

APLICACIÓN DE LA PROGRAMACIÓN GENÉTICA EN LA PREDICCIÓN DEL FRACASO EMPRESARIAL EN PEQUEÑAS Y MEDIANAS EMPRESAS

Autores:

Ángel Beade¹
Manuel Rodríguez¹
José Santos²

¹ Departamento de Empresa, Universidade da Coruña – Cátedra AECA-Abanca, IESIDE, Campus de Elviña, s/n 15071 A Coruña, España

² Departamento de Ciencias de la Computación y Tecnologías de la Información – Centro de Investigación en Tecnologías de la Información (CITIC), Universidade da Coruña, Campus de Elviña, s/n 15071 A Coruña, España

Área temática:

Nuevas tecnologías y contabilidad.

Research Workshop 2:

Digitalización e inteligencia artificial: implicaciones en la gestión.

Palabras clave:

Fracaso empresarial, modelos de predicción, inteligencia artificial, programación genética, PYMES gallegas.

APLICACIÓN DE LA PROGRAMACIÓN GENÉTICA EN LA PREDICCIÓN DEL FRACASO EMPRESARIAL EN PEQUEÑAS Y MEDIANAS EMPRESAS

RESUMEN

Con este trabajo se pretende evaluar la idoneidad de la Programación Genética como método para la creación de modelos de predicción del fracaso empresarial en el horizonte de 1 a 4 años. El trabajo se centra en las pequeñas y medianas empresas ubicadas en la Comunidad Autónoma de Galicia.

Los modelos obtenidos para los años 1, 2, 3 y 4 previos al fracaso se muestran eficientes en los distintos contrastes a los que se han sometido, alcanzando porcentajes en línea con los obtenidos en otros estudios con otras técnicas.

1. INTRODUCCIÓN

La empresa es una unidad básica de la economía de un país y las consecuencias negativas que provoca la desaparición de una empresa se dejan sentir en la sociedad en general, en factores como las tasas de crecimiento del PIB, la fuerza laboral, la inversión y la distribución de la renta, entre otras.

Por su parte, las pequeñas y medianas empresas (PYME) son la tipología de negocio dominante en las economías occidentales. De su importancia baste resaltar el hecho de que las empresas con menos de 250 asalariados suponían en España, en diciembre de 2019, un empleo generado de 10,5 millones de personas.

Consecuencia de lo antedicho, el tema del fracaso empresarial y la posibilidad de su predicción ha ido tomando gran relevancia. De forma general, los modelos actuales presentan buenas estimaciones del fracaso únicamente para horizontes temporales no superiores a dos años (Altman y col., 2015, p. 1). Por ello, la precisión y el horizonte temporal son los dos desafíos de la predicción del fallo. Es deseable, no solo la precisión de la predicción a corto plazo, sino que dicha predicción tenga el mayor horizonte posible con una precisión suficiente para que dicha predicción sea realmente útil.

El problema se ha abordado con diferentes metodologías, tanto paramétricas como no paramétricas, que en general comparten una elevada capacidad para identificar tensiones financieras subyacentes y cuantificar el riesgo de insolvencia, sin que destaque una de ellas como la más adecuada.

En esta línea, el objetivo del presente trabajo es modelar la predicción de insolvencia en cuatro situaciones temporales diferentes: a 1, 2, 3 y 4 años previos al fracaso, pero aportando una visión diferente en aras a una mayor evidencia empírica utilizando la técnica de Inteligencia Artificial (IA) de la Programación Genética (PG), más incipiente como técnica y novedosa en cuanto a su aplicación en el contexto del pronóstico del fallo empresarial.

PG es una de las técnicas englobadas en el campo o rama de Computación Evolutiva (CE), técnicas que, simulando la evolución natural, permiten obtener de modo automático soluciones a un determinado problema, típicamente de optimización o búsqueda. En el caso de PG, las soluciones corresponden a posibles programas que generalmente se representan en forma de árbol, programas que utilizan una serie de primitivas u operadores pre-especificados. Esos programas/árboles serán los que determinen la predicción a categorización del fracaso de una empresa, partiendo de la

información de diferentes ratios financieras. La ventaja de PG, frente a otras técnicas de IA, es la interpretabilidad del árbol o programa optimizado.

Se trata, por tanto, de proveer de una herramienta útil para la toma de decisiones de los distintos usuarios de la información contable que además pueda proporcionar esa interpretabilidad. Esta es la principal razón de su aplicación en el presente trabajo.

2. ESTADO DEL ARTE

En los últimos años (desde la década de los noventa), las técnicas de IA mediante redes neuronales y árboles de decisión, conocidas también como técnicas no paramétricas, se han ido aplicando al estudio del fracaso empresarial. Serrano y Martín (Serrano Cinca y Martín del Brío, 1993) fueron unos de los pioneros en su aplicación y estudiaron el fracaso en bancos que había sido estudiado por Laffarga y col. (Laffarga Briones y col., 1987), bajo el modelado de redes neuronales basadas en el perceptrón multicapa y los mapas auto-organizados. Las redes neuronales se mostraron como un excelente predictor a corto plazo, con índices de aciertos cercanos al 94% en el caso del perceptrón multicapa y resultados más moderados para los mapas auto-organizados (aprendizaje no supervisado, sin requerir ejemplos categorizados a priori). Sin embargo, no está demostrado el predominio de estas técnicas, pero son un interesante campo de desarrollo por su mayor flexibilidad (Mateos-Ronco y col., 2011).

De igual manera, se encuentran los árboles de decisión, que integran tres componentes básicos: los nodos de decisión, las ramas y los nodos terminales; donde *“el árbol se construye mediante la partición determinista de la muestra inicial N (incluida en el nodo inicial) en varios subconjuntos descendientes (definidos en nodos hijos) a través de algún test lógico, conectándose los distintos nodos mediante ramas”* (Florez-Lopez, 2007). Este tipo de herramientas enmarca una gran variedad de modelos, dependiendo de su arquitectura; un amplio panorama sobre este tipo de metodología puede encontrarse en: Kumar y Ravi (2007); Sanz y Brio (1997) y Sosa Sierra (2011), entre otros. Dentro de este grupo cabe incluir el método de participaciones iterativas, basado en el desarrollo de un árbol de clasificaciones binarias (empresas sanas y fracasadas), que en cada nodo clasifica en dos grupos mediante una función univariante, con una ratio como variable independiente. Uno de los inconvenientes que presenta esta técnica es que no permite establecer comparaciones entre empresas que estén en el mismo grupo porque todas tienen la misma clasificación.

Sánchez (2005) analizó la crisis bancaria española durante el periodo 1977-1985, con una aplicación empírica de las redes neuronales artificiales, a partir de ratios financieras y contrastando con los métodos tradicionales de predicción discriminante (Análisis discriminante, Regresión *Probit* y *Logit*). En la muestra de estimación, las redes neuronales ofrecieron una fiabilidad del 100%, mientras la regresión *Logit* acertó el 95%, el análisis discriminante el 92,5% y el análisis *Probit* el 90%. Sin embargo, en opinión del autor, el modelado con redes neuronales en su nivel explicativo presenta carencias, ya que la capa oculta actúa como una caja negra y no es sencillo determinar el

significado de los pesos sinápticos, por ello no considera las redes de neuronas artificiales como un instrumento sustitutivo, sino complementario con los otros métodos estadísticos multivariantes.

La modelización mediante la aplicación de técnicas de computación evolutiva (dentro de las que se encuentra la PG) no es la más habitual en la predicción del fracaso empresarial, a pesar de los excelentes resultados que han obtenido las mismas en multitud de campos. De cualquier forma, existe un número limitado de trabajos previos que utilizan la PG:

- McKee y Lensbergn estudian el fracaso empresarial sobre un conjunto de empresas estadounidenses en el periodo 1991-1997. Utilizan la PG y utilizan las variables – ratios financieras – derivados de un estudio previo con *Rough sets*. Comparando los resultados, concluyen que los obtenidos con PG son mejores que los obtenidos con *Rough sets* (McKee y Lensberg, 2002).
- Salcedo-Sanz y col. estudian el fracaso en empresas españolas de seguros de no vida en el periodo 1983-1994, en un modelo con horizonte temporal 5 años. Utilizan ratios financieras y comparan los resultados de PG con los obtenidos con los métodos de Máquinas de soporte vectorial (SVM) y *Rough sets* (Salcedo-Sanz y col., 2005).
- Lensberg y col. centran su estudio en empresas noruegas en el periodo 1993-1998. Utilizan ratios financieras, pero añaden a la información contable otro tipo de información (que denominan factores de riesgo de fraude), en algún caso cualitativa (Lensberg y col., 2006).
- Alfaro-Cid y col. estudian un conjunto de empresas españolas en el periodo 1998-2000 para implementar modelos con horizonte temporal de 3 y 4 años. Abordan el problema de la utilización de una base de datos muy desequilibrada (el número de empresas fracasadas es muy inferior al de empresas no fracasadas) y con fuertes carencias de datos. Incluye información diferente a la obtenida de balance y cuentas de resultados, léase: tamaño, edad, etc. Asimismo, utiliza un amplio conjunto de funciones (Alfaro-Cid y col., 2007).
- Lilia García-Almanza y col. se centran en el sector bancario y aplican una técnica basada en PG y denominada *Evolving Decision Rules* (EDR) (Lilia Garcia-Almanza y col., 2010).
- Divsalar y col. utilizan otras técnicas basadas en PG. En concreto, utilizan dos variantes: *Gene expression programming* (GEP) y *Multi-expression programming* (MEP), aplicándolas en un conjunto de empresas iraníes (Divsalar y col., 2012).

Este trabajo pretende diferenciarse en varios aspectos:

- Abordar 4 modelos con horizontes temporales diferentes (1, 2, 3 y 4 años previos al fracaso).
- Restringir la información de entrada a la meramente contable (balance y cuenta de resultados públicos) contrastando la suficiencia de dicha información.
- Conjunto de funciones limitado en aras de aumentar la interpretabilidad (ver 3.5.3.1)
- Generalización, por medio de la búsqueda de modelos aplicables a diferentes sectores y tamaños de empresa.

3. DISEÑO DEL MODELO DE PREDICCIÓN

3.1. OBJETIVO

De una forma muy sucinta, el objetivo principal del presente trabajo es modelar mediante PG la predicción de insolvencia usando información financiera de diferentes años previos de la empresa bajo estudio. En concreto, se busca modelar la predicción de insolvencia en cuatro situaciones temporales diferentes: a uno, dos, tres y cuatro años del fracaso.

3.2. POBLACIÓN

El trabajo se basa en el estudio empírico relativo a la mortalidad de las pymes gallegas en el período 2003 – 2015. Para la modelización, se dispone de una población total de 3.413 empresas (419 fracasadas y 2.994 no fracasadas) con información contable del periodo 2003 a 2014 (este último año de forma parcial). La información contable de las empresas se ha obtenido de la base de datos SABI (Sistema de Análisis de Balances Ibéricos) de la compañía Informa, SA. Por otra parte, se ha utilizado información del Registro Público Concursal y de empresas especializadas en informes empresariales para obtener la información legal específica de su situación de fracaso.

3.2.1. Criterio de categorización de la empresa

La noción de pyme se entiende referida al criterio del Boletín Oficial del Estado del 12 de octubre 2013, que ha publicado la circular del Banco de España que recoge la modificación del concepto de pyme para las entidades de crédito y que, a partir de ese momento, se basará en la Recomendación 2003/361/CE sobre la definición de microempresas, pequeñas y medianas empresas, publicada en el Diario Oficial de la Unión Europea L 124, p. 36, de 20 de mayo de 2003.

En el caso de este trabajo - en aras a tener una población manejable - se aplicaron conjuntamente los siguientes criterios:

1. Plantilla (medida en unidades de trabajo anual) igual o mayor a 10 y menor que 250.
2. Y al menos una de las siguientes condiciones:

- Volumen de negocios anual mayor a 2 y menor o igual a 50 millones de euros.
- Balance general anual mayor a 2 y menor o igual a 43 millones de euros.

3.2.2. Sectores excluidos

Debido a las peculiaridades de algunos sectores que, entre otras razones, utilizan criterios contables o de valoración específicos, se ha optado por excluir a los mismos del estudio a fin de no alterar la interpretación de las ratios y distorsionar los resultados. Se ha procurado minimizar el número de sectores a excluir. Los sectores excluidos han sido los siguientes:

- CNAE 2009 (Sólo códigos primarios): 41 - Construcción de edificios, 42 - Ingeniería civil, 43 - Actividades de construcción especializada, 64 - Servicios financieros, excepto seguros y fondos de pensiones, 65 - Seguros, reaseguros y fondos de pensiones, excepto Seguridad Social obligatoria, 66 - Actividades auxiliares a los servicios financieros y a los seguros.

3.2.3. Horizonte temporal

El estudio contempla las observaciones de empresas fracasadas y no fracasadas desde 2003 hasta 2013 (entendiendo por observación la dupla: dato contable, ejercicio económico). El plazo ha sido escogido por dos motivos básicos: 1) Evaluar el comportamiento con una fuerte crisis financiera dentro del periodo contemplado y 2) por la oportunidad de poder comparar resultados con otros estudios previos de este equipo.

De las 37.543 observaciones posibles (3.413 empresas en 11 ejercicios económicos), 31.379 son utilizables (el resto se corresponden – en su inmensa mayoría - con datos inexistentes).

3.2.4. Criterio de fracaso

En el contexto de este trabajo, el concepto de fracaso está relacionado con la declaración legal de suspensión de pagos o quiebra, y se refiere principalmente a

situaciones reguladas que contemplan diferentes manifestaciones, de acuerdo a la normativa del sector y del país estudiado.

Siendo múltiples las causas de fracaso, la declaración legal de suspensión de pagos o quiebra en investigación es, a falta de confirmar el evento, el concepto más utilizado en los estudios de fracaso empresarial por su objetividad y concreción, además del más fácilmente aplicable en función de la información pública disponible.

3.2.5. Fuente de datos

Como se ha indicado, la información contable para configurar el conjunto de empresas del trabajo se ha obtenido de la base de datos SABI (Sistema de Análisis de Balances Ibéricos) de la compañía Informa, SA. De dicha base se han obtenido el balance y la cuenta de resultados de las pymes gallegas seleccionadas desde 2003 a 2013. Es preciso señalar que en el antedicho intervalo temporal se han producido relevantes modificaciones en los datos que las empresas están obligadas a facilitar (básicamente, se sustituyó el Plan General de Contabilidad vigente desde 1991, por el Plan General de Contabilidad en vigor desde 31 de diciembre de 2007. A su vez, este último se modificó en setiembre de 2010 y con efecto en los ejercicios a partir de 1 de enero de 2010).

Las modificaciones antedichas provocan que parte del detalle que la información contable facilita no pueda utilizarse (por responder a criterios diferentes, no homogéneos) dado que el trabajo utiliza series homogeneizadas de datos. Por ello se ha trabajado a nivel de grandes masas patrimoniales.

3.3. VARIABLES EXPLICATIVAS

Las variables explicativas del modelo son las ratios financieras calculadas a partir de los datos aportados por las cuentas anuales de las empresas. Las ratios a utilizar se han elegido atendiendo a la relevancia comunicada en la literatura y a su presencia en los modelos predictivos contrastados en trabajos previos. En el anexo 1 se recogen las 59 ratios utilizadas.

3.4. SELECCIÓN DE LA MUESTRA

Los criterios utilizados para conformar la muestra en este trabajo han sido básicamente los siguientes:

- La muestra se ha conformado con un 50% de empresas fracasadas y un 50% de empresas no fracasadas como forma de abordar el hecho de una población totalmente desequilibrada (las empresas fracasadas suponen el 12,2% del total).
- Las empresas fracasadas se han elegido de forma aleatoria.
- Las empresas no fracasadas se han elegido buscando el emparejamiento con las fracasadas siempre que fuese posible. El emparejamiento se ha realizado con la siguiente prelación:
 1. Ejercicio contable.
 2. Tamaño.
 3. Sector de actividad, a nivel código de la Clasificación Nacional de Actividades Económicas (CNAE).
- Por otra parte, se ha buscado que las mismas empresas configuren las muestras de los modelos 1, 2, 3 y 4.

En función de las anteriores restricciones y al hecho de que muchas empresas fracasadas dejaron de presentar cuentas anuales en el año previo a su fracaso, la muestra se ha configurado con 214 empresas (107 fracasadas y 107 no fracasadas).

3.5. LA PROGRAMACIÓN GENÉTICA

La programación genética (*genetic programming* en inglés) es una técnica de computación evolutiva que resuelve problemas automáticamente sin requerir que el usuario conozca o especifique previamente la forma o la estructura de la solución. A nivel más abstracto, es un método sistemático e independiente del dominio para conseguir que los ordenadores resuelvan problemas de forma automática partiendo de un conocimiento de alto nivel sobre lo que “se necesita hacer” (Poli y col., 2008, p. 1).

Los métodos de CE simulan en ordenador el proceso biológico de la evolución natural, con la selección natural como su motor. Se mantiene una población de genotipos o individuos que codifican soluciones a un determinado problema. Las soluciones se

evalúan y se les asigna una calidad, aptitud o *fitness*, en función de cuán bien resuelven el problema. De modo general, esas soluciones se someten a diferentes “operadores genéticos”, que modifican los genotipos de esas soluciones, operadores como cruce del material genético entre 2 individuos o la mutación de parte del material genético de una solución. Con los operadores genéticos se definen nuevos individuos y con el operador de selección se determina cuáles individuos pasan a formar parte de la siguiente generación. El operador de selección debe respetar la selección natural Darwiniana, en el sentido de que a mayor calidad mayor probabilidad de pasar el material genético de un individuo a la siguiente generación. Este proceso iterado a lo largo de diferentes generaciones obtendrá soluciones que progresivamente irán resolviendo mejor el problema concreto que codifican. Este proceso general se concreta en diferentes algoritmos en CE, como los clásicos Algoritmos Genéticos, Estrategias Evolutivas y Programación Genética.

Lo distintivo en PG es que se va evolucionando una población de “programas” de ordenador. En otras palabras, generación a generación, PG transforma – estocásticamente - poblaciones de programas en otras nuevas y, posiblemente mejores, poblaciones de programas (Poli y col., 2008, p. 2).

Originalmente, la PG fue concebida por John Koza (Koza, 1992) para evolucionar programas de ordenador, pero de esa idea original de “evolucionar programas” (entendido como secuencia de instrucciones para realizar una tarea) se ha ido ampliando al esquema de “evolucionar modelos” (entendido como representación simplificada de las relaciones entre las distintas variables involucradas en un problema) (Carmona Suárez y Fernández Galán, 2019, p. 125). Es esa capacidad de la PG para poder diseñar modelos de forma automática la que le confiere una gran versatilidad y le permite abordar problemas no fácilmente abordables por otras técnicas de CE.

3.5.1. Algoritmo

Los pasos básicos de PG se muestran en el siguiente pseudo-código (Poli y col., 2008, p. 4).

- 1: Crear aleatoriamente una población de programas a partir del conjunto disponible de primitivas.

- 2: **Repetir (en cada generación).**

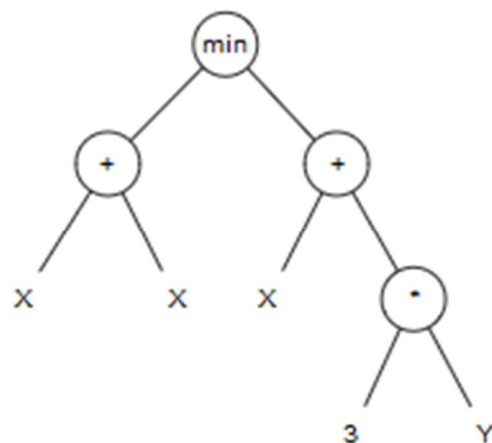
- 3: Ejecutar cada programa y evaluar su aptitud (*fitness*).
- 4: Seleccionar uno o dos programas de la población con una probabilidad basada en su aptitud para participar en las operaciones genéticas.
- 5: Crear nuevos programas individuales aplicando operadores genéticos con probabilidades previamente especificadas.
- 6: **Hasta** encontrar una solución aceptable o que se produzca otra condición de parada (p.ej.: alcanzar un número máximo de generaciones).
- 7: Devolver el mejor individuo.

3.5.2. Representación

La diferencia básica de la PG con otros paradigmas de computación evolutiva radica en el esquema de representación utilizado, que en el caso de PG son árboles sintácticos. A continuación, se muestra como ejemplo el árbol sintáctico correspondiente a la expresión:

$$\min (x+x, x+3*y).$$

Figura 1: Representación en PG



3.5.3. Pasos previos

3.5.3.1. Terminales y Funciones

En PG las variables, las funciones sin argumento y las constantes y variables del programa (en el ejemplo anterior: x , y , 3) son hojas y se denominan terminales.

Por otra parte, PG precisa de un conjunto de funciones (en el ejemplo anterior: $+$, $*$, \min) que son nodos interiores y que variarán en función del dominio del problema a resolver. Las funciones pueden ser de cualquier tipo (aritméticas, comparadores, condicionales, booleanas, etc.). No es extraño encontrar conjuntos de terminales y/o funciones que incluyen elementos diseñados expresamente para resolver problemas en un dominio específico.

Al conjunto de terminales y funciones se le denomina primitivas. Para que la PG sea efectiva en la resolución de problemas, el conjunto de funciones y terminales debe cumplir dos propiedades: cerradura y suficiencia (Koza, 1992, p. 82).

De forma muy sucinta, la propiedad de cerradura (*closure* en inglés) requiere que cualquier árbol que se construya con el conjunto de primitivas debe ser correcto. Por su parte, la propiedad de suficiencia (*sufficiency* en inglés) indica que la solución al problema debe poder ser especificada con el conjunto de primitivas indicado.

Las funciones que se usarán serán básicamente:

- operaciones aritméticas (suma, resta, multiplicación, división).
- condicionales (if-then-else).
- comparadores (mayor que, menor que).
- booleanas (y, o, no).

3.5.3.2. Función de evaluación

El conjunto de primitivas define indirectamente el espacio de búsqueda de PG, que vendrá dado por todos los programas (árboles) que puedan obtenerse combinando los elementos del conjunto de primitivas de todas las formas posibles. Sin embargo, falta dirigir la búsqueda hacia las zonas y/o elementos prometedores de ese espacio. Esa es la tarea encomendada a la función de evaluación de la aptitud. La función de evaluación es el mecanismo principal (a veces único) para dar una idea de los requisitos de alto nivel de un problema a la PG (Poli y col., 2008, p. 24)

La función de evaluación puede adoptar múltiples formas. En el caso del problema que nos ocupa, debemos tener en mente que vamos a evolucionar programas que determinen el fracaso empresarial. Las variables de entrada serán las ratios consideradas (Anexo 1) y las funciones antedichas (ver 3.5.3.1), mientras que el nodo en la jerarquía más alta del árbol categoriza a la empresa. Por tanto, la función de aptitud de cada árbol de la población genética estará determinada por cuán bien realiza ese árbol/programa la clasificación de las empresas. Y esa medida lógicamente será alguna de las ampliamente usadas en problemas de clasificación. En concreto, se han utilizado las siguientes medidas como posibles: Error cuadrático medio (*Mean squared error* o *MSE*), error cuadrático medio acotado (*Bounded mean squared error*), Pearson R^2 y puntos de penalización (*Penalty score*).

3.5.3.3. Parámetros

El siguiente paso previo es definir una serie de parámetros que controlarán la ejecución de PG. Los más relevantes son:

- El tamaño de la población.
- El número de generaciones.
- La profundidad máxima de un árbol.
- Las probabilidades de los operadores genéticos (cruce y mutación).
- La forma de inicializar la población.

No existe ninguna regla fija para determinar a priori el valor óptimo de cada parámetro para un problema dado. De hecho, en multitud de ocasiones son las limitaciones ajenas las que influyen en los valores fijados (p.ej.: el tamaño de la población, el número máximo de ejecuciones y/o la profundidad máxima de un árbol, que a menudo vienen condicionados por el tiempo de evaluación de la aptitud, dado que se evalúa individuo a individuo y ello puede conducir a tiempos de ejecución no aceptables o, simplemente, imposibles de abordar).

Se han utilizado parámetros estándar, muchos indicados por la plataforma software utilizada (*HeuristicLab*, comentado posteriormente), mientras otros se han establecido por un proceso clásico de barrido (determinando los mejores valores de un parámetro entre un conjunto finito mientras los demás se mantienen a un valor prefijado estándar).

3.5.3.4. Finalización y solución

Generalmente, el criterio de finalización es alcanzar el máximo número de generaciones, pero pueden incluirse otros o sustituirse el antedicho (p.ej.: alcanzar un determinado valor en la función de evaluación o aptitud). Asimismo, la solución presentada suele coincidir con el individuo que presenta el mejor valor para la función de evaluación de la aptitud.

En este trabajo el criterio de finalización ha sido el máximo número de generaciones y la solución presentada o seleccionada se elige priorizando los verdaderos positivos en una solución equilibrada (ver apartado 4).

3.5.4. Inicializando la población

El primer paso del algoritmo antedicho es crear aleatoriamente una población de programas a partir del conjunto disponible de primitivas.

3.5.5. Selección

Como en otros algoritmos evolutivos, los operadores genéticos en PG se aplican a los individuos que se seleccionan de forma probabilística en función de su aptitud (medida previamente de acuerdo con la función de evaluación utilizada). Existen diferentes métodos para la selección dentro de PG, pero quizás el más utilizado sea el de Torneo (*Tournament*). De cualquier forma, es frecuente utilizar otros métodos de selección (*Diversity, Generalized Rank, etc.*).

En Torneo, se escoge aleatoriamente de la población a un número dado de individuos (fijado como parámetro). Se comparan entre ellos y el que mejor aptitud tiene se escoge como “padre”.

3.5.6. Recombinación y Mutación

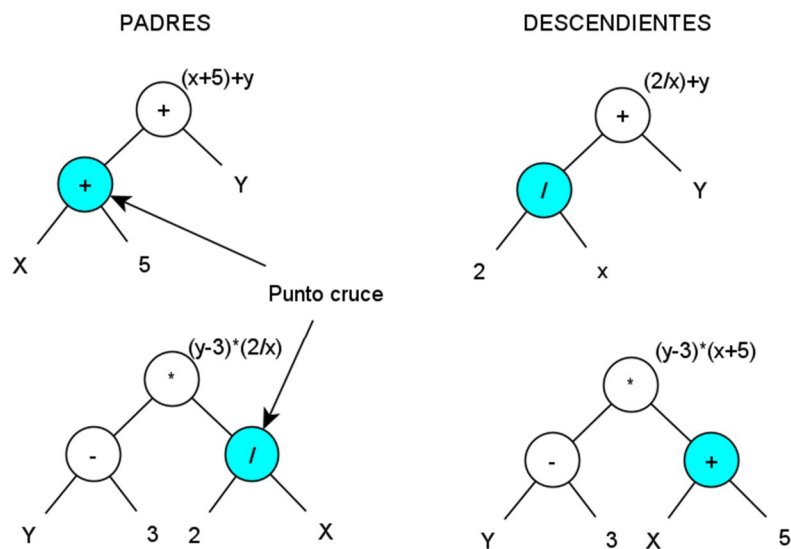
La recombinación en PG, como en otros algoritmos evolutivos, consiste en intercambiar el material genotípico entre dos individuos. El operador de recombinación estándar en PG es el *cruce de subárboles* (*subtree crossover* en inglés) (Carmona Suárez y

Fernández Galán, 2019, p. 137). Se utilizan dos versiones del cruce de subárboles, en la primera de ellas, a partir de dos padres se genera un único descendiente. En la segunda, a partir de dos padres se generan dos descendientes. El funcionamiento es sencillo.

En la versión de un descendiente: dados dos árboles padres, se elige aleatoriamente un nodo en cada uno de ellos. Entonces, se crea el descendiente reemplazando el subárbol que comienza en el punto de cruce del primer padre con el árbol cuyo nodo inicial es el punto de cruce del segundo padre (Poli y col., 2008, p. 16).

En la versión más usual de dos descendientes: dados dos árboles padres, se elige aleatoriamente un nodo en cada uno de ellos. Dicho nodo se convierte en nodo raíz del subárbol correspondiente a intercambiar. Finalmente, los subárboles intercambian su posición en los nodos padre y los árboles resultantes constituyen los dos hijos (Carmona Suárez y Fernández Galán, 2019, p. 137), como se ilustra en la siguiente figura:

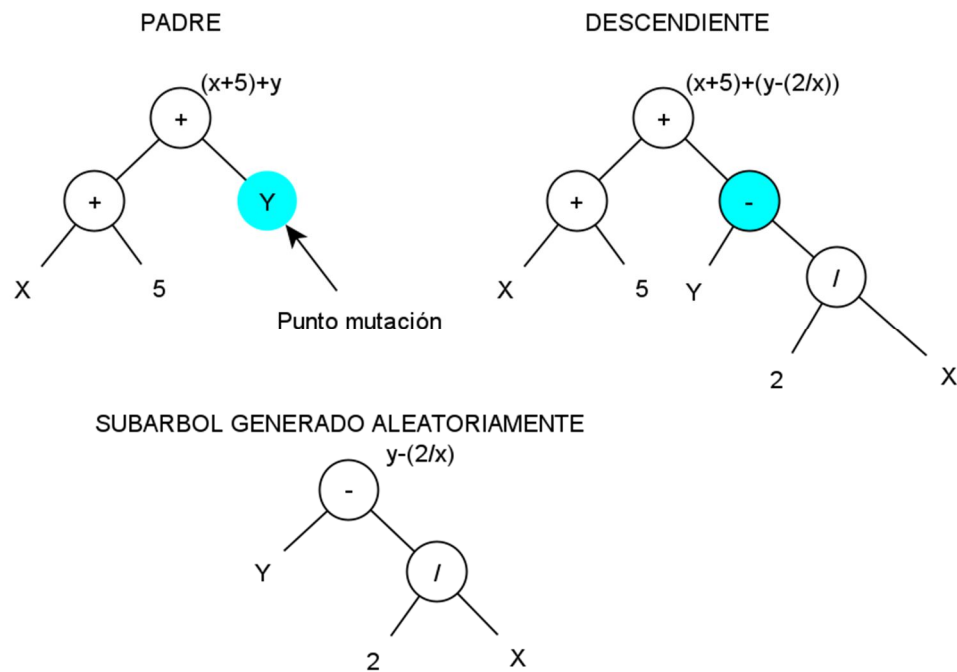
Figura 2: Recombinación



La mutación consiste en la obtención de un nuevo individuo a partir de otro en el que se realiza una alteración aleatoria. La mutación de subárbol es la operación de mutación estándar en PG. Dado un árbol padre, se elige un nodo de forma aleatoria. Dicho nodo se constituye en el nodo raíz del subárbol correspondiente a reemplazar. A continuación,

se construye un nuevo árbol independiente de forma aleatoria. Finalmente se reemplaza el subárbol seleccionado con el nuevo árbol creado y el árbol resultante constituye el nuevo individuo mutado (Carmona Suárez y Fernández Galán, 2019, p. 138), como se ilustra en la siguiente figura (Poli y col., 2008, p. 17).

Figura 3: Mutación



Existe un tercer operador denominado *reproducción* que opera sobre un único individuo, seleccionado previamente, haciendo una copia exacta (clonación) en la descendencia. Se utiliza con distintos fines: preservar al mejor, o mejores individuo(s) para la próxima generación (elitismo), disminuir el coste computacional, etc. (Carmona Suárez y Fernández Galán, 2019, p. 137).

3.6. SOFTWARE

El software utilizado para implementar el problema en PG ha sido *HeuristicLab* (Wagner y col., 2014). Se puede descargar en su página web: <https://dev.heuristiclab.com/trac.fcgi/>

Este software tiene una serie de características relevantes para este trabajo:

- De una parte, contiene una serie de algoritmos básicos relacionados con técnicas de computación evolutiva (algoritmos genéticos, programación genética, etc.).
- Un conjunto de problemas que utilizan dichos algoritmos en diferentes campos y que pueden ser reutilizados (p.ej.: resolución del problema del viajante, 11-bit multiplexer, clasificación simbólica para la detección del cáncer de mama, etc.).
- Permite una muy completa y relativamente sencilla parametrización.

3.7. PROCESO

De forma sucinta se ha seguido el siguiente proceso para cada uno de los modelos de predicción del fracaso:

- Se modeliza en *HeuristicLab* como un problema de Programación Genética - Clasificación simbólica.
- *HeuristicLab* permite dividir el conjunto de entrenamiento (*training*) en dos subconjuntos: aptitud (*fitness*) y validación (*validation*). El primero (aptitud) es el que se usa en la progresión del proceso evolutivo. Esta división en dos subconjuntos va a permitir obtener, en cada ejecución, dos soluciones, una que corresponde al individuo que mejor evaluación tiene en el conjunto total de entrenamiento y otra que corresponde al mejor individuo en el subconjunto de validación. Este es el objetivo de dicho subconjunto de validación: permitir detectar soluciones que deberían tener un buen comportamiento al generalizar en el conjunto de test, o, dicho de otro modo, impedir el sobre-ajuste (*overfitting*). En este trabajo se ha optado por este enfoque, utilizando un conjunto de entrenamiento de 214 observaciones (ver 3.4), de las cuales 144 se han destinado al subconjunto de aptitud y 70 al subconjunto de validación. Con ello se pretende evitar el sobre-ajuste y obtener soluciones que generalicen bien.
- Asimismo, se le facilita a *HeuristicLab* el conjunto de test, compuesto por las observaciones restantes, a fin de que el software aplique a dicho conjunto los modelos obtenidos a partir del conjunto de entrenamiento y el conjunto de validación y obtenga las métricas oportunas.
- Las opciones básicas que, por defecto, han sido establecidas en el algoritmo utilizado por este trabajo, son las siguientes (usando los nombres de *HeuristicLab*):

○ Clase positiva	1 (empresas fallidas)
○ Evaluator	Mean squared error
○ ModelCreator	Accuracy Maximizing Thresholds
○ SolutionCreator	Probabilistic Tree Creator
○ MaximunDepth	10
○ MaximunLength	100
○ Analyzer	Multi Analyzer
○ Crossover	Subtree Swapping Crossover - Prob. 90%
○ Elites	1
○ MaximunGeneration	100
○ MutationProbability	15%
○ Mutator	Multi Symbolic Expression Tree Manipulator
○ PopulationSize	1000
○ Selector	Tournament Group 5

- Se procedió al ajuste, uno a uno, de los diferentes parámetros en busca del que mejor se adapta al problema en cuestión (*ceteris paribus*), mediante el barrido de valores de parámetros anteriormente comentado. Los parámetros más relevantes fueron los siguientes:
 - Profundidad máxima del árbol.
 - Tamaño de la población.
 - Número máximo de generaciones.
 - Forma de inicialización de la población.
 - Selector.
 - Conjunto de funciones a utilizar (Gramática).
 - Evaluador (función de evaluación de aptitud).
 - Probabilidad de cruce.
 - Probabilidad de mutación.
- Determinado el perfil a utilizar para abordar el problema, se realizan experimentos de los que se extraen las mejores soluciones.

4. RESULTADOS

Los modelos que se presentan en este apartado han sido escogidos por su capacidad de generalizar en el conjunto de test. Para tal elección se han realizado 3.000 ejecuciones de PG y, como se explicó en 3.7, cada ejecución selecciona 2 modelos: los que mejor evaluación tienen sobre el conjunto de entrenamiento y de validación. A posteriori, se realiza otra selección/filtrado, pues nos quedamos con aquel modelo/programa que mejor generaliza en el conjunto de test.

En el campo de la inteligencia artificial una matriz de confusión es una herramienta que permite la visualización del desempeño de un algoritmo que se emplea en aprendizaje supervisado.

Figura 4: Matriz de confusión

		Real o Target	
		Clase Negativa 0	Clase Positiva 1
Clase predicha	Clase Negativa 0	<i>Verdaderos negativos TN</i>	<i>Falsos negativos FN</i>
	Clase Positiva 1	<i>Falsos Positivos FP</i>	<i>Verdaderos positivos TP</i>

Existen diferentes métricas para escoger un modelo, en función del objetivo del problema propuesto y del uso que se vaya a hacer del mismo. Teniendo en cuenta las diferentes categorías de la matriz de confusión (verdaderos positivos: TP, verdaderos negativos: TN, falsos positivos: FP, falsos negativos: FN), tres métricas habituales son:

- Porcentaje de aciertos (*accuracy*) que sería igual a: $(TP+TN)/(TP+TN+FP+FN)$.
- *F1 score*, que en su formulación general sería igual a: $(1+\beta^2)TP/((1+\beta^2)TP + \beta^2FN + FP)$. Si $\beta=1$ se obtiene la media armónica, aunque es habitual utilizar los valores de $\beta=2$ o $\beta=0.5$ para hacer más énfasis en los falsos negativos o en los falsos positivos respectivamente.
- El coeficiente de correlación de Matthews (MCC).

Las ventajas de MCC sobre *accuracy* y *F1 score* las explica Davide Chicco en su artículo "Ten quick tips for machine learning in computational biology" (Chicco, 2017) y Chicco y Giuseppe Jurman en su artículo "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation" (Chicco y Jurman, 2020).

El coeficiente de correlación de Matthews es más informativo que F1 score y *accuracy* en la evaluación de los problemas de clasificación binaria, porque tiene en cuenta los coeficientes de equilibrio de las cuatro categorías de la matriz de confusión. Su valor oscila entre -1 y +1, siendo -1 su peor valor y +1 el mejor.

La fórmula de MCC es la siguiente:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Debido a que MCC trabaja con los valores absolutos de TP, TN, FP y FN, es sensible al hecho de que el conjunto de test se encuentre desequilibrado, pues las observaciones de clase 0 (en este caso, empresas no fracasadas) son muchas más que las observaciones de clase 1 (empresas fracasadas). Por ello, la incidencia en MCC de la mejora marginal en la correcta clasificación de una observación de clase 0 y la correspondiente a la de una observación de clase 1, es totalmente diferente.

Cualquiera de esas métricas podría utilizarse en este problema, pero el criterio elegido (para seleccionar los modelos finales) ha sido arbitrario y orientado a visualizar la flexibilidad de la PG. Se buscan modelos “equilibrados” que prioricen el porcentaje de verdaderos positivos (*True positive rate, recall o sensitivity*) pero sin renunciar a un elevado porcentaje de verdaderos negativos (*True negative rate o specificity*). En concreto se busca el mayor porcentaje de verdaderos positivos siempre y cuando la diferencia entre verdaderos positivos y verdaderos negativos no supere el 10% del porcentaje de verdaderos positivos. Se adopta la posición de un inversor, en la que, debido a las pérdidas que produce, se supone una mayor sensibilidad por tener modelos que como prioridad tengan la detección de los verdaderos positivos (empresas fracasadas), aunque haciendo concesiones al equilibrio con la detección de verdaderos negativos (empresas no fracasadas).

A continuación, se presentan los resultados de los diferentes modelos en el conjunto de entrenamiento (*fitness+validation*). Recuérdese que la elección de modelos que se presenta se ha regido por su capacidad de generalización (comportamiento en el conjunto de test), como se ha explicado previamente.

Tabla 1: Porcentaje de aciertos en el conjunto de entrenamiento

CONJUNTO DE ENTRENAMIENTO
(fitness + validation)

<i>porcentaje de aciertos</i>	<i>Fracasadas</i>	<i>No fracasadas</i>	<i>Total</i>
Modelo año 1	93,46%	92,52%	92,99%
Modelo año 2	80,35%	83,24%	81,79%
Modelo año 3	79,44%	78,50%	78,97%
Modelo año 4	63,55%	76,64%	70,09%

A continuación, se presentan los resultados de los diferentes modelos en el conjunto de test.

Tabla 2: Porcentaje de aciertos en el conjunto de test

CONJUNTO DE TEST

<i>porcentaje de aciertos</i>	<i>Fracasadas</i>	<i>No fracasadas</i>	<i>Total</i>
Modelo año 1	89,72%	83,05%	83,07%
Modelo año 2	82,95%	77,09%	77,14%
Modelo año 3	81,47%	75,40%	75,45%
Modelo año 4	78,88%	72,69%	72,75%

Nótese que el haber optado por seleccionar los modelos por su capacidad de generalización en el conjunto de test hace que sea posible seleccionar modelos en los que el porcentaje de acierto en el conjunto de entrenamiento es menor que el correspondiente en el conjunto de test (p.ej.: modelo año 4: porcentaje aciertos empresas fracasadas en conjunto de entrenamiento 63,55% y en conjunto de test 78,88%).

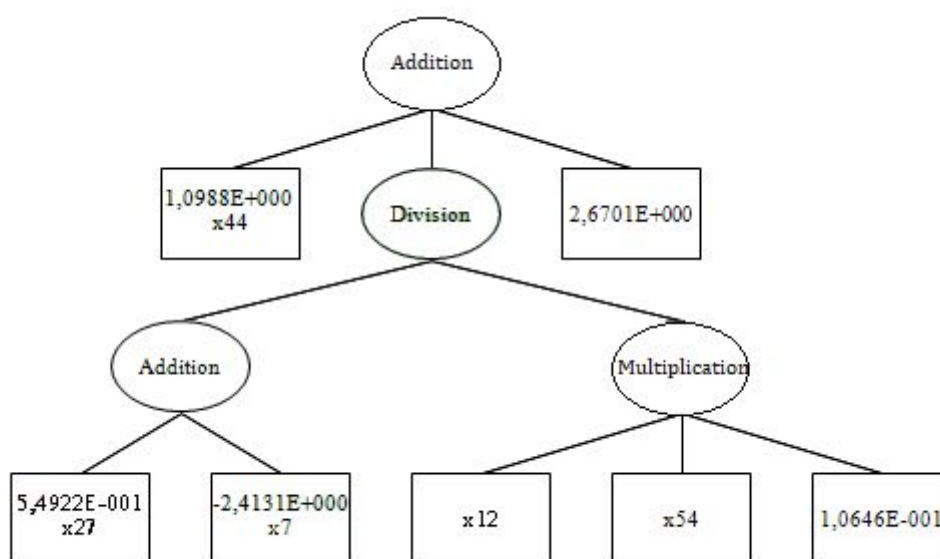
Los resultados de los principales estudios, con horizonte temporal de 1 a 3 años, pueden encontrarse en la recopilación de los mismos realizada en 2011 por Du Jardin y Séverin (Du Jardin y Séverin, 2011, p. 702) . Puede observarse que, a pesar de la restricción auto-impuesta de priorizar los verdaderos positivos y no desequilibrar el modelo, con lo que se deteriora el porcentaje total de aciertos (pues la población de empresas no fracasadas es mucho mayor que la de fracasadas), los resultados obtenidos con PG se pueden considerar muy buenos, al encontrarse de media en el percentil 75% para los modelos 1, 2 y 3, que son comparables con el estudio antedicho.

Por otra parte, los resultados obtenidos están en consonancia con los de trabajos previos de PG, concretamente el de Alfaro y col. para horizontes temporales de 3 y 4 años (Alfaro-Cid y col., 2007) y el de Lensberg y col. (Lensberg y col., 2006) para horizonte temporal 1 año, aunque ambos manejen información adicional a la meramente contable.

Aunque con un criterio de fracaso empresarial ligeramente diferente al de este trabajo, pues al concepto de fracaso utilizado en este trabajo, vinculado a la declaración legal de suspensión de pagos o quiebra, se añadía el relacionado con la situación patrimonial precursora del fracaso futuro, referida principalmente a situaciones que pudieran incurrir en quiebra técnica y disminución de resultados, los resultados obtenidos son equiparables a los obtenidos previamente por este equipo aplicando redes de neuronas artificiales (Beade y col., 2016).

Se ha comentado que una de las ventajas de la PG es la interpretabilidad del modelo. En ese sentido, a título de ejemplo, en la siguiente figura se muestra el árbol correspondiente a la solución del modelo 4 antedicho.

Figura 5: Representación gráfica modelo 4



En ella, la correspondencia con las ratios sería la siguiente:

- x7 = Gastos financieros / Exigible total
- x12 = Exigible a largo / Exigible total
- x27 = Cash flow / Exigible total
- x44 = Importe neto cifra de ventas / Activo circulante
- x54 = Fondos propios / Activo total

El umbral utilizado para la clasificación es 2,3. Es decir, es el umbral que se utiliza en el nodo más alto de la jerarquía (con la operación *Addition*) a la hora de la clasificación entre fracasada/no fracasada.

No únicamente es importante la interpretabilidad en el sentido antedicho. El entorno de PG facilita las ratios relevantes e inmediatamente es factible calcular el impacto de cada una de ellas en la valoración final y, con ello, en la clasificación. Es fácil ver un análisis de sensibilidad de las variaciones de las ratios y su impacto final. Por ejemplo, nótese la importancia, en el modelo de 4 años previos al fracaso, de las ratios concernientes al endeudamiento, su coste y su estructura (Fondos propios / Activo total no deja de ser un complemento de Exigible total / Pasivo total).

5. CONCLUSIONES

Los resultados del trabajo desarrollado muestran la ventaja de la PG frente a otras metodologías, básicamente en los siguientes aspectos.

- La interpretabilidad. Las soluciones obtenidas por PG son perfectamente interpretables.
- Los resultados obtenidos son perfectamente comparables, cuando no mejores que los de otras técnicas, tanto en modelos a corto plazo como a largo plazo. Ello, a pesar de las restricciones auto-impuestas en este estudio. En concreto:
 - Abordar intencionadamente un periodo en el que se sitúa una fuerte crisis financiera, que hizo fracasar multitud de empresas de forma inesperada.
 - Restringir el conjunto de datos a aquellos que son inmediatamente accesibles y habituales (balance y cuenta de resultados).
 - Restringir la gramática en aras de una interpretabilidad más comprensible. Algunos autores, como Alfaro-Cid y col. (2007) utilizan conjuntos de funciones que incluyen el coseno, el logaritmo y la exponencial.
 - Generalización, por medio de la búsqueda de modelos aplicables a diferentes sectores y tamaños de empresa. Los modelos obtenidos son aplicables al conjunto de PYMES, cualquiera que sea su tamaño y/o su sector de actividad (con excepción de los sectores excluidos indicados en 3.2.2).
 - En la selección de soluciones se adopta la posición de un inversor, en la que se supone una mayor sensibilidad por tener modelos que como prioridad tengan la detección de los verdaderos positivos (empresas fracasadas que el modelo detecta como tales, $TP/(TP+FN)$) y, consecuentemente, minimizar el porcentaje de falsos negativos (empresas fracasadas que el modelo no detecta como tales, $FN/(TP+FN)$) por las pérdidas potenciales que éstos conllevan para el inversor. Todo ello, haciendo concesiones al equilibrio con la detección de verdaderos negativos (empresas no fracasadas). Esta postura perjudica el porcentaje total de aciertos al ser muy minoritaria la clase positiva (empresas fracasadas).

6. BIBLIOGRAFIA

- Alfaro-Cid, E., Sharman, K., y Esparcia-Alcazar, A. (2007). A genetic programming approach for bankruptcy prediction using a highly unbalanced database. En M. Giacobini (Ed.), *Applications of Evolutionary Computing, Proceedings* (Vol. 4448, pp. 169-+).
- Altman, E., Iwanicz-Drozdowska, M., Laitinen, E., y Suvas, A. (2015). Financial and Non-Financial Variables as Long-Horizon Predictors of Bankruptcy. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2669668>
- Beade, A., Santos, J., y Rodríguez López, M. (2016). Análisis temporal de predicción de la insolvencia empresarial en pymes gallegas mediante modelos conexionistas. *VII Jornadas de AECA (Asociación Española de Contabilidad y Administración de Empresas) sobre "Valoración, Financiación y Gestión de Riesgo"*. Vigo.
- Carmona Suárez, E. J., y Fernández Galán, S. (2019). *Fundamentos de la computación evolutiva*. Marcombo.
- Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *Biodata Mining*, 10, 35. <https://doi.org/10.1186/s13040-017-0155-3>
- Chicco, D., y Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *Bmc Genomics*, 21(1). <https://doi.org/10.1186/s12864-019-6413-7>
- Divsalar, M., Roodsaz, H., Vahdatinia, F., Norouzzadeh, G., y Behrooz, A. H. (2012). A Robust Data-Mining Approach to Bankruptcy Prediction. *Journal of Forecasting*, 31(6), 504-523. <https://doi.org/10.1002/for.1232>
- Du Jardin, P., y Séverin, E. (2011). Predicting corporate bankruptcy using a self-organizing map: An empirical study to improve the forecasting horizon of a financial failure model. *Decision Support Systems*, 51(3), 701-711. <https://doi.org/10.1016/j.dss.2011.04.001>

- Florez-Lopez, R. (2007). Análisis de los determinantes del riesgo de crédito en presencia de carteras de bajo incumplimiento. Una propuesta de aplicación. *Revista europea de dirección y economía de la empresa*, 16(2), 71–92.
- Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. The MIT Press.
- Kumar, P. R., y 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.
- Laffarga Briones, J., Martín Marín, J. L., y Vázquez Cueto, M. J. (1987). Predicción de la crisis bancaria en España: Comparación entre el análisis logit y el análisis discriminante. *Cuadernos de Ciencias Económicas y Empresariales*, 18, 49-57.
- Lensberg, T., Eilifsen, A., y McKee, T. E. (2006). Bankruptcy theory development and classification via genetic programming. *European Journal of Operational Research*, 169(2), 677-697. <https://doi.org/10.1016/j.ejor.2004.06.013>
- Lilia Garcia-Almanza, A., Alexandrova-Kabadjova, B., y Martinez-Jaramillo, S. (2010). Understanding bank failure: A close examination of rules created by Genetic Programming. En *2010 IEEE Electronics, Robotics and Automotive Mechanics Conference (CERMA 2010)* (pp. 34-39). IEEE Computer Soc. <https://doi.org/10.1109/CERMA.2010.14>
- Mateos-Ronco, A., del Mar Marín-Sánchez, M., Marí-Vidal, S., y Seguí-Mas, E. (2011). Los modelos de predicción del fracaso empresarial y su aplicabilidad en cooperativas agrarias. *CIRIEC-España, Revista de Economía Pública, Social y Cooperativa*, 70, 179–208.
- McKee, T. E., y Lensberg, T. (2002). Genetic programming and rough sets: A hybrid approach to bankruptcy classification. *European Journal of Operational Research*, 138(2), 436-451. Scopus. [https://doi.org/10.1016/S0377-2217\(01\)00130-8](https://doi.org/10.1016/S0377-2217(01)00130-8)

- Poli, R., Langdon, W. B. (William B.), McPhee, N. F., y Koza, J. R. (2008). *A field guide to genetic programming*. [S.I.]: [Lulu Press], lulu.com.
<http://archive.org/details/AFieldGuideToGeneticProgramming>
- Salcedo-Sanz, S., Fernandez-Villacanas, J. L., Segovia-Vargas, M. J., y Bousono-Calzon, C. (2005). Genetic programming for the prediction of insolvency in non-life insurance companies. *Computers & Operations Research*, 32(4), 749-765.
<https://doi.org/10.1016/j.cor.2003.08.015>
- Sánchez, J. de A. (2005). Comparativa de métodos de predicción de la quiebra: Redes neuronales artificiales vs. métodos estadísticos multivariantes. *Partida doble*, 168, 105-113.
- Sanz, B. M. A., y Brió, B. d. (1997). Redes Neuronales y sistemas borrosos. *Ra-Ma. Madrid*.
- Serrano Cinca, C., y Martín del Brío, B. (1993). Predicción de la quiebra bancaria mediante el empleo de redes neuronales artificiales. *Revista española de financiación y contabilidad*, 22(74), 153–176.
- Sosa Sierra, M. D. C. (2011). Inteligencia artificial en la gestión financiera empresarial. *Revista científica Pensamiento y gestión*, 23.
- Wagner, S., Kronberger, G., Beham, A., Kommenda, M., Scheibenpflug, A., Pitzer, E., Vonolfen, S., Kofler, M., Winkler, S., Dorfer, V., y Affenzeller, M. (2014). *Advanced Methods and Applications in Computational Intelligence* (R. Klempous, J. Nikodem, W. Jacak, & Z. Chaczko, Eds.; Vol. 6, pp. 197–261). Springer.
http://link.springer.com/chapter/10.1007/978-3-319-01436-4_10

7. ANEXOS

Anexo 1: Ratios financieras utilizadas en el análisis

REFERENCIA	RATIO	MEDIDA FINANCIERA
ACT01	Gastos financieros / Valor agregado	ACTIVIDAD
ACT02	Gastos personal / Activo fijo	ACTIVIDAD
ACT03	Gastos personal + Dotación amortización inmovilizado / Valor agregado	ACTIVIDAD
ACT04	Ingresos explotación / Gastos explotación	ACTIVIDAD
ACT05	Valor agregado / Importe neto cifra de ventas	ACTIVIDAD
APL01	Gastos financieros / EBIT	APALANCAMIENTO
APL02	Gastos financieros / Exigible total	APALANCAMIENTO
APL03	Gastos financieros / Resultados explotación	APALANCAMIENTO
APL04	Resultado ejercicio / Exigible total	APALANCAMIENTO
END01	Exigible total / Fondos propios	ENDEUDAMIENTO
END02	Fondos propios - Resultado ejercicio / Exigible a corto	ENDEUDAMIENTO
END03	Exigible a largo / Exigible total	ENDEUDAMIENTO
EST01	Capital suscrito / Fondos propios	ESTRUCTURA
EST02	Activo circulante / Activo total	ESTRUCTURA
EST03	Dotación amortización inmovilizado / Inmovilizado	ESTRUCTURA
EST04	Capital circulante / Activo total	ESTRUCTURA
EST05	Capital circulante / Exigible total	ESTRUCTURA
EST06	Capital circulante / Importe neto cifra de ventas	ESTRUCTURA
EST07	Otros activos líquidos / Activo total	ESTRUCTURA
EST08	Resultado ejercicio / Capital circulante	ESTRUCTURA
EST09	Medida descomposición del activo	ESTRUCTURA
LIQ01	Cash flow operativo / Activo total	LIQUIDEZ
LIQ02	Cash flow operativo / Exigible total	LIQUIDEZ
LIQ03	Cash flow operativo / Exigible a corto	LIQUIDEZ
LIQ04	Cash flow operativo / Importe neto cifra de ventas	LIQUIDEZ
LIQ05	Cash flow / Activo total	LIQUIDEZ
LIQ06	Cash flow / Exigible total	LIQUIDEZ
LIQ07	Cash flow / Exigible a corto	LIQUIDEZ
LIQ08	Cash flow / Importe neto cifra de ventas	LIQUIDEZ
LIQ09	Otros activos líquidos / Exigible a corto	LIQUIDEZ
LIQ10	Existencias / Exigible a corto	LIQUIDEZ
LIQ11	Existencias + Deudores / Exigible a corto	LIQUIDEZ
LIQ12	Intervalo sin crédito	LIQUIDEZ
LIQ13	Deudores / Exigible a corto	LIQUIDEZ
REN01	EBIT / Activo total	RENTABILIDAD
REN02	EBIT / Importe neto cifra de ventas	RENTABILIDAD
REN03	Resultado ejercicio / Importe neto cifra de ventas	RENTABILIDAD
REN04	Resultado ejercicio - Deudores - Existencias / Activo total	RENTABILIDAD
REN05	Resultado ejercicio / Activo total	RENTABILIDAD
REN06	Resultado ejercicio / Fondos propios	RENTABILIDAD
ROT01	Activo circulante - Existencias / Importe neto cifra de ventas	ROTACION
ROT02	Existencias / Importe neto cifra de ventas	ROTACION
ROT03	Importe neto cifra de ventas / Ingresos de explotación	ROTACION
ROT04	Importe neto cifra de ventas / Activo circulante	ROTACION
ROT05	Importe neto cifra de ventas / Inmovilizado	ROTACION
ROT06	Importe neto cifra de ventas / Activo total	ROTACION
ROT07	Importe neto cifra de ventas / Capital circulante	ROTACION
ROT08	Otros activos líquidos / Importe neto cifra de ventas	ROTACION
SOL01	Activo circulante - Existencias / Exigible a corto	SOLVENCIA
SOL02	Activo circulante / Exigible total	SOLVENCIA
SOL03	Activo circulante / Exigible a corto	SOLVENCIA
SOL04	Inmovilizado / Fondos propios	SOLVENCIA
SOL05	Exigible total / Activo total	SOLVENCIA

Anexo 1: Continuación

REFERENCIA	RATIO	MEDIDA FINANCIERA
SOL06	Fondos propios / Activo total	SOLVENCIA
SOL07	Fondos propios / Inmovilizado	SOLVENCIA
SOL08	Exigible a corto / Activo total	SOLVENCIA
SOL09	Resultados ordinarios antes impuestos / Exigible a corto	SOLVENCIA
TES01	Tesorería / Exigible a corto	TESORERIA
TES02	Tesorería / Importe neto cifra de ventas	TESORERIA