sensibilidad y especificidad y, con ello, reduciendo los errores del Tipo I y Tipo II. Con este propósito se ha utilizado una muestra de 5.903 empresas del sector sanitario privado español obtenidas de las base de datos de Sistema de Análisis de Balances Ibéricos (SABI). Los resultados muestran que los coeficientes para el modelo predictivo obtenido con los AG presenta mayor exactitud, sensibilidad y especificidad que el propuestos originalmente por Altman para empresas privadas tanto con los datos test como con todos los datos de la muestra. De esta forma, se pueden realizar predicciones con mayor exactitud, reduciendo el Error Tipo I y el Error Tipo II.

Palabras Clave: Empresas sanitarias, concurso de acreedores, Z-Altman, predicción de bancarrota, Algoritmos Genéticos

2.2 Abstract

This paper presents a methodology to optimize, through the Z'-Score of Altman for private companies, the prediction of bankruptcy, in private companies of the Spanish health sector. The proposed method consists in the application of the genetic algorithms (GA) to find the coefficients of the formula of the chain of ratios proposed by Altman in its formula for private companies which optimize the prediction, maximizing the sensitivity and specificity and, reducing the Type I and Type II errors. For this purpose a sample of 5903 companies of the Spanish private health sector has been used which have been obtained from the database of the Iberian Balance sheet Analysis System (SABI). The results show that the predictive model obtained with the GA presents greater accuracy, sensitivity and specificity than that proposed by Altman for private companies both with the test data and with all records in the database. Thus, predictions can be made with more accuracy, reducing Type I Error and Type II Error.

Keywords: Health companies, bankruptcy, Z-Altman, bankruptcy prediction, Genetic Algorithms

2.3 Introducción y Revisión de la Literatura

La bancarrota es un evento crítico que puede causar grandes pérdidas a los prestamistas, accionistas, acreedores y naciones, por ello, el desarrollo de modelos de predicción de bancarrota ha sido uno de los principales temas de investigación en el área de las finanzas (Kim y Kang, 2012). Realizar predicciones precisas es una cuestión vital para los profesionales del sector, ya que esta información les proporciona indicios de lo que puede suceder con suficiente antelación (Ko y Lin, 2006).

La literatura de la predicción de la bancarrota empieza en la década de 1930. FitzPatrick (1932) comparó 13 ratios de empresas en bancarrota y empresas en no-bancarrota (activas) y descubrió que las empresas activas poseían ratios más favorables. Hasta mediados de la década de los años 60, las investigaciones se basaban en análisis univariable. El más reconocido fue el de Beaver (1966), dicho estudio comparaba la media de 30 ratios de 79 empresas en bancarrota y no-bancarrota de 30 sectores industriales. El autor encontró que un número determinado de ratios podían discriminar entre bancarrota y no-bancarrota con hasta cinco años de antelación. Dicho autor sugirió la posibilidad de que múltiples ratios considerados al mismo tiempo podrían aumentar el poder predictivo y, de esta forma, empezó la evolución de los modelos predictivos de bancarrota. Altman (1968), publicó el primer análisis multivariable, utilizando el análisis discriminante múltiple (ADM) para desarrollar un modelo predictivo. Este modelo utilizaba cinco ratios financieros para predecir la bancarrota de empresas manufactureras. El modelo “Z-Score”, como es llamado, predice la bancarrota si la puntuación de la empresa cae dentro de un rango de valores. Por su parte, Ohlson (1980) utilizó un modelo logit para superar las desventajas asociadas a los modelos ADM. Su estudio utilizaba 9 variables, de las cuales, ocho eran ratios financieros. Zmijewski (1984) propuso una nueva fórmula utilizando un modelo probit. Este estudio utilizaba tres ratios financieros y la exactitud de este modelo con los datos de su muestra fue del 99%. Posteriormente, Altman (2000) adaptó su fórmula propuesta en 1968 para el sector de las empresas privadas, cambiando la variable “*Market Value of Equity*” (capitalización bursátil) por “*Book values of equity*” (patrimonio neto).

Por otro lado, cabe mencionar que hasta 1990, las técnicas estadísticas multivariables más usadas para la predicción de la bancarrota eran el ADM (e.g. Altman 1968,1983), el logit (e.g. Ohlson, 1980) y el probit (e.g. Zmijewski, 1984). Sin embargo,

estas técnicas deben cumplir unos supuestos como la linealidad, la normalidad e independencia de las observaciones, que no siempre se cumplen en el ámbito financiero (Deakin, 1972). En consecuencia, dichos métodos pueden tener limitaciones de validez y efectividad (Shin y Lee, 2002).

Desde finales de los años 80, las técnicas de inteligencia artificial, como las redes neuronales (RN) y los AG, han demostrado ser menos vulnerables a las suposiciones comentadas previamente (Chung y Tam, 1993; Messier y Hansen, 1988) y han sido aplicadas con éxito en la predicción de la bancarrota (e.g. Angelini *et al.*, 2008; Atiya, 2001; Paliwal y Kumar, 2009). El primer artículo del uso de RN en la predicción de la bancarrota fue realizado por (Odom y Sharda, 1990). Doce años después Shin y Lee (2002) utilizaron los AG para predecir la bancarrota en 528 empresas manufactureras de tamaño medio.

La Tabla 2.1 (Altman, 1968; Altman, 2000; Aziz *et al.*, 1988; Bauer y Agarwal, 2014; Beaver, 1966; Becchetti y Sierra, 2003; Beynon y Peel, 2001; Blanco-Oliver *et al.*, 2015; Charitou *et al.*, 2004; Chi y Tang, 2006; Cho *et al.*, 2009; Chung *et al.*, 2016; Cielen *et al.*, 2004; Cultrera y Brédart, 2016; Dambolena y Khoury, 1980; Divsalar y Roodsaz, 2012; du Jardin, 2017; Dugan y Zavgren, 1989; Gu y Gao, 1999; Jeong *et al.*, 2012; Joo y Jin, 2000; Khan, 1985; Kim *et al.*, 2016; Kim y Gu, 2006; Kim & Kang, 2012; Laitinen y Laitinen, 2000; Lennox, 1999; Li y Miu, 2010; Mousavi *et al.*, 2015; Nouri y Soltani, 2016; Ohlson, 1980; Pompe y Bilderbeek, 2005; Shumway, 2001; Wang *et al.*, 2014; Wilson & Summers, 2002; Xu y Zhang, 2009) muestra las principales ratios financieros utilizados y sus referencias.

Tabla 2.1 Ejemplos de ratios financieros, sus categorías y referencias

Categorías	Variables y definiciones	Estudios Previos
I. CAPITAL STRUCTURE VARIABLE	Long Term Debt to Total Asset	Beaver 1966; Ohlson 1980; Gu y Gao, 2000; Mousavi et al, 2015
	Short and Long Term Debt to Equity	Chi et al 2006; Zmijewski,1984
	Short y Long Term Debt to Total Assets	Beaver 1966; Laitinen et al 2000; Lennox, 1999,
	Cash-Holdings to Total Assets	Dugan et al 1989; Laitinen et al 2000
	Equity Share of Total Assets	Altman,1968; Cielen et al 2004; Charitou et al , 2004; Pantalone y Platt, 1987; Pompe et al 2005; Shumway,2001
	Current Liabilities to Total Assets	Blanco-Oliver et al, 2015; Gu y Gao 2000; Jeong et al, 2012; Lo, 1985; Pantalone y Platt ,1987; Zmijewski,1984
	Total liabilities to total assets	Divsalar et al, 2012, Mousavi et al, 2015; Xu y Zhang, 2009

Capítulo 2: Optimización de la predicción de problemas financieros en empresas sanitarias privadas españolas aplicando Algoritmos Genéticos

	Total liabilities to current assets	Blanco-Oliver et al, 2015
	Total liabilities to working capital	Mousavi et al, 2015
	Quick assets to total assets	Divsalar et al, 2012
	Quick assets to current assets	Blanco-Oliver et al, 2015
	Quick assets to current liabilities	Du Jardin, 2017
	Equity to total assets	Cultrera y Brédart, 2016
	Current assets to total assets	Kim y Kang, 2012; Wang et al, 2014; Zhao et al, 2017
	Current debt to total debt	Wang et al, 2014
	Retained earnings to total assets	Blanco-Oliver et al, 2015; Li y Miu, 2010; Kim y Kang, 2012
	Current Liabilities to total capital	Cho et al, 2009
	Current ratio	Blanco-Oliver et al, 2015; Cultrera y Brédart, 2016; Chung et al, 2016; Wang et al, 2014; Xu y Zhang, 2009; Zhao et al, 2017
	Debt-total assets	Chung et al, 2016; Kim y Kang, 2012
	Total borrowings to total assets	Kim et al, 2016
II. LIQUIDITY RATIOS	Quick Ratio	Altman, 1968; Cielen et al 2004; Charitou et al , 2004; Chung et al, 2016; Nam y Jinn ,2001; Kim y Kam, 2016; Shumways, 2001
	Working capital to total asset	Nouri y Soltari, 2016; Li y Miu, 2010; Mousavi et al, 2015; Xu y Zhang, 2009; Zhao et al, 2017
	Working capital to sales	Zhao et al, 2017
	Cash to current debt	Du Jardin, 2017; Wang et al, 2014; Zhao et al, 2017
	Cash to current assets	Du Jardin, 2017
	Cash to total assets	Blanco-Oliver et al, 2015; Du Jardin, 2017, Zhao et al, 2017
	Cash to Long y Short Term Debt	Dugan y Zavgren 1989; Charitou et al , 2004
III. PROFITABILITY RATIOS	Return on Equity	Altman, 1968; Pompe et al 2005; Shumways, 2001
	Return on Assets	Becchetti et al 2003; Chung et al, 2016; Dambolena y Khoury 1980; Kim et al, 2016; Nouri y Soltari, 2016; Ohlson, 1980; Zmijewski, 1984
	Net Profit Margin	Beaver 1966; Kim et al 2006
	Net Profit to total assets	Zhao et al, 2017
	Net Profit to current assets	Zhao et al, 2017
	Gross Profit Margin	Cielen et al 2004; Kim et al 2006
	Gross Profit to sales	Zhao et al, 2017
	Profit Margin	Becchetti et al 2003; Pompe et al 2005
	Net profit to liabilities	Zhao et al, 2017
	Net incomen to current assets	Mousavi et al, 2015
	Net incomen to total assets	Bauer y Agarwal, 2014; Xu y Zhang, 2009
	Net incomen to capital	Mousavi et al, 2015
	Net incomen to sales	Jeong et al, 2012
Ordinary income to sales	Jeong et al, 2012; Kim y Kang, 2012	
Ordinary income to total assets	Kim y Kang, 2012	

Capítulo 2: Optimización de la predicción de problemas financieros en empresas sanitarias privadas españolas aplicando Algoritmos Genéticos

	Ordinary income to capital	Kim y Kang, 2012
	Return on capital employment	Nouri y Soltari, 2016
	Profit to sales	Cho et al, 2009
	Operating profit to sales	Cho et al, 2009
IV. CASHFLOW	Cash flow to Sales	Beaver 1966; Wang et al, 2014
	Cash flow to Total Assets	Aziz et al. , 1988; Pompe et al 2005; Young et al, 2005
	Cash flow to total debt	Cultrera y Brédart, 2016
	Operating cash flow to sales	Wang et al, 2014
	Operating cash flow to total debt	Kim y Kang, 2012; Wang et al, 2014
	Operating cash flow to total assets	Kim et al, 2016; Wang et al, 2014
	Cash flow to long y short term debt	Beaver 1966; Chung et al, 2016
	Sales to Total Assets	Altman 1968; ; Li y Miu, 2010
	Sales to current assets	Divsalar et al, 2012
	EBITDA to Total Assets	Altman 1968; Cultrera y Brédart, 2016; Du Jardin, 2017; Gu y Gao, 2000; Young et al 2005
	EBITDA to sales	Kim et al, 2016
	EBIT to total assets	Mousavi et al, 2015
	EBITDA to interest expenses	Kim y Kang, 2012
	Retained Earnings to Total Assets	Altman 1968
Financial Expenses to Sales	Becchetti et al 2003	
V. DEVELOPMENT AND FIRM	Change in Sales from Last Year	Arshad 1985; Pompe et al 2005
	Change in Total Assets from Last Year	Arshad 1985; Becchetti et al 2003
	Firms Age	Beaver 1966; Kim et al 2006; Altman, 2000
	Firm Size (Natural Logarithm of sales)	Cielen et al 2004; Kim et al 2006
	Firm Size (Natural Logarithm of total assets)	Blanco-Oliver et al, 2015; Beynon, y Peel, 2001; Kim y Kang, 2012; Ohlson, 1980; Wilson y Summer, 2002
	Growth rate of tangible assets	Cho et al, 2009
	Growth rate of total assets	Jeong, et a, 2012
	Growth rate of current assets	Cho et al, 2009
VI. TURNOVER	Accounts Receivable to Sales	Beaver 1966
	Inventory to Sales	Beaver 1966; Du Jardin, 2017; Kim y Kang, 2012; Mousavi et al, 2015; Young et al 2005
	Inventory to Total Assets	Cielen et al 2004
	Current assets to sales	Du Jardin, 2017; Wang et al, 2014
	Current liabilities to sales	Du Jadin, 2017
	Receivables to sales	Du Jadin, 2017
	Sales to total assets	Du Jadin, 2017
	Asset turnover ratio	Chung et al, 2016; Jeong et al, 2012; Nouri y Soltari, 2016
	Inventory to turnover ratio	Nouri y Soltari, 2016
	Turnover of managerial assets	Cho et al, 2009
VII. MARKET VARIABLES	Market value to total debt	Nouri y Soltari, 2016
	Market value of total assets	Bauer y Agarwal, 2014

Capítulo 2: Optimización de la predicción de problemas financieros en empresas sanitarias privadas españolas aplicando Algoritmos Genéticos

	Market value of equity to total assets	Bauer y Agarwal, 2014; Li y Miu, 2010
	Stock price	Nouri y Soltari, 2016
	Stock return	Nouri y Soltari, 2016
VIII. OTROS	Productivity of capital	Kim et al, 2016
	Interest expenses-gross profit	Chung et al, 2016
	Interest expenses to sales	Cho et al, 2009
	Financial expenses to total debt	Kim y Kang, 2012
	Capital employed to total liabilities	Blanco-Oliver et al, 2015

Cabe mencionar que la predicción de la bancarrota ha sido utilizada en multitud de sectores y países (e.g. empresas manufactureras (Altman, 1968) y no manufactureras (Altman, 2000), empresas de tamaño medio y pequeñas italianas (Gordini, 2014), empresas de Korea (Shin y Lee, 2002). Sin embargo, no existe un modelo específico para el sector sanitario privado español. Por ello, este estudio propone la creación de un modelo ad-hoc en el que utilizando las mismas ratios que Altman (2000) tenga unos coeficiente que permitan predecir mejor la salud financiera de las citadas empresas con dos años de antelación (du Jardin, 2015; Hennawy y Morris, 1983), utilizando para ello la técnica de los AG.

Para observar la importancia de este sector en España, cabe mencionar que en el año 2015, el gasto de la sanidad española supuso el 9% del Producto Interior Bruto nacional (PIB) (OECD/EU, 2016), el 2,7% si se hace referencia solo a la sanidad privada, siendo este dato superior respecto al de países de nuestro entorno. También se puede observar como la colaboración entre el sector público y privado va en aumento y parece ser el camino que se seguirá en un futuro cercano. Del mismo modo, se pretende que la industria farmacéutica evolucione hacia el desarrollo de nuevos modelos de colaboración y con mayor desarrollo de servicios (Deloitte, 2014). En la Tabla 2.2 se muestra el gasto sanitario respecto del PIB en países de la UE.

Con el propósito citado anteriormente, el trabajo, además de la presente introducción, se ha organizado de la siguiente forma. La sección 2 presenta la metodología. El apartado 3 muestra los resultados y el epígrafe 4 las conclusiones y limitaciones.

Tabla 2.2 Gasto Sanitario respecto del PIB en países de la UE

País	Gasto Sanitario Total respecto PIB	Gasto Sanitario Público respecto PIB	Gasto Sanitario Privado respecto PIB
España	9,0 %	6,3 %	2,7 %
Francia	11,0 %	8,6 %	2,4 %
Italia	9,0 %	6,8 %	2,2 %
Reino Unido	9,7 %	7,7 %	2,0 %
Suecia	11,1 %	9,3 %	1,8 %
Alemania	11,1 %	9,4 %	1,7 %
Total UE	9,9 %	7,8 %	2,1 %

Fuente: OECD/EU (2016)

2.4 Metodología

2.4.1 Datos

Los datos se obtuvieron de la base de datos SABI (Sistema de Análisis de Balances Ibéricos) (Bureau Van Dijk, 2017). La muestra estuvo formada por aquellas empresas cuya última información disponible se encontraba entre los años 2007 y 2015. Los CNAEs (Clasificación Nacional de Actividades Económicas) utilizados para seleccionar las empresas del sector privado sanitario español fueron los siguientes: 2110, 2120, 3250, 4646, 4773, 4774, 7211, 7219, 7490, 8610, 8621, 8622, 8623 y 8690. Se incluyeron un total de 7.462 empresas activas y 168 empresas en concurso de acreedores (bancarota). Posteriormente se seleccionaron aquellas empresas que tenían los cinco ratios propuestos por (Altman, 2000) correctamente informados (véase fórmula 2.1). Ello hizo que la muestra final quedase de la siguiente forma: 5.824 empresas activas y 79 empresas en concurso de acreedores. En la Tabla 2.3 se muestra la estadística descriptiva de la muestra desglosada por CNAE y estado

Tabla 2.3 Estadística Descriptiva de la muestra utilizada

CNAE (Descripción)	Empresas activas n (%)	Empresas concurso n (%)
2110: Fabricación de productos farmacéuticos en base	68 (1,17)	2 (2,53)
2120: Fabricación de especialidades farmacéuticas	147 (2,52)	0 (0)
3250: Fabricación de instrumentos y suministros médicos y odontológicos	160 (2,75)	4 (5,06)
4646: Comercio al por mayor de productos farmacéuticos	767 (13,17)	4 (5,06)
4773: Comercio al por menor de productos farmacéuticos en establecimientos especializados	150 (2,58)	0 (0)
4774: Comercio al por menor de artículos médicos y ortopédicos en establecimientos especializados	525 (9,01)	2 (2,53)
7211: Investigación y desarrollo experimental en biotecnología	37 (0,64)	1 (1,27)
7219: Otra investigación y desarrollo experimental en ciencias naturales y técnicas	153 (2,63)	9 (11,39)

Capítulo 2: Optimización de la predicción de problemas financieros en empresas sanitarias privadas españolas aplicando Algoritmos Genéticos

7490: Otras actividades profesionales, científicas y técnicas n.c.o.p.	1256 (21,57)	35 (44,3)
8610: Actividades hospitalarias	206 (3,54)	1 (1,27)
8621: Actividades de medicina general	237 (4,07)	5 (6,33)
8622: Actividades de medicina especializada	541 (9,29)	4 (5,06)
8623: Actividades odontológicas	695 (11,93)	3 (3,8)
8690: Otras actividades sanitarias	882 (15,14)	9 (11,39)
Total	5824 (100)	79 (100)

Fuente: elaboración propia

2.4.2 Técnica de los Algoritmos Genéticos

Los AG simulan la teoría de la evolución propuesta por DarwinGordini, 2014. Fueron desarrollados por Holland (1975) y son procesos estocásticos robustos (Kim y Kang, 2012; Shin y Lee, 2002) que pueden usarse para resolver problemas de búsqueda y optimización a través de una función *fitness* o función objetivo. Los AG realizan el proceso de búsqueda en cuatro pasos: inicialización, selección, cruce y mutación (Shin y Lee, 2002).

En el proceso de inicialización, una población de individuos, llamados cromosomas, se distribuyen de forma aleatoria por el rango de búsqueda definido previamente para cada coeficiente y son evaluados a través de la función *fitness* (Gordini, 2014).

El proceso de selección se basa en el principio de supervivencia, es decir, aquellos cromosomas que presenten valores más óptimos en la función *fitness*, serán seleccionados para crear la siguiente generación (Gordini, 2014). En el proceso de cruce, se seleccionan dos individuos con valores óptimos en la función *fitness* y se unen de manera que el descendiente contiene información parcial de cada uno de sus progenitores, de esta forma, cabe esperar, que el descendiente mejore el resultado de la función *fitness* obtenido por sus progenitores (Shin y Lee, 2002).

La mutación en los AG es un mecanismo donde de forma aleatoria se seleccionan miembros de la población y también de forma aleatoria se cambia alguno de sus genes. De esta forma, se desplazan valores numéricos hacia zonas del espacio de búsqueda que no pueden ser alcanzadas por los otros operadores genéticos (Shin y Lee, 2002).

La Tabla 2.4 muestra los parámetros utilizados para controlar el proceso de los AG (Kim *et al.*, 2016).

Tabla 2.4 Parámetros utilizados en los AG

Descripción	Mínimo	Máximo
Rango de búsqueda X1	0,01	1,00
Rango de búsqueda X2	0,01	1,00
Rango de búsqueda X3	2,00	3,50
Rango de búsqueda X4	3,50	9,00
Rango de búsqueda X5	2,00	4,50
Punto de Corte	4,00	5,50
Población inicial	1.000	
Tasa de cruce	0,6	
Tasa de mutación	0,1	
Número máximo de iteraciones	80	

Fuente: elaboración propia

2.4.3 Datos no balanceados

Los datos no balanceados ocurren cuando la proporción de una de las categorías es extremadamente pequeña comparada con la otra categoría, como sucede en este análisis, donde el número de empresas activas es muy superior al número de empresas en concurso de acreedores. En esta situación, el modelo de clasificación tiende a predecir a las empresas en la clase mayoritaria, de tal forma que se aumenta la exactitud (*accuracy*) del estudio a costa de ignorar a la clase minoritaria. Para resolver este problema, se ha utilizado la técnica de infra-muestrear la categoría mayoritaria de tal forma que ambas categorías estén igualmente representadas (Kim *et al.*, 2016).

2.4.4 Análisis estadístico

A continuación se detalla el procedimiento utilizado para procesar los datos.

Para resolver el problema de los datos no balanceados, se infra-muestra la categoría mayoritaria, seleccionando al azar el mismo número de empresas activas que empresas en concurso de acreedores, que son 79.

Por lo tanto, se parte de un conjunto de datos con 158 registros, al cual, se le aplica la técnica de “Validación cruzada dejando uno fuera” (Leave-one-out cross-validation, LOOCV (Kohavi, 1995), es decir, 157 registros van a formar parte del fichero de entrenamiento (training) y sólo uno va a formar parte del fichero de validación (test).

Capítulo 2: Optimización de la predicción de problemas financieros en empresas sanitarias privadas españolas aplicando Algoritmos Genéticos

Al fichero de entrenamiento, se le aplica la técnica de los AG donde se obtienen un determinado número de soluciones.

Por una parte, estas soluciones son acumuladas para calcular los coeficientes definitivos al final del proceso. Por otra, se calculan las medias de estas soluciones para obtener una única solución. Con ésta, se procesa el registro del fichero de validación y se construye la matriz de confusión (véase Tabla 2.5) del modelo de AG para los datos Test. Para comparar los resultados obtenidos con este modelo, AG, se procesa este mismo registro utilizando la fórmula Z'-Score de Altman (véase fórmula 2.1) y de esta forma se construye su correspondiente matriz de confusión.

Este proceso se ejecuta 158 veces, de tal forma, que cada registro ha formado parte una única vez del fichero de validación. Una vez ejecutado este conjunto de datos, se vuelve a generar de forma aleatoria otro conjunto de 158 registros, el cual es procesado de igual forma que el primero de ellos.

Con el objetivo de obtener unos resultados robustos y no debidos al azar, este proceso se repite 100 veces. Una vez terminado el proceso, se calculan las medias de los coeficientes de todas las soluciones obtenidas para obtener los coeficientes definitivos.

La Figura 2.1 muestra gráficamente el proceso que se acaba de detallar.

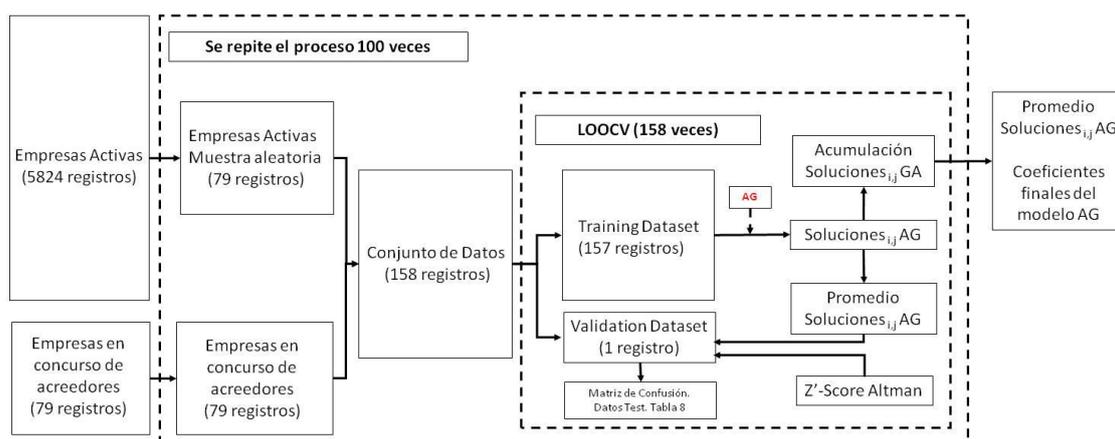


Figura 2.1. Proceso de Infra-muestreo con LOOCV y AG

El programa estadístico utilizado ha sido R Core (2017). Los paquetes utilizados han sido: GA (Scrucca, 2013), fmsb (Nakazawa, 2015), pROC (Robin *et al.*, 2011).

2.4.5 Medidas de rendimiento

Las medidas de rendimiento utilizadas en este trabajo se extraen a partir de una matriz de confusión (Kim *et al.*, 2016) como la que aparece en la Tabla 2.5, donde los verdaderos positivos (VP) indican el número de empresas en concurso clasificadas correctamente, los falsos negativos (FN), el de empresas en concurso clasificadas erróneamente, los verdaderos negativos (VN) el de empresas activas clasificadas correctamente y, los falsos positivos (FP), el de empresas activas clasificadas erróneamente.

Tabla 2.5. Matriz de Confusión

Clase Actual	Predicción	
	Concurso Acreedores	Activa
Concurso Acreedores	Verdaderos Positivos (VP)	Falsos Negativos (FN)
Activa	Falsos Positivos (FP)	Verdaderos Negativos (VN)

La función *fitness* utilizada en los AG se ha definido como: $(0,30 * \text{Sensibilidad} + 0,70 * \text{Especificidad})$, donde la sensibilidad representa la proporción de clasificados como VP frente al total que realmente eran positivos, $VP / (VP + FN)$ y la especificidad es la de los clasificados como VN, frente a los que realmente eran negativos $VN / (VN + FP)$ (Altman y Bland, 1994). Para la elección de estos pesos, se realizaron distintas pruebas, siendo estos los que finalmente arrojaban un mejor resultado. Hay que reseñar, que se dio más peso a la especificidad debido a los datos no balanceados existentes en la población.

También se han utilizado otras medias como (M.-Y. Chen, 2011):

- Exactitud (*accuracy*): $(VP + VN) / (VP + FP + FN + VN)$, su utilización se ha realizado teniendo en cuenta que esta medida no es la más adecuada para evaluar modelos con datos no balanceados, ya que se ve muy afectada por la clase mayoritaria.
- Media geométrica (G-Mean): $\sqrt{\text{Sensibilidad} * \text{Especificidad}}$, dando la misma importancia a la sensibilidad y a la especificidad (Barandela *et al.*, 2003).
- Área bajo la curva ROC, la curva ROC representa de forma gráfica el porcentaje de empresas activas clasificadas incorrectamente (1 - Especificidad) en el eje de abscisas y el porcentaje de empresas en concurso clasificadas correctamente (Sensibilidad) en el eje de ordenadas

(Barandela *et al.*, 2003). Para evaluar el rendimiento de ambos modelos se utilizó el área bajo la curva ROC (AUROC), que equivale a la probabilidad de que si se eligiesen al azar a dos empresas, una en situación de concurso y otra activa, la prueba las clasificase a ambas correctamente. Cuanto más cerca de 1 sea el AUROC, mayor poder predictivo tendrá el modelo (Kim y Kang, 2012).

- F1-Score representa la media armónica entre la Sensibilidad y el Valor Predictivo Positivo ($VP / (VP + FP)$):

$$(2 * VPP * Sensibilidad) / (VPP + Sensibilidad)$$

- Matthews correlation coefficient (MCC) o phi coefficient es una medida que tiene en cuenta a los VP, VN, FP y FN y se puede utilizar tanto en datos balanceados como no balanceados:

$$(VP * VN) - (FP * FN) / \sqrt{(VP + FP) * (VP + FN) * (VN + FP) * (VN + FN)}$$

Finalmente, para evaluar el poder predictivo del método obtenido a través de los AG, se comparan los resultados obtenidos a través de este método, con los resultantes de aplicar la fórmula Z'-Score propuesta por Altman para empresas privadas (Altman, 2000), la cual se detalla a continuación:

$$Z' = 0.717 * X1 + 0.847 * X2 + 3.107 * X3 + 0.42 * X4 + 0.998 * X5 \quad (\text{fórmula 2.1})$$

Donde,

- $X1 = (\text{Activo Circulante} - \text{Pasivo Liquido}) / \text{Total Activo}$
- $X2 = \text{Reservas} / \text{Total Activo}$
- $X3 = \text{EBIT} / \text{Total Activo}$
- $X4 = \text{Patrimonio Neto} / \text{Deudas Totales}$
- $X5 = \text{Ingresos de Explotación} / \text{Total Activo}$

Los valores umbrales que estableció el autor fueron los siguientes: puntuaciones del valor Z' superiores a 2,90, indican que la empresa está en una zona segura. Si la puntuación está entre 1,23 y 2,90, la empresa está en una zona gris y si la puntuación está por debajo de 1,23, entonces la empresa está en riesgo de bancarrota. En este trabajo, se clasificó a la empresa en estrés financiero si su valor Z' era inferior a 1,23.

Teniendo en cuenta que la fórmula Z'-Score de Altman presenta buen rendimiento para predecir la bancarrota tanto con un año de antelación, como con dos años de

antelación (Altman, 2000), se seleccionaron los datos del último año disponible para comprobar si las empresas estaban en concurso de acreedores o no y por otra parte, se seleccionaron los datos de dos años antes para utilizarlas como variables predictivas.

2.5 Resultados

La Tabla 2.6 muestra la estadística descriptiva de los datos de la muestra desglosados por estado para las cinco variables estudiadas. Se puede observar como las medias y medianas son superiores en las cinco ratios en las empresas activas en relación a las empresas en concurso, siendo las diferencias más grandes en las variables X4 y X5.

Tabla 2.6. Estadística Descriptiva de los registros de la muestra

Variable	Empresas Activas (n= 5824)		Empresas en Concurso (n=79)	
	Media (Dt)	Mediana (P25 – P75)	Media (Dt)	Mediana (P25 – P75)
X1	0,24 (0,32)	0,23 (0,04 – 0,45)	0,12 (0,36)	0,1 (-0,01 – 0,36)
X2	0,31 (0,27)	0,26 (0,08 – 0,5)	0,21 (0,24)	0,11 (0,04 – 0,31)
X3	0,08 (0,18)	0,05 (0,01 - 0,13)	-0,01 (0,19)	0,02 (-0,03 – 0,04)
X4	0,47 (0,33)	0,48 (0,26 – 0,71)	0,22 (0,31)	0,19 (0,07 – 0,39)
X5	1,67 (1,7)	1,31 (0,77 – 2,1)	1,12 (1,24)	0,74 (0,27 – 1,6)

La Tabla 2.7 muestra los coeficientes Z'-Score de Altman (2000) y la estadística descriptiva de los coeficientes obtenidos a través del proceso de los AG. Se puede observar que los valores de la media y la mediana obtenidos mediante el proceso de los AG son bastante similares, lo que implica la robustez del modelo construido. Por otra parte, observando los percentiles P25 y P75 se puede concluir que los rangos de búsqueda descritos en la Tabla 2.4 han sido correctos.

Tabla 2.7. Coeficientes Z'-Score de Altman y estadística descriptiva de las soluciones obtenidas en el proceso AG

Coeficiente de la variable	Z'-Score	AG		
	Altman	Media (Dt)	Mediana (P ₂₅ – P ₇₅)	Min-Max
X1	0.717	0,44 (0,24)	0,42 (0,24-0,62)	0,01-1,00
X2	0.847	0,51 (0,24)	0,52 (0,33-0,69)	0,01-1,00
X3	3.107	2,79 (0,34)	2,79 (2,54-3,04)	2,00-3,50
X4	0.42	7,41 (1,13)	7,77 (6,84-8,21)	3,50-9,00
X5	0.998	2,91 (0,47)	2,87 (2,57-3,15)	2,00-4,50
Punto de Corte	1.23	4,72 (0,29)	4,72 (4,53-4,91)	4,00-5,50

El número de soluciones encontradas a lo largo del proceso es: 551.346

La fórmula 2.2 refleja el modelo resultante definitivo obtenido a través del proceso de los AG y que se propone en este trabajo como idóneo para evaluar a las empresas del sector médico privado en España:

$$AG = 0.436 * X1 + 0.510 * X2 + 2.79 * X3 + 7.407 * X4 + 2.908 * X5 \quad (\text{fórmula 2.2})$$

y cuyo punto de corte es: 4,715, es decir:

- Si $AG > 4,715$: Zona Segura, la previsión es una empresa activa
- Si $AG \leq 4,715$: Zona de Riesgo, la previsión es una empresa en concurso de acreedores

Las principales diferencias entre los coeficientes de ambos modelos se encuentran en las variables X4 y X5 (véase fórmula 2.1 y 2.2). Así, al analizar la Tabla 2.7, se observa que las mayores diferencias entre las medias y medianas, se encuentra precisamente en esas variables.

La Tabla 2.8 muestra los datos de la matriz de confusión obtenidos durante el proceso de ejecución de los datos test (detallado en el apartado 2.4 Análisis Estadístico). En la parte inferior, aparecen los datos obtenidos con la ejecución del modelo AG (véase fórmula 2.2) y con el modelo Z'-Score (véase fórmula 2.1) utilizando los 5.903 registros.

La Tabla 2.9 muestra los resultados de las medidas de rendimiento calculadas con las matrices de confusión de la tabla anterior (véase Tabla 2.8), donde se puede apreciar cierta mejoría en el modelo AG respecto del modelo Z'-Score en todas las medidas propuestas. En esta tabla, también aparecen los resultados de comparar el poder predictivo del modelo Z'-Score de Altman (véase fórmula 2.1) con el modelo desarrollado con AG (véase fórmula 2.2) con los 5.903 registros. Se puede comprobar cómo la sensibilidad mejora 7,6 puntos, la especificidad 2,04 puntos y el área bajo la curva ROC 3,6 puntos. Por otro lado, cabe mencionar que si bien es verdad que hay medidas que tienen el mismo valor, por ejemplo, el Valor Predictivo Negativo (VPN) (99,38% vs 99,24%), no se debe obviar que tanto la Precisión o VPP, como el VPN están afectadas por la prevalencia (porcentaje de empresas en situación de concurso). Como la prevalencia es muy baja (1,34%), entonces el VPP es muy bajo y el VPN es muy elevado.

Capítulo 2: Optimización de la predicción de problemas financieros en empresas sanitarias privadas españolas aplicando Algoritmos Genéticos

Tabla 2.8. Matriz de confusión de los modelos con los datos test y con todos los registros

Datos	Modelo	Clase Actual	Predicción		Total
			Concurso Acred.	Activa	
Datos Test	AG	Concurso Acred.	4.453 (56,37)	3.447 (43,63)	7.900
		Activa	1.130 (14,30)	6.770 (85,70)	7.900
	Z'-Score Altman	Concurso Acred.	4.200 (53,16)	3.700 (46,84)	7.900
		Activa	1.212 (15,34)	6.688 (84,66)	7.900
Todos los registros	AG	Concurso Acred.	48 (60,76)	31 (39,24)	79
		Activa	845 (14,51)	4.979 (85,49)	5.824
	Z'-Score Altman	Concurso Acred.	42 (53,16)	37 (46,84)	79
		Activa	964 (16,55)	4.860 (83,45)	5.824

Tabla 2.9. Comparación de las Medidas de Rendimiento con los datos test y todos los registros

Medidas de Rendimiento	Datos Test			Todos los registros		
	AG	Z'-Score	Altman	AG	Z'-Score	Altman
Prevalencia	50,00	50,00	1,34	1,34	53,16	53,16
Sensibilidad (Recall)	56,37	53,16	60,76	60,76	83,45	83,45
Especificidad	85,70	84,66	85,49	85,49	83,04	83,04
Exactitud (Accuracy)	71,03	68,91	85,16	85,16	4,17	4,17
Valor Predictivo Positivo (Precisión)	79,76	77,61	5,38	5,38	99,24	99,24
Valor Predictivo Negativo	66,26	64,38	0,72	0,72	0,67	0,67
G-Mean	0,70	0,67	0,10	0,10	0,08	0,08
F1-Score	0,66	0,63	0,15	0,15	0,11	0,11
MCC (phi coeficient)	0,44	0,40	0,08	0,08	0,05	0,05
Índice Kappa	0,42	0,38	0,761	0,761	0,725	0,725
Área Bajo la Curva ROC (AUROC)	-	-	0,761	0,761	0,725	0,725

En la Figura 2.2 se muestra la comparación de curvas ROC, en ella se puede apreciar el mayor poder predictivo del nuevo modelo.

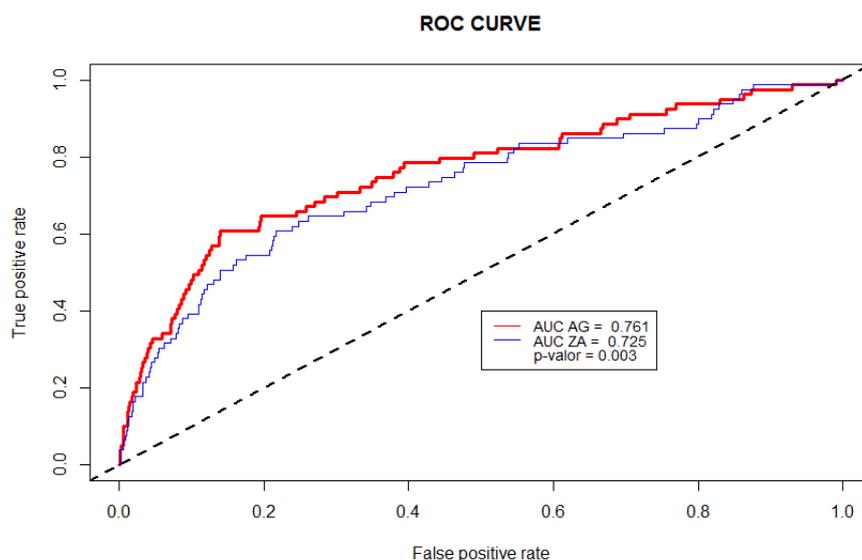


Figura 2.2 Comparación Curvas ROC entre el modelo AG y Z'-Score Altman

2.6 Discusión

Debido a la cada vez más importante presencia de las empresas privadas en el sector sanitario, se ha propuesto la creación de un modelo predictivo que sirva para dar estabilidad financiera a los *stakeholders* del mencionado sector. La principal contribución de este estudio es la elaboración de una fórmula ad-hoc para predecir la bancarrota con dos años de antelación en las empresas del sector sanitario privado. Ésta se construyó partiendo de la fórmula Z'-Score de Altman para empresas privadas y utilizando AG.

A la luz de los resultados obtenidos, se observa que la nueva fórmula mejoran ligeramente las predicciones obtenidas a través del modelo propuesto por Altman (2000). A través de la utilización de una media ponderada de la sensibilidad y especificidad, se mejoró la predicción de la situación financiera de las empresas, y se consiguió reducir el Error Tipo I y II. No debe pasarse por alto que la ocurrencia de errores tipo I en este caso puede acarrear problemas como, por ejemplo, la concesión de créditos a una empresa en concurso de acreedores, lo que conllevaría la dificultad de poder recuperar el capital principal y sus respectivos intereses por parte de los stakeholders. Por su parte, el Error Tipo II puede suponer la no concesión de créditos a empresas solventes, y la creación de un ambiente de incertidumbre debido a la falta de credibilidad por parte de los acreedores (Etheridge *et al.*, 2000).

También se puede afirmar que los resultados obtenidos (exactitud 85,16%) están en la misma línea que varios modelos predictivos comentados previamente (e.g. Odom y Sharda, 1990, 81,81% y Shin y Lee, 2002, 80,8%). Se obtiene un resultado similar en la evolución de la fórmula Z-Score de Altman propuesta por Almamy *et al.* (2016), donde se logra una exactitud del 82.9%. En relación al área bajo la curva ROC, la mejoría observada en este artículo es similar a la obtenida por Altman *et al.* (2010) en su comparativa entre modelos con información financiera y un segundo modelo donde se añaden variables con información no financiera (e.g. 0,80 vs 0,74 o 0,75 vs 0,71).

Finalmente, señalar que las limitaciones de este estudio radican en que el uso de los AG no garantiza que se encuentre la solución óptima del problema. Estas soluciones dependerán del rango de búsqueda aplicado para el cálculo de los coeficientes, el tamaño de la población, la ratio de cruce y la ratio de mutación aplicados en el modelo. En cuanto a futuras líneas de investigación, cabe destacar que podría utilizarse otras técnicas de

infra-muestreo para tratar datos no balanceados, como la validación cruzada de k-iteraciones o técnicas de agrupamiento como el algoritmo de agrupamiento de k-medias o jerárquico.

2.7 Referencias

- AEAS. (2014). XIII Encuesta de suministro de agua potable y saneamiento en España.
- Alam, P., Booth, D., Lee, K., & Thordarson, T. (2000). The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study. *Expert Systems with Applications*, 18(3), 185-199. [https://doi.org/10.1016/S0957-4174\(99\)00061-5](https://doi.org/10.1016/S0957-4174(99)00061-5)
- Almamy, J., Aston, J., & Ngwa, L. N. (2016). An evaluation of Altman's Z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence from the UK. *Journal of Corporate Finance*, 36, 278-285. <https://doi.org/10.1016/j.jcorpfin.2015.12.009>
- Altman. (2000). Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA ® Models Background. *Stern School of Business, New York University*, 1-54.
- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 1 : sensitivity and specificity. *Bmj*, 308(6943), 1552. <https://doi.org/10.1258/phleb.2012.012J05>
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Altman, E. I. (1983). *Corporate financial distress : a complete guide to predicting, avoiding, and dealing with bankruptcy* (3rd Editio). Wiley.
- Altman, E. I., Sabato, G., & Wilson, N. (2010). The Value of non-financial information in small-sized enterprise risk management. *The Journal of Credit Risk*, 6(2), 1-33.
- Angelini, E., di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance*, 48(4), 733-755. <https://doi.org/10.1016/j.qref.2007.04.001>
- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, 34(5), 502-508. <https://doi.org/10.1111/opo.12131>
- Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks*, 12(4), 929-935. <https://doi.org/10.1109/72.935101>
- Aziz, A., Emanuel, D., & Lawson, G. (1988). *Journal of management studies. Journal of Management Studies* (Vol. 25). Blackwell Publishers.
- Barandela, R., Sánchez, J. S., García, V., & Rangel, E. (2003). Strategies for learning in class imbalance problems. *Pattern Recognition*, 36(3), 849-851. [https://doi.org/10.1016/S0031-3203\(02\)00257-1](https://doi.org/10.1016/S0031-3203(02)00257-1)
- Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 40, 432-442. <https://doi.org/10.1016/j.jbankfin.2013.12.013>
- Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Source Journal of Accounting Research Empirical Research in Accounting: Selected Studies*, 4, 71-111.

- Becchetti, L., & Sierra, J. (2003). Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of Banking & Finance*, 27(11), 2099-2120. [https://doi.org/10.1016/S0378-4266\(02\)00319-9](https://doi.org/10.1016/S0378-4266(02)00319-9)
- Beynon, M. J., & Peel, M. J. (2001). Variable precision rough set theory and data discretisation: an application to corporate failure prediction. *Omega*, 29(6), 561-576. [https://doi.org/10.1016/S0305-0483\(01\)00045-7](https://doi.org/10.1016/S0305-0483(01)00045-7)
- Blanco-Oliver, A., Irimia-Dieguez, A., Oliver-Alfonso, M., & Wilson, N. (2015). Improving Bankruptcy Prediction in Micro-Entities by Using Nonlinear Effects and Non-Financial Variables, (2), 144-167.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Bureau Van Dijk. (2017). SABI. Sistemas de Análisis de Balances Ibéricos.
- Charitou, A., Neophytou, E., & Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, 13(3), 465-497. <https://doi.org/10.1080/0963818042000216811>
- Chen, M.-Y. (2011). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers and Mathematics with Applications*, 62, 4514-4524. <https://doi.org/10.1016/j.camwa.2011.10.030>
- Chen, M.-Y. (2013). A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. *Information Sciences*, 220, 180-195. <https://doi.org/10.1016/j.ins.2011.09.013>
- Chen, N., Ribeiro, B., Vieira, A., & Chen, A. (2013). Clustering and visualization of bankruptcy trajectory using self-organizing map. *Expert Systems With Applications*, 40, 385-393. <https://doi.org/10.1016/j.eswa.2012.07.047>
- Chi, L.-C., & Tang, T.-C. (2006). Bankruptcy Prediction: Application of Logit Analysis in Export Credit Risks. *Australian Journal of Management*, 31(1), 17-27. <https://doi.org/10.1177/031289620603100102>
- Cho, S., Kim, J., & Bae, J. K. (2009). An integrative model with subject weight based on neural network learning for bankruptcy prediction. *Expert Systems with Applications*, 36(1), 403-410. <https://doi.org/10.1016/j.eswa.2007.09.060>
- Chung, C.-C., Chen, T.-S., Lin, L.-H., Lin, Y.-C., & Lin, C.-M. (2016). Bankruptcy Prediction Using Cerebellar Model Neural Networks. *International Journal of Fuzzy Systems*, 18(2), 160-167. <https://doi.org/10.1007/s40815-015-0121-5>
- Chung, H.-M. M., & Tam, K. Y. (1993). A Comparative Analysis of Inductive-Learning Algorithms. *Intelligent Systems in Accounting, Finance and Management*, 2(1), 3-18. <https://doi.org/10.1002/j.1099-1174.1993.tb00031.x>
- Ciampi, F., & Gordini, N. N. (2013). Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises. *Journal of Small Business Management*, 51(1), 23-45. <https://doi.org/10.1111/j.1540-627X.2012.00376.x>
- Ciampi, F., Vallini, C., Gordini, N., & Benvenuti, M. (2009). Are Credit Scoring Models Able to Predict Small Enterprise Default? Statistical Evidence from Italian Small Enterprises.
- Cielen, A., Peeters, L., & Vanhoof, K. (2004). Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research*, 154(2), 526-532. [https://doi.org/10.1016/S0377-2217\(03\)00186-3](https://doi.org/10.1016/S0377-2217(03)00186-3)
- Cultrera, L., & Brédart, X. (2016). Bankruptcy prediction: the case of Belgian SMEs. *Review of Accounting and Finance*, 15(1), 101-119. <https://doi.org/10.1108/RAF-06-2014-0059>

- Dambolena, I. G., & Khoury, S. J. (1980). American Finance Association Ratio Stability and Corporate Failure. *Source: The Journal of Finance*, 35(4), 1017-1026.
- Deakin, E. B. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 10(1), 167. <https://doi.org/10.2307/2490225>
- Deloitte. (2014). El sector sanitario en España, situación actual y perspectivas de futuro.
- Díaz-Uriarte, R., & Alvarez De Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7:3. <https://doi.org/10.1186/1471-2105-7-3>
- Dietrich, J. R., & Kaplan, R. S. (1982). Empirical Analysis of the Commercial Loan Classification Decision. *Source: The Accounting Review*, 57(1), 18-38.
- Divsalar, M., & Roodsaz, H. (2012). A Robust Data-Mining Approach to Bankruptcy Prediction. *Journal of Forecasting*, 523(March 2011), 504-523.
- Dong, Y.-S. D. Y.-S., & Han, K.-S. H. K.-S. (2004). A comparison of several ensemble methods for text categorization. *IEEE International Conference on Services Computing, 2004. (SCC 2004). Proceedings. 2004*, 0-3. <https://doi.org/10.1109/SCC.2004.1358033>
- du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications*, 75, 25-43. <https://doi.org/10.1016/j.eswa.2017.01.016>
- Du Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, 242(1), 286-303. <https://doi.org/10.1016/j.ejor.2014.09.059>
- Dugan, M. T., & Zavgren, C. (1989). How a Bankruptcy Model Could Be Incorporated as an Analytical Procedure. *The CPA journal*, 94-65.
- Eom, J.-H., Kim, S.-C., & Zhang, B.-T. (2008). AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction. *Expert Systems with Applications*, 34(4), 2465-2479. <https://doi.org/10.1016/j.eswa.2007.04.015>
- Etheridge, H. L., Sriram, R. S., & Hsu, H. Y. K. (2000). A Comparison of Selected Artificial Neural Networks that Help Auditors Evaluate Client Financial Viability. *Decision Sciences*, 31(2), 531-550. <https://doi.org/10.1111/j.1540-5915.2000.tb01633.x>
- Fan, A., & Palaniswami, M. (2000). Selecting bankruptcy predictors using a support vector machine approach. En *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium* (pp. 354-359 vol.6). IEEE. <https://doi.org/10.1109/IJCNN.2000.859421>
- Fedorova, E., Gilenko, E., & Dovzhenko, S. (2013). Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications*, 40(18), 7285-7293. <https://doi.org/10.1016/j.eswa.2013.07.032>
- Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.
- FitzPatrick, P. (1932). *A comparison of the ratios of successful industrial enterprises with those of failed companies*. [Washington].
- Fritsch, S., & Guenther, F. (2010). neuralnet: Training of neural networks. *The R Journal*, 2(1), 30-38. <https://doi.org/10.1109/SP.2010.25>
- González-Gómez, F., García-Rubio, M. A., & Gonz Alez-Martínez, J. (2014). Beyond the publicprivate controversy in urban water management in Spain. *Utilities Policy*, 31, 1-9. <https://doi.org/10.1016/j.jup.2014.07.004>
- Gordini, N. A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy, 41 *Expert Systems with Applications* § (2014). Elsevier Ltd.

- <https://doi.org/10.1016/j.eswa.2014.04.026>
- Gu, Z., & Gao, L. (1999). A multivariate model for predicting business failures of hospitality firms. *Tourism and Hospitality Research*, 2(1).
- Guerrini, A., Romano, G., Ferretti, S., Fibbi, D., & Daddi, D. (2016). A Performance Measurement Tool Leading Wastewater Treatment Plants toward Economic Efficiency and Sustainability. *Sustainability*, 8(12), 1250. <https://doi.org/10.3390/su8121250>
- Hachicha, N., Jarboui, B., & Siarry, P. (2011). A fuzzy logic control using a differential evolution algorithm aimed at modelling the financial market dynamics. *Information Sciences*, 181(1), 79-91. <https://doi.org/10.1016/j.ins.2010.09.010>
- Hadipuro, W. (2010). Indonesia ' s Water Supply Regulatory Framework : Between Commercialisation and Public Service ? *Water Alternatives*, 3(3), 475-491.
- Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10), 993-1001. <https://doi.org/10.1109/34.58871>
- Hennawy, R. H. A., & Morris, R. C. (1983). The significance of base year in developing failure prediction models. *Journal of Business Finance & Accounting*, 10(2), 209-223. <https://doi.org/10.1111/j.1468-5957.1983.tb00424.x>
- Holland, J. H. (1975). *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press.
- Hornik, K., Buchta, C., & Zeileis Wirtschaftsuniversität Wien, A. W. (2009). Open-Source Machine Learning: R Meets Weka. <https://doi.org/10.1007/s00180-008-0119-7>
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2016). A Practical Guide to Support Vector Classification.
- Hua, Z., Wang, Y., Xu, X., Zhang, B., & Liang, L. (2007). Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications*, 33(2), 434-440. <https://doi.org/10.1016/j.eswa.2006.05.006>
- Jeong, C., Min, J. H., & Kim, M. S. (2012). A tuning method for the architecture of neural network models incorporating GAM and GA as applied to bankruptcy prediction. *Expert Systems with Applications*, 39(3), 3650-3658. <https://doi.org/10.1016/j.eswa.2011.09.056>
- Jones, F. L. (1987). Current Techniques in Bankruptcy Prediction. *Journal of Accounting Literature*, 6, 131-164.
- Joo, H. N., & Jin, T. (2000). Bankruptcy Prediction: Evidence from Korea Listed Companies During the IMF Crisis. *Journal of International Financial Management and Accounting*, 11(3), 195-217.
- Kaski, S., Sinkkonen, J., & Peltonen, J. (2001). Bankruptcy analysis with self-organizing maps in learning metrics. *IEEE Transactions on Neural Networks*, 12(4), 936-947. <https://doi.org/10.1109/72.935102>
- Khan, A. M. (1985). Analyzing financial statements for managerial performance measurement and bankruptcy prediction. *Engineering Management International*, 3(3), 165-174. [https://doi.org/10.1016/0167-5419\(85\)90003-1](https://doi.org/10.1016/0167-5419(85)90003-1)
- Kim, H.-J. J., Jo, N.-O. O., & Shin, K.-S. S. (2016). Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. *Expert Systems with Applications*, 59, 226-234. <https://doi.org/10.1016/j.eswa.2016.04.027>
- Kim, H., & Gu, Z. (2006). Predicting Restaurant Bankruptcy: A Logit Model in

- Comparison with a Discriminant Model. *Journal of Hospitality & Tourism Research*, 30(4), 474-493. <https://doi.org/10.1177/1096348006290114>
- Kim, M.-J. J., & Kang, D.-K. K. (2012). Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction. *Expert Systems with Applications*, 39(10), 9308-9314. <https://doi.org/10.1016/j.eswa.2012.02.072>
- Kittler, J., Hatef, M., Duin, R. P. W., & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 226-239. <https://doi.org/10.1109/34.667881>
- Ko, P.-C., & Lin, P.-C. (2006). An evolution-based approach with modularized evaluations to forecast financial distress. *Knowledge-Based Systems*, 19(1), 84-91. <https://doi.org/10.1016/j.knosys.2005.11.006>
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence*, 2, 1137-1143. <https://doi.org/10.1067/mod.2000.109031>
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59-69. <https://doi.org/10.1007/BF00337288>
- Laitinen, E. K., & Laitinen, T. (2000). Bankruptcy prediction Application of the Taylor's expansion in logistic regression. *International Review of Financial Analysis*, 9, 327-349.
- Lee, C.-H., & Yu, Y.-H. (2012). Characteristics of public-private partnerships for municipal wastewater treatment in Taiwan. *Journal of the Chinese Institute of Engineers*, 35(2), 245-258. <https://doi.org/10.1080/02533839.2012.638535>
- Lennox, C. (1999). Identifying failing companies: a re-evaluation of the logit, probit and DA approaches. *Journal of Economics and Business*, 51(4), 347-364. [https://doi.org/10.1016/S0148-6195\(99\)00009-0](https://doi.org/10.1016/S0148-6195(99)00009-0)
- Li, M. Y. L., & Miu, P. (2010). A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach. *Journal of Empirical Finance*, 17(4), 818-833. <https://doi.org/10.1016/j.jempfin.2010.04.004>
- Liaw, A., Wiener, M., & Andy Liaw, M. (2015). Breiman and Cutler's Random Forests for Classification and Regression Description Classification and regression based on a forest of trees using random inputs.
- Lin, W.-Y., Hu, Y.-H., & Tsai, C.-F. (2012). Machine Learning in Financial Crisis Prediction: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4), 421-436. <https://doi.org/10.1109/TSMCC.2011.2170420>
- López Iturriaga, F. J., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems with Applications*, 42(6), 2857-2868. <https://doi.org/10.1016/j.eswa.2014.11.025>
- Martin, D. (1977). Early warning of bank failure. *Journal of Banking & Finance*, 1(3), 249-276. [https://doi.org/10.1016/0378-4266\(77\)90022-X](https://doi.org/10.1016/0378-4266(77)90022-X)
- Mekonnen, M. M., & Hoekstra, A. Y. (2016). Four billion people facing severe water scarcity. *Science Advances*, 2(2), e1500323-e1500323. <https://doi.org/10.1126/sciadv.1500323>
- Messier, W. F., & Hansen, J. V. (1988). Inducing Rules for Expert System Development: An Example Using Default and Bankruptcy Data. *Management Science*, 34(12), 1403-1415. <https://doi.org/10.1287/mnsc.34.12.1403>
- Meyer, D. (2001). Support Vector Machines * The Interface to libsvm in package e1071, 13(9).
- Michael, S., Georgios, D., Nikolaos, M., & Constantin, Z. (1999). A Fuzzy Knowledge-

- based Decision Aiding Method for the Assessment of Financial Risk: The Case of Corporate Bankruptcy Prediction. *European Symposium on intelligent techniques*, (February).
- Min, J. H., & Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4), 603-614. <https://doi.org/10.1016/j.eswa.2004.12.008>
- Mousavi, M. M., Ouenniche, J., & Xu, B. (2015). Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. *International Review of Financial Analysis*, 42, 64-75. <https://doi.org/10.1016/j.irfa.2015.01.006>
- Nakazawa, M. (2015). fmsb: Functions for medical statistics book with some demographic data. *R package version 0.5.2*, 1-40.
- Nilsson, N. J. (1965). *Learning machines : foundations of trainable pattern-classifying systems*. McGraw-Hill.
- Nouri, B. A., & Soltani, M. (2016). Designing a Bankruptcy Prediction Model Based on Account, Market and Macroeconomic Variables, 9(1), 125-147.
- Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *International Joint Conference on Neural Networks*, 2, 163-168. <https://doi.org/10.1109/IJCNN.1990.137710>
- OECD/EU. (2016). Health at a Glance: Europe 2016 State of Health in the EU Cycle. <https://doi.org/10.1787/9789264265592-en>
- Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), 109-131. <https://doi.org/10.2307/2490395>
- ONU. (2015). *World Population Prospects*. New York.
- ONU. (2017). Informe Mundial sobre el Desarrollo de los Recursos Hídricos de las Naciones Unidas 2017: Las aguas residuales: el recurso desaprovechado, resumen ejecutivo; 2017.
- Oza, N. C., & Tumer, K. (2007). Classifier ensembles: Select real-world applications. <https://doi.org/10.1016/j.inffus.2007.07.002>
- Package «hybridEnsemble». (2015).
- Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems With Applications*, 36, 2-17. <https://doi.org/10.1016/j.eswa.2007.10.005>
- Pompe, P. P. M., & Bilderbeek, J. (2005). The prediction of bankruptcy of small- and medium-sized industrial firms. *Journal of Business Venturing*, 20(6), 847-868. <https://doi.org/10.1016/j.jbusvent.2004.07.003>
- PWC. (2014). *La gestión del agua en España, análisis de la situación actual del sector y retos futuros*.
- R Core, T. (2017). R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ranawana, R., & Palade, V. (2006). Multi-Classifer Systems: Review and a roadmap for developers. *International Journal of Hybrid Intelligent Systems*, 3(1), 35-61. <https://doi.org/10.3233/HIS-2006-3104>
- 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. <https://doi.org/10.1016/j.ejor.2006.08.043>
- Ravi Kumar, P., Ravi, V., Kumar, P. R., & 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.

- <https://doi.org/10.1016/j.ejor.2006.08.043>
- Rijsberman, F. R. (2006). Water scarcity: Fact or fiction? *Agricultural Water Management*, 80(1), 5-22. <https://doi.org/10.1016/j.agwat.2005.07.001>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., ... Müller Markus. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12(4116), 1-77. <https://doi.org/10.1126/science.182.4116.990>
- Rokach, L. (2009). Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography. *Computational Statistics & Data Analysis*, 53(12), 4046-4072. <https://doi.org/10.1016/j.csda.2009.07.017>
- Sato, T., Qadir, M., Yamamoto, S., Endo, T., & Zahoor, A. (2013). Global, regional, and country level need for data on wastewater generation, treatment, and use.
- Scrucca, L. (2013). GA : A Package for Genetic Algorithms in R. *Journal of Statistical Software*, 53(4), 1-37. <https://doi.org/10.18637/jss.v053.i04>
- Serrano-Cinca, C. (1996). Self organizing neural networks for financial diagnosis. *Decision Support Systems*, 17, 227-238.
- Shin, K.-S., Lee, T. S., & Kim, H.-J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, 28(1), 127-135. <https://doi.org/10.1016/j.eswa.2004.08.009>
- Shin, K.-S., & Lee, Y.-J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications*, 23(3), 321-328. [https://doi.org/10.1016/S0957-4174\(02\)00051-9](https://doi.org/10.1016/S0957-4174(02)00051-9)
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*, 74(1), 101-124. <https://doi.org/10.1086/209665>
- Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, 10(5), 335-347. [https://doi.org/10.1016/0167-8655\(89\)90037-8](https://doi.org/10.1016/0167-8655(89)90037-8)
- Tay, F. E. H., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29, 309-317.
- Tin Kam Ho, Hull, J. J., & Srihari, S. N. (1994). Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1), 66-75. <https://doi.org/10.1109/34.273716>
- Tsai. (2008). Financial decision support using neural networks and support vector machines. *Expert Systems*, 25(4), 380-393. <https://doi.org/10.1111/j.1468-0394.2008.00449.x>
- Tsai, Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing Journal*, 24, 977-984. <https://doi.org/10.1016/j.asoc.2014.08.047>
- Tsai, & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34(4), 2639-2649. <https://doi.org/10.1016/j.eswa.2007.05.019>
- Vapnik, V. N. (1998). *Statistical learning theory*. Wiley.
- 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. <https://doi.org/10.1016/j.eswa.2013.09.033>
- Wehrens, R., & Buydens, L. M. C. (2007). Self- and super-organizing maps in R: The kohonen package. *Journal of Statistical Software*, 21(5), 1-19. <https://doi.org/10.18637/jss.v021.i05>
- West, D., Dellana, S., & Qian, J. (2005). Neural network ensemble strategies for financial decision applications. *Computers & Operations Research*, 32(10), 2543-2559.

- <https://doi.org/10.1016/j.cor.2004.03.017>
- Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, 18(1), 30-55. <https://doi.org/10.1007/s10618-008-0116-z>
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Comprehensive R Archive Network (CRAN).
- Wilcox, J. W. (1973). A Prediction of Business Failure Using Accounting Data. *Journal of Accounting Research*, 11, 163. <https://doi.org/10.2307/2490035>
- Wilson, N., & Summers, B. (2002). Trade credit terms offered by small firms: Survey evidence and empirical analysis. *Journal of Business Finance and Accounting*, 29(3-4), 317-351. <https://doi.org/10.1111/1468-5957.00434>
- World Economic Forum. (2015). *Global Risks 2015. 10th Edition*. Geneva.
- Xie, Y., Li, X., Ngai, E. W. T., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36(3), 5445-5449. <https://doi.org/10.1016/j.eswa.2008.06.121>
- Xu, M., & Zhang, C. (2009). Bankruptcy prediction: The case of Japanese listed companies. *Review of Accounting Studies*, 14(4), 534-558. <https://doi.org/10.1007/s11142-008-9080-5>
- Yazdanfar, D., & Nilsson, M. (2008). The bankruptcy determinants of Swedish SMEs. *Institute for Small Business & Entrepreneurship*, (November), 1-14.
- Zadeh, L. a. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- Zhang, Z. (2016). Variable selection with stepwise and best subset approaches. *Annals of translational medicine*, 4(7), 136. <https://doi.org/10.21037/atm.2016.03.35>
- Zhong, L., Mol, A. P. J., & Fu, T. (2008). Public-private partnerships in China's urban water sector. *Environmental Management*, 41(6), 863-877. <https://doi.org/10.1007/s00267-008-9070-1>
- Zhou, L., Lai, K. K., & Yen, J. (2012). Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation. *International Journal of Systems Science*, 45(July 2015), 1-13. <https://doi.org/10.1080/00207721.2012.720293>
- Zibanezhad, E., Foroghi, D., & Monadjemi, A. (2011). Applying decision tree to predict bankruptcy. *Proceedings - 2011 IEEE International Conference on Computer Science and Automation Engineering, CSAE 2011*, 4, 165-169. <https://doi.org/10.1109/CSAE.2011.5952826>
- Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22(1), 59-82. <https://doi.org/10.2307/2490859>

Capítulo 3: Predicción del estrés financiero en empresas del sector privado del agua, una aproximación desde el punto de vista de la Inteligencia Artificial

3.1 Resumen

Existe una conciencia creciente sobre el uso del agua, los problemas que acarrea la escasez de recursos hídricos y la importancia de realizar una gestión integral del ciclo del agua. Así, la Agenda 21 la calificada como un bien económico y social. Por otro lado, debe tenerse en cuenta que la gestión de este recurso esencial se encuentra, cada vez más, en manos de empresas privadas. En este sentido conocer la salud financiera de estas empresas es de gran importancia para evitar los problemas económicos, sociales y ambientales que puede causar la entrada en quiebra de este tipo de empresas, por ello, es fundamental que éstas tengan una buena posición financiera. Por este motivo, en el presente trabajo se plantea establecer un modelo de predicción basado en inteligencia artificial (artificial intelligence, AI) y, para ello, se ha realizado la comparación del rendimiento para el problema mencionado de cinco técnicas de inteligencia artificial: a. Redes Neuronales (NN), b. *Support Vector Machine* (SVM), c. *Support Vector Machine Optimizado* (SVM-T), árboles de decisión (DT) y *Random Forest* (RF). Los resultados empíricos demuestran que la técnica que mejores predicciones realiza con los datos de validación es SVM-T, seguido de RF.

Palabras Claves: estrés financiero, empresas suministradoras de agua, Z-Altman, inteligencia artificial

3.2 Abstract

There is a growing awareness of water use, the problems of scarce water resources and the importance of comprehensive water cycle management. Thus, Agenda 21 is classified as an economic and social good. On the other hand, it must be taken into account that the management of this essential resource is increasingly in the hands of private companies. In this sense knowing the financial health of these companies is of great importance to avoid the economic, social and environmental problems that can cause the bankruptcy of this type of companies, therefore, it is fundamental that they have a good financial position. For this reason, the present work proposes to establish a prediction model based on artificial intelligence (AI) and, for this, the performance of the mentioned problem of five artificial intelligence techniques has been performed: a. Neural Networks (NN), b. Support Vector Machine (SVM), c. Support Optimized Vector Machine (SVM-T), decision trees (DT) and Random Forest (RF). The empirical results demonstrate that the technique that best predictions performs with validation data is SVM-T, followed by RF.

Keywords: financial stress, water supply companies, Z-Altman, artificial intelligence

3.3 Introducción

La población mundial pasará de 7.300 millones de habitantes a unos 11.200 millones en el año 2100 (ONU, 2015). Resulta evidente que este crecimiento poblacional puede convertirse en una seria amenaza para la humanidad si no viene acompañado por la generación de recursos suficientes para su abastecimiento. Entre estos recursos, qué duda cabe, el agua es uno de los elementos fundamentales para el mantenimiento de la vida en el planeta. Prácticamente ya nadie duda de la importancia del agua y de ésta como recurso clave para el desarrollo de cualquier sociedad.

Según Mekonnen y Hoekstra (2016), anualmente a nivel mundial existe suficiente agua dulce para satisfacer las necesidades actuales, pero las variaciones espaciales y temporales de la demanda y disponibilidad de agua son grandes, lo que lleva a la escasez de este líquido en varias partes del mundo durante épocas específicas del año. Por su parte, Rijsberman (2006) asegura que para el año 2025 hasta el 60% de la población mundial puede sufrir escasez física de agua. Quizá por este mismo motivo, el Foro Económico Mundial señala las crisis hídricas como el mayor riesgo mundial en términos de impacto potencial (World Economic Forum, 2015).

Hasta finales del siglo XX, la preocupación por el acceso al agua estuvo muy reducida a determinadas zonas geográficas de gran escasez y no se planteó como un problema de escala global. En la actualidad, el agua se concibe como un recurso básico para la vida humana que ha de ser gestionado como un bien económico escaso de creciente valor. En España las dificultades de abastecimiento y restricciones que periódicamente padecen determinadas poblaciones, una regulación fragmentada a lo largo del territorio nacional y las polémicas interregionales en torno a la asignación del recurso, provocan que el agua se convierta en un verdadero problema que afecta a sectores completos de nuestra economía y comienza a determinar las condiciones esenciales de vida de amplias zonas de población (PWC, 2014).

En los años 80, se inició en el Reino Unido y Estados Unidos una privatización de monopolios que antes estaban en manos públicas. Esta ola de privatización también se extendió a los servicios públicos locales y algunos países facilitaron la participación privada en la gestión de los servicios de agua urbanos. En 1989, Inglaterra y Gales decidieron la privatización completa de esta industria. Hoy en día, en Francia, Chile o

República Checa, las empresas privadas suministran servicios de agua a una proporción significativa de la población (González-Gómez *et al.*, 2014).

España no ha sido inmune a este cambio y a partir de 1985, la legislación española ha permitido la privatización de los servicios urbanos del agua. Como consecuencia, particularmente desde la última década del siglo XX, ha habido una marcada expansión de la privatización en este sector, llegando ya en 2012 a una situación en la que la gestión privada (a través de empresas mixtas o privadas) supone ya el 50%, generando unos ingresos de 5.202 millones de euros procedentes del suministro de agua (AEAS, 2014).

El debate de la gestión pública o privada del agua, ha sido importante en las últimas décadas. Los partidarios de la privatización la defienden como la mejor manera de lograr una gestión más eficiente de los recursos y ofrecer un servicio de mejor calidad. Por el contrario, los detractores argumentan que la gestión pública es la mejor opción para asegurar el acceso universal a una necesidad primaria y básica, así como para evitar que las empresas privadas, en un monopolio natural, abusen de su posición dominante (González-Gómez *et al.*, 2014). Independientemente de la citada controversia, lo que resulta indudable, es que la presencia de las empresas privadas, cada vez es mayor en sector del agua español.

España, a pesar de ser la primera productora de agua desalada en Europa, América y Oriente Medio, no cuenta con un organismo regulador técnico independiente que aporte predictibilidad y estabilidad al sector, a pesar del elevado número de administraciones públicas con competencias en esta materia (PWC, 2014).

A la falta de un ente regulador óptimo, se le añade el hecho de que al precio del agua no se le repercute el coste real del servicio y que existe un déficit en las inversiones a realizar por el sector, lo que conlleva sustanciales pérdidas de agua en la red de distribución e ineficiencia de las infraestructuras (AEAS, 2014). Todo ello genera una situación que puede conducir que estas empresas lleguen al colapso financiero, lo que podría conllevar problemas de desabastecimiento y complejos procesos de rescate. Un ejemplo claro de una situación como la que se plantea es el de Indonesia donde, a pesar de ser uno de los países más ricos desde un punto de vista hídrico, sus empresas de suministro de agua se vieron en graves problemas financieros a raíz de la crisis de 1998 (Hadipuro, 2010).

Los métodos para predecir la bancarrota en las empresas, se han convertido en un asunto de alto interés desde la década de 1960, desde entonces se han propuesto distintas herramientas de predicción. Un mayor énfasis en el desarrollo de modelos predictivos, podría ser tomado como un indicador del grado de desarrollo y robustez de la economía de un país. También ha contribuido al desarrollo de los citados modelos y realizar esfuerzos para proporcionar una mejor visión y predicción de los eventos de quiebra, los altos costes individuales, económicos y sociales inherentes a los fracasos corporativos (M.-Y. Chen, 2011).

Los modelos de predicción de bancarrota pueden clasificarse en dos grandes categorías: técnicas estadísticas convencionales e IA. Beaver (1966) fue pionero en los métodos estadísticos, seguido por Altman (1968), que aplicó el análisis discriminante múltiple (ADM), posteriormente, Ohlson (1980) utilizó un modelo logit y Zmijewski, (1984) un modelo probit. Con la intención de obtener unos modelos predictivos más precisos, diversas técnicas de IA como NN, SVM, *fuzzy logic*, algoritmos genéticos, DT o RF, se han aplicado con éxito para la predicción de la bancarrota.

Ante todo ello, el presente trabajo tiene como objetivo mejorar el conocimiento desde un punto de vista financiero de un sector imprescindible para el desarrollo humano, como es el sector hídrico. Por tanto, se pretende plantear un modelo que se ajuste a las características de estas empresas y permita predecir con un porcentaje de acierto elevado aquellas empresas que en un horizonte temporal de dos años tengan un alto riesgo de presentar problemas financieros severos. Este tipo de predicciones ya se realizan en otras empresas y sectores (e.g. empresas que cotizan en la bolsa de Taiwan (M.-Y. Chen, 2011), empresas de tamaño medio manufactureras de Korea (Shin y Lee, 2002)), sin embargo no existe modelo específico para el sector del agua.

Otro de los objetivos que se plantean, es realizar una comparativa entre dos modelos con distinto número de variables predictivas, para comprobar cuál de ellos presenta un mejor rendimiento (en el apartado Análisis Estadístico y en la Tabla 3.2, que se comentan con posterioridad, aparecen detallados como se han creado estos dos modelos).

Este artículo se ha organizado de la siguiente forma. La sección 2 presenta la revisión de la literatura, la sección 3 la metodología (técnicas empleadas, medidas de

rendimiento y datos empleados). La sección 4 muestra los resultados y la sección 5 la discusión y limitaciones.

3.4 Revisión de la Literatura

La grave crisis financiera que ha venido afectando a la gran mayoría de países en los últimos años, ha generado una cada vez mayor preocupación por el riesgo financiero y ha relanzado el interés por trabajos que analizan las causas del fracaso empresarial, con el claro objetivo de establecer un método fiable que permita detectar y paliar, con la suficiente anticipación, los graves efectos que suelen generar este tipo de situaciones. Es evidente que detectar el fracaso empresarial resulta necesario para garantizar la estabilidad de los stakeholders, los mercados financieros y la prosperidad económica general (Jones, 1987). Aunque resulta evidente la importancia de la identificación temprana de los problemas financieros, en el caso concreto del sector del agua adquiere mayor significación al tratarse de un bien básico y necesario (PWC, 2014), de esta forma, se podrían evitar problemas de desabastecimiento o complejos procesos de rescate del servicio por parte de la administración pública.

La literatura de la predicción de la bancarrota comienza en la década de 1930. FitzPatrick (1932) comparó trece ratios de empresas en bancarrota y empresas en no-bancarrota (activas) y descubrió que las empresas activas poseían ratios más favorables. Hasta mediados de la década de los años 60, las investigaciones se basaban en análisis univariable. El más reconocido fue el de Beaver (1966), que encontró que un número determinado de ratios podían discriminar entre bancarrota y no-bancarrota con hasta cinco años de antelación. Dicho autor sugirió la posibilidad de que múltiples ratios considerados al mismo tiempo podrían aumentar el poder predictivo y, de esta forma, empezó la evolución de los modelos predictivos de bancarrota. Altman (1968) fue el primero en publicar un análisis multivariable. Este autor utilizó el análisis discriminante múltiple (ADM) para desarrollar un modelo predictivo usando cinco ratios financieros para predecir la bancarrota de empresas manufactureras.

Por su parte, Ohlson (1980) utilizó un modelo logit para superar las desventajas asociadas a los modelos ADM. Este estudio utilizaba 9 variables, de las cuales, ocho eran ratios financieros. Zmijewski (1984) desarrolló una nueva fórmula utilizando un modelo

Capítulo 3: Predicción del estrés financiero en empresas del sector privado del agua, una aproximación desde el punto de vista de la Inteligencia Artificial

probit. Este estudio utilizaba tres ratios financieros y la exactitud de este modelo con los datos de su muestra fue del 99%. Altman (2000) adaptó su fórmula propuesta en 1968 para el sector de las empresas privadas, cambiando la variable “*Market Value of Equity*” (capitalización bursátil) por “*Book values of equity*” (patrimonio neto).

La Tabla 3.1 (Altman, 1968; Altman, 2000; Aziz *et al.*, 1988; Bauer y Agarwal, 2014; Beaver, 1966; Becchetti y Sierra, 2003; Beynon y Peel, 2001; Blanco-Oliver *et al.*, 2015; Charitou *et al.*, 2004; Chi y Tang, 2006; Cho *et al.*, 2009; Chung *et al.*, 2016; Cielen *et al.*, 2004; Cultrera y Brédart, 2016; Dambolena y Khoury, 1980; Divsalar y Roodsaz, 2012; du Jardin, 2017; Dugan y Zavgren, 1989; Gu y Gao, 1999; Jeong *et al.*, 2012; Joo y Jin, 2000; Khan, 1985; Kim *et al.*, 2016; Kim y Gu, 2006; Kim y Kang, 2012; Laitinen y Laitinen, 2000; Lennox, 1999; Li y Miu, 2010; Mousavi *et al.*, 2015; Nouri y Soltani, 2016; Ohlson, 1980; Pompe y Bilderbeek, 2005; Shumway, 2001; Wang *et al.*, 2014; Wilson y Summers, 2002; Xu y Zhang, 2009) muestra las principales ratios financieros utilizados y sus referencias.

Tabla 3.1 Ejemplos de ratios financieros, sus categorías y referencias

Categorías	Variables y definiciones	Estudios Previos
I. CAPITAL STRUCTURE VARIABLE	Long Term Debt to Total Asset	Beaver 1966; Ohlson 1980; Gu y Gao, 2000; Mousavi et al, 2015
	Short and Long Term Debt to Equity	Chi et al 2006; Zmijewski,1984
	Short y Long Term Debt to Total Assets	Beaver 1966; Laitinen et al 2000; Lennox, 1999,
	Cash-Holdings to Total Assets	Dugan et al 1989; Laitinen et al 2000
	Equity Share of Total Assets	Altman,1968; Cielen et al 2004; Charitou et al , 2004; Pantalone y Platt, 1987; Pompe et al 2005; Shumway,2001
	Current Liabilities to Total Assets	Blanco-Oliver et al, 2015; Gu y Gao 2000; Jeong et al, 2012; Lo, 1985; Pantalone y Platt ,1987; Zmijewski,1984
	Total liabilities to total assets	Divsalar et al, 2012, Mousavi et al, 2015; Xu y Zhang, 2009
	Total liabilities to current assets	Blanco-Oliver et al, 2015
	Total liabilities to working capital	Mousavi et al, 2015
	Quick assets to total assets	Divsalar et al, 2012
	Quick assets to current assets	Blanco-Oliver et al, 2015
	Quick assets to current liabilities	Du Jardin, 2017
	Equity to total assets	Cultrera y Brédart, 2016
	Current assets to total assets	Kim y Kang, 2012; Wang et al, 2014; Zhao et al, 2017
	Current debt to total debt	Wang et al, 2014
	Retained earnings to total assets	Blanco-Oliver et al, 2015; Li y Miu, 2010; Kim y Kang, 2012
Current Liabilities to total capital	Cho et al, 2009	

Capítulo 3: Predicción del estrés financiero en empresas del sector privado del agua, una aproximación desde el punto de vista de la Inteligencia Artificial

	Current ratio	Blanco-Oliver et al, 2015; Cultrera y Brédart, 2016; Chung et al, 2016; Wang et al, 2014; Xu y Zhang, 2009; Zhao et al, 2017
	Debt-total assets	Chung et al, 2016; Kim y Kang, 2012
	Total borrowings to total assets	Kim et al, 2016
II. LIQUIDITY RATIOS	Quick Ratio	Altman,1968; Cielen et al 2004; Charitou et al , 2004; Chung et al, 2016; Nam y Jinn ,2001; Kim y Kam, 2016; Shumways, 2001
	Working capital to total asset	Nouri y Soltari, 2016; Li y Miu, 2010; Mousavi et al, 2015; Xu y Zhang, 2009; Zhao et al, 2017
	Working capital to sales	Zhao et al, 2017
	Cash to current debt	Du Jardin, 2017; Wang et al, 2014; Zhao et al, 2017
	Cash to current assets	Du Jardin, 2017
	Cash to total assets	Blanco-Oliver et al, 2015; Du Jardin, 2017, Zhao et al, 2017
	Cash to Long y Short Term Debt	Dugan y Zavgren 1989; Charitou et al , 2004
III. PROFITABILITY RATIOS	Return on Equity	Altman, 1968; Pompe et al 2005; Shumways, 2001
	Return on Assets	Becchetti et al 2003; Chung et al, 2016; Dambolena y Khoury 1980; Kim et al, 2016; Nouri y Soltari, 2016; Ohlson, 1980; Zmijewski, 1984
	Net Profit Margin	Beaver 1966; Kim et al 2006
	Net Profit to total assets	Zhao et al, 2017
	Net Profit to current assets	Zhao et al, 2017
	Gross Profit Margin	Cielen et al 2004; Kim et al 2006
	Gross Profit to sales	Zhao et al, 2017
	Profit Margin	Becchetti et al 2003; Pompe et al 2005
	Net profit to liabilities	Zhao et al, 2017
	Net incomen to current assets	Mousavi et al, 2015
	Net incomen to total assets	Bauer y Agarwal, 2014; Xu y Zhang, 2009
	Net incomen to capital	Mousavi et al, 2015
	Net incomen to sales	Jeong et al, 2012
	Ordinary income to sales	Jeong et al, 2012; Kim y Kang, 2012
	Ordinary income to total assets	Kim y Kang, 2012
	Ordinary income to capital	Kim y Kang, 2012
	Return on capital employment	Nouri y Soltari, 2016
Profit to sales	Cho et al, 2009	
Operating profit to sales	Cho et al, 2009	
IV. CASHFLOW	Cash flow to Sales	Beaver 1966; Wang et al, 2014
	Cash flow to Total Assets	Aziz et al. , 1988; Pompe et al 2005; Young et al, 2005
	Cash flow to total debt	Cultrera y Brédart, 2016
	Operating cash flow to sales	Wang et al, 2014
	Operating cash flow to total debt	Kim y Kang, 2012; Wang et al, 2014
	Operating cash flow to total assets	Kim et al, 2016; Wang et al, 2014

Capítulo 3: Predicción del estrés financiero en empresas del sector privado del agua, una aproximación desde el punto de vista de la Inteligencia Artificial

	Cash flow to long y short term debt	Beaver 1966; Chung et al, 2016
	Sales to Total Assets	Altman 1968; ; Li y Miu, 2010
	Sales to current assets	Divsalar et al, 2012
	EBITDA to Total Assets	Altman 1968; Cultrera y Brédart, 2016; Du Jardin, 2017; Gu y Gao, 2000; Young et al 2005
	EBITDA to sales	Kim et al, 2016
	EBIT to total assets	Mousavi et al, 2015
	EBITDA to interest expenses	Kim y Kang, 2012
	Retained Earnings to Total Assets	Altman 1968
	Financial Expenses to Sales	Becchetti et al 2003
V. DEVELOPMENT AND FIRM	Change in Sales from Last Year	Arshad 1985; Pompe et al 2005
	Change in Total Assets from Last Year	Arshad 1985; Becchetti et al 2003
	Firms Age	Beaver 1966; Kim et al 2006; Altman, 2000
	Firm Size (Natural Logarithm of sales)	Cielen et al 2004; Kim et al 2006
	Firm Size (Natural Logarithm of total assets)	Blanco-Oliver et al, 2015; Beynon, y Peel, 2001; Kim y Kang, 2012; Ohlson, 1980; Wilson y Summer, 2002
	Growth rate of tangible assets	Cho et al, 2009
	Growth rate of total assets	Jeong, et a, 2012
	Growth rate of current assets	Cho et al, 2009
VI. TURNOVER	Accounts Receivable to Sales	Beaver 1966
	Inventory to Sales	Beaver 1966; Du Jardin, 2017; Kim y Kang, 2012; Mousavi et al, 2015; Young et al 2005
	Inventory to Total Assets	Cielen et al 2004
	Current assets to sales	Du Jardin, 2017; Wang et al, 2014
	Current liabilities to sales	Du Jadin, 2017
	Receivables to sales	Du Jadin, 2017
	Sales to total assets	Du Jadin, 2017
	Asset turnover ratio	Chung et al, 2016; Jeong et al, 2012; Nouri y Soltari, 2016
	Inventory to turnover ratio	Nouri y Soltari, 2016
	Turnover of managerial assets	Cho et al, 2009
VII. MARKET VARIABLES	Market value to total debt	Nouri y Soltari, 2016
	Market value of total assets	Bauer y Agarwal, 2014
	Market value of equity to total assets	Bauer y Agarwal, 2014; Li y Miu, 2010
	Stock price	Nouri y Soltari, 2016
	Stock return	Nouri y Soltari, 2016
VIII. OTROS	Productivity of capital	Kim et al, 2016
	Interest expenses-gross profit	Chung et al, 2016
	Interest expenses to sales	Cho et al, 2009
	Financial expenses to total debt	Kim y Kang, 2012
	Capital employed to total liabilities	Blanco-Oliver et al, 2015

Hasta 1990, las técnicas estadísticas multivariantes más usadas para la predicción de la bancarrota eran el ADM (e.g. E. I. Altman, 1968, 1983), el logit (e.g. Ohlson, 1980) y el probit (e.g. Zmijewski, 1984). Sin embargo, estas técnicas deben cumplir unos supuestos como la linealidad, la normalidad y la independencia de las observaciones, que no siempre se cumplen en el ámbito financiero (Hua *et al.*, 2007). En consecuencia, dichos métodos pueden tener limitaciones de validez y efectividad (Shin y Lee, 2002).

Desde finales de los años 80, las técnicas de AI, que son menos vulnerables a las suposiciones comentadas previamente (Chung y Tam, 1993; Messier y Hansen, 1988), han sido aplicadas con éxito en la predicción de la bancarrota (e.g. Angelini *et al.*, 2008; Atiya, 2001; Paliwal y Kumar, 2009). Fan y Palaniswami, 2000 fueron los primeros en utilizar SVMs en problemas de predicción de bancarrota. La técnica SVM fue comparada con las técnicas “linear discriminant analysis” (LDA), redes neuronales (NN) y “learning vector quantization” (LVQ) con un fichero de empresas australianas. SVM obtuvo los mejores resultados con una exactitud del 70%, seguido por las NN con un 68%, seguido de LVQ con un 63% y seguido de LDA con un 63% (Atiya, 2001).

Uno de los primeros estudios en aplicar los DT a la predicción de la bancarrota fue Zibanezhad *et al.* (2011), donde se clasificaban correctamente el 94,5% de los datos de entrenamiento y el 90% de los datos de validación.

Breiman (2001) propuso los RF y utilizó el árbol de clasificación y regresión para realizar la selección de características. En el ámbito financiero, RF ha sido implementado con éxito para la detección de fraudes de tarjetas de crédito (Whitrow *et al.*, 2009) y la predicción de la pérdida de clientes de los bancos (Xie *et al.*, 2009). También ha sido utilizado por López Iturriaga y Sanz (2015) para estudiar la predicción de la bancarrota de los bancos norteamericanos comerciales. La predicción de RF era del 88,46% frente al 92,31% de las NN para los bancos activos y del 86,54% y 94,23% para los bancos no activos respectivamente.

3.5 Metodología

3.5.1 Datos

Los datos se obtuvieron de la base de datos SABI (Sistema de Análisis de Balances Ibéricos) (Bureau Van Dijk, 2017). Se seleccionaron aquellas empresas cuyo CNAE (Clasificación Nacional de Actividades Económicas) era el 3600 (Captación, depuración y distribución de agua) y, además, se encontraban en situación “activa”. Se incluyeron un total de 583 empresas y se seleccionaron las que tenían todos los ratios informados. Ello hizo que la muestra final quedase en 270 empresas, 111 se encontraban en estrés financiero y 159 que no lo estaban. Para definir si la empresa se encontraba dicha situación, se utilizó la fórmula Z'-Score de Altman para empresas privadas (Altman, 2000). Esta elección se debió a que Altman es uno de los autores más citados en relación a la predicción de la bancarrota (Almamy *et al.*, 2016):

$$Z' = 0.717 * X1 + 0.847 * X2 + 3.107 * X3 + 0.420 * X4 + 0.998 * X5 \quad (\text{fórmula 3.1})$$

donde:

X1: Working Capital/Total Assets

X2: Retained Earnings/Total Assets

X3: EBIT/Total Assets

X4: Book Value Equity/Total liabilities

X5: Sales/Total Assets

Los valores umbrales que estableció el autor fueron los siguientes: puntuaciones del valor Z' superiores a 2,90, indican que la empresa está en una zona segura. Si la puntuación está entre 1,23 y 2,90, la empresa está en una zona gris y si la puntuación está por debajo de 1,23, entonces la empresa está en riesgo de bancarrota. En nuestro caso, se clasificó a la empresa en estrés financiero si su valor Z' era inferior a 1,23.

Los datos para predecir el estrés financiero corresponden a dos años antes respecto al momento donde se evaluó la fórmula Z'-Score de Altman para determinar si la empresa presentaba estrés financiero o no.

3.5.2 Técnicas de Inteligencia Artificial utilizadas

A continuación se detallan sucintamente las técnicas de IA utilizadas en el presente trabajo.

3.5.2.1 Redes Neuronales (NN)

Las NN ofrecen un paradigma computacional inspirado en las redes neuronales biológicas del sistema nervioso humano (Kumar *et al.*, 2007). Contienen unidades de procesamiento de información similares a las neuronas disponibles en el cerebro humano (Tsai *et al.*, 2014). Las NN pueden aprender con la experiencia y tomar decisiones basadas en éstas. Los nodos y las conexiones a través de estos son análogos a las neuronas del cerebro y sus correspondientes sinapsis. Una red neuronal consiste en nodos neuronales que están vinculados a nodos ponderados. Todas las ponderaciones de una NN, son modificados constantemente para corregir o reducir errores hasta que el error total de todos los datos de entrenamiento alcanza el límite de tolerancia definido previamente. El modelo de NN más común es el perceptrón multicapa, el cual está formado por una capa de entrada que contiene un número determinado de nodos sensoriales, equivalentes a las variables independientes, una o más capas ocultas de nodos computacionales y una capa de salida correspondiente a las variables dependientes. Cada interconexión está asociada con un peso que es ajustado durante la fase de entrenamiento (Tsai, 2008).

Para seleccionar el número de capas ocultas y el número de nodos en cada capa, se utilizó la técnica de los AG para minimizar el error obtenido a través de la NN. Para el modelo 1 (véase apartado análisis estadístico y Tabla 3.2), se obtuvo que el número de nodos óptimos en la primera capa oculta era de 12 y 11 nodos en la segunda. En el modelo 2, se obtuvo que el número óptimo de nodos en la primera capa oculta era de 14 y 5 nodos ocultos en la segunda capa.

3.5.2.2 Support Vector Machine

Vapnik, 1998 fue el primero en utilizar la técnica SVM para realizar clasificaciones binarias. Utilizan un modelo lineal para implementar límites de clases no lineales a través de algunos vectores de entrada de mapeo no lineales en un espacio de entidad de alta dimensionalidad. El modelo lineal construido en el nuevo espacio puede representar un límite de decisión no lineal en el espacio original. En el nuevo espacio, se construye un hiperplano de separación óptimo. Así, SVM se conoce como el algoritmo que encuentra un tipo especial de modelo lineal, el hiperplano de margen máximo. Este hiperplano da la máxima separación entre las clases de decisión. Los ejemplos de entrenamiento más cercanos al hiperplano de margen máximo se llaman vectores de soporte. Todos los demás ejemplos de entrenamiento son irrelevantes para definir los límites de la clase binaria.

SVM es lo suficientemente simple como para ser analizado matemáticamente, ya que se puede demostrar que corresponde a un método lineal en un espacio de características de alta dimensional no linealmente relacionado con el espacio de entrada. En este sentido, SVM puede servir como una alternativa prometedora que combina las fortalezas de los métodos estadísticos convencionales, que son más teóricos y fáciles de analizar, y de los métodos de aprendizaje automático, libres de distribución y robustos (Min y Lee, 2005).

En este estudio, la función kernel utilizada fue la “función radial basis”. Además, hay dos parámetros asociados a la “función radial basis” que juegan un papel muy importante en el desarrollo de la ejecución del SVM: C (cost) y γ (gamma). C es el parámetro de penalización del término de error. Y γ (gamma) es el coeficiente de la función kernel, que en este caso es la “función radial basis”. La incorrecta selección de estos parámetros pueden causar problemas de “overfitting” o “underfitting” (Hsu *et al.*, 2016; Tay y Cao, 2001).

Para comprobar la importancia de la correcta selección de estos parámetros, y teniendo en cuenta que no se sabe de antemano cuales van a ser los valores más óptimos para un problema dado (Hsu *et al.*, 2016), se proponen dos técnicas de SVM:

1. El valor C era de 100 y el parámetro gamma de 1. Este modelo corresponde con la técnica SVM.

2. El valor C se seleccionaba entre los valores 0,1, 1, 10 ó 100 y el valor de gamma entre 0,5, 1 ó 2. Este modelo corresponde con la técnica SVM-T.

3.5.2.3 Decision Tree J48

El árbol de decisión toma la forma de una estructura de árbol inversa, es decir, de arriba a abajo, donde se van desglosando los datos para crear las ramas. Un árbol de decisión se construye con nodos, donde cada nodo denota una prueba en una variable, y cada rama que sale del nodo representa un resultado de la prueba. El nodo superior de un árbol es el nodo raíz con el que mayor discriminación (o ganancia de información) se realiza. Después del nodo raíz, la variable que tenga mayor poder discriminatorio, se elige como la prueba para el siguiente nodo. Este proceso continúa hasta que todos los atributos se comparan o no hay atributos restantes en los que las muestras puedan tener más ramificaciones (Tsai *et al.*, 2014). En este estudio, se utiliza el modelo J48, que corresponde con el árbol de decisión C4.5 implementado en la herramienta WEKA de minería de datos (Waikato Environment for Knowledge Analysis) y que permite tratar datos categóricos.

3.5.2.4 Random Forest

El *Random Forest* es un algoritmo de clasificación elaborado por Leo Breiman que utiliza un conjunto de árboles de clasificación (decisión tree). Cada uno de los árboles de clasificación se construye utilizando una muestra de bootstrap de los datos, y en cada división el conjunto de variables candidatas es un subconjunto aleatorio del total de las variables (Díaz-Uriarte y Alvarez De Andrés, 2006). RF trata de crear diferentes árboles de decisión para obtener la mejor clasificación entre diferentes clases de datos de acuerdo con la variable dependiente. Esta técnica tiene mejores resultados en términos de precisión de clasificación al construir un gran número de árboles en lugar de un solo árbol.

El RF es un algoritmo de clasificación bien adaptado para los datos de microarrays: muestra un excelente rendimiento, incluso cuando la mayoría de las variables predictivas son ruidosas. También puede utilizarse cuando el número de

variables es mucho mayor que el número de observaciones y en los problemas que involucran más de dos categorías en la variable dependiente (Díaz-Uriarte y Alvarez De Andrés, 2006).

3.5.2.5 Técnica de los Algoritmos Genéticos

Los AG simulan la teoría de la evolución propuesta por Darwin (Gordini, 2014). Fueron desarrollados por Holland (1975) y son procesos estocásticos robustos que pueden usarse para selección binaria de variables o para resolver problemas de búsqueda y optimización a través de una función *fitness* o función objetivo. Esta función, es una función de un cromosoma cuyo valor siempre se utiliza para medir el rendimiento de un grupo de parámetros asociados al cromosoma correspondiente y que siempre se debe maximizar. La función de *fitness* se puede elegir en función de las necesidades de los objetivos de negocio (Zhou *et al.*, 2012).

Los AG realizan el proceso de búsqueda en cuatro pasos: inicialización, selección, cruce y mutación (Shin y Lee, 2002).

En el proceso de inicialización, una población de individuos, llamados cromosomas, se distribuyen de forma aleatoria por el rango de búsqueda definido previamente para cada variable y son evaluados a través de la función *fitness* (Gordini, 2014).

El proceso de selección se basa en el principio de supervivencia de los individuos mejor adaptados, es decir, aquellos cromosomas que presenten valores más óptimos en la función *fitness*, serán seleccionados para la crear la siguiente generación (Gordini, 2014). En el proceso de cruce, se seleccionan dos individuos con valores óptimos en la función *fitness* y se unen de manera que el descendiente contiene información parcial de cada uno de sus progenitores, de esta forma, cabe esperar, que el descendiente mejore el resultado de la función *fitness* obtenido por sus progenitores (Shin y Lee, 2002).

La mutación en los AG es un mecanismo donde de forma aleatoria se seleccionan miembros de la población y se cambia alguno de sus genes. De esta forma, se evitan máximos locales, ya que se desplazan valores numéricos hacia zonas del espacio de

búsqueda que no pueden ser alcanzadas por los otros operadores genéticos (Shin y Lee, 2002).

En este análisis, se han utilizado los AG para selección binaria de variables a través de modelos de regresión logística (Scrucca, 2013), en el primer modelo se minimizaba el error “AIC” y en el segundo el error “BIC”. Las variables que sean seleccionadas en este paso, serán las que se utilicen posteriormente con las técnicas de IA. Siedlecki y Sklansky (1989) fueron los primeros que propusieron la utilización de los AG para la selección de variables.

Por otra parte, tal y como se ha comentado previamente, también se han utilizado los AG para seleccionar el número de capas y nodos óptimos en la técnica NN, pero no se han utilizado como una técnica predictiva.

3.5.3 Análisis estadístico

A través de una revisión de la bibliografía (Altman, 1968; Ciampi *et al.*, 2009; Ciampi y Gordini, 2013; Fedorova *et al.*, 2013; Yazdanfar y Nilsson, 2008; Zmijewski, 1984), se seleccionaron inicialmente 48 ratios financieros, en la Tabla 3.2 se representa la estadística descriptiva de estas ratios.

Tabla 3.2 Estadística descriptiva de las ratios financieras y selección de variables

Variable		Media (Dt)	Mediana (P25 – P75)	Min - Max
Pasivo/PN	R01	8,3 (88)	1,47 (0,51 - 2,96)	(-25,28 - 1442,77)
PC/ Pasivo	R02	0,59 (0,27)	0,61 (0,37 - 0,8)	(0,04 - 1)
RAIT	R03	2364,09 (13197,92)	88,5 (9 - 465,75)	(-6825 - 166175)
RE	R04	0,03 (0,1)	0,03 (0,01 - 0,06)	(-0,72 - 0,4)
Ingresos explotación/ Total activo	R05	0,7 (0,61)	0,57 (0,28 - 0,91)	(0,01 - 3,71)
Reservas/Total activo	R06	0,22 (0,27)	0,13 (0,02 - 0,33)	(0 - 2,05)
CC/Total activo	R07	0,06 (0,29)	0,07 (-0,07 - 0,2)	(-1,42 - 0,93)
PN/Pasivo	R08	1,61 (3,92)	0,59 (0,25 - 1,43)	(-0,9 - 45,78)
Res/AT	R09	0,01 (0,09)	0,01 (0 - 0,04)	(-0,74 - 0,29)
PT/AT	R10	0,65 (0,62)	0,63 (0,41 - 0,8)	(0,02 - 9,58)
AC/PC	R11	2,09 (4,1)	1,23 (0,77 - 1,95)	(0 - 52,05)
Deudas CP/PN	R12	1,06 (11,05)	0,14 (0,02 - 0,46)	(-3,3 - 181,1)
Tesorería/Total activo	R13	0,08 (0,12)	0,03 (0,01 - 0,09)	(0 - 0,96)
EBITDA/Total activo	R14	0,06 (0,1)	0,06 (0,02 - 0,1)	(-0,69 - 0,46)
EBITDA/Gastos financieros	R15	381,01 (4399,29)	4,72 (1,71 - 14,09)	(-4427,1 - 60345,77)
Cash Flow/Deuda total	R16	0,34 (1,06)	0,09 (0,03 - 0,28)	(0 - 12,47)
Total deuda/Total deuda + PN	R17	0,65 (0,62)	0,63 (0,41 - 0,8)	(0,02 - 9,58)
Test acido	R18	1,58 (2,42)	1 (0,6 - 1,5)	(0 - 21,88)
Interest/Turn over	R19	0,06 (0,26)	0,02 (0,01 - 0,05)	(0 - 4,1)
Equity/Long term material asset (PN/IM)	R20	15,13 (61,25)	1,54 (0,8 - 6,9)	(-47,12 - 692,35)
Net financial position /Turn over	R21	0,41 (1,3)	0,12 (0,03 - 0,31)	(0 - 14,74)
Long term assests/n ² employees	R22	-16,35 (42,52)	-4,88 (-9,87 - -1,63)	(-403,25 - -0,02)
Interes charges/bank loan	R23	9,79 (119,75)	0,06 (0,03 - 0,11)	(0 - 1866,18)
Quick Ratio (AC-exis/PC)	R24	1,59 (2,42)	1,01 (0,61 - 1,5)	(0 - 21,88)
RF	R25	0,06 (1,03)	0,05 (0 - 0,12)	(-8,99 - 12,74)
Activos explotación	R26	56182,89 (301148,57)	2779,5 (684,25 - 15318)	(17 - 3368912)

Capítulo 3: Predicción del estrés financiero en empresas del sector privado del agua, una aproximación desde el punto de vista de la Inteligencia Artificial

Resultado explotación	R27	1528,62 (9774,43)	56 (5 - 386,25)	(-12798 - 116564)
Rexplo	R28	0,03 (0,12)	0,03 (0,01 - 0,07)	(-0,72 - 0,46)
Cash Flow	R29	3212,68 (13930,93)	148,98 (20,84 - 1099,8)	(-6448,96 - 163784)
Cash Flow/AT	R30	0,05 (0,09)	0,05 (0,02 - 0,09)	(-0,72 - 0,34)
Cash Flow/PN	R31	0,22 (1,37)	0,15 (0,05 - 0,29)	(-8,14 - 19,98)
Ebitda/Ventas	R32	0,01 (1,67)	0,12 (0,05 - 0,22)	(-26,72 - 1,07)
Ventas/ANC	R33	3,66 (12,32)	0,9 (0,36 - 2,02)	(0 - 127,84)
Ventas/AC	R34	2,55 (12,01)	1,54 (0,93 - 2,4)	(0 - 197,63)
Ventas/Pasivo	R35	1,33 (1,53)	0,88 (0,46 - 1,74)	(0 - 15,06)
Ventas/CC	R36	14,02 (217,32)	1,11 (-2,8 - 4,08)	(-334,53 - 3459,21)
CC (AC-PC)	R37	3133,63 (48265,72)	87,5 (-328,5 - 548,25)	(-139073 - 583363)
NOF	R38	2099,3 (26700,13)	72 (-100,75 - 521,5)	(-92103 - 367079)
Tesor	R39	0,71 (1,75)	0,16 (0,04 - 0,61)	(0 - 14,55)
EBITDA	R40	2654,56 (12784,68)	174,5 (27 - 914)	(-12419 - 140602)
SOLV	R41	2,61 (3,92)	1,59 (1,25 - 2,43)	(0,1 - 46,78)
Act. Extrf	R42	13056,91 (96199,71)	25,5 (0 - 1446,5)	(0 - 1468053)
Pasnocial	R43	23304,73 (108130,92)	930,5 (201,5 - 6429)	(4 - 1306749)
Solvextrf	R44	1,12 (5,24)	0,03 (0 - 0,54)	(0 - 63,93)
CBF	R45	1,33 (5,11)	0,99 (0,83 - 1,15)	(-7,98 - 80,98)
PASNC/EBITDA	R46	7,18 (35,11)	3,67 (1,56 - 8,74)	(-161,73 - 364,74)
Gastos Personal/Activo total	R47	0,19 (0,26)	0,12 (0,06 - 0,23)	(0 - 2,88)
Gastos Personal/Importe neto cifra negocios	R48	0,93 (10,04)	0,25 (0,16 - 0,36)	(0,01 - 165)

^a Las 10 ratios seleccionadas para los 2 modelos están en negrita.

Se realizó una regresión logística con las ratios propuestas en la Tabla 3.2 donde se utilizó la técnica de los AG (empleando la clasificación binaria, `type="binary"`) para seleccionar aquellas ratios que minimizasen el error “*Akaike information criterion*” (AIC), por ello, la función *fitness* utilizada fue “-AIC” (en los AG, la función *fitness* siempre se debe maximizar).

En este modelo, fueron seleccionados 20 ratios. Teniendo en cuenta que algunos estudios han demostrado que las técnicas de IA pueden ver afectado su rendimiento por problemas de multicolinealidad (e.g. Eom *et al.*, 2008, Dong y Han, 2004), se analizó el factor de inflación de la varianza (*variance information factors*, VIF) obtenido al realizar la regresión logística sobre las 20 variables seleccionadas. De estas 20 variables, 10 de ellas presentaban valores superiores a 5 (Field *et al.*, 2012; Kim y Kang, 2012) por lo que se procedió a excluirlas del modelo, de esta forma se garantiza un modelo de 10 variables sin problemas de multicolinealidad. Estas 10 variables (R01, R05, R06, R07, R08, R15, R18, R23, R25, R33) forman el modelo 1.

A continuación, se volvió a repetir el proceso de selección de variables mediante los AG, pero en esta ocasión el error a minimizar fue el “*Bayesian information criteria*” (BIC), debido a que este error es más restrictivo que el AIC en cuanto al número de variables seleccionadas (Zhang, 2016). En esta ocasión, el número de variables seleccionadas fue de 4, y todas ellas presentaban valores inferiores a 5 en el estadístico VIF. Estas 4 variables (R05, R06, R07, R08) forman el modelo 2. En la Tabla 3.3 se

Capítulo 3: Predicción del estrés financiero en empresas del sector privado del agua, una aproximación desde el punto de vista de la Inteligencia Artificial

puede observar las variables seleccionadas con los modelos de AG y sus correspondientes valores VIF.

Una vez establecidos ambos modelos, se utilizaron las técnicas de IA comentadas previamente: NN, SVM, SVM-T, DT y RF para establecer cuál de las técnicas y modelos empleados presentaba mejor rendimiento.

Tabla 3.3 Variables explicativas del estrés financiero incluidas en el modelo 1 y del modelo 2

Variable		Selección Minimizar Error AIC (modelo 1)	VIF	Selección Minimizar Error BIC (modelo 2)	VIF
Pasivo/PN *	R01	1	2.86	0	-
RAIT	R03	1	46.77	0	-
RE	R04	1	21.92	0	-
Ingresos explotación/ Total activo **	R05	1	2.04	1	1.13
Reservas/Total activo **	R06	1	1.20	1	1.03
CC/Total activo **	R07	1	1.44	1	1.02
PN/Pasivo **	R08	1	1.30	1	1.15
EBITDA/Gastos financieros *	R15	1	1.39	0	-
Test acido *	R18	1	1.26	0	-
Interes charges/bank loan *	R23	1	1.01	0	-
RF *	R25	1	3.40	0	-
Activos explotación	R26	1	62.38	0	-
Resultado explotación	R27	1	151.04	0	-
Rexplo	R28	1	20.42	0	-
Cash Flow	R29	1	67.23	0	-
Cash Flow/PN	R31	1	5.64	0	-
Ventas/ANC *	R33	1	1.49	0	-
NOF	R38	1	5.35	0	-
EBITDA	R40	1	258.46	0	-
Pasnocial	R43	1	124.79	0	-

* Variables seleccionadas en el modelo 1. ** Variables seleccionadas en el modelo 1 y 2

Con la finalidad de obtener unos resultados robustos, y no debidos al azar, el proceso de creación de ficheros de entrenamiento y validación, sus correspondientes ejecuciones y predicciones se repitió 1.000 veces tanto para el modelo 1 como para el modelo 2). El resultado final se calculó como el valor promedio obtenido en las citadas 1.000 ejecuciones.

El programa estadístico utilizado ha sido R Core, 2017. Los paquetes utilizados han sido: GA (Scrucca, 2013), neuralnet (Fritsch y Guenther, 2010), e1071(Meyer, 2001), randomForest (Liaw *et al.*, 2015), RWeka (Hornik *et al.*, 2009), gplot2(Wickham, 2009).

3.5.4 Medidas de Rendimiento

Para comprobar el rendimiento del proceso se utilizó una matriz de confusión (Kim *et al.*, 2016) como la mostrada en la Tabla 3.4. Los Verdaderos Positivos (VP) indican el número de empresas en situación de estrés correctamente clasificadas. Los Falsos Positivos (FP) indican el número de empresas sin estrés financiero incorrectamente clasificadas. Los Falsos Negativos (FN) indican el número de empresas en estrés financiero incorrectamente clasificadas y los Verdaderos Negativos (VN) indican el número de empresas sin estrés correctamente clasificadas. La Sensibilidad se define como $VP/(VP + FN)$, la especificidad como $VN / (VN + FP)$, la exactitud (*accuracy*) es la proporción de empresas correctamente clasificadas, i.e., $(VP + VN) / (VP + FN + FP + VN)$ (Altman y Bland, 1994).

Tabla 3.4 Matriz de confusión

Actual	Predicted	
	Estrés	No-estrés
Estrés	Verdaderos Positivos (VP)	Falsos Negativos (FN)
No-estrés	Falsos Positivos (FP)	Verdaderos Negativos (VN)

3.6 Resultados

Tal y como se indica en la metodología, se utilizaron 5 técnicas de IA para procesar los datos: a. NN, b. SVM, c. SVM-T, d. J48 y e. RF y dos modelos predictivos con distintos ratios. Se utilizó un porcentaje del 75-25% para crear los ficheros de entrenamiento (training) y validación (test). El objetivo era evaluar el rendimiento de las distintas técnicas predictivas aplicadas en el modelo 1 y en el modelo 2 para comprobar cuál es la mejor técnica y modelo. Se realizaron 1.000 iteraciones y en cada una de ellas se hizo una selección aleatoria de los registros de la base de datos para crear el fichero de entrenamiento (training dataset) y el fichero de validación (test dataset). Los resultados del fichero de validación de cada una de las 1.000 iteraciones, se iban acumulando en una matriz de confusión (véase Tabla 3.4) para cada una de las cinco técnicas aplicadas a cada modelo. Al finalizar el proceso, se calcularon la sensibilidad, especificidad y exactitud con los datos acumulados en cada una de las matrices de confusión. La Tabla 3.5 muestra las medidas de rendimiento obtenidas con el fichero de validación (test).

Hay que destacar, por encima del resto, el buen rendimiento de las técnicas SVM-T y RF. La técnica más exacta es el SVM-T, siendo superior el modelo 2 respecto del modelo 1 (85,82% vs 84,40%). Seguidas de la técnica RF, en este caso, la exactitud es ligeramente superior en el modelo 1 (84,48 % vs 84,37%). Sin embargo, en el resto de técnicas, los porcentajes de exactitud son más elevados en el modelo 2. También hay que destacar la mejora en el rendimiento de la técnica SVM-T respecto de la técnica SVM. Esto demuestra la importancia de la selección de los parámetros C (cost) y γ (gamma) tal y como se ha comentado previamente (véase Apartado 3.5.2.2).

Tabla 3.5 Medidas de Rendimiento de los datos de validación

Medidas Rendimiento	Prevalencia	Sensibilidad	Especificidad	Exactitud
MOD1-NN	40,99	75,50	82,05	79,37
MOD1-SVM	40,99	68,29	84,70	77,97
MOD1-SVM-T	40,99	79,11	88,08	84,40
MOD1-J48	40,99	72,72	82,09	78,25
MOD1-RF	40,99	78,60	88,57	84,48
Mod1-Promedio	40,99	74,84	85,10	80,89
MOD2-NN	40,99	76,69	82,38	80,04
MOD2-SVM	40,99	79,53	82,29	81,16
MOD2-SVM-T	40,99	82,87	87,87	85,82
MOD2-J48	40,99	76,34	82,41	79,92
MOD2-RF	40,99	79,10	88,03	84,37
Mod2-Promedio	40,99	78,91	84,60	82,26

Observando la sensibilidad y especificidad, se puede observar que la especificidad es superior a la sensibilidad en todas las técnicas y modelos, lo que indica que ambos modelos son muy específicos y por lo tanto, van a ser muy útiles para predecir aquellas empresas que presenten riesgo financiero.

Si observamos los promedios de las medidas de rendimiento, se puede observar que la sensibilidad del modelo 2 es superior al modelo 1 (78,91 % vs 85,13%). La especificidad del modelo 1 es ligeramente superior al modelo 2 (85,10% vs 84,60%) y la exactitud del modelo 2 es superior al modelo 1 (82,26% vs 80,89%). Por lo que el modelo 2 es más exacto y sensible y el modelo 1 es ligeramente más específico.

La Tabla 3.6 muestra las comparaciones dos a dos de la exactitud entre todas las técnicas con la corrección de Bonferroni (Armstrong, 2014). Se puede observar que no hay diferencias significativas.

Tabla 3.6 Comparaciones pareadas de la exactitud con la corrección de Bonferroni. Test dataset a (p-valor)

	Mod 1 NN	Mod 1 SVM	Mod 1 SVM-T	Mod 1 J48	Mod 1 RF	Mod2 NN	Mod 2 SVM	Mod 2 SVM-T	Mod 2 J48
Mod 1 - SVM	1.00	-	-	-	-	-	-	-	-
Mod 1 - SVM-T	1.00	1.00	-	-	-	-	-	-	-
Mod 1- J48	1.00	1.00	1.00	-	-	-	-	-	-
Mod 1 - RF	1.00	1.00	1.00	1.00	-	-	-	-	-
Mod 2 - NN	1.00	1.00	1.00	1.00	1.00	-	-	-	-
Mod 2 - SVM	1.00	1.00	1.00	1.00	1.00	1.00	-	-	-
Mod 2 - SVM-T	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	-
Mod 2 - J48	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-
Mod 2 - RF	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

^a Los cálculos se han realizado en términos porcentuales, no con los valores brutos obtenidos en la matriz de confusión del proceso ejecutado.

Además, la Figura 3.1 muestra que el modelo 2, (con 4 variables), presenta mayor exactitud que el modelo 1 (con 10 variables), exceptuando en la técnica RF que el comportamiento es muy similar en ambos modelos. Esto confirma el hecho de que el error “AIC” tiende a sobreestimar el número de variables a seleccionar (Zhang, 2016).

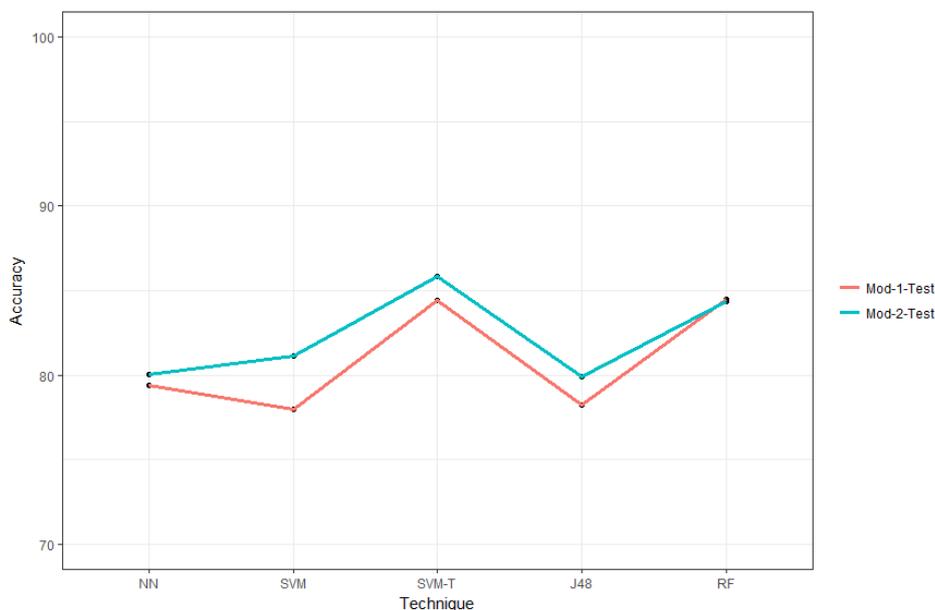


Figura 3.1. Comparación de la Exactitud entre modelos y técnicas

3.7 Discusión

Teniendo en cuenta que cada vez hay mayor porcentaje de empresas privadas gestionando un recurso imprescindible como es el agua y dada la falta de modelos predictivos financieros para este sector, se han propuesto la creación de dos modelos que sirvan para dar estabilidad financiera.

La principal contribución de este estudio es la elaboración de dos modelos predictivos que presentan un gran rendimiento para predecir el estrés financiero con dos

años de antelación en estas empresas. También se puede observar que el modelo 2 es más sensible que el modelo 1 y que el modelo 1 es ligeramente más específico que el modelo 2, siendo la exactitud superior en el modelo 2. Ello implica que un modelo predictivo con mayor número de variables, no garantiza que se realicen mejores predicciones.

La otra gran aportación está relacionada con las técnicas empleadas. Así, sin lugar a dudas, la técnica que mejores predicciones realiza es la SVM-T, seguida de RF, siendo ambas técnicas más específicas que sensibles. Como consecuencia de los buenos resultados obtenidos con estas dos técnicas, se puede predecir con suficiente antelación las empresas que se encuentran en un riesgo elevado de entrar en bancarrota. Ello supone que tanto los *stakeholders* como los gestores de la propia empresa, cuentan con tiempo suficiente para tomar las medidas oportunas. Con esto se puede reducir la probabilidad de problemas posteriores de insolvencias con el consecuente abandono de la prestación del servicio, con las consecuencias tan drásticas que conllevarían esta situación a las personas y empresas que reciben el servicio.

Por otra parte, aquellas empresas que no presenten riesgo financiero, seguirán teniendo la confianza de los acreedores y podrán obtener con menor dificultad la financiación para realizar las inversiones necesarias en infraestructuras, eficiencia, innovación y tener la confianza de quien los contrate.

Por otra parte, el principal inconveniente de técnicas como las NN, o RF es que son cajas negras y es muy difícil explicar cómo realizan sus predicciones. En cuanto a futuras líneas de investigación, cabe destacar que podrían utilizarse técnicas de “*classifier ensembles*” como *bagging* o *boosting* junto a técnicas de inteligencia artificial como las NN o SVM, con la intención de mejorar las predicciones y reducir los errores o el uso de otras técnicas de IA como *fuzzy logic* o mapas auto-organizados.

3.8 Referencias

- AEAS. (2014). XIII Encuesta de suministro de agua potable y saneamiento en España.
- Alam, P., Booth, D., Lee, K., & Thordarson, T. (2000). The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study. *Expert Systems with Applications*, 18(3), 185-199. [https://doi.org/10.1016/S0957-4174\(99\)00061-5](https://doi.org/10.1016/S0957-4174(99)00061-5)
- Almamy, J., Aston, J., & Ngwa, L. N. (2016). An evaluation of Altman's Z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence

- from the UK. *Journal of Corporate Finance*, 36, 278-285.
<https://doi.org/10.1016/j.jcorpfin.2015.12.009>
- Altman. (2000). Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA ® Models Background. *Stern School of Business, New York University*, 1-54.
- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 1 : sensitivity and specificity. *Bmj*, 308(6943), 1552. <https://doi.org/10.1258/phleb.2012.012J05>
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.
<https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Altman, E. I. (1983). *Corporate financial distress : a complete guide to predicting, avoiding, and dealing with bankruptcy* (3rd Editio). Wiley.
- Altman, E. I., Sabato, G., & Wilson, N. (2010). The Value of non-financial information in small-sized enterprise risk management. *The Journal of Credit Risk*, 6(2), 1-33.
- Angelini, E., di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance*, 48(4), 733-755.
<https://doi.org/10.1016/j.qref.2007.04.001>
- Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, 34(5), 502-508. <https://doi.org/10.1111/opo.12131>
- Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks*, 12(4), 929-935.
<https://doi.org/10.1109/72.935101>
- Aziz, A., Emanuel, D., & Lawson, G. (1988). *Journal of management studies. Journal of Management Studies* (Vol. 25). Blackwell Publishers.
- Barandela, R., Sánchez, J. S., García, V., & Rangel, E. (2003). Strategies for learning in class imbalance problems. *Pattern Recognition*, 36(3), 849-851.
[https://doi.org/10.1016/S0031-3203\(02\)00257-1](https://doi.org/10.1016/S0031-3203(02)00257-1)
- Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 40, 432-442. <https://doi.org/10.1016/j.jbankfin.2013.12.013>
- Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Source Journal of Accounting Research Empirical Research in Accounting: Selected Studies*, 4, 71-111.
- Becchetti, L., & Sierra, J. (2003). Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of Banking & Finance*, 27(11), 2099-2120.
[https://doi.org/10.1016/S0378-4266\(02\)00319-9](https://doi.org/10.1016/S0378-4266(02)00319-9)
- Beynon, M. J., & Peel, M. J. (2001). Variable precision rough set theory and data discretisation: an application to corporate failure prediction. *Omega*, 29(6), 561-576.
[https://doi.org/10.1016/S0305-0483\(01\)00045-7](https://doi.org/10.1016/S0305-0483(01)00045-7)
- Blanco-Oliver, A., Irimia-Dieguez, A., Oliver-Alfonso, M., & Wilson, N. (2015). Improving Bankruptcy Prediction in Micro-Entities by Using Nonlinear Effects and Non-Financial Variables, (2), 144-167.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
<https://doi.org/10.1023/A:1010933404324>
- Bureau Van Dijk. (2017). SABI. Sistemas de Análisis de Balances Ibéricos.
- Charitou, A., Neophytou, E., & Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, 13(3), 465-497.
<https://doi.org/10.1080/0963818042000216811>
- Chen, M.-Y. (2011). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers*

- and Mathematics with Applications*, 62, 4514-4524.
<https://doi.org/10.1016/j.camwa.2011.10.030>
- Chen, M.-Y. (2013). A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. *Information Sciences*, 220, 180-195. <https://doi.org/10.1016/j.ins.2011.09.013>
- Chen, N., Ribeiro, B., Vieira, A., & Chen, A. (2013). Clustering and visualization of bankruptcy trajectory using self-organizing map. *Expert Systems With Applications*, 40, 385-393. <https://doi.org/10.1016/j.eswa.2012.07.047>
- Chi, L.-C., & Tang, T.-C. (2006). Bankruptcy Prediction: Application of Logit Analysis in Export Credit Risks. *Australian Journal of Management*, 31(1), 17-27. <https://doi.org/10.1177/031289620603100102>
- Cho, S., Kim, J., & Bae, J. K. (2009). An integrative model with subject weight based on neural network learning for bankruptcy prediction. *Expert Systems with Applications*, 36(1), 403-410. <https://doi.org/10.1016/j.eswa.2007.09.060>
- Chung, C.-C., Chen, T.-S., Lin, L.-H., Lin, Y.-C., & Lin, C.-M. (2016). Bankruptcy Prediction Using Cerebellar Model Neural Networks. *International Journal of Fuzzy Systems*, 18(2), 160-167. <https://doi.org/10.1007/s40815-015-0121-5>
- Chung, H.-M. M., & Tam, K. Y. (1993). A Comparative Analysis of Inductive-Learning Algorithms. *Intelligent Systems in Accounting, Finance and Management*, 2(1), 3-18. <https://doi.org/10.1002/j.1099-1174.1993.tb00031.x>
- Ciampi, F., & Gordini, N. N. (2013). Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises. *Journal of Small Business Management*, 51(1), 23-45. <https://doi.org/10.1111/j.1540-627X.2012.00376.x>
- Ciampi, F., Vallini, C., Gordini, N., & Benvenuti, M. (2009). Are Credit Scoring Models Able to Predict Small Enterprise Default? Statistical Evidence from Italian Small Enterprises.
- Cielen, A., Peeters, L., & Vanhoof, K. (2004). Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research*, 154(2), 526-532. [https://doi.org/10.1016/S0377-2217\(03\)00186-3](https://doi.org/10.1016/S0377-2217(03)00186-3)
- Cultrera, L., & Brédart, X. (2016). Bankruptcy prediction: the case of Belgian SMEs. *Review of Accounting and Finance*, 15(1), 101-119. <https://doi.org/10.1108/RAF-06-2014-0059>
- Dambolena, I. G., & Khoury, S. J. (1980). American Finance Association Ratio Stability and Corporate Failure. *Source: The Journal of Finance*, 35(4), 1017-1026.
- Deakin, E. B. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 10(1), 167. <https://doi.org/10.2307/2490225>
- Deloitte. (2014). El sector sanitario en España, situación actual y perspectivas de futuro.
- Díaz-Uriarte, R., & Alvarez De Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7:3. <https://doi.org/10.1186/1471-2105-7-3>
- Dietrich, J. R., & Kaplan, R. S. (1982). Empirical Analysis of the Commercial Loan Classification Decision. *Source: The Accounting Review*, 57(1), 18-38.
- Divsalar, M., & Roodsaz, H. (2012). A Robust Data-Mining Approach to Bankruptcy Prediction. *Journal of Forecasting*, 523(March 2011), 504-523.
- Dong, Y.-S. D. Y.-S., & Han, K.-S. H. K.-S. (2004). A comparison of several ensemble methods for text categorization. *IEEE International Conference on Services Computing, 2004. (SCC 2004). Proceedings. 2004*, 0-3. <https://doi.org/10.1109/SCC.2004.1358033>
- du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction.

- Expert Systems with Applications*, 75, 25-43.
<https://doi.org/10.1016/j.eswa.2017.01.016>
- Du Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, 242(1), 286-303.
<https://doi.org/10.1016/j.ejor.2014.09.059>
- Dugan, M. T., & Zavgren, C. (1989). How a Bankruptcy Model Could Be Incorporated as an Analytical Procedure. *The CPA journal*, 94-65.
- Eom, J.-H., Kim, S.-C., & Zhang, B.-T. (2008). AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction. *Expert Systems with Applications*, 34(4), 2465-2479.
<https://doi.org/10.1016/j.eswa.2007.04.015>
- Etheridge, H. L., Sriram, R. S., & Hsu, H. Y. K. (2000). A Comparison of Selected Artificial Neural Networks that Help Auditors Evaluate Client Financial Viability. *Decision Sciences*, 31(2), 531-550. <https://doi.org/10.1111/j.1540-5915.2000.tb01633.x>
- Fan, A., & Palaniswami, M. (2000). Selecting bankruptcy predictors using a support vector machine approach. En *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium* (pp. 354-359 vol.6). IEEE.
<https://doi.org/10.1109/IJCNN.2000.859421>
- Fedorova, E., Gilenko, E., & Dovzhenko, S. (2013). Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications*, 40(18), 7285-7293. <https://doi.org/10.1016/j.eswa.2013.07.032>
- Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.
- FitzPatrick, P. (1932). *A comparison of the ratios of successful industrial enterprises with those of failed companies*. [Washington].
- Fritsch, S., & Guenther, F. (2010). neuralnet: Training of neural networks. *The R Journal*, 2(1), 30-38. <https://doi.org/10.1109/SP.2010.25>
- González-Gómez, F., García-Rubio, M. A., & Gonz Alez-Martínez, J. (2014). Beyond the public-private controversy in urban water management in Spain. *Utilities Policy*, 31, 1-9. <https://doi.org/10.1016/j.jup.2014.07.004>
- Gordini, N. A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy, 41 *Expert Systems with Applications* § (2014). Elsevier Ltd.
<https://doi.org/10.1016/j.eswa.2014.04.026>
- Gu, Z., & Gao, L. (1999). A multivariate model for predicting business failures of hospitality firms. *Tourism and Hospitality Research*, 2(1).
- Guerrini, A., Romano, G., Ferretti, S., Fibbi, D., & Daddi, D. (2016). A Performance Measurement Tool Leading Wastewater Treatment Plants toward Economic Efficiency and Sustainability. *Sustainability*, 8(12), 1250.
<https://doi.org/10.3390/su8121250>
- Hachicha, N., Jarboui, B., & Siarry, P. (2011). A fuzzy logic control using a differential evolution algorithm aimed at modelling the financial market dynamics. *Information Sciences*, 181(1), 79-91. <https://doi.org/10.1016/j.ins.2010.09.010>
- Hadipuro, W. (2010). Indonesia ' s Water Supply Regulatory Framework : Between Commercialisation and Public Service ? *Water Alternatives*, 3(3), 475-491.
- Hansen, L. K., & Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10), 993-1001.
<https://doi.org/10.1109/34.58871>
- Hennawy, R. H. A., & Morris, R. C. (1983). The significance of base year in developing failure prediction models. *Journal of Business Finance & Accounting*, 10(2), 209-

223. <https://doi.org/10.1111/j.1468-5957.1983.tb00424.x>
- Holland, J. H. (1975). *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press.
- Hornik, K., Buchta, C., & Zeileis Wirtschaftsuniversität Wien, A. W. (2009). Open-Source Machine Learning: R Meets Weka. <https://doi.org/10.1007/s00180-008-0119-7>
- Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2016). A Practical Guide to Support Vector Classification.
- Hua, Z., Wang, Y., Xu, X., Zhang, B., & Liang, L. (2007). Predicting corporate financial distress based on integration of support vector machine and logistic regression. *Expert Systems with Applications*, 33(2), 434-440. <https://doi.org/10.1016/j.eswa.2006.05.006>
- Jeong, C., Min, J. H., & Kim, M. S. (2012). A tuning method for the architecture of neural network models incorporating GAM and GA as applied to bankruptcy prediction. *Expert Systems with Applications*, 39(3), 3650-3658. <https://doi.org/10.1016/j.eswa.2011.09.056>
- Jones, F. L. (1987). Current Techniques in Bankruptcy Prediction. *Journal of Accounting Literature*, 6, 131-164.
- Joo, H. N., & Jin, T. (2000). Bankruptcy Prediction: Evidence from Korea Listed Companies During the IMF Crisis. *Journal of International Financial Management and Accounting*, 11(3), 195-217.
- Kaski, S., Sinkkonen, J., & Peltonen, J. (2001). Bankruptcy analysis with self-organizing maps in learning metrics. *IEEE Transactions on Neural Networks*, 12(4), 936-947. <https://doi.org/10.1109/72.935102>
- Khan, A. M. (1985). Analyzing financial statements for managerial performance measurement and bankruptcy prediction. *Engineering Management International*, 3(3), 165-174. [https://doi.org/10.1016/0167-5419\(85\)90003-1](https://doi.org/10.1016/0167-5419(85)90003-1)
- Kim, H.-J. J., Jo, N.-O. O., & Shin, K.-S. S. (2016). Optimization of cluster-based evolutionary undersampling for the artificial neural networks in corporate bankruptcy prediction. *Expert Systems with Applications*, 59, 226-234. <https://doi.org/10.1016/j.eswa.2016.04.027>
- Kim, H., & Gu, Z. (2006). Predicting Restaurant Bankruptcy: A Logit Model in Comparison with a Discriminant Model. *Journal of Hospitality & Tourism Research*, 30(4), 474-493. <https://doi.org/10.1177/1096348006290114>
- Kim, M.-J. J., & Kang, D.-K. K. (2012). Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction. *Expert Systems with Applications*, 39(10), 9308-9314. <https://doi.org/10.1016/j.eswa.2012.02.072>
- Kittler, J., Hatef, M., Duin, R. P. W., & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 226-239. <https://doi.org/10.1109/34.667881>
- Ko, P.-C., & Lin, P.-C. (2006). An evolution-based approach with modularized evaluations to forecast financial distress. *Knowledge-Based Systems*, 19(1), 84-91. <https://doi.org/10.1016/j.knosys.2005.11.006>
- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence*, 2, 1137-1143. <https://doi.org/10.1067/mod.2000.109031>
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59-69. <https://doi.org/10.1007/BF00337288>
- Laitinen, E. K., & Laitinen, T. (2000). Bankruptcy prediction Application of the Taylor's

- expansion in logistic regression. *International Review of Financial Analysis*, 9, 327-349.
- Lee, C.-H., & Yu, Y.-H. (2012). Characteristics of public–private partnerships for municipal wastewater treatment in Taiwan. *Journal of the Chinese Institute of Engineers*, 35(2), 245-258. <https://doi.org/10.1080/02533839.2012.638535>
- Lennox, C. (1999). Identifying failing companies: a re-evaluation of the logit, probit and DA approaches. *Journal of Economics and Business*, 51(4), 347-364. [https://doi.org/10.1016/S0148-6195\(99\)00009-0](https://doi.org/10.1016/S0148-6195(99)00009-0)
- Li, M. Y. L., & Miu, P. (2010). A hybrid bankruptcy prediction model with dynamic loadings on accounting-ratio-based and market-based information: A binary quantile regression approach. *Journal of Empirical Finance*, 17(4), 818-833. <https://doi.org/10.1016/j.jempfin.2010.04.004>
- Liaw, A., Wiener, M., & Andy Liaw, M. (2015). Breiman and Cutler’s Random Forests for Classification and Regression Description Classification and regression based on a forest of trees using random inputs.
- Lin, W.-Y., Hu, Y.-H., & Tsai, C.-F. (2012). Machine Learning in Financial Crisis Prediction: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4), 421-436. <https://doi.org/10.1109/TSMCC.2011.2170420>
- López Iturriaga, F. J., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems with Applications*, 42(6), 2857-2868. <https://doi.org/10.1016/j.eswa.2014.11.025>
- Martin, D. (1977). Early warning of bank failure. *Journal of Banking & Finance*, 1(3), 249-276. [https://doi.org/10.1016/0378-4266\(77\)90022-X](https://doi.org/10.1016/0378-4266(77)90022-X)
- Mekonnen, M. M., & Hoekstra, A. Y. (2016). Four billion people facing severe water scarcity. *Science Advances*, 2(2), e1500323-e1500323. <https://doi.org/10.1126/sciadv.1500323>
- Messier, W. F., & Hansen, J. V. (1988). Inducing Rules for Expert System Development: An Example Using Default and Bankruptcy Data. *Management Science*, 34(12), 1403-1415. <https://doi.org/10.1287/mnsc.34.12.1403>
- Meyer, D. (2001). Support Vector Machines * The Interface to libsvm in package e1071, 13(9).
- Michael, S., Georgios, D., Nikolaos, M., & Constantin, Z. (1999). A Fuzzy Knowledge-based Decision Aiding Method for the Assessment of Financial Risk: The Case of Corporate Bankruptcy Prediction. *European Symposium on intelligent techniques*, (February).
- Min, J. H., & Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4), 603-614. <https://doi.org/10.1016/j.eswa.2004.12.008>
- Mousavi, M. M., Ouenniche, J., & Xu, B. (2015). Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. *International Review of Financial Analysis*, 42, 64-75. <https://doi.org/10.1016/j.irfa.2015.01.006>
- Nakazawa, M. (2015). fmsb: Functions for medical statistics book with some demographic data. *R package version 0.5.2*, 1-40.
- Nilsson, N. J. (1965). *Learning machines : foundations of trainable pattern-classifying systems*. McGraw-Hill.
- Nouri, B. A., & Soltani, M. (2016). Designing a Bankruptcy Prediction Model Based on Account, Market and Macroeconomic Variables, 9(1), 125-147.
- Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction.

- International Joint Conference on Neural Networks*, 2, 163-168.
<https://doi.org/10.1109/IJCNN.1990.137710>
- OECD/EU. (2016). Health at a Glance: Europe 2016 State of Health in the EU Cycle.
<https://doi.org/10.1787/9789264265592-en>
- Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy.
Journal of Accounting Research, 18(1), 109-131. <https://doi.org/10.2307/2490395>
- ONU. (2015). *World Population Prospects*. New York.
- ONU. (2017). Informe Mundial sobre el Desarrollo de los Recursos Hídricos de las Naciones Unidas 2017: Las aguas residuales: el recurso desaprovechado, resumen ejecutivo; 2017.
- Oza, N. C., & Tumer, K. (2007). Classifier ensembles: Select real-world applications.
<https://doi.org/10.1016/j.inffus.2007.07.002>
- Package «hybridEnsemble». (2015).
- Paliwal, M., & Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems With Applications*, 36, 2-17.
<https://doi.org/10.1016/j.eswa.2007.10.005>
- Pompe, P. P. M., & Bilderbeek, J. (2005). The prediction of bankruptcy of small- and medium-sized industrial firms. *Journal of Business Venturing*, 20(6), 847-868.
<https://doi.org/10.1016/j.jbusvent.2004.07.003>
- PWC. (2014). *La gestión del agua en España, análisis de la situación actual del sector y retos futuros*.
- R Core, T. (2017). R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ranawana, R., & Palade, V. (2006). Multi-Classifer Systems: Review and a roadmap for developers. *International Journal of Hybrid Intelligent Systems*, 3(1), 35-61.
<https://doi.org/10.3233/HIS-2006-3104>
- 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. <https://doi.org/10.1016/j.ejor.2006.08.043>
- Ravi Kumar, P., Ravi, V., Kumar, P. R., & 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.
<https://doi.org/10.1016/j.ejor.2006.08.043>
- Rijsberman, F. R. (2006). Water scarcity: Fact or fiction? *Agricultural Water Management*, 80(1), 5-22. <https://doi.org/10.1016/j.agwat.2005.07.001>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., ... Müller Markus. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12(4116), 1-77.
<https://doi.org/10.1126/science.182.4116.990>
- Rokach, L. (2009). Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography. *Computational Statistics & Data Analysis*, 53(12), 4046-4072. <https://doi.org/10.1016/j.csda.2009.07.017>
- Sato, T., Qadir, M., Yamamoto, S., Endo, T., & Zahoor, A. (2013). Global, regional, and country level need for data on wastewater generation, treatment, and use.
- Scrucca, L. (2013). GA : A Package for Genetic Algorithms in R. *Journal of Statistical Software*, 53(4), 1-37. <https://doi.org/10.18637/jss.v053.i04>
- Serrano-Cinca, C. (1996). Self organizing neural networks for financial diagnosis. *Decision Support Systems*, 17, 227-238.
- Shin, K.-S., Lee, T. S., & Kim, H.-J. (2005). An application of support vector machines

- in bankruptcy prediction model. *Expert Systems with Applications*, 28(1), 127-135. <https://doi.org/10.1016/j.eswa.2004.08.009>
- Shin, K.-S., & Lee, Y.-J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications*, 23(3), 321-328. [https://doi.org/10.1016/S0957-4174\(02\)00051-9](https://doi.org/10.1016/S0957-4174(02)00051-9)
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*, 74(1), 101-124. <https://doi.org/10.1086/209665>
- Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, 10(5), 335-347. [https://doi.org/10.1016/0167-8655\(89\)90037-8](https://doi.org/10.1016/0167-8655(89)90037-8)
- Tay, F. E. H., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29, 309-317.
- Tin Kam Ho, Hull, J. J., & Srihari, S. N. (1994). Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1), 66-75. <https://doi.org/10.1109/34.273716>
- Tsai. (2008). Financial decision support using neural networks and support vector machines. *Expert Systems*, 25(4), 380-393. <https://doi.org/10.1111/j.1468-0394.2008.00449.x>
- Tsai, Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. *Applied Soft Computing Journal*, 24, 977-984. <https://doi.org/10.1016/j.asoc.2014.08.047>
- Tsai, & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34(4), 2639-2649. <https://doi.org/10.1016/j.eswa.2007.05.019>
- Vapnik, V. N. (1998). *Statistical learning theory*. Wiley.
- 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. <https://doi.org/10.1016/j.eswa.2013.09.033>
- Wehrens, R., & Buydens, L. M. C. (2007). Self- and super-organizing maps in R: The kohonen package. *Journal of Statistical Software*, 21(5), 1-19. <https://doi.org/10.18637/jss.v021.i05>
- West, D., Dellana, S., & Qian, J. (2005). Neural network ensemble strategies for financial decision applications. *Computers & Operations Research*, 32(10), 2543-2559. <https://doi.org/10.1016/j.cor.2004.03.017>
- Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, 18(1), 30-55. <https://doi.org/10.1007/s10618-008-0116-z>
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Comprehensive R Archive Network (CRAN).
- Wilcox, J. W. (1973). A Prediction of Business Failure Using Accounting Data. *Journal of Accounting Research*, 11, 163. <https://doi.org/10.2307/2490035>
- Wilson, N., & Summers, B. (2002). Trade credit terms offered by small firms: Survey evidence and empirical analysis. *Journal of Business Finance and Accounting*, 29(3-4), 317-351. <https://doi.org/10.1111/1468-5957.00434>
- World Economic Forum. (2015). *Global Risks 2015. 10th Edition*. Geneva.
- Xie, Y., Li, X., Ngai, E. W. T., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36(3), 5445-5449. <https://doi.org/10.1016/j.eswa.2008.06.121>
- Xu, M., & Zhang, C. (2009). Bankruptcy prediction: The case of Japanese listed companies. *Review of Accounting Studies*, 14(4), 534-558.

- <https://doi.org/10.1007/s11142-008-9080-5>
- Yazdanfar, D., & Nilsson, M. (2008). The bankruptcy determinants of Swedish SMEs. *Institute for Small Business & Entrepreneurship*, (November), 1-14.
- Zadeh, L. a. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.
[https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- Zhang, Z. (2016). Variable selection with stepwise and best subset approaches. *Annals of translational medicine*, 4(7), 136. <https://doi.org/10.21037/atm.2016.03.35>
- Zhong, L., Mol, A. P. J., & Fu, T. (2008). Public-private partnerships in China's urban water sector. *Environmental Management*, 41(6), 863-877.
<https://doi.org/10.1007/s00267-008-9070-1>
- Zhou, L., Lai, K. K., & Yen, J. (2012). Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation. *International Journal of Systems Science*, 45(July 2015), 1-13.
<https://doi.org/10.1080/00207721.2012.720293>
- Zibanezhad, E., Foroghi, D., & Monadjemi, A. (2011). Applying decision tree to predict bankruptcy. *Proceedings - 2011 IEEE International Conference on Computer Science and Automation Engineering, CSAE 2011*, 4, 165-169.
<https://doi.org/10.1109/CSAE.2011.5952826>
- Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22(1), 59-82.
<https://doi.org/10.2307/2490859>

Conclusiones

4.1 Conclusiones

El presente estudio analiza la importancia de disponer de herramientas predictivas financieras válidas en sectores de vital importancia, desde un punto de vista tanto económico, como social y ambiental, como son el sector de recogida y tratamiento de aguas residuales, el sector sanitario privado y el sector del suministro del agua.

Estas herramientas predictivas, construidas de forma ad-hoc para cada uno de los sectores, pueden resultar de utilidad para conocerlos mejor y, de esta forma, proporcionar estabilidad financiera a la propia empresa, a otras organizaciones del sector, clientes, proveedores, bancos, administración pública, etc., ya que les permitirían realizar predicciones de la bancarrota con un alto grado de validez con dos años de antelación.

Para plantear unos modelos predictivos validos, se ha utilizado una metodología que se basa en las principales técnicas clasificación asociadas a la IA, como son NN, FL, DT, SVM, RF o SOM. La elección de este tipo de técnicas en lugar de las tradicionales como logit o probit se debe a que , según Tsai (2008) presentan un mejor rendimiento para predecir.

Además, para aportar mayor validez, se han creado diversos modelos predictivos para poder comparar qué modelo y técnica son los que presentan mejor rendimiento. Dichos modelos han sido creados mediante la clasificación binaria de selección de variables de los AG, que también es una técnica de IA y mediante el cálculo del área bajo la curva ROC (AUCROC).

La simbiosis entre las técnicas de IA aplicadas y los diversos modelos obtenidos con los AG y AUCROC, han dado como resultado unos excelentes modelos predictivos, que en algún caso han llegado a presentar una exactitud superior al 95%.

En consecuencia, con los modelos propuestos sería posible establecer con un nivel alto de fiabilidad, el estado financiero de las empresas, lo que indudablemente, sería de gran interés, como ya se ha mencionado, para los clientes, entidades financieras, administraciones públicas, etc.

En el modelo predictivo asociado a las empresas de recogida y tratamiento de aguas residuales, se utilizaban cuatro técnicas de IA, a. SVM, b. CE, c. FL y d. SOM,

aplicados a tres modelos construidos mediante diferentes procedimientos, el primero de ellos con cuatro variables, el segundo con tres y el tercero con seis. Se observa que la técnica que realiza predicciones más válidas es SOM, siendo el primer modelo más exacto y específico que el segundo, mientras que el segundo modelo es más sensible y presenta una mayor media geométrica. El tercer modelo es el que peor rendimiento presenta.

En relación a las empresas privadas del sector sanitario español, y utilizando los cinco ratios financieros propuestos por Altman (2000) para empresas privadas, se utilizaron los AG con la finalidad de optimizar la predicción, maximizando la sensibilidad y especificidad y de esta forma, reducir el error Tipo I y Tipo II. Los resultados obtenidos con los nuevos coeficientes mejoran a las estimaciones realizadas con la fórmula Z'-Score de Altman. Por tanto, se puede concluir, que se aporta un modelo predictivo creado ad-hoc para las empresas privadas que presenta un mejor rendimiento que el modelo propuesto por Altman para empresas privadas.

En el modelo predictivo asociado a las empresas de suministro de agua, se proponían 5 técnicas de IA, a. NN, b. SVM, c. SVM optimizado, d. DT, e. RF. Por otra parte, y al igual que para el tratamiento de aguas residuales, se implementaron dos modelos predictivos, uno de ellos con cuatro variables y otro con diez. La técnica que mejor rendimiento presentó fue SVM optimizado (superior al 85%), mientras que la técnica RF, presentó un rendimiento ligeramente inferior. En este caso, se puede observar que un modelo compuesto de 4 ratios, realiza mejores predicciones que el que tenía 10.

4.2 Limitaciones y futuras líneas de investigación

En relación a las limitaciones de este estudio, cabe señalar que éstas radican principalmente en que el uso de los AG no garantiza que se encuentre la solución óptima del problema. Estas soluciones dependerán del rango de búsqueda aplicado para el cálculo de los coeficientes, el tamaño de la población, la ratio de cruce y la ratio de mutación aplicados en el modelo. Otra limitación, puede estar relacionada con el reducido tamaño muestral del número de empresas del sector sanitario cuyo estado era “concurso de acreedores” y que tenían sus datos financieros informados en la base de datos SABI.

En cuanto a futuras líneas de investigación, cabe destacar que podrían introducirse variables no-financieras para comprobar si éstas pueden aumentar el poder predictivo. También podrían utilizarse otras técnicas asociadas a la IA, como son las *hybrid ensemble* o técnicas de *classifier ensemble* como el *bagging* o *boosting*, deep learning o redes neuronales convolucionales. Por otra parte, se podría comparar el rendimiento de las técnicas utilizadas frente a las técnicas estadísticas convencionales con el objetivo de verificar si se cumple lo que aseveran autores como Tsai (2008) o Tsai & Wu (2008), los cuales consideran más exactas las técnicas basadas en IA. Además, también se podrían aplicar las técnicas y modelos empleados en este estudio a otros sectores empresariales por comprobar si los resultados son similares, o si hay otras variables con mayor poder predictivo que las utilizadas en este estudio. Otro campo de investigación, sería utilizar las técnicas de IA para realizar estudios en relación a la detección de fraude contable.

En el caso particular de las empresas del sector sanitario privado, se podría comprobar la validez del modelo propuesto a lo largo del tiempo y plantear los ajustes necesarios que deberían realizarse, en caso de que su rendimiento vaya disminuyendo a medida que éste avanza.

4.3 Referencias

- Altman. (2000). Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA ® Models Background. *Stern School of Business, New York University*, 1-54.
- Tsai. (2008). Financial decision support using neural networks and support vector machines. *Expert Systems*, 25(4), 380-393. <https://doi.org/10.1111/j.1468-0394.2008.00449.x>
- Tsai, & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34(4), 2639-2649. <https://doi.org/10.1016/j.eswa.2007.05.019>

