



Doble Grado en Ingeniería Informática y Administración y Dirección
de Empresas (2021/2022)

Identificación de empresas zombis PYMES mediante redes neuronales convolutivas

Presentado por: Sumit Kumar Jethani Jethani

DNI: 43303840-S

Tutorizado por: Agustín Sánchez Medina y Félix Blázquez Santana

Las Palmas de Gran Canaria, a 27 de enero de 2022

SOLICITUD DE DEFENSA DE TRABAJO DE FIN DE TÍTULO

D/D^a **Sumit Kumar Jethani Jethani**, autor/a del Trabajo de Fin de Título: **Identificación de empresas zombis PYMES mediante redes neuronales convolutivas**, correspondiente a la titulación **Doble Grado en Ingeniería Informática y Administración y Dirección de Empresas**, en colaboración con la empresa/proyecto (indicar en su caso) _____

SOLICITA

que se inicie el procedimiento de defensa del mismo, para lo que se adjunta la documentación requerida, haciendo constar que

☒ se autoriza / ☐ no se autoriza, la grabación en audio de la exposición y turno de preguntas.

Asimismo, con respecto al registro de la propiedad intelectual/industrial del TFT, declara que:

☐ Se ha iniciado o hay intención de iniciarlo (defensa no pública).
☒ No está previsto.

Y para que así conste firma la presente. (fecha en firma electrónica)

El/la estudiante

Fdo.: _____

A rellenar y firmar **obligatoriamente** por el/la/los/las tutor/a/es/as.

En relación a la presente solicitud, se informa: (firmar donde corresponda)

Positivamente (en caso de detección de copia, esta firma quedará invalidada)

Negativamente (la justificación en caso de informe negativo deberá incluirse en el TFT05)

DIRECTORA DE LA ESCUELA DE INGENIERÍA INFORMÁTICA

Agradecimientos

Me gustaría aprovechar estas primeras líneas y de manera muy breve, para mostrar agradecimientos hacia aquellas personas que me han apoyado en la realización de este trabajo de fin de titulación.

Así, en primer lugar, agradecer a mis tutores académicos: Agustín Sánchez Medina y Félix Blázquez Santana, que en todo momento me han apoyado y han sabido resolver las numerosas dudas que me han surgido en la realización de este trabajo. En segundo lugar, a mi familia que en todo momento ha creído en mí. Y por último, a todos mis compañeros, en especial a Ricardo Arbelo Guillén, que han formado parte de esta etapa de mi vida en la Universidad de Las Palmas de Gran Canaria.

Abstract

Las empresas zombis pueden suponer un alto riesgo para la economía de un país así como para las empresas sanas de la misma. Para poder lidiar con ellas, es indispensable su identificación temprana. Utilizando información financiera disponible en *SABI* sobre las pymes del sector turístico español (2017-2019), en el presente estudio se aporta una red neuronal convolutiva (CNN) que permite la identificación temprana de dichas empresas. Además, se demuestra que las CNN presentan una mayor eficacia (AUC de 0,87) en la predicción de las zombis que las metodologías tradicionales. Por último, el estudio también contribuye a la teoría de las empresas zombis en España, demostrando que el 8,1% de las pymes de hostelería son identificadas como zombis bajo el criterio principal del estudio.

Palabras claves: Empresas zombis; Pymes; Inteligencia Artificial; Redes neuronales convolutivas (CNN); España

Abstract

Zombie enterprises may involve a high risk to a country's economy as well as to the healthy companies in the same. In order to deal with them, early identification is essential. Using financial information available in *SABI* on SME's in the Spanish tourism sector (2017-2019), this study provides a convolutional neural network (CNN) that allows the early identification of these companies. Furthermore, CNN's are shown to have a higher efficiency (AUC of 0.87) in predicting zombies than traditional methodologies. Finally, the study also contributes to the theory of zombie companies in Spain, showing that 8.1% of hotel/restaurant SME's are identified as zombies under the main criteria of the study.

Keywords: Zombie enterprises; SME; Artificial Intelligence; Convolutional Neural Networks (CNN); Spain

Guía de lectura

El presente estudio consiste en un trabajo final de la doble titulación en Ingeniería Informática y Administración y Dirección de Empresas. Es por ello, que a continuación se enumeran las secciones en las que el tribunal de Administración y Dirección de Empresas deben prestar especial atención para la correcta evaluación del trabajo. No obstante, se les recomienda la lectura completa del presente estudio.

Secciones para el tribunal de Administración y Dirección de Empresas:

- 1 - Introducción
- 2 - Competencias, Normativa y Plan de Trabajo
- 4 - Marco teórico
 - 4.1 - Marco teórico sobre empresas zombis
- 5 - Recursos
 - 5.1 - Gestión de datos
- 6 - Desarrollo CRISP-DM
 - 6.2 - Comprensión de los datos
 - 6.3 - Preparación de los datos
 - 6.5 - Evaluación
- 7 - Conclusiones
- 8 - Futuras investigaciones

Para el tribunal de Ingeniería Informática, se requiere la lectura completa para entender todas las etapas de la metodología utilizada en este estudio.

Índice de contenidos

1. Introducción	10
2. Competencias, Normativa y Plan de Trabajo	14
2.1. Competencias.....	14
2.2. Legislación y Normativa	14
2.3. Plan de Trabajo	15
3. Metodología	15
3.1. Comprensión del negocio (<i>Business Understanding</i>)	17
3.2. Comprensión de los datos (<i>Data Understanding</i>)	18
3.3. Preparación de los datos (<i>Data Preparation</i>)	18
3.4. Modelado (<i>Modeling</i>)	19
3.5. Evaluación (<i>Evaluation</i>)	19
3.6. Despliegue (<i>Deployment</i>)	20
4. Marco teórico	20
4.1. Marco teórico sobre empresas zombis.....	20
4.1.1. Definiciones	¡Error! Marcador no definido.
4.1.2. Características	22
4.1.3. Identificación.....	26
4.1.4. Criterios para la identificación de pymes zombis en España	30
4.2. Marco teórico sobre inteligencia artificial.....	33
4.2.1. Redes neuronales convolutivas	33
4.2.2. Redes neuronales convolutivas en el ámbito financiero	38
5. Recursos.....	41
5.1. Gestión de datos.....	41
5.2. Lenguaje y entorno de programación	42
5.3. Kaggle.....	43
5.4. Librerías para redes neuronales convolutivas.....	43
5.5. Otras librerías	45
6. Desarrollo CRISP-DM.....	46
6.1. Comprensión del negocio	46
6.2. Comprensión de los datos	48
6.2.1. Identificación de las pymes zombis españolas	50
6.2.2. Variables para la predicción de pymes zombis	53
6.3. Preparación de los datos	55
6.3.1. Selección y tratamiento de los datos faltantes	56
6.3.2. Elaboración de ratios propias	58
6.3.3. Generación de las imágenes en escala de grises.....	62

6.4. Modelado	65
<u>6.4.1. Selección de las técnicas de modelado</u>	66
<u>6.4.2. Generación del plan de prueba</u>	68
<u>6.4.3. Construcción de los modelos.....</u>	69
<u>6.4.4. Evaluación de los modelos</u>	71
<u>6.5. Evaluación.....</u>	76
<u>6.6. Despliegue</u>	77
7. Conclusiones	78
8. Futuras investigaciones	78
9. Bibliografía	80
Anexo.....	90

Índice de tablas

Tabla 1: Estimación inicial del plan de trabajo	15
Tabla 2: Horas reales consumidas	15
Tabla 3: Artículos con mayores citaciones para identificación de zombis	26
Tabla 4: Artículos que han empleado CNN en el ámbito financiero	40
Tabla 5: Librerías redes neuronales	44
Tabla 6: Variables para la identificación de zombis	51
Tabla 7: Identificación de pymes zombis en España	52
Tabla 8: Partidas contables <i>SABI</i>	54
Tabla 9: Variables para predicción de zombis	60
Tabla 10: Estadística descriptiva para pymes zombis	61
Tabla 11: Estadística descriptiva para pymes no zombis	61
Tabla 12: Información sobre <i>datasets</i> generados	65
Tabla 13: Métricas modelos	74
Tabla A1: Abreviaturas empleadas en <i>el</i> anexo	90
Tabla A2: Primeros estudios sobre empresas zombis	92
Tabla A3: Estudios recientes sobre empresas zombis.....	94
Tabla A4: Partidas de las variables de predicción.....	97

Índice de figuras

Figura 1: Niveles de la metodología <i>CRISP-DM</i>	16
Figura 2: Fases del modelo <i>CRISP-DM</i>	17
Figura 3: Campos de la Inteligencia Artificial	33
Figura 4: Modelo matemático de una neurona artificial	34
Figura 5: Perceptrón Multicapa (con varias capas ocultas).....	35
Figura 6: Ejemplo de una CNN.....	36
Figura 7: Operación de convolución	37
Figura 8: Ejemplos de filtros 3x3	37
Figura 9: Arquitectura CNN (típica)	38
Figura 10: Tendencia de librerías para redes neuronales	45
Figura 11: Criterios pymes.....	48
Figura 12: Pymes españolas en 2019	49
Figura 13: Proporción de pymes en hostelería (2019)	50
Figura 14: Estrategia de búsqueda modificada	51
Figura 15: Top 10 partidas <i>SABI</i> con datos <i>n.d.</i>	55
Figura 16: Imágenes con variables iniciales	63
Figura 17: Imágenes con nuevas variables.....	64
Figura 18: Imágenes con el método correlacionado.....	65
Figura 19: Operación de convolución 1x1	66
Figura 20: Módulo <i>Inception</i>	67
Figura 21: Desbalanceo en los datos	69
Figura 22: Pesos de las clases	71
Figura 23: Matriz de confusión binaria	72
Figura 24: Curva AUC-ROC.....	73
Figura 25: Curva AUC-ROC <i>model_dsa1-logreg_1</i>	75
Figura 26: Curva AUC-ROC <i>model_dsa2-logreg_2</i>	76
Figura 27: Curva AUC-ROC <i>model_dsa3-logreg_3</i>	76
Figura A1: Arquitectura GoogLeNet	98

1. Introducción

Una de las primeras definiciones sobre las empresas zombis fue propuesta por el economista Edward J. Kane (1987). Este las definió como, empresas que deberían haber quebrado debido a su baja eficiencia y falta de rentabilidad, pero que lograron sobrevivir gracias al apoyo externo del sector gubernamental o bancario.

De acuerdo con Kane (1987) y otros autores (*e.g.* Chang *et al.*, 2021; Fu *et al.*, 2019; Jiang *et al.*, 2017; Nurmi *et al.*, 2020; Wu *et al.*, 2021), este tipo de empresas logran mantenerse vivas en el mercado gracias al apoyo de los gobiernos en forma de subvenciones, y al apoyo de los bancos (Ahearne & Shinada, 2005; Hoshi, 2006; Jiang *et al.*, 2017), los cuales suelen estar estipulados en forma de créditos bancarios subvencionados (Carreira *et al.*, 2021; Dai *et al.*, 2019; Wu *et al.*, 2021).

En la literatura de las empresas zombis existen evidencias que demuestran que estas reciben un trato preferencial sobre las no zombis en el momento de obtener subsidios gubernamentales y un mejor acceso a los mercados financieros (Shen & Chen, 2017; Wang & Zhu, 2021), lo cual genera graves distorsiones en la asignación de recursos con respecto a las empresas sanas. Dicha preferencia se debe a los efectos negativos que estas podrían desencadenar en la economía de un país si llegasen a quebrar. En este sentido, los dos efectos que destacan son el aumento del desempleo (Adalet McGowan *et al.*, 2018; Dai *et al.*, 2019; Du & Li, 2019; He *et al.*, 2020; Liu *et al.*, 2019; Peek & Rosengren, 2003; Shen & Chen, 2017) y la reducción de los ingresos fiscales del gobierno (Dai *et al.*, 2019; Liu *et al.*, 2019).

Además, es de resaltar que, la existencia de empresas zombis en un sector, en primer lugar, distorsiona los patrones normales de creación y destrucción de empresas en el mismo (Caballero *et al.*, 2008). Esta distorsión deprime la productividad de la industria (Adalet McGowan *et al.*, 2018; Caballero *et al.*, 2008; Carreira *et al.*, 2021) al mantener empresas ineficientes a expensas de nuevos participantes potenciales más productivos. En segundo lugar, generan una distorsión en la asignación de los recursos del sector (Adalet McGowan *et al.*, 2018; Geng *et al.*, 2021; He *et al.*, 2020; Papworth, 2013; Shen & Chen, 2017; Wang & Zhu, 2021; Yu *et al.*, 2021) evitando que dichos recursos (*e.g.*, capital, mano de obra, etc) se redesplieguen en otras actividades o sectores más productivos. Asimismo, es de destacar que un sector que contenga empresas zombis verá disminuida la creación de empleo (Caballero *et al.*, 2008; Hoshi, 2006; Tan *et al.*, 2016); afectando de manera negativa a las empresas sanas del mismo.

En este sentido, Caballero *et al.* (2008) y Adalet McGowan *et al.* (2018) mencionan que el crecimiento de la inversión y el empleo para las empresas no zombis disminuye a medida que aumenta el porcentaje de los zombis en el sector. Además, entre otros posibles efectos de estas empresas ineficientes sobre otras más productivas destacan: la disminución de la tasa de beneficios (Shen & Chen, 2017); el desplazamiento de las oportunidades de crecimiento, por una mayor proporción de préstamos a prestatarios en dificultades y de baja productividad (Storz *et al.*, 2003); la reducción de la voluntad de entrada de los nuevos participantes, el aumento del riesgo de salida de las empresas normales y el debilitamiento de los incentivos a la innovación (Geng *et al.*, 2021); el aumento de los costes de financiación (Yu *et al.*, 2021) y, por último, las mayores restricciones financieras, debido a que atan abundante capital financiero en el mercado (Wang & Zhu, 2021).

Con respecto a las características de estas empresas, varios autores las asocian como empresas con altos niveles de apalancamiento (*e.g.* Blažková & Dvoutě, 2020; Chang *et al.*, 2021; Dai *et al.*, 2019; El Ghoul *et al.*, 2021; Hoshi, 2006; Lam *et al.*, 2017; Storz *et al.*, 2003). Estas además, generan una rentabilidad económica negativa (Dai *et al.*, 2019; De la Garza Cárdenas *et al.*, 2019; Goto & Wilbur, 2019; Hoshi, 2006; Lam *et al.*, 2017b) y son significativamente menos productivas que las empresas no zombis (Carreira *et al.*, 2021). Por otro lado, cabe destacar que, en la literatura académica revisada, no se ha encontrado un acuerdo entre los distintos autores respecto al tamaño de éstas. Así, Dai *et al.* (2019) y Urionabarrenetxea *et al.* (2018) mencionan que las empresas grandes son más propensas a convertirse en zombis. En la misma línea, Fang *et al.* (2020) y Carreira *et al.* (2021) proponen que, cuando los activos de una organización aumentan también lo hace la probabilidad de ser zombi. De manera contraria, Blažková & Dvoutě (2020) indican que las zombis son empresas pequeñas y sociedades con responsabilidades limitadas. También, Hoshi (2006) y Chang *et al.* (2021) aseveran que, cuanto mayor es el tamaño de la empresa, menor es la probabilidad de ser zombi. Sin embargo, sí que parece haber una opinión consensuada sobre la edad de estas empresas, de manera que cuanto mayor es su edad mayor es la probabilidad de ser considerada zombi (*e.g.* Chang *et al.*, 2021; Dai *et al.*, 2019; Lam *et al.*, 2017).

Asimismo, varios estudios realizados sobre estas empresas en China demuestran que, las empresas estatales son más propensas de convertirse en zombi (*e.g.* Chang *et al.*, 2021; Dai *et al.*, 2019; Lam *et al.*, 2017; Liu *et al.*, 2019), caracterizándose por una menor capacidad productiva (Liu *et al.*, 2019; Shen & Chen, 2017). Esta situación puede venir derivada por una

particularidad de dicho país debido a que las empresas estatales, generalmente guardan mejores relaciones con el gobierno (Wang & Zhu, 2021).

Como se puede observar, son muchos los autores que han investigado sobre las empresas zombis, sin embargo muy pocos de ellos se han centrado de manera exclusiva en su formato pyme. Entre ellos, Imai (2016) analiza hasta qué punto existían las zombis entre las pymes japonesas, proponiendo para ello una medida de identificación basada en los criterios de Caballero *et al* (2008) y Fukuda & Nakamura (2011). Además, examina los comportamientos de endeudamiento e inversión de estas, llegando a la conclusión de que la disminución del valor de la garantía de los préstamos (tierras de la empresa), no condujeron a una contracción crediticia entre las pymes zombis, sino que dieron lugar al fenómeno de los préstamos evergreen entre estas empresas. Así, dichos préstamos lograron incrementar la inversión de las zombis, pero las cuales no fueron necesariamente productivas. Posteriormente, en el año 2019, Goto & Wilbur también analizaron la situación zombi entre las pymes japonesas. Estos autores llegaron a la conclusión de que, existían muchas empresas zombis entre las pymes analizadas, aumentando la proporción de éstas a medida que disminuye el tamaño de las empresas; observando, además, que la probabilidad de salida del estado zombi es mayor para las pymes.

En el contexto europeo destaca el estudio realizado por Carreira *et al.* (2021), quienes analizaron la recuperación y salida de las zombis entre las pymes de Portugal, así como los determinantes de estas transiciones. Así, el estudio confirma que las empresas zombis son significativamente menos productivas que las no zombis, que los sectores con mayores porcentajes de zombis tienen menores niveles de productividad agregada y que la reducción y reestructuración de la deuda, son cruciales para mejorar la recuperación de dichas empresas.

Por último, haciendo referencia al contexto español, Urionabarrenetxea *et al.* (2018) identificaron las empresas zombis españolas con mayor riesgo (caracterizadas por tener patrimonio neto negativo), basándose para ello en el índice EZ (índice de zombis extremos). No obstante, otros autores también han puesto en relevancia la existencia de estas empresas en España. Así, Adalet McGowan *et al.* (2018) mencionan que la prevalencia de las empresas zombis y el porcentaje del capital invertido en ellas aumentó durante el periodo del 2007 hasta 2013. A la misma conclusión llegan Storz *et al.* (2003) quienes analizando el periodo desde 2010 hasta 2014, indican que la mayor proporción de empresas zombis se da en Portugal, seguida de España. Sin embargo, no existen estudios que se hayan centrado únicamente en las

pymes españolas, lo cual es una característica diferenciadora de este estudio al emplear este tipo de empresas.

Teniendo en cuenta todo lo mencionado, resulta interesante, como principal objetivo, aportar una herramienta que permita identificar e predecir de manera temprana este tipo de empresas para mitigar los efectos negativos que generan. Para ello y como aportación principal del presente trabajo, se ha utilizado una metodología no empleada hasta ahora en la bibliografía revisada para la predicción de las empresas zombis. La gran mayoría de estudios de la literatura se basan en el empleo de regresiones para la identificación de estas empresas. Sin embargo, no se han encontrado evidencias del uso de la inteligencia artificial para su predicción. Por lo tanto, este estudio va más allá del resto y pretende la predicción de pymes zombis en España, mediante el uso de la inteligencia artificial, en concreto usando el aprendizaje profundo (*Deep Learning*). Además se empleará las técnica de redes neuronales convolutivas (CNN) sobre imágenes que representan diferentes variables formuladas a partir de los estados contables de la empresa. Ya previamente, Hosaka (2019) empleo la misma técnica para la predicción de empresas en bancarrota corporativa, sobre imágenes que representaban las ratios financieras de la empresa, logrando con ella un excelente rendimiento.

De esta manera, los objetivos principales de este estudio son los siguientes: (1) como prueba de concepto, desarrollar y analizar si las redes neuronales convolutivas permiten la identificación e predicción temprana de estas empresas y (2) aportar, a través de la revisión bibliográfica, una mayor luz a la escasa literatura existente sobre pymes zombis en España.

Finalmente, el presente documento se organiza de la siguiente manera. En la sección 2 se comentan las competencias, la legislación y normativa, y el plan de trabajo a seguir para este proyecto. La sección 3 recoge la metodología empleada en el trabajo. En la sección 4 se aborda el marco teórico, tanto sobre las empresas zombis como sobre la inteligencia artificial (redes neuronales). Posteriormente, en la sección 5 se mencionan los recursos empleados para la elaboración del presente trabajo. En la sección 6 se procede con el desarrollo de las distintas fases de la metodología seleccionada. Y por último, la sección 7 y 8 se recogen las conclusiones y futuras investigaciones a realizar, respectivamente.

2. Competencias, Normativa y Plan de Trabajo

2.1. Competencias

El presente estudio consiste en un trabajo final de la doble titulación en Ingeniería Informática y Administración y Dirección de Empresas de la Universidad de Las Palmas de Gran Canaria. Esto supone que se deben cumplir con unas competencias básicas adquiridas a lo largo de ambas carreras. Es por ello, que a continuación se enumeran siguiendo un formato numérico las competencias cubiertas por este trabajo para ambos grados.

De este modo, para el grado en Administración y Dirección de Empresas se consideran alcanzadas las siguientes competencias **genéricas**: CG1, CG2, CG3, CG5, CG6, CG7, CG9, CG10, CG11, CG20, CG23 y CG27, y **específicas**: CE1, CE8, CE10 y CE15. Así, dichas competencias pueden ser localizadas en la página web de la Facultad de Economía, Empresa y Turismo (FEET, 2021).

Por otra parte, para el grado en Ingeniería Informática se consideran alcanzadas las siguientes competencias **generales**: G1, G2, G3 y G5, **formación básica**: FB01, FB03, FB05 y FB06, **comunes**: CII07 y CII015, y **específicas**: CP05 y TI01. Dichas competencias pueden ser localizadas en la página web de la Universidad de Las Palmas de Gran Canaria (ULPGC, 2021b).

2.2. Legislación y Normativa

Para la elaboración de la memoria del presente trabajo se han seguido las directrices establecidas bajo los siguientes documentos oficiales:

- *Reglamento General de los Trabajos de Fin de Grado y de Fin de Máster de las Titulaciones Oficiales de la Universidad de Las Palmas de Gran Canaria* (ULPGC, 2021a).
- *Reglamento para la realización y evaluación de Trabajos de Fin de Título* de la Facultad de Economía, Empresa y Turismo (ULPGC, 2017).
- *Normativa de desarrollo de las especificidades de los Trabajos de Fin de Título* de la Escuela de Ingeniería Informática (ULPGC, 2020).

Además, la utilización de los datos contables en este proyecto se ampara bajo el *Real Decreto Legislativo 1/2010, de 2 de julio, por el que se aprueba el texto refundido de la Ley de Sociedades de Capital* (GOBIERNO DE ESPAÑA, 2021a), que en su artículo 281 sobre la

publicidad del depósito expone que “cualquier persona podrá obtener información del Registro Mercantil de todos los documentos depositados”.

Por último, el software desarrollado en el presente estudio se ampara bajo el *Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia* (GOBIERNO DE ESPAÑA, 2021b).

2.3. Plan de Trabajo

Para la realización de cualquier proyecto, sea del índole que sea, es necesario disponer previamente de una planificación temporal inicial. Así, esta debe reflejar cómo se distribuyen las horas iniciales estimadas entre las distintas fases del proyecto. De este modo, la [Tabla 1](#) representa el plan de trabajo previsto para el presente proyecto.

Tabla 1: Estimación inicial del plan de trabajo

Fases	Duración estimada (horas): ADE	Duración estimada (horas): II
Comprensión del negocio	80	20
Comprensión de los datos	5	5
Preparación de los datos	5	5
Modelado	20	80
Evaluación	10	10
Despliegue	30	30
Total	150	150

Fuente: Elaboración propia del autor

Sin embargo, a pesar de una planificación inicial rigurosa finalmente se han necesitado aproximadamente 25 horas adicionales para la finalización del proyecto ([Tabla 2](#)). Esto se debe al tiempo extra consumido en las fases de comprensión y preparación de los datos, así como en el despliegue final del proyecto.

Tabla 2: Horas reales consumidas

Fases	Duración estimada (horas): ADE	Duración estimada (horas): II
Comprensión del negocio	80	20
Comprensión de los datos	5	5 + 10
Preparación de los datos	5	5 + 10
Modelado	20	80
Evaluación	10	10
Despliegue	30 + 2,5	30 + 2,5
Total	152,5	172,5

Fuente: Elaboración propia del autor

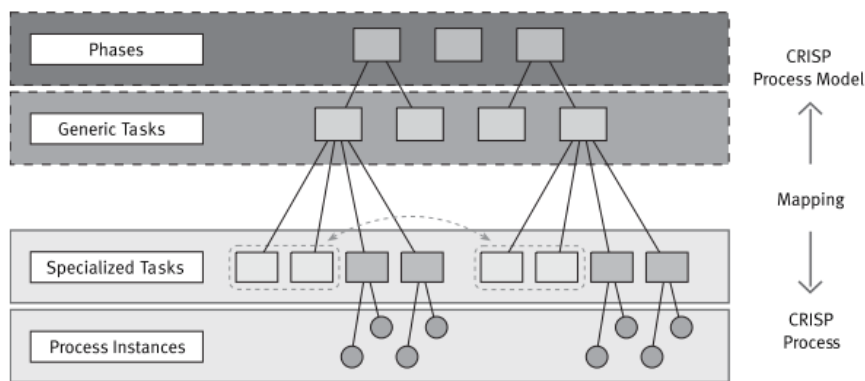
3. Metodología

La metodología que se emplea en este trabajo es la referida como *CRISP-DM* (*Cross Industry Standard Process for Data Mining*), la cual consiste en un modelo compuesto por seis fases que

describen de forma natural el ciclo de vida de un proyecto de ciencia de datos (Data Science Process Alliance, 2021).

Sin embargo, antes de proceder con la descripción de dichas fases, es esencial mencionar los cuatro niveles en los que se divide esta metodología. Así, *CRISP-DM* se describe en términos de un modelo de proceso jerárquico, que consiste en conjuntos de tareas descritas en cuatro niveles de abstracción (Chapman *et al.*, 2000). Estos niveles, de general a específico, son: fases, tareas genéricas, tareas especializadas e instancias de procesos ([Figura 1](#)).

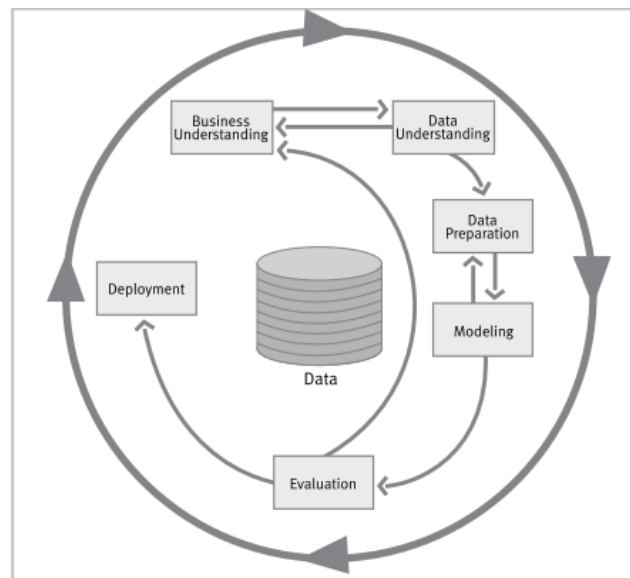
Figura 1: Niveles de la metodología *CRISP-DM*



Fuente: Extraído de Chapman *et al.* (2000)

Con respecto a las distintas fases que componen la metodología a utilizar, estas se encuentran recogidas en la [Figura 2](#). Así, tal y como mencionan Chapman *et al.* (2000), dicha secuencia de fases no es rígida, sino que se puede avanzar y retroceder en ellas. Además, el resultado de cada fase determina que fase, o tarea en concreto de una fase, debe realizarse como la siguiente. Finalmente, el círculo exterior de la [Figura 2](#), simboliza la naturaleza cíclica de un proyecto de minería de datos. Es decir, este no finaliza una vez que se implanta una solución (Chapman *et al.*, 2000).

Figura 2: Fases del modelo CRISP-DM



Fuente: Extraído de Chapman *et al.* (2000)

A continuación, se procede con la explicación de las distintas fases de la metodología *CRISP-DM*, tal y como se plantea en la *Guía oficial de la metodología CRISP-DM* (Chapman *et al.*, 2000).

3.1. Comprensión del negocio (*Business Understanding*)

Esta fase inicial se centra en la comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial. Posteriormente, dicho conocimiento se convierte en un problema de minería de datos y en un plan preliminar diseñado para alcanzar los objetivos propuestos. De este modo, esta fase se compone de las siguientes tareas:

- **Determinación de los objetivos del proyecto:** se plantean los objetivos que el proyecto pretende alcanzar, además de establecer los criterios que miden el éxito de este.
- **Evaluación de la situación actual:** se realiza una investigación sobre los recursos disponibles (*e.g.* software, hardware, datos, etc), limitaciones, suposiciones y otros factores que deben tenerse en cuenta para determinar los objetivos de análisis de datos y el plan de proyecto.
- **Determinación de los objetivos de la minería de datos:** además de plantear los objetivos del proyecto, también se definen estos desde un punto de vista de la minería de datos.
- **Elaboración del plan del proyecto:** se describe el plan previsto para alcanzar los objetivos de la minería de datos y, por tanto, lograr los objetivos del proyecto. Dicho

plan especifica los pasos a realizar durante el resto del proyecto, incluyendo la selección inicial de herramientas y técnicas.

3.2. Comprensión de los datos (*Data Understanding*)

Esta fase se enfoca en identificar, recopilar y analizar los conjuntos de datos que pueden ayudar a lograr los objetivos del proyecto. Así, también se compone de cuatro tareas:

- **Recopilación de datos iniciales:** se accede a los datos que figuran en los recursos anteriormente definidos y (si es necesario) se cargan estos en alguna herramienta de análisis de datos que se emplee en el proyecto.
- **Descripción de los datos:** se describen los datos que se han obtenido en la tarea anterior. Es decir, se especifica el formato y la cantidad de estos, la identidad de sus campos y cualquier otra característica que se haya descubierto. Además, se evalúa si los datos obtenidos satisfacen los requisitos pertinentes.
- **Exploración de los datos:** se procede con una exploración profunda de los datos, mediante su visualización e identificación de las relaciones entre estos a través de técnicas básicas de estadística descriptiva.
- **Verificación de la calidad de los datos:** se examina la calidad de estos planteando preguntas como: ¿los datos son completos? ¿los datos son correctos o contienen errores y, si los hay, qué tan comunes son? ¿faltan elementos en los datos? En caso afirmativo, ¿cómo se representan, dónde se producen y qué frecuencia tienen?

3.3. Preparación de los datos (*Data Preparation*)

Esta fase abarca todas las actividades necesarias para la construcción de los *datasets* que se suministrarán como entradas a las herramientas de modelización. Así, entre sus principales tareas destacan:

- **Selección de los datos:** se determinan los datos a incluir e excluir, así como el motivo de dichas decisiones.
- **Limpieza de los datos:** se eleva la calidad de los datos al nivel requerido por las técnicas de análisis seleccionadas. Ello implica, entre otros, la selección de subconjuntos limpios de datos, inserción de valores predeterminados adecuados o la estimación de los datos faltantes.

- **Construcción de los datos:** se derivan atributos o registros completamente nuevos de los datos originales. Además, en esta tarea también se pueden transformar los valores de los atributos existentes en el *dataset*.
- **Integración de los datos:** son métodos gracias a los cuales se combina la información de distintas tablas para la creación de nuevos registros.
- **Formateo de los datos:** consisten en modificaciones sintácticas realizadas a los datos que no cambian su significado, pero que pueden ser requeridas por las herramientas de modelado.

3.4. Modelado (*Modeling*)

En esta fase se seleccionan y aplican diversas técnicas de modelización, así como la calibración de sus parámetros hasta alcanzar los valores óptimos. Además, algunas de estas técnicas requieren un formato específico de los datos, por lo tanto, es bastante habitual retroceder a la fase de preparación de estos. Las tareas que componen esta fase son:

- **Selección de las técnicas de modelado:** documentación de la técnica de modelado a utilizar en el proyecto (*e.g.* regresiones, redes neuronales, etc.)
- **Generación del plan de prueba:** descripción del plan previsto para el entrenamiento, prueba y evaluación de los modelos. Un aspecto crítico de este plan, es determinar cómo se divide el *dataset* en conjuntos de entrenamiento, validación y test.
- **Construcción del modelo:** consiste en la ejecución de la técnica de modelado seleccionada sobre el *dataset* preparado de las fases anteriores, para generar uno o más modelos. Asimismo, se enumeran y justifican la selección de los parámetros de estos.
- **Evaluación del modelo:** generalmente múltiples modelos compiten entre ellos. Es por ello, que en esta tarea se enumeran las cualidades de los distintos modelos generados (*e.g.* en términos de exactitud, precisión, etc), para posteriormente establecer una clasificación entre estos. Asimismo, también se revisan los ajustes de parámetros y se repiten las tareas de construcción y evaluación (si son necesarios) hasta encontrar el mejor modelo posible.

3.5. Evaluación (*Evaluation*)

Antes de proceder con la implantación definitiva del modelo, es importante revisar los pasos seguidos en su creación y estar seguro de que cumple con los objetivos iniciales del proyecto.

Además, es clave determinar si algún aspecto del proyecto no se ha tenido en cuenta lo suficiente. De esta manera, las tareas que componen la presente fase son:

- **Evaluación de los resultados:** la evaluación de la fase anterior se refiere a los aspectos técnicos del modelo, como su precisión y generalidad. Sin embargo, esta tarea evalúa el grado en que un modelo cumple con los objetivos del proyecto. Así, al final de esta se aprueban uno o varios modelos que satisfagan los criterios de éxito del proyecto.
- **Proceso de revisión:** se revisa el trabajo realizado. ¿Se ha pasado algo por alto? ¿Se han ejecutado correctamente todos los pasos? Finalmente, se resumen los resultados y se corrige lo que sea necesario.
- **Determinación de los siguientes pasos:** en función de los resultados de evaluación y del proceso de revisión, se determina si se procede con la implantación, una nueva iteración o el inicio de un proyecto nuevo.

3.6. Despliegue (*Deployment*)

La construcción del modelo no suele ser el final del proyecto. Así, dependiendo de los requisitos, la fase de despliegue puede ser tan sencilla como generar un informe o tan compleja como implementar este proceso de manera repetida. Por lo tanto, esta fase final se compone de las siguientes cuatro tareas (Data Science Process Alliance, 2021):

- **Plan de despliegue:** se desarrolla y documenta el plan de despliegue a seguir para el modelo seleccionado.
- **Planificación de la supervisión y mantenimiento:** se elabora el plan de seguimiento y mantenimiento para el modelo desplegado.
- **Elaboración del informe final:** consiste en la redacción de un informe final del proyecto. Así, dicho informe puede ser desde un resumen del proyecto, hasta una presentación en la que se muestran los resultados obtenidos.
- **Revisión del proyecto:** se realiza una retrospectiva del proyecto sobre aspectos que han salido bien, podrían haber salido mejor y mejoras de cara al futuro.

4. Marco teórico

4.1. Marco teórico sobre empresas zombis

Esta sección se divide en cuatro partes. En primer lugar, se revisa la literatura de las empresas zombis con el fin de determinar las distintas definiciones propuestas por los autores. A

continuación, se citan las características, tanto financieras y económicas como organizativas, de estas empresas. Posteriormente, se detallan los criterios más empleados para su detección. Y por último, se explican los criterios seguidos en este estudio para la identificación de las pymes zombis españolas.

4.1.1. Concepto

Las empresas zombis han existido en las principales economías del mundo, tales como: Japón (Ahearne & Shinada, 2005; Caballero *et al.*, 2008; Goto & Wilbur, 2019; Hoshi, 2006), China (Dai *et al.*, 2019; Dong *et al.*, 2020; Fang *et al.*, 2020; He *et al.*, 2020; Lam *et al.*, 2017b; Liu *et al.*, 2019; Tan *et al.*, 2016), Reino Unido (Papworth, 2013), España (Urionabarrenetxea *et al.*, 2018), y en otros países europeos como: República Checa (Blažková & Dvoutě, 2020), Portugal (Carreira *et al.*, 2021) y Finlandia (Nurmi *et al.*, 2020).

En la revisión de la literatura centrada en esta temática realizada, no se ha encontrado una única definición de consenso que las identifique. En este sentido, que son varios los autores (*e.g.* Caballero *et al.*, 2008; Hoshi, 2006; Kane, 1987; Tan *et al.*, 2016; Urionabarrenetxea *et al.*, 2018), que han aportado definiciones propias. Sin embargo, es de resaltar la definición propuesta por el economista Kane (1987), al ser uno de los primeros autores en referirse a ellas: son empresas que irían a la bancarrota debido a sus bajos ingresos pero que logran sobrevivir con el apoyo externo de los gobiernos o del sector financiero.

Por su parte, es de destacar el caso de las zombis en Japón, las cuales han sido ampliamente señaladas como un factor clave en el lento crecimiento económico del país, en el período comúnmente conocido como las "décadas perdidas" (1990) (Ahearne & Shinada, 2005; Caballero *et al.*, 2008; Peek & Rosengren, 2003). Es por dicha razón que, varios estudios en este país se han centrado en identificar y otorgar una definición sobre éstas. Así, Caballero *et al.* (2008) las delimita como, empresas de baja productividad que saldrían del mercado en ausencia de la ayuda de los bancos. En líneas similares, Hoshi (2006) las define como, empresas insolventes y con poca esperanzas de recuperación, pero que evitan el fracaso gracias a una gran variedad de apoyo de sus acreedores. No obstante, Japón no es el único país que se ha visto comprometido por la presencia de las zombis, entre ellos también se encuentra China. Varios estudios de dicho país las definen como, empresas insolventes que permanecen en funcionamiento (*e.g.* Dai *et al.*, 2019; Fang *et al.*, 2020; He *et al.*, 2020; Tan *et al.*, 2016), gracias al apoyo financiero externo que reciben, o como empresas ineficientes que dependen de transfusiones externas y carecen de su propia rentabilidad (Dong *et al.*, 2020; Liu *et al.*, 2019).

Por otra parte, si nos enfocamos en el contexto europeo de estas, destaca el estudio de Papworth (2013) en Reino Unido, quien las define como: “*empresas que sólo son capaces de pagar los intereses de sus préstamos, pero tienen pocas esperanzas de pagar el capital, y que sólo son capaces de continuar [...] por la tolerancia de las instituciones financieras*”. Otra definición en el mismo contexto, tal vez más apartada del resto de los autores, es la aportada por el Banco de Finlandia (2020) al establecer que las zombis no son empresas verdaderamente angustiadas, sino empresas con ingresos temporalmente bajos en relación a los pagos de los intereses.

Finalmente, centrándonos en el contexto del presente estudio, es decir España, es de vital importancia hacer referencia al trabajo llevado a cabo por los autores Urionabarrenetxea *et al.* (2018) por ser uno de los pocos trabajos detectados sobre zombis en España. Dicho estudio, en primer lugar, delimita a estas como empresas poco competitivas y en gran medida inviables. Además expone que, aquellas empresas con capital negativo y que siguen haciendo negocios a pesar de haber perdido todo su capital, se pueden considerar como un tipo especial de zombis, en concreto, zombis de mayor riesgo.

Como se puede observar, en la literatura académica revisada de las zombis existen numerosas definiciones propuestas por distintos autores. Sin embargo, los dos países que más se han visto comprometidos por la presencia de estas empresas y por ende más hincapié han hecho en su definición son: Japón e China. Es por ello, que el presente trabajo basándose en numerosos estudios de tales países, concretamente, Ahearne & Shinada (2005); Caballero *et al.* (2008); Dai *et al.* (2019); Fang *et al.* (2020); Goto & Wilbur (2019); He *et al.* (2020); Hoshi (2006); Liu *et al.* (2019) y Tan *et al.* (2016), define a las zombis como: **empresas insolventes e ineficientes que permanecen en funcionamiento gracias al apoyo financiero externo que reciben, bien sea en forma de subvenciones del gobierno o créditos bancarios subvencionados.**

4.1.2. Características

La primera característica que destaca sobre las zombis en las definiciones anteriores, es su supervivencia en el mercado gracias a los apoyos externos que reciben. Entre estos apoyos, varios autores (*e.g.* Chang *et al.*, 2021; Fu *et al.*, 2019; Jiang *et al.*, 2017; Nurmi *et al.*, 2020; Wu *et al.*, 2021) señalan que, los subsidios gubernamentales juegan un papel fundamental en el sustento de dichas empresas. Por otra parte, Ahearne & Shinada (2005) indican que, estas empresas ineficientes se mantienen en parte gracias al apoyo financiero de los bancos. Así, en líneas similares a estos, Hoshi (2006) concluye que los bancos juegan un papel importante en

el apoyo de éstas. Además, autores como Carreira *et al.* (2021), Dai *et al.* (2019) y Wu *et al.* (2021) señalan que dichas ayudas por parte de los bancos suelen formalizarse en forma de créditos bancarios subvencionados.

Sin embargo, dicha característica no queda del todo clara, es decir, ¿porque los gobiernos o los bancos ofrecen apoyos a estas empresas en problemas para que sigan existiendo en el mercado? La respuesta a tal pregunta se encuentra en dos principales razones identificadas en su literatura. La primera de ellas está relacionada con las políticas del gobierno de un país. Así, la resolución de las zombis consiste en un proceso doloroso y costoso para la economía de cualquier país (Dai *et al.*, 2019), pudiendo generar consecuencias como: el aumento del desempleo (Adalet McGowan *et al.*, 2018; Dai *et al.*, 2019; Du & Li, 2019; He *et al.*, 2020; Liu *et al.*, 2019; Peek & Rosengren, 2003; Shen & Chen, 2017) especialmente para el caso de las empresas grandes o empresas que cuentan con una gran intensidad de mano de obra o en los tiempos de crisis; la reducción de los ingresos fiscales del gobierno (Dai *et al.*, 2019; Liu *et al.*, 2019); distorsión de la estabilidad social y política (Du & Li, 2019; He *et al.*, 2020; Liu *et al.*, 2019; Peek & Rosengren, 2003) y efectos negativos sobre el desarrollo económico de una región (Liu *et al.*, 2019). Mientras que la segunda, tiene que ver con las intenciones de los bancos: para cubrir sus préstamos morosos (Dai *et al.*, 2019; Peek & Rosengren, 2003); mantener las ratios de capital que les son exigidas (Peek & Rosengren, 2003) y para continuar sus relaciones con determinadas empresas u organizaciones (Adalet McGowan *et al.*, 2018).

En este sentido, numerosos estudios han investigado la facilidad con la que estas empresas acceden a dichos apoyos externos. Así, Shen & Chen (2017) llegan a la conclusión de que las empresas zombis tienen un mejor acceso a los mercados financieros, reciben más subsidios gubernamentales que las empresas saludables y disfrutan de tasas impositivas más bajas. En líneas similares, Wang & Zhu (2021) mencionan que las zombis pueden conseguir con mayor facilidad la protección del gobierno y capital crediticio a un coste bajo. Dicha circunstancia genera en el sector que posea empresas zombis, una asignación ineficiente de los recursos (Adalet McGowan *et al.*, 2018; He *et al.*, 2020; Shen & Chen, 2017; Wang & Zhu, 2021; Yu *et al.*, 2021). Además, la presencia de zombis hace que las empresas sanas del sector se vean afectadas en cuestiones de inversión y empleo (Adalet McGowan *et al.*, 2018; Caballero *et al.*, 2008); beneficios (Shen & Chen, 2017); oportunidades de crecimiento (Storz *et al.*, 2003); innovación (Geng *et al.*, 2021); costes de financiación (Yu *et al.*, 2021) y restricciones financieras (Wang & Zhu, 2021).

Con respecto a las características financieras de estas, Papworth (2013) las asocia con empresas altamente endeudadas que generan ingresos mínimos para pagar los intereses de sus préstamos, pero no la cuantía principal de estos. En este sentido, varios estudios de la literatura revisada coinciden en que las zombis presentan un alto grado de apalancamiento (*e.g.* Blažková & Dvouletý, 2020; El Ghouli *et al.*, 2021; Hoshi, 2006; Lam *et al.*, 2017a; Storz *et al.*, 2003). De igual manera, Chang *et al.* (2021) y Dai *et al.* (2019) proponen que cuanto mayor es el apalancamiento de la empresa, mayor es la probabilidad de convertirse en zombi. Otra medida financiera que está relacionada con las mencionadas anteriormente es la rentabilidad financiera (ROE) de la empresa. De este modo, He *et al.* (2020) señalan que una empresa con menor ROE (debido a sus menores beneficios netos) presenta mayores probabilidades de convertirse en una zombi.

En lo que respecta a las características económicas de éstas empresas, varios autores (*e.g.* Dai *et al.*, 2019; De la Garza Cárdenas *et al.*, 2019; Goto & Wilbur, 2019; Hoshi, 2006; Lam *et al.*, 2017) coinciden en la existencia de una relación negativa, entre la rentabilidad económica (ROA) y la probabilidad de ser una empresa zombi. Así, Blažková & Dvouletý (2020) observan que antes de que una empresa se convierta en zombi sus beneficios disminuyen y se acumulan pérdidas del año anterior. Además, estas empresas son significativamente menos productivas que las no zombis (Carreira *et al.*, 2021) y poseen una capacidad productiva inferior (Liu *et al.*, 2019; Shen & Chen, 2017).

Otra variable abordada en la literatura de estas empresas es la cuota de mercado. De la Garza Cárdenas *et al.* (2019) llegan a la conclusión de que, esta ejerce una relación positiva en la aparición de las empresas zombis, ya que a medida que una empresa acapara mayor cuota de mercado, tiene mayor probabilidad de convertirse en zombi. El fundamento de esto según los autores se debe a que, una alta concentración de mercado, que representa un mercado poco competido, hace que las empresas estancadas en él se mantengan, ya que no son obligadas a realizar cambios en su estrategia empresarial y no se ven forzadas a plantear la posibilidad de abandonar el mercado.

Con respecto a las características organizativas de las zombis, autores como Dai *et al.* (2019), Chang *et al.* (2021) y Lam *et al.* (2017) señalan que la edad de una empresa se asocia de manera positiva con la probabilidad de ser zombi. Coincidiendo en parte con estos, Blažková & Dvouletý (2020) encuentran un patrón no lineal invertido en forma de U, advirtiendo que las zombis eran probablemente empresas de mediana edad.

En cuanto al tamaño de estas, no se ha encontrado unanimidad de criterio entre los distintos autores de la literatura académica revisada. Por un lado, Urionabarrenetxea *et al.* (2018) indican que las empresas grandes presentan un mayor riesgo de convertirse en zombi. En este sentido, los autores mencionan que dichas empresas tienden a ser respaldadas por los gobiernos debido al impacto negativo que su fracaso podría tener en la economía en su conjunto (Urionabarrenetxea *et al.*, 2018). Además, estas también cuentan con el apoyo de la sociedad, ya que se da por sentado que una empresa solo puede crecer en grande si funciona bien (Urionabarrenetxea *et al.*, 2018). A una conclusión similar llegan los autores Dai *et al.* (2019), quienes señalan que las empresas grandes son más propensas a convertirse en zombi. Además, teniendo en cuenta que los activos de una empresa constituyen una forma alternativa para medir su tamaño, Carreira *et al.* (2021) y Fang *et al.* (2020) señalan que estos están positivamente relacionados con la probabilidad de ser zombi.

Por otro lado, y de manera totalmente contraria a los anteriores, Chang *et al.* (2021) señalan que cuanto mayor es el tamaño de una empresa menor es la probabilidad de que se convierta en una zombi. A la misma conclusión llega Hoshi (2006), quien además aporta que entre las empresas pequeñas, es más probable que las empresas relativamente grandes se conviertan en zombis. En este aspecto y de manera más directa, Blažková & Dvouletý (2020) señalan que las zombis son sociedades pequeñas y con responsabilidades limitadas.

Entre otras posibles características que han sido estudiadas sobre estas empresas, podemos destacar su relación negativa con la responsabilidad ambiental corporativa (Han *et al.*, 2019) y con la responsabilidad social corporativa (Han *et al.*, 2020). En relación a esta última, los autores concluyen que las zombis se asocian negativamente con la publicación de informes de RSC (responsabilidad social corporativa) y entre las empresas que publican dichos informes, las zombis cumplen (CSRF) en menor medida con sus promesas (CSRD).

Una última particularidad a señalar de estas es su tipo de propiedad. Así, varios estudios chinos (*e.g.* Chang *et al.*, 2021; Dai *et al.*, 2019; Lam *et al.*, 2017; Liu *et al.*, 2019), exponen que las empresas estatales son más propensas a convertirse en zombis. No obstante, esto se debe a las estrechas relaciones de dichas empresas con el gobierno chino (Wang & Zhu, 2021), las cuales le facilitan el apoyo del gobierno y de las entidades financieras.

En resumen, se puede destacar que las empresas zombis **dependen de apoyos externos**, ya sean por parte del gobierno o bancos, para lograr sobrevivir en el mercado. Además, en la

mayoría de los casos, se tratan de **empresas endeudadas**, es decir **apalancadas**, con **rentabilidades económicas (ROA) bajas** y normalmente **empresas de mediana-larga edad**.

4.1.3. Identificación

En la revisión bibliográfica sobre las empresas zombis, son varios los estudios (*e.g.* Adalet McGowan *et al.*, 2018; Caballero *et al.*, 2008; Fukuda & Nakamura, 2011; Imai, 2016; Schivardi *et al.*, 2017; Storz *et al.*, 2003; Urionabarrenetxea *et al.*, 2018; Zhu & Fan, 2016) que han propuesto diferentes criterios para su identificación. Pero, sin lugar a duda, el criterio más citado (véase [Tabla 3](#)) es el de Caballero, Hoshi y Kashyap (2008). De esta manera, en la literatura existen estudios que emplean de manera idéntica dicho criterio (*e.g.* Fang *et al.*, 2020; Jiang *et al.*, 2017) o bien una modificación de este, para la identificación de las zombis (*e.g.* Fukuda & Nakamura, 2011; Hoshi, 2006).

Tabla 3: Artículos con mayores citaciones para identificación de zombis

Nombre del artículo	Autor/es	Año	Revista	Nº de citas
Zombie lending and depressed restructuring in Japan	Caballero, Ricardo J. Hoshi, Takeo Kashyap, Anil K.	2008	American Economic Review	Scopus: 408
Why did 'zombie' firms recover in Japan?	Fukuda, Shin ichi Nakamura, Jun ichi	2011	World Economy	Wiley Online Library: 76
A panel study of zombie SMEs in Japan: Identification, borrowing and investment behavior	Imai, Kentaro	2016	Journal of the Japanese and International Economies	Scopus: 28
Quantity and Characteristics of Zombie Firms in China	Zhu, He Fan, H E	2016	Journal of Beijing Technology & Business University	Scopus: 6
The walking dead? Zombie firms and productivity performance in OECD countries	Adalet McGowan, Müge Andrews, Dan Millot, Valentine	2018	Economic Policy	Web of Science: 21
Living with zombie companies: Do we know where the threat lies?	Urionabarrenetxea, Sara Garcia-Merino, Jose Domingo San-Jose, Leire Retolaza, Jose Luis	2018	European Management Journal	Science Direct: 8

Así, Caballero *et al.* (2008) (en adelante CHK), identifican como zombis a aquellas empresas cuyos pagos de intereses (R) son inferiores al pago mínimo de intereses (R^*) que se le exige a la empresa por su deuda. Dicho límite inferior de intereses se estima, aplicando los tipos que el prestatario más solvente habría pagado por su deuda al principio del periodo. Sobre este aspecto, los autores mencionan que las empresas en apuros deben haber recibido un alivio sustancial de los intereses y, por lo tanto, sus pagos de intereses deben haber sido inferiores a los de las

empresas sanas. Es por ello, que para proceder con la identificación de zombies, se calcula en primer lugar el límite inferior hipotético para los pagos de intereses, de la siguiente manera:

$$R_{i,t}^* = r_{s,t-1} BS_{i,t-1} + \left(\frac{1}{5} \sum_{j=1}^5 r_{l,t-j} \right) BL_{i,t-1} + r_{cb_{\min \text{ over last 5 years},t}} Bonds_{i,t-1} \quad (1)$$

Donde $R_{i,t}^*$ es el pago mínimo de intereses exigido a cada empresa en el año t . $BS_{i,t}$, $BL_{i,t}$ y $Bonds_{i,t}$ son los préstamos bancarios a corto plazo (menos de un año), préstamos bancarios a largo plazo (más de un año) y bonos en circulación (incluyendo los bonos convertibles y bonos con garantía), respectivamente, de la empresa i al final del año t . Y, $r_{s,t}$, $r_{l,t}$ y $r_{cb_{\min \text{ over last 5 years},t}}$ son el tipo medio de interés preferente a corto plazo en el año t , el tipo medio de interés preferente a largo plazo en el año t y el tipo de cupón mínimo observado de cualquier bono corporativo convertible emitido en los últimos cinco años antes de t .

Posteriormente, para poder clasificar a las empresas como zombies o no zombies, se comparan los pagos de intereses realizados por estas con el hipotético límite calculado anteriormente. Y a continuación, se divide esta diferencia por el importe del endeudamiento total al principio del periodo ($B_{i,t-1}$):

$$x_{i,t} = \frac{R_{i,t} - R_{i,t}^*}{B_{i,t-1}} \quad (2)$$

Donde $B_{i,t-1} = BS_{i,t-1} + BL_{i,t-1} + Bonds_{i,t-1} + CP_{i,t-1}$. Y donde, $CP_{i,t-1}$ es el importe de los efectos comerciales en circulación de la empresa i al principio del período t . De esta manera, si $x_{i,t}$ es menor que 0, entonces la empresa se considera como una zombie, en el año t .

Como puede observarse, CHK identifican a las zombies basándose en un criterio simple. Sin embargo, varios autores (*e.g.* Carreira *et al.*, 2021; Fukuda & Nakamura, 2011; Han *et al.*, 2020; Hoshi, 2006) han señalado que este se trata de una medida ruidosa, con errores de tipo uno y dos. Así, en primer lugar, dicho criterio puede identificar erróneamente a las empresas sanas como zombies ya que, las mejores empresas suelen disfrutar de tipos de intereses inferiores a los tipos preferentes empleados por CHK (Shen & Chen, 2017). Además, el criterio también puede abarcar a aquellas empresas sanas que decidan amortizar voluntariamente una parte importante de su deuda (Hoshi, 2006).

En segundo lugar, existe la posibilidad de que el criterio no pueda identificar aquellas zombies que reciben apoyo de otras formas que no sean la concesión de intereses o la condonación de la deuda. Por ejemplo, algunos bancos pueden conceder nuevos préstamos (con tipos de interés

no especialmente bajos) a empresas en dificultades para que estas puedan hacer frente a los intereses de los préstamos existentes (*evergreening*) (Hoshi, 2006). Esto permite a dichas empresas evitar, o al menos retrasar, su quiebra. Además, también ayuda a los bancos para evitar o retrasar un nuevo aumento de sus préstamos morosos declarados ya que, de esta forma el banco no está obligado a dotar provisiones por las pérdidas de los mismos, lo que reduciría sus beneficios y, por tanto, su capital (Peek & Rosengren, 2003).

Por tales razones, el criterio de Fukuda & Nakamura (2011) (en adelante FN), se presenta como el segundo con mayor número de citas en la [Tabla 3](#) y procede a mejorar el criterio CHK. Específicamente, éstos autores introducen dos criterios adicionales para la identificación de las zombis: el criterio de rentabilidad y el criterio de los préstamos “*evergreen*”. Así, de acuerdo con su primer criterio, todas aquellas empresas cuyos beneficios antes de intereses e impuestos (EBIT) superen los pagos de intereses hipotéticos sin riesgo de CHK, son excluidas de la categoría de zombis. Al excluir estas empresas rentables se reduce la probabilidad de que el criterio CHK las identifique como zombis. Con respecto a su segundo criterio, aquellas empresas cuyo EBIT sea menor que R^* en el período t , la deuda externa total sea más de la mitad de sus activos totales en el período $t-1$, y los préstamos hayan aumentado en el período t , son categorizadas como zombis en el período t . El hecho de clasificar estas últimas como zombis, reduce la probabilidad de que el criterio CHK identifique erróneamente las empresas poco saludables como no zombis.

Más adelante, Imai (2016) en su investigación sobre las pymes zombis en Japón, plantea un nuevo método para la identificación de estas basándose en los criterios de CHK y FN. Este en su estudio advierte que, el método de FN tampoco es del todo adecuado para la identificación de tales empresas. Puesto que, aquellas empresas sanas que experimenten disminuciones temporales en sus beneficios podrían ser clasificadas como zombis, y viceversa. Es por ello, que Imai (2016) procede a modificar el criterio de rentabilidad de FN (sin alterar el criterio de los préstamos *evergreen*) de la siguiente manera:

$$\sum_{m=0}^T EBIT_{i,t-m} - R_{i,t-m}^* < 0 \quad (3)$$

Para esta ecuación, según indica Imai (2016), el límite de tres años ($T=3$) se considera una norma relativamente estable para identificar a las zombis. De esta manera, todas aquellas empresas que satisfacen la condición dinámica recogida en la [ecuación 3](#) para tres años consecutivos, cumplen con el criterio de rentabilidad modificado de FN.

Por lo tanto el criterio de Imai (2016) cuya metodología es la tercera más citada, indica que, todas aquellas pymes que cumplen con el criterio de rentabilidad modificado de FN y cuyos R son inferiores a R^* , o que puedan obtener un nuevo préstamo, son consideradas como zombies.

Otros criterios citados en la literatura académica revisada, aunque en menor medida que los anteriores, son los de Adalet McGowan *et al.* (2018) y Zhu & Fan, (2016). Por un lado, Adalet McGowan *et al.* (2018) investigan hasta qué punto las zombies estaban reprimiendo los resultados de la productividad laboral en países miembros de la OECD. Para ello, definen a estas como: empresas que tienen diez o más años y su ratio de cobertura de intereses (ICR) es inferior a uno para tres periodos consecutivos. Así, dicha ratio la calculan de la siguiente manera:

$$ICR(3) = \frac{EBIT}{\text{Gastos de intereses}} < 1 \quad (4)$$

Además, los autores señalan que la restricción de la edad se postula para distinguir las verdaderas empresas zombies de las jóvenes start-ups innovadoras. Por otro lado, Zhu & Fan (2016) proponen el criterio de los beneficios reales para identificar a estas empresas ineficientes en China. Según su método, si la diferencia entre el beneficio neto de la empresa y el beneficio no recurrente es inferior a cero, la empresa es categorizada como una zombie.

Por último, centrándonos en el contexto español, es de destacar el trabajo de los autores Urionabarrenetxea *et al.* (2018), los cuales identifican las empresas zombies con mayores riesgos (empresas con patrimonio neto negativo). Para ello, proponen el índice EZ (índice de zombies extremos), que permite la clasificación de estas en función de los riesgos y cambios en el tiempo. Además dicho índice, se formula teniendo en cuenta las cuatro dimensiones del problema de los zombies extremos: extensión, contagio, signos de recuperación e inmediatez. De este modo, con el índice EZ, los autores tratan de explicar la persistencia de las empresas zombies en el tiempo, el efecto que tienen sobre otras empresas sanas, la posibilidad de que se recuperen y la urgencia del problema.

Así, y en resumen de lo visto, podemos concluir que los estudios más destacados para la identificación de las zombies son los citados en la [Tabla 3](#). Además, en la literatura académica revisada, la gran mayoría de los trabajos que proponen la identificación de estas empresas, emplean como criterios los utilizados por CHK, FN y Imai, bien sean en su versión original o una combinación de estos (véanse Tablas [A1](#), [A2](#) y [A3](#)).

4.1.4. Criterios para la identificación de pymes zombis en España

En vista de las definiciones, características y las distintas medidas utilizadas en la literatura de las empresas zombis, esta sección explica los criterios finalmente seleccionados para la identificación de las pymes zombis en España.

De este modo, el presente estudio se basa en el criterio propuesto por Imai (2016) debido a las siguientes razones:

- Consiste en uno de los primeros en plantear y lograr la identificación de zombis en su formato pyme, logro que justamente pretende alcanzar este estudio.
- Corrige los errores de los criterios CHK y FN utilizando un periodo más amplio para la identificación de estas empresas. Así, dicha modificación ayuda al criterio eludir el problema de identificar erróneamente a las empresas sanas como zombis si estas experimentan descensos temporales en sus beneficios, y viceversa (Imai, 2016). Además, este criterio presenta mayores niveles de precisión que los de CHK y FN.
- Por último, consiste en uno de los criterios con mayor número de citas en *Scopus* ([Tabla 3](#)), cuando se realizan búsquedas relacionados con la identificación de las empresas zombis. Adicionalmente, es de destacar que el criterio seleccionado ha sido empleado por varios autores de la literatura de las empresas zombis (*e.g.* Du & Li, 2019; Geng *et al.*, 2021; Goto & Wilbur, 2019; Wu *et al.*, 2021).

Por dichas razones, a continuación se plantean los dos criterios que se utilizarán en este estudio para la identificación de las pymes zombis españolas.

El primer criterio, y el que se considera como principal del estudio, consiste en cumplir las siguientes dos condiciones de Imai (2016), ligeramente modificadas, para que la empresa sea considerada zombi en el periodo t :

- **Condición 1:** $\sum_{m=0}^{T=2} EBIT_{i,t-m} - R_{i,t-m}^* < 0$ (5)
- **Condición 2:** $DRT_{i,t} > DRT_{i,t-1} \text{ O } R_{i,t} < R_{i,t}^*$ (6)

Donde $EBIT_{i,t}$, $R_{i,t}$, $R_{i,t}^*$ y $DRT_{i,t}$ son, los beneficios antes de intereses e impuestos de la empresa i , el tipo de interés efectivo pagado por la empresa, el pago mínimo de intereses exigido a la empresa y la deuda remunerada total de la empresa, respectivamente, al final del periodo t .

Tal y como se puede observar, las [ecuaciones 5](#) y [6](#) se asemejan bastante al criterio original de Imai (2016) explicado en la sección [4.1.3](#). Así, con la primera condición, la cual no es más que

el criterio de rentabilidad de FN adaptado para tres años, se evita el problema de identificar a las empresas sanas como zombies si estas experimentan descensos temporales en sus beneficios y viceversa. Mientras que con la segunda, se logra mantener el criterio de los préstamos “*evergreen*”, para no incurrir en los problemas del criterio CHK mencionados en la misma sección.

De esta manera, en el presente estudio **se identifican como pymes zombies todas aquellas empresas cuyo EBIT sea menor que R^* por tres periodos consecutivos y que satisfagan con al menos una de las reglas de la segunda condición.**

Sin embargo, debida a la limitación de datos en los estados contables de las pymes españolas, no es posible el cálculo de R y R^* , tal y como se procede en la literatura de estas empresas. Es por ello, que el trabajo actual se basa en el empleo de medidas *proxys* para intentar estimar el valor original de dichas variables. Así, en la literatura de las zombies, R consiste en el pago de intereses reales por parte de la empresa, sin embargo, en este estudio se emplea como *proxy* el tipo de interés efectivo pagado por la empresa. Dicho tipo se calcula de la siguiente manera:

$$\frac{\text{Gastos Financieros}_{i,t}}{\text{DRT}_{i,t-1}} \quad (7)$$

Donde $\text{DRT}_{i,t} = \text{DR}_{i,t}^{c/p} + \text{DR}_{i,t}^{l/p}$, y donde, $\text{DR}_{i,t}^{c/p}$ y $\text{DR}_{i,t}^{l/p}$ son la deuda remunerada a corto y largo plazo, respectivamente, a final del periodo t .

Por otro lado, para el cálculo de $R_{i,t}^*$ se utiliza la siguiente ecuación:

$$rs_{t-1} \text{DR}_{i,t-1}^{c/p} + \left(\frac{1}{5} \sum_{j=1}^5 rl_{t-j} \right) \text{DR}_{i,t-1}^{l/p} \quad (8)$$

Donde rs_t y rl_t son, el tipo de interés medio anual a corto y largo plazo, respectivamente, aplicado por las instituciones financieras monetarias residentes en la Unión Económica y Monetaria (UEM) sobre los préstamos y créditos concedidos a las sociedades no financieras. Además, dichos tipos medios se calculan con la ayuda de los tipos mensuales publicados en la página oficial del Banco de España (BANCO DE ESPAÑA, 2021).

Tal y como se puede comprobar, la [ecuación 8](#) consiste en una versión simplificada de la [ecuación 1](#) del criterio original de CHK. Sin embargo, dicha versión no es obra del autor del informe, sino que ya previamente numerosos autores (*e.g.* Adalet McGowan *et al.*, 2018; Dai *et al.*, 2019; De la Garza Cárdenas *et al.*, 2019; Tan *et al.*, 2016) han empleado está en sus respectivos estudios debido a la limitación de datos.

No obstante, este trabajo difiere del resto en las variables que se escogen para el cálculo de rs_t y rl_t . Así, por ejemplo, varios estudios (*e.g.* Dai *et al.*, 2019; De la Garza Cárdenas *et al.*, 2019; Tan *et al.*, 2016) que también utilizan esta versión simplificada, se ciñen a la definición original y toman como rs_t y rl_t , el tipo medio de interés preferente a corto y largo plazo. Mientras que estudios como el de los autores Adalet McGowan *et al.* (2018) siguiendo líneas similares a la nuestra, calculan dichas variables como medias anuales de las tasas mensuales.

Así, de acuerdo con la literatura académica revisada, lo ideal para el cálculo de R^* sería disponer de los tipos de intereses preferentes, ya que multiplicando estos por el saldo de la deuda viva al principio del periodo se obtendrían el pago mínimo de intereses que la empresa debe afrontar en el periodo actual. Sin embargo, no hay manera de conocer tales tipos preferentes con los datos disponibles y, es por ello, que se emplean los tipos de intereses medios como medida *proxy*.

Además, tal y como se ha definido el criterio a emplear, este presenta un problema grave en su segunda condición que puede distorsionar el número de zombies identificados. Dicho problema consiste en que, al comparar R con R^* , lo que realmente se está comparando es una medida relativa con otra no relativa. Por lo tanto, si no se corrige este error, la segunda regla de la segunda condición se cumplirá en la gran mayoría de los casos, identificando un número mayor de zombies. Es por ello, que teniendo en cuenta el criterio CHK y varios estudios de la literatura (*e.g.* De la Garza Cárdenas *et al.*, 2019; Han *et al.*, 2019, 2020; Tan *et al.*, 2016; Yu *et al.*, 2021), se procede a dividir $R_{i,t}^*$ por la deuda remunerada total al principio del periodo t , pero únicamente cuando se compara con $R_{i,t}$.

El segundo criterio que se plantea tiene como fin único el aportar a la literatura de las pymes zombies un criterio sencillo, en comparación al primero, para su identificación. Así, dicho criterio se basa en las mismas condiciones que el principal pero difiere en el cálculo de $R_{i,t}^*$, que se realiza con la siguiente ecuación:

$$rs_{t-1} DR_{i,t-1}^{c/p} + rl_{t-1} DR_{i,t-1}^{l/p} \quad (9)$$

Como se puede observar, la [ecuación 9](#) difiere ligeramente de la [ecuación 8](#), ya que en esta no se considera una media de los intereses a largo plazo de los últimos cinco años, sino que se toma directamente el interés medio anual a largo plazo del periodo inmediatamente anterior. Además, en este criterio también es necesario dividir $R_{i,t}^*$ cuando se compara con $R_{i,t}$, para no incurrir en los errores ya comentados. Por último, es de destacar que esta forma de calcular R^*

se asemeja bastante a la empleada por los autores Du & Li (2019), con la única diferencia de que estos emplean el tipo de interés mínimo de los préstamos a corto y largo plazo.

Así, y en conclusión de lo visto, a lo largo de este trabajo se empleará el primer criterio debido a su alto grado de aceptación en la literatura de las zombis. No obstante, en la sección [6.2.1](#) se emplean ambos criterios de manera separada, con el fin de comparar el número de pymes zombis identificadas.

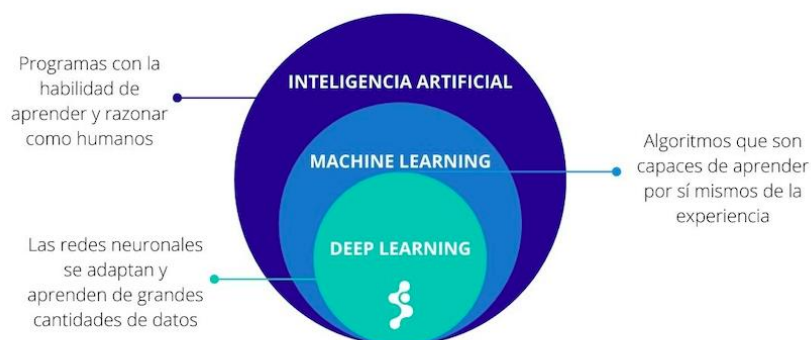
4.2. Marco teórico sobre inteligencia artificial

Antes de proceder con la ejecución de la prueba de concepto, es necesario precisar de qué se entiende por las redes neuronales convolutivas (CNN). Es por ello, que esta sección comienza con una introducción básica sobre las redes neuronales artificiales (ANN). Posteriormente, se aporta una posible definición de las CNN. Y, por último, se realiza un breve recorrido de estudios que han empleado dichas redes en cuestiones del ámbito financiero.

4.2.1. Redes neuronales convolutivas

Las ANN forman parte del campo del aprendizaje profundo (*Deep Learning*), siendo este un paradigma de la inteligencia artificial y del aprendizaje automático (*Machine Learning*) (Emmert-Streib *et al.*, 2020). Así, el aprendizaje profundo no es más que un subconjunto del aprendizaje automático ([Figura 3](#)), el cual suele estar enfocado en la creación de las ANN, es decir, sistemas que imitan al cerebro humano, adaptándose y aprendiendo a partir de grandes cantidades de datos (IBERDROLA, 2020). Dichas redes han sido empleadas para el reconocimiento del lenguaje humano, la identificación de imágenes o para realizar predicciones, entre otros posibles casos.

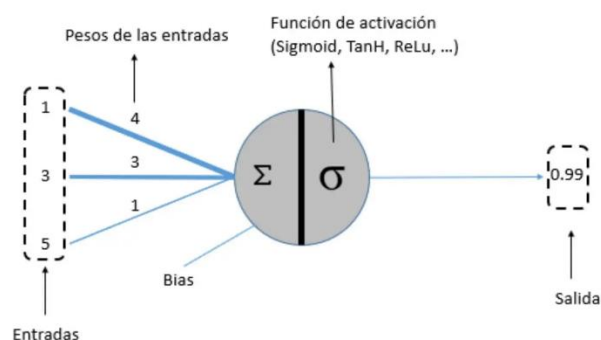
Figura 3: Campos de la Inteligencia Artificial



Fuente: Extraído de Juan Pérez & Cantero (2021)

Particularmente, las ANN son modelos matemáticos motivados por el funcionamiento del cerebro humano. La entidad básica de cualquier ANN es el modelo de una neurona, también conocida como perceptrón. Dicho modelo propone, como se puede observar en la [Figura 4](#), que las entradas se suman de forma ponderada mediante unos pesos. De este modo, si dicha suma supera un umbral o *bias*, esta neurona genera un cierto valor de salida (Guerra Artal, 2021). Además, esta salida puede estar limitada por una función de activación la cual determina si una neurona debe activarse o no en función de su relevancia para la predicción del modelo (Verdence, 2020).

Figura 4: Modelo matemático de una neurona artificial

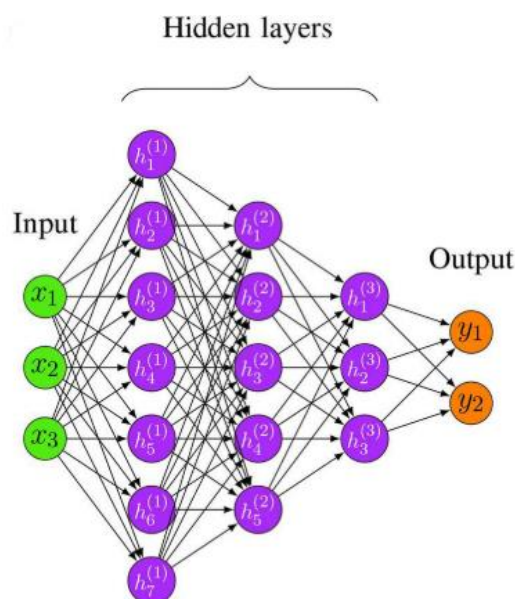


Fuente: Extraído de García-Olalla Olivera (2019)

No obstante, para poder construir redes neuronales, es necesario que estas neuronas se conecten entre ellas formando capas ([Figura 5](#)). Así, se pueden clasificar estas capas en tres posibles tipos (Calvo, 2018):

- **Capa de entrada:** conecta la red con el exterior. En ella, cada neurona se corresponde con una de las variables de entrada a la red. Además, en esta capa no se produce ningún tipo de procesamiento de los datos.
- **Capas ocultas:** una vez que la información de entrada se encuentra en la red, esta se someterá a diversas operaciones en estas capas. Así, se tratan de neuronas cuyas entradas son los resultados de la función de activación de las neuronas de la capa anterior. Además, las salidas de las neuronas de esta capa (limitadas normalmente por una función de activación) pasarán como entradas a la siguiente capa.
- **Capa de salida:** conecta las capas ocultas con la salida de la red que proporciona los resultados.

Figura 5: Perceptrón Multicapa (con varias capas ocultas)



Fuente: Extraído de Emmert-Streib *et al.* (2020)

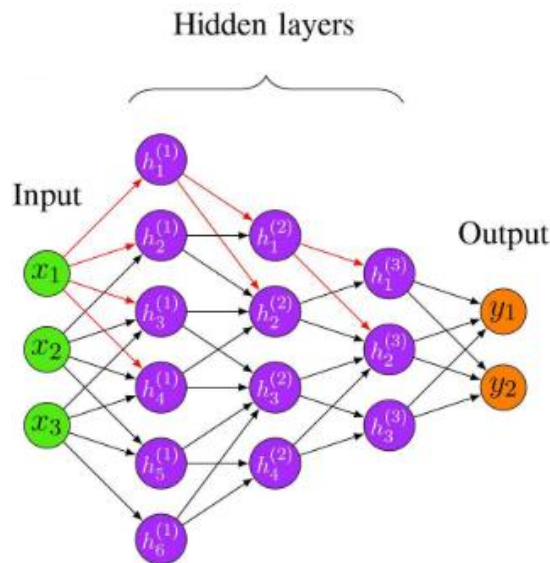
Por otra parte, lo que se observa en la [Figura 5](#) es también una de las arquitecturas más simples en las redes neuronales, conocidas como el perceptrón multicapa. Dichas redes están formadas por las capas de entrada, salida y por una o varias capas ocultas. Así, en una ANN típica todas las salidas de la capa anterior están conectadas a todas y cada una de las entradas de la capa siguiente (capas totalmente conectadas), donde cada conexión constituye un parámetro de la red ([Figura 5](#)). Por lo tanto, el principal objetivo de una ANN es ajustar los pesos de cada una de estas conexiones para que el valor de salida estimado por la red se asemeje cada vez más al real. Sin embargo, este tipo de redes pueden generar un gran número de parámetros (Emmert-Streib *et al.*, 2020).

Tras esta introducción muy básica sobre las ANN, nos centramos en el concepto y funcionamiento de las CNN.

Así, estas consisten en un tipo de ANN en las que sus capas imitan al córtex visual del ojo humano para identificar distintas características en las entradas, que en definitiva hacen que pueda “ver” e identificar objetos (Ignacio Bagnato, 2018). Debido a su aplicación habitual sobre matrices bidimensionales o tridimensionales, estas han demostrado ser muy útiles para tareas de visión artificial, como la identificación y clasificación de imágenes, detección de objetos y reconocimiento de rostros (Prabhu, 2018). Concretamente consisten en una variación del perceptrón multicapa ([Figura 5](#)) y están compuestas por: **las capas convolutivas** y **las capas totalmente conectadas**.

Las capas convolutivas forman una parte esencial de las CNN. En ellas, cada neurona está conectada solo a un subconjunto de salidas de la capa anterior ([Figura 6](#)). Gracias a esto, se reduce de manera considerable el número total de parámetros de estas redes en comparación con las totalmente conectadas (Emmert-Streib *et al.*, 2020). Además, estas capas se caracterizan porque en ellas se realiza una operación matemática denominada como **convolución**. Así, la [Figura 7](#) muestra cómo se realiza dicha operación sobre una entrada bidimensional.

Figura 6: Ejemplo de una CNN



Las flechas rojas señalan como las neuronas de una capa solo están conectadas a un subconjunto de neuronas de la siguiente capa

Fuente: Extraído de Emmert-Streib *et al.* (2020)

Además, si se supone que la entrada de la operación de convolución consiste en una imagen en escala de grises (en realidad sus píxeles), dicha operación no hace más que alterar sus píxeles para generar una nueva imagen.

Cada píxel de la nueva imagen se calcula situando una matriz de números: filtro o *kernel*, sobre la imagen original. Posteriormente se realiza un producto escalar de los píxeles que correspondan de la imagen con los valores del kernel ([Figura 7](#)). De esta manera, si se desplaza dicho filtro por todos los píxeles de la imagen original se genera lo que se conoce como un mapa de característica. Dicho mapa indica una cierta característica (*e.g.* cambios de contraste, texturas, superficies planas, etc.) de la imagen original que ha sido detectada por el filtro. Por lo tanto, cada vez que el filtro detecta en la imagen original la característica para el que han sido programado sus valores, esto se refleja en el mapa con un píxel blanco. En la [Figura 8](#) se recogen algunos ejemplos de filtros 3x3 con sus respectivos valores y la característica que detectarían en una imagen.

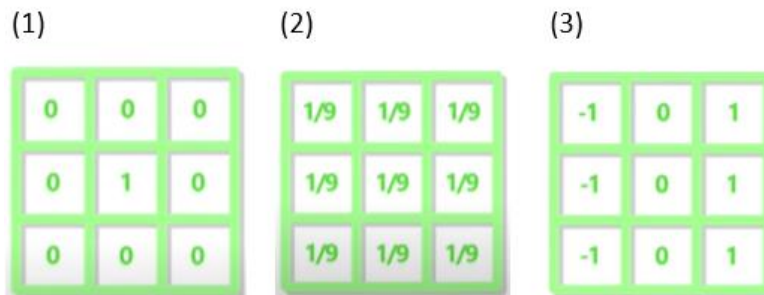
Figura 7: Operación de convolución

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} * \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 2 & 2 \\ 0 & 3 & 2 & 3 \\ 3 & 1 & 5 & 4 \\ 0 & 3 & 2 & 3 \end{pmatrix}$$

Dato de entrada bidimensional de tamaño 6x6 sobre el que se aplica un filtro de 3x3 para generar un mapa de características de 4x4

Fuente: Extraído de Emmert-Streib *et al.* (2020)

Figura 8: Ejemplos de filtros 3x3



(1) Filtro que genera una copia exacta de la imagen original; (2) Filtro que genera una imagen nueva desenfocada; (3) Filtro que detecta bordes verticales en una imagen.

Fuente: Elaboración propia del autor

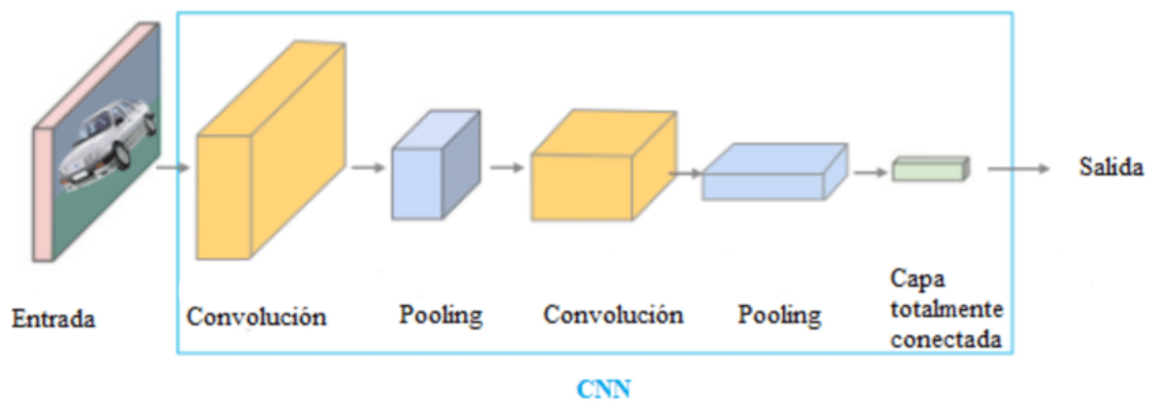
De esta manera, la operación de convolución sobre una imagen puede detectar diferentes características según como sean los valores de este. Sin embargo, dichos valores no se ajustan manualmente, sino que es la propia red quien los irá aprendiendo poco a poco para realizar mejor la tarea para la que ha sido programada. Tal que, aprender estos filtros para detectar patrones es la principal tarea de una CNN. Además, es posible aplicar varios filtros sobre una misma imagen para identificar diferentes características presentes en ella.

En resumen, entra una imagen a la red, se le aplican una serie de convoluciones y se generan un conjunto de mapas de características. No obstante, la operación de convolución por sí sola no puede detectar características complejas en una imagen, más que bordes y patrones muy simples. ¿Entonces como una CNN puede identificar estas características complejas? El verdadero potencial de este tipo de redes se encuentra en que dichas operaciones de convolución se realizan de manera secuencial. Es decir, estas capas realizan detecciones de características sobre detecciones previas de las capas anteriores. Esto permite reconocer patrones cada vez más complejos de la imagen y hacen que la operación de convolución sea cada vez más potente.

Es por lo dicho, que el diseño de una arquitectura convolutiva se representa habitualmente como un embudo. En ella la imagen inicial se va comprimiendo espacialmente, es decir su resolución

va disminuyendo (véase [Figura 9](#)). Dicha comprensión se logra insertando capas de *pooling* entre las capas convolutivas las cuales tienen como objetivo reducir la dimensión de la entrada con algún método de agrupación preestablecido. Lo dicho genera como resultado entradas más pequeñas al conservar la mayor cantidad de información posible (Emmert-Streib *et al.*, 2020). Además, al mismo tiempo que la imagen se comprime, su grosor va aumentando, es decir, el número de mapa de características va en aumento.

Figura 9: Arquitectura CNN (típica)



Fuente: Extraído de Andrade Carrera *et al.* (2021)

En conclusión, las CNN están compuestas por capas convolutivas y capas totalmente conectadas. En las primeras se realiza la extracción de las características, donde la información cada vez se hace más elaborada y compleja a medida que atraviesa las distintas capas convolutivas (Guerra Artal, 2021). Mientras que las segundas, llevan a cabo el proceso de clasificación propiamente dicho, donde se detecta la combinación de características que conforman el objeto o patrón a determinar (Guerra Artal, 2021).

4.2.2. Redes neuronales convolutivas en el ámbito financiero

Una vez delimitado el concepto de las CNN se realiza un breve recorrido sobre estudios que han empleado este tipo de redes en cuestiones del ámbito financiero. Para ello se ha elaborado la [Tabla 4](#), la cual se encuentra dividida respecto al dato de entrada a la CNN y por la temática general de dichos estudios. En este sentido, se han identificado estudios que predicen desde el precio de las acciones hasta el destino de una empresa (*e.g.* bancarrota, fracaso empresarial, etc). Además, las CNN, como adelantábamos en la sección anterior, han resultado ser muy útiles en tareas de visión artificial (Prabhu, 2018), pero esto no descarta su uso sobre otras entradas como: series temporales, vectores densos o datos de panel.

Así, centrando el foco sobre estudios que no se basan en las entradas tradicionales (imágenes), son varios los autores que han empleado estas redes para la predicción del precio de las acciones (e.g. Ding *et al.*, 2015; Gunduz *et al.*, 2017; Tsantekidis *et al.*, 2017). Así, Ding *et al.* (2015) utilizan información de las noticias mensuales para modelar la influencia de estas en el movimiento del precio de las acciones. Dichos autores llegan a la conclusión de que su modelo obtiene mayores beneficios que otros tradicionales. Por otra parte, Tsantekidis *et al.* (2017) basándose en series temporales de alta frecuencia predicen los movimientos de precios de las acciones con una precisión media de entre 65% y 67%. Con respecto a la temática de predicción del destino de una empresa, destaca el estudio de Alam *et al.* (2021), quienes convierten datos de panel en puntos grassmannianos con el fin de predecir el fracaso empresarial. Así, el modelo propuesto por estos alcanza una exactitud de 91.2% para discriminar entre las empresas fracasadas y las no fracasadas.

Tabla 4: Artículos que han empleado CNN en el ámbito financiero

Entrada CNN	Temática general	Nombre artículo	Autor/es	Año	Medidas
Imágenes	Trading	Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach	Sezer, Omer Berat Ozbayoglu, Ahmet Murat	2018	Exactitud: - Acciones: 58% - Fondos cotizados: 62%
		Encoding candlesticks as images for pattern classification using convolutional neural networks	Chen, Jun Hao Tsai, Yun Cheng	2020	Exactitud (*): - Datos simulación: 92.42% - Datos reales: 90.7%
	Predicción destino empresarial	Bankruptcy prediction using imaged financial ratios and convolutional neural networks	Hosaka, Tadaaki	2019	Exactitud (empresas bancarota) (*): - Método aleatorio: entre 77-90% - Método correlacionado: entre 77-92%
Otros (series temporales, vectores densos, datos de panel)	Predicción precio de las acciones	Deep learning for event-driven stock prediction	Ding, Xiao Zhang, Yue Liu, Ting Duan, Junwen	2015	Exactitud: - Índice S&P500: 64% - Acciones individuales: 65%
		Forecasting stock prices from the limit order book using convolutional neural networks	Tsantekidis, Avraam Passalis, Nikolaos Tefas, Anastasios Kanniainen, Juho Gabbouj, Moncef Iosifidis, Alexandros	2017	Precisión (*): entre 65-67%, para diferentes horizontes temporales
	Predicción destino empresarial	Corporate Failure Prediction: An Evaluation of Deep Learning vs Discrete Hazard Models	Alam, Nurul Gao, Junbin Jones, Stewart	2021	Exactitud: 91.2% para discriminar entre empresas fracasadas y no fracasadas

(*) Valores medios

Fuente: Elaboración propia del autor

En cuanto a los estudios que emplean las imágenes como dato de entrada a la CNN, se detecta que el ámbito del *trading* financiero ha comenzado a explotar las cualidades de estas redes de manera significativa. Así, el trading consiste en la compra y venta de activos financieros de distintos tipos como: acciones, divisas, materias primas y criptomonedas, entre otros (Caro Mora, 2021), con el fin de generar ganancias. Además, suele apoyarse en diferentes tipos de análisis y estrategias (Fernández Fernández, 2018) para operar en los mercados financieros. En este sentido, Sezer & Ozbayoglu (2018) proponen un algoritmo de trading basado en CNN que predice numerosas estrategias de este ámbito con una alta precisión. Para ello, convierten las series temporales en imágenes 2D, que posteriormente se suministran como entrada a la CNN. En líneas similares, Chen & Tsai (2020) plantean un enfoque de dos pasos para el reconocimiento de los patrones de velas (tipo de gráfico en el *trading*) mediante las CNN. Así, los autores llegan a la conclusión de que su modelo es capaz de detectar tales patrones con una exactitud media de 90.7%.

Por último, centrando el foco en la temática que es de interés para este estudio, es decir, trabajos que utilizan imágenes para la predicción del destino de una empresa, es fundamental hacer referencia al estudio de Hosaka (2019). Esto se debe a que, dicho estudio y el actual se basan en una metodología bastante parecida, con la única diferencia de que el presente estudio identifica a las pymes zombis y el de Hosaka (2019) a empresas en bancarrota. Así, dicho estudio aplica las CNN para la predicción de las empresas en bancarrota, que en la mayoría de los casos consiste en un problema de clasificación de dos clases. Para ello, en primer lugar, se transforman las distintas partidas contables de una empresa en diferentes tipos de ratios financieras. Posteriormente tales ratios se representan como píxeles específicos de una imagen en escala de grises utilizando el método aleatorio o correlacionado. Finalmente, estas imágenes se suministran como entrada a la CNN cuyo propósito principal es la predicción de las empresas mencionadas. De esta manera, el modelo propuesto por Hosaka (2019) alcanza unas tasas de identificación para las empresas en bancarrota de entre 77% y 92%.

Como se ha visto, las CNN han permitido obtener rendimientos generalmente satisfactorios en los diferentes ámbitos analizados en esta sección. Sin embargo, lo que es de interés al actual estudio es el empleo de estas redes sobre entradas que consisten en imágenes. En este sentido, destaca el estudio de Hosaka (2019) que con una metodología muy similar a la nuestra logra detectar a empresas en bancarrota con unas altas tasas de identificación.

Por último, es de recalcar que en la literatura académica revisada sobre las zombis, no se ha encontrado evidencia alguna del uso de las CNN para su predicción. Esto como bien mencionábamos en la introducción, consiste en uno de los aspectos diferenciadores de este trabajo. Además, el hecho de emplear este tipo de redes para la predicción de las pymes zombis en España otorga al presente trabajo la segunda característica diferenciadora. En definitiva, el actual estudio puede considerarse como uno de los primeros, por no decir el primero, que propone la predicción de las pymes zombis en España a través de las redes neuronales convolutivas.

5. Recursos

5.1. Gestión de datos

En este trabajo, se emplea información proveniente de los diferentes estados contables (*e.g.* balance, cuenta de resultados, etc) de las pymes españolas. Dicha información se extrae de la herramienta conocida como *SABI: Sistema de Análisis de Balances Ibéricos*. Así, esta consiste

en una herramienta web para el análisis financiero y marketing estratégico elaborada por la empresa *INFORMA* en colaboración con *Bureau Van Dijk*. Además, dicha herramienta permite el manejo sencillo y rápido de la información general y cuentas anuales de más de 2,7 millones de empresas españolas y de aproximadamente 800.000 empresas portuguesas (INFORMA, 2021).

Por otro lado, para el análisis bibliográfico realizado en este estudio, es necesario contar con artículos de alta calidad y fiabilidad. Es por ello, que la mayoría de dichos artículos han sido seleccionados y extraídos de la herramienta online denominada como *Scopus*, a cargo de la empresa *ELSEVIER*. Esta consiste en una base de datos de citaciones y resúmenes seleccionados por expertos (Elsevier, 2021).

5.2. Lenguaje y entorno de programación

El lenguaje de programación seleccionado para el desarrollo de este trabajo es *Python*. Además, su versión más reciente es la 3.10, pero como el desarrollo del trabajo se comenzó desde principios de septiembre de 2021, se ha seleccionado finalmente la versión 3.9. Ambas versiones son perfectamente compatibles con las librerías utilizadas en este trabajo.

Python es de uno de los lenguajes más populares hoy en día. Esto es debido a la gran facilidad que ofrece a los desarrolladores tanto para programar como para la comprensión del código (Malik, 2019a). Consiste en un lenguaje interpretado, de alto nivel, orientado a objetos y con semántica dinámica (python, 2021). Además, posee estructuras de datos de alto nivel combinadas con la tipificación y vinculación dinámica (python, 2021). Asimismo, permite organizar sus códigos en módulos y paquetes lo que fomenta su modularidad y reutilización.

Otro de los aspectos positivos a mencionar de este lenguaje, es que cuenta con numerosas librerías ya desarrolladas por su gran comunidad (Malik, 2019a). En este sentido, existen librerías tanto para el desarrollo web, análisis de datos o cálculo numérico, como para los diferentes ámbitos de la inteligencia artificial. Por último, *Python* permite un desarrollo rápido de las aplicaciones frente a otros lenguajes.

Como entorno de programación del presente trabajo se ha seleccionado *Jupyter Notebook*. Este consiste en una aplicación web para crear y compartir documentos que contienen códigos junto con diferentes tipos de visualizaciones, ecuaciones matemáticas, texto y otros medios enriquecedores (jupyter, 2021). De esta manera, los *notebooks* de *jupyter* consisten en una gran alternativa para escribir e iterar sobre un código en Python (Velotio Technologies, 2019).

Además, estos *notebooks* son cada vez una opción más popular para proyectos de ciencia de datos, modelado estadístico o aprendizaje automático.

5.3. Kaggle

Las CNN entrenan de forma considerablemente acelerada en un equipo informático que cuenta con una unidad de procesamiento gráfica (GPU). Sin embargo, el autor del presente informe no dispone de tales recursos *hardware* en su equipo local.

Es por tal motivo, que se recurre a la plataforma externa conocida como *Kaggle* (kaggle, 2022) que proporciona a sus usuarios acceso gratuito a las GPU TESLA P100 de NVIDIA. Así, estas aceleran de manera significativa el entrenamiento de los modelos de aprendizaje profundo. Asimismo, esta plataforma también permite la ejecución de *notebooks* de *jupyter* en un entorno computacional en la nube, además de compartir diferentes *datasets* con su comunidad.

Por último, todos los *notebooks* y *datasets* del presente estudio se entregan en dos formatos:

- Como enlaces de *Kaggle* que ya disponen de toda la preparación previa para la ejecución correcta de dichos *notebooks* (**se recomienda esta alternativa**). De esta manera, el único paso que debe realizar el usuario es la habilitación de la GPU al momento de ejecutar el *notebook principal* del estudio.
- Como ficheros comprimidos para su ejecución en el equipo local. Sin embargo, esta alternativa requiere la realización de minuciosas modificaciones que se indican en los mismos *notebooks*. Además, no se recomienda esta alternativa si no se dispone de una GPU en el equipo local.

5.4. Librerías para redes neuronales convolutivas

Tal y como se ha comentado, el objetivo principal del estudio es el desarrollo de una CNN que permita la predicción de las pymes zombis en España. Sin embargo, previamente se ha elaborado una tabla comparativa con las distintas librerías disponibles en Python que permiten la construcción de estas redes ([Tabla 5](#)).

De este modo, para la elaboración de dicha tabla se han seleccionado las librerías: ***TensorFlow***, ***Pytorch*** y ***Keras***. Todas estas bibliotecas son empleadas por la comunidad científica e informática y gozan de bastante popularidad. No obstante, *Keras* es la que mejor tendencia presenta de las tres para el desarrollo de las redes neuronales ([Figura 10](#)).

Con respecto al nivel de API de estas librerías, *TensorFlow* a pesar de proporcionar *Keras* como un *framework* que facilita el trabajo, no siempre es sencilla de utilizar. De manera similar, *Pytorch* también presenta niveles bajos de API, al trabajar directamente con expresiones sobre matrices, las cuales hacen que su uso sea complicado y poco legible. Por el lado contrario, *Keras* presenta un nivel alto de API y es capaz de ejecutarse sobre *TensorFlow*, *CNTK* y *Theano*. Además, está se trata de la librería más popular para el desarrollo de las CNN debido a su arquitectura simple, concisa y legible. Así, *Keras* permite a los desarrolladores crear redes neuronales en muy pocos pasos.

Sin embargo, el rendimiento de *Keras* es relativamente inferior a las librerías de bajo nivel: *TensorFlow* y *Pytorch*. Además, estas últimas permiten trabajar con modelos de alto rendimiento, grandes *datasets* que requieren una ejecución rápida, redes neuronales complejas y proporcionan unas capacidades de depuración mejor que las librerías de alto nivel.

Por último, con respecto a la documentación disponible sobre las CNN, es de destacar que todas ellas presentan abundante tutoriales y documentos oficiales en su página web dejando a *Pytorch*, que se queda algo escasa.

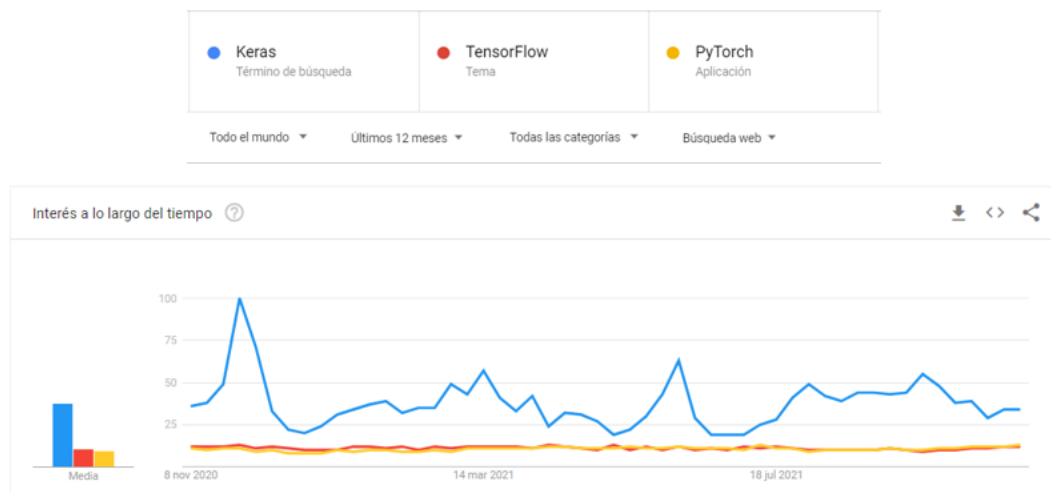
Tabla 5: Librerías redes neuronales

Características/Librerías	Pytorch	TensorFlow	Keras
Nivel de API	Baja	Baja/Alta	Alta
Arquitectura	Compleja y poco legible	No fácil de usar	Simple, concisa y legible
Depuración	Ofrece buenas capacidades	Difícil	No suele requerir depuración (redes simples)
Popularidad	Popular (3ª)	Popular (2ª)	Popular (1ª)
Velocidad	Rápida y alto rendimiento	Rápida y alto rendimiento	Lenta y bajo rendimiento
Conjunto de datos (<i>Datasets</i>)	Grandes	Grandes	Pequeños
Documentación	Algo escasa	Abundante	Abundante
Escrita en	C, C++, Cuda, Python	C++, Cuda, Python	Python
Especialidad	Aprendizaje profundo y procesamiento de lenguaje natural	Redes neuronales y programación de flujo de datos en una amplia gama de tareas	Experimentación rápida con redes neuronales

Fuente: Elaborado a partir de Terra (2021)

En resumen, teniendo en cuenta todas las características mencionadas, se procede a seleccionar la librería *Keras* (ejecutada sobre *TensorFlow*) para el desarrollo de la CNN del presente estudio. Así, dicha selección se debe por, su arquitectura simple, concisa y legible, su numerosa documentación disponible sobre dichas redes y su alto nivel de API. Asimismo, otros de los factores que favorecen su selección, es su uso previo por parte del autor en proyectos de índole personal.

Figura 10: Tendencia de librerías para redes neuronales



Fuente: Elaboración propia del autor

5.5. Otras librerías

A continuación se mencionan el resto de librerías utilizadas en este trabajo, junto con una breve descripción de estas. Así, dichas librerías junto con las mencionadas en el apartado anterior sirven como apoyo al objetivo principal del trabajo. Estas librerías son:

- **Matplotlib:** biblioteca completa en *Python* para crear visualizaciones estáticas, animadas e interactivas (matplotlib, 2021), normalmente en dos dimensiones. Dichas visualizaciones pueden abarcar desde diagramas de diferentes tipos hasta imágenes en distintos formatos. Asimismo, esta librería permite la manipulación de ciertas características de las imágenes.
- **NumPy:** librería fundamental para la computación científica en *Python* (NumPy, 2021), de código abierto y que plantea los vectores y matrices como su estructura de datos básica (Malik, 2019b). Además, esta puede ser empleada para realizar operaciones matemáticas sobre dichos vectores y matrices, como rutinas trigonométricas, estadísticas y algebraicas (Malik, 2019b).
- **Pandas:** librería de análisis y manipulación de datos de código abierto, rápida, potente, flexible y fácil de usar, construida sobre el lenguaje de programación *Python* (pandas, 2021).
- **Pillow:** biblioteca para el procesamiento de imágenes en *Python* y que admite formatos de archivo como: BMP, PNG, JPEG y TIFF (pillow, 2022).
- **Scikit-learn:** librería de aprendizaje automático en *Python* que ofrece funcionalidades para el aprendizaje supervisado y no supervisado (scikit-learn, 2022). Además, proporciona herramientas para el desarrollo, selección y evaluación de modelos, así

como muchas otras utilidades, incluido el preprocesamiento de datos (Myrianthous, 2021).

6. Desarrollo CRISP-DM

6.1. Comprensión del negocio

En esta fase se determinan los objetivos a alcanzar desde el punto de vista del proyecto como de la minería de datos. Posteriormente, se evalúa la situación actual del proyecto (*e.g.* recursos, limitaciones, suposiciones, etc) para finalmente proponer una planificación a la que este debe ceñirse.

De este modo, esta sección comienza determinando el objetivo principal de la minería de datos para este proyecto. Dicho objetivo ayudará a conducir al estudio hacia sus objetivos finales y consiste en:

- La elaboración de una CNN capaz de predecir con datos contables de periodos anteriores, aquellas pymes españolas que pasarán al estado zombi en un futuro cercano.

Una vez definido dicho objetivo, se procede con la fijación de los objetivos del proyecto. Así, es bastante habitual confundir estos últimos con los primeros. En este sentido, los objetivos del proyecto se describen en términos generales mientras que los primeros presentan un carácter más técnico. De este modo, estos objetivos son:

- La puesta en marcha de una prueba de concepto en la que se desarrolla y analiza si las CNN permiten la identificación temprana de las pymes zombis en el sector turístico español.
- La aportación, a través de la revisión bibliográfica, de una mayor luz a la escasa literatura existente sobre pymes zombis en España.

Así, llegados a este punto, han quedado delimitados los objetivos de la minería de datos y del proyecto. Ahora bien, en ningún momento se han citado los criterios que medirán el éxito del proyecto. Es por ello, que la próxima tarea de la fase de comprensión del negocio consiste en la determinación de dichos criterios.

Además, tal y como se comentó en la sección [4.4.2](#), no existe evidencia en la literatura de las empresas zombis del empleo de este tipo de redes para su predicción, lo cual supone aspectos favorables como desfavorables para el presente estudio. Por el lado favorable, hace que este

estudio sea uno de los primeros, por no decir el primero, en intentar la predicción de dichas empresas mediante las CNN. Mientras que por el otro, supone que el estudio no pueda tomar como referencia unas determinadas métricas y por ende establecer un criterio formal que permita la medición del éxito de este.

Pero, como el establecimiento de los criterios consiste en una parte fundamental de la metodología *CRISP-DM* y que en la literatura de las zombis se ha observado una alta tendencia del empleo de las regresiones logarítmicas (*e.g.* Goto & Wilbur, 2019; Hoshi, 2006; Lam *et al.*, 2017a; Tan *et al.*, 2016) para su identificación, se proponen los siguientes criterios que miden el éxito del proyecto:

- La comparación de las diferentes métricas obtenidas en la predicción de pymes zombis mediante las CNN y las regresiones logísticas binarias. Así, se espera que las primeras métricas superen de manera clara a las obtenidas por las metodologías tradicionales, es decir las regresiones.
- La identificación de las pymes zombis en el sector turístico español. En este aspecto, se confía que los criterios de la sección [4.1.4](#) permitirán la identificación de las pymes zombis guardando una alta semejanza entre ellos.

Con respecto a los recursos que se emplean en el presente estudio, estos ya han sido cubiertos de manera exhaustiva en la sección [5](#). Aunque, es de resaltar la enorme trascendencia de la base de datos *SABI*, ya que sin esta no sería posible la realización de la presente investigación. Además, el estudio debe cumplir con la legislación y normativa mencionada en la sección [2.2](#) y presenta como requerimientos básicos el alcance de las competencias citadas en la sección [2.1](#).

En cuanto a las limitaciones que han sido detectadas a estas alturas del proyecto destaca la no disponibilidad en *SABI* de los tipos de intereses a los que fueron concedidos los créditos a las empresas no financieras. Dicha información, como se ha visto en la sección [4.1.4](#), es fundamental para el cálculo de la variable R^* . Es por ello, que se ha recurrido a una fuente externa fiable: tipos de intereses publicados por el Banco de España, para la posible aplicación de los criterios seleccionados en el estudio.

Por último, en la sección [2.3](#) se detalló la planificación a seguir para este proyecto junto con la estimación en horas de sus diferentes etapas. Estas etapas, como se puede comprobar en la [Tabla 1](#), no son más que las diferentes fases de la metodología *CRISP-DM*.

6.2. Comprensión de los datos

En esta fase se identifican, recopilan y analizan los conjuntos de datos que pueden ayudar a lograr los diferentes objetivos mencionados en la fase de comprensión del negocio. Sin embargo, antes de proceder con dicha recopilación de datos es necesario aclarar algunas características sobre estos.

Así, en primer lugar y como se ha venido insistiendo, el presente estudio se centra en la predicción de las zombis en el tejido de las pymes españolas. Sin embargo, en ningún momento se han citado las características que discriminan a estas empresas respecto a las demás. De este modo, es fundamental comprender tales características para la correcta elaboración de las diferentes condiciones de búsquedas en *SABI*.

Por lo tanto, siguiendo los criterios propuestos por la Comisión Europea (2016) (véase [Figura 11](#)) definimos a estas como: empresas cuyo número de trabajadores anual es menor a 250 empleados y que presentan un volumen de negocio igual o menor a 50 millones de euros o cuyo balance general es igual o menor a 43 millones de euros.

Figura 11: Criterios pymes

Categoría de empresa	Efectivos: Unidades de trabajo anual (UTA)	Volumen de negocios anual	Balance general anual
Medianas	< 250	≤ 50 millones EUR	≤ 43 millones EUR
Pequeñas	< 50	≤ 10 millones EUR	≤ 10 millones EUR
Microempresas	< 10	≤ 2 millones EUR	≤ 2 millones EUR

Fuente: Extraído de Comisión Europea (2016)

Como se puede apreciar en la [Figura 11](#), la definición de pymes propuesta por la Comisión Europea (2016) abarca tres posibles categorías de empresas: medianas, pequeñas y microempresas. Este estudio solo se centra en las primeras dos y por lo tanto, excluye a las microempresas. El motivo de dicha exclusión se debe a que estas son menos profesionalizadas, es decir, su personal no suele tener especializaciones técnicas y sus procesos productivos generalmente no están automatizados (Peña, 2020); sus trabajadores suelen ser familiares o personas unidas por lazos cercanos (asesorias, 2021); poseen un radio de acción geográfico muy limitado y generalmente se dedican a nichos de mercado muy concretos (asesorias, 2021). Además, las microempresas se comportan de forma diferente en comparación a las empresas




de categorías superiores y con frecuencia se constituyen para el autoempleo. Asimismo, estas aparecen y desaparecen del mercado con una mayor frecuencia que las pequeñas y medianas empresas.

Una vez definidos los criterios que identifican a las pymes, otro punto a aclarar, es el sector en el que se basa el presente estudio. Esto último es de vital importancia, ya que si se seleccionan las pymes de todos los sectores disponibles, se teme a que las diferencias observadas entre las zombis y no zombis no sean debidas a su idiosincrasia sino a la mayor abundancia de las zombis en un determinado sector (Rojo Ramírez *et al.*, 2011). Es por tal motivo, que el trabajo actual se centra en el sector turístico español, concretamente, en el *grupo I* de la *Clasificación Nacional de Actividades Económicas (CNAE) 2009* (Vida Barea, 2021) conocido como “Hostelería”. Así, dicho grupo está compuesto por las actividades (solo códigos primarios): servicios de alojamiento (55) y servicios de comidas y bebidas (56).

Además, para el estudio solo se han seleccionado las pymes que se encuentran activas y que consisten en sociedades anónimas o limitadas, excluyendo de esta manera: comunidad de bienes, sociedades civiles, asociaciones, unión temporal de empresas, cooperativas, entidades no residentes y otras forma jurídicas. De igual modo, se excluyen todos los datos contables del año 2020 debido a que este fue un año marcado por la pandemia generada por el coronavirus y que puede llegar a alterar los resultados de la presente investigación.

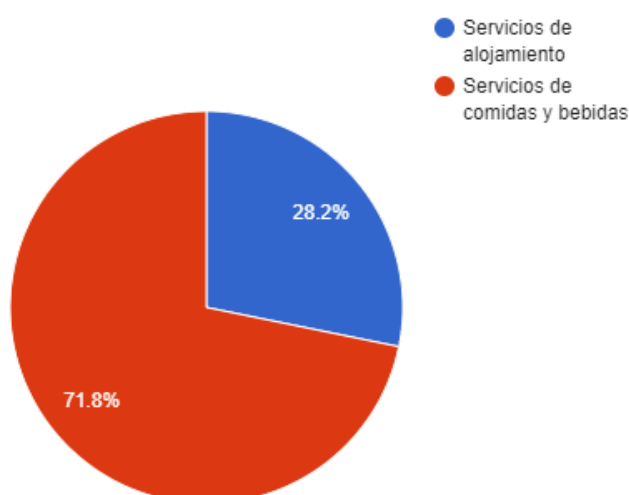
A continuación, se procede con la realización de la búsqueda de las pymes españolas en *SABI*. De esta manera, la [Figura 12](#) muestra el número total de pymes disponibles (13.792) para el año 2019 y que pertenecen a las actividades (55) y (56). Además, como se puede observar en la [Figura 13](#), el número de pymes es relativamente mayor en los servicios de comidas y bebidas (71,8%) que en los de alojamiento (28,2%).

Figura 12: Pymes españolas en 2019

ESTRATEGIA DE BÚSQUEDA		 Guardar	 Imprimir	 Borrar todas las etapas
<input checked="" type="checkbox"/>	1. Región/País: España			1.826.745
<input checked="" type="checkbox"/>	2. Tipos de empresas: Empresas			1.826.729
<input checked="" type="checkbox"/>	3. CNAE 2009(Sólo códigos primarios): 55 - Servicios de alojamiento, 56 - Servicios de comidas y bebidas			118.447
<input checked="" type="checkbox"/>	4. Estados España: Activa			919.442
<input checked="" type="checkbox"/>	5. Forma jurídica España: Sociedad anónima, Sociedad limitada			1.808.267
<input checked="" type="checkbox"/>	6. Número empleados: 2019, max=249			581.580
<input checked="" type="checkbox"/>	7. Total activo (A + B) (mil EUR): 2019, max=43.000			850.658
<input checked="" type="checkbox"/>	8. Importe neto de la cifra de negocios (mil EUR): 2019, max=50.000			706.908
<input checked="" type="checkbox"/>	9. Número empleados: 2019, min=10			129.903
<input checked="" type="checkbox"/>	10. Total activo (A + B) (mil EUR): 2019, min=2.000.001			130.180
<input checked="" type="checkbox"/>	11. Importe neto de la cifra de negocios (mil EUR): 2019, min=2.000.001			85.594
 Búsqueda booleana <input type="text" value="1 Y 2 Y 3 Y 4 Y 5 Y 6 Y (7 O 8) Y (9 O (10 Y 11))"/>		<input type="button" value="Actualizar"/>	 TOTAL : 13.792	

Fuente: Elaboración propia del autor

Figura 13: Proporción de pymes en hostelería (2019)



Fuente: Elaboración propia del autor

El próximo paso sería separar las zombis y no zombis siguiendo los criterios de la sección [4.1.4](#) junto con las condiciones expuestas en la [Figura 12](#). Sin embargo, en *SABI* no es posible calcular la variable R^* ya que esta depende de información externa a dicha base de datos. Por tal motivo, en este punto se hace necesaria la división de la fase actual en dos secciones bien diferenciadas.

Así, en la primera de ellas, se presenta un *notebook* en *Python* que basándose en los criterios de la sección [4.1.4](#) identifica las pymes zombis en el año 2019. Mientras que la segunda, tomando como entrada los resultados del *notebook* anterior descarga de *SABI* todas aquellas variables que finalmente se usarán para la predicción de las pymes zombis.

6.2.1. Identificación de las pymes zombis españolas

En esta primera sección de la fase de comprensión de los datos, se aporta el *notebook* llamado: **Etiquetado de las pymes españolas** que etiqueta a una pyme como zombi (1) o no zombi (0), para el periodo 2019.

Para la correcta ejecución de dicho *notebook*, es indispensable descargarse de *SABI* las variables que figuran en la [Tabla 6](#). No obstante, dicho trabajo ya ha sido realizado por parte del autor del informe y se encuentra recogido en el Excel denominado como: **pymes_iniciales**. Además, los criterios seleccionados en este estudio no solo se basan en los datos contables del periodo 2019 sino que también dependen de la información de los periodos 2018, 2017 y 2016, respectivamente. Es por ello, que en la [Tabla 6](#) también se recogen los años necesarios para cada una de las variables que permitirán la identificación de las pymes zombis.

Tabla 6: Variables para la identificación de zombis

Variables	Años
EBIT	2019, 2018 y 2017
Deudas a corto plazo	2019, 2018, 2017 y 2016
Deudas a largo plazo	2019, 2018, 2017 y 2016
Deudas con empresas del grupo y asociadas a corto plazo	2019, 2018, 2017 y 2016
Deudas con empresas del grupo y asociadas a largo plazo	2019, 2018, 2017 y 2016
Gastos financieros	2019

Fuente: Elaboración propia del autor

Asimismo, en este punto se modifica ligeramente la estrategia de búsqueda en *SABI* con respecto a la observada en la [Figura 12](#), incorporándole la condición de que las empresas seleccionadas dispongan de cuentas anuales para los periodos: 2016, 2017, 2018 y 2019 ([Figura 14](#)). Esto nos deja con un total de 10.907 pymes que se le suministrarán como entrada al *notebook* que identifica a las zombis.

Figura 14: Estrategia de búsqueda modificada

ESTRATEGIA DE BÚSQUEDA		Añadir un criterio de búsqueda	Guardar	Borrar todas las etapas
<input checked="" type="checkbox"/> 1. Región/País: España	1.826.744			
<input checked="" type="checkbox"/> 2. Tipos de empresas: Empresas	1.826.728			
<input checked="" type="checkbox"/> 3. CNAE 2009(Sólo códigos primarios): 55 - Servicios de alojamiento, 56 - Servicios de comidas y bebidas	118.447			
<input checked="" type="checkbox"/> 4. Estados España: Activa	919.389			
<input checked="" type="checkbox"/> 5. Forma jurídica España: Sociedad anónima, Sociedad limitada	1.808.266			
<input checked="" type="checkbox"/> 6. Años con cuentas disponibles: 2016, 2017, 2018, 2019	604.452			
<input checked="" type="checkbox"/> 7. Número empleados: 2019, max=249	582.404			
<input checked="" type="checkbox"/> 8. Total activo (A + B) (mil EUR): 2019, max=43.000	852.030			
<input checked="" type="checkbox"/> 9. Importe neto de la cifra de negocios (mil EUR): 2019, max=50.000	707.996			
<input checked="" type="checkbox"/> 10. Número empleados: 2019, min=10	130.030			
<input checked="" type="checkbox"/> 11. Total activo (A + B) (mil EUR): 2019, min=2.000.001	130.428			
<input checked="" type="checkbox"/> 12. Importe neto de la cifra de negocios (mil EUR): 2019, min=2.000.001	85.682			
Búsqueda booleana	1 Y 2 Y 3 Y 4 Y 5 Y 6 Y 7 Y (8 O 9) Y (10 O (11 Y 12))	Actualizar		TOTAL : 10.907

Fuente: Elaboración propia del autor

De este modo, el *notebook* comienza con la carga de las pymes y los tipos de intereses desde los Excel: **pymes_iniciales** y **tipos_intereses**, respectivamente. En este sentido, una característica que resalta entre la mayoría de las pymes es la alta cantidad de datos faltantes (*n.d.*) en las deudas con empresas del grupo y asociadas. Esto se debe a la sencilla razón de que, no todas las empresas pertenecen a un grupo y por ende no disponen de dicho tipo de deuda. Es por ello, que en el momento de calcular la deuda total a corto y largo plazo dichos registros se sustituyen por el valor 0.

Esta misma situación se repite con la variable de los gastos financieros representando aproximadamente el 15% de la muestra. Sin embargo, para este caso no es correcto sustituir el valor *n.d.* por un 0 ya que se podría cometer cierto error en el cálculo de la variable R y por consiguiente en los criterios para ser considerado zombi. Por dicha razón, se procede con la eliminación de todos los registros que presentan el valor *n.d.* en la variable de los gastos financieros.

Al margen de los gastos financieros, otra variable que es relevante en las condiciones para ser considerado zombi es el EBIT. Así, si no se dispone el valor de dicha variable para los periodos 2017, 2018 y 2019, no es posible determinar con certeza de que la empresa sea zombi o no zombi en el año 2019. Es por ello, que también se eliminan las empresas que no presentan dicha variable.

Esto nos deja con un total de 9.204 pymes para las que el *notebook* calcula la variable R y R*, y aplica los criterios comentados en la sección 4.1.4. De esta manera, tras la ejecución completa del *notebook*, se obtienen los resultados que figuran en la [Tabla 7](#).

Como se puede observar, el criterio principal del estudio identifica una mayor cantidad de zombis, exactamente 189, que el alternativo. El motivo de dicha diferencia se centra en el cálculo de R*, ya que el criterio principal considera una media de los últimos cinco periodos para los intereses a largo plazo, mientras que el alternativo toma directamente el interés medio a largo plazo del periodo inmediatamente anterior.

Tabla 7: Identificación de pymes zombis en España

Criterios/Pymes	Pymes Zombis	Pymes No Zombis
Principal	753	8.451
Alternativo	564	8.640
Grado de semejanza entre el criterio alternativo y principal	100%	

Fuente: Elaboración propia del autor

No obstante, un aspecto positivo a resaltar de ambos criterios es su grado de semejanza. Es decir, las 564 pymes que han sido detectadas como zombis por el criterio alternativo lo son también por el principal. Esto demuestra que ambos criterios son válidos para identificar a las zombis, con la única consecuencia de que el criterio alternativo identifica una proporción menor de dichas empresas. En lo que resta del estudio, debido a su alto grado de aceptación y utilización en la literatura de las zombis, **se empleará el criterio principal**.

Así, se concluye que el 8,2% de las pymes de hostelería son consideradas zombis, bajo el criterio que se utiliza en este estudio. Sin embargo, para una mayor facilidad, se puede utilizar el criterio alternativo que identifica una proporción (6.1%) no tan desorbitada a la del criterio principal. Además, llegados a este punto, se da por satisfecho el segundo criterio de éxito del proyecto por la alta semejanza observada entre los métodos aplicados.

6.2.2. Variables para la predicción de pymes zombis

En esta última sección de la fase de comprensión de los datos, se procede con la extracción desde *SABI* de todas aquellas partidas contables que permitirán la elaboración de las distintas ratios financieras. De este modo, serán tales ratios las que finalmente compondrán la imagen en escala de grises para cada una de las empresas de la muestra. Posteriormente dichas imágenes se suministrarán como entrada a la CNN, cuya tarea principal será determinar si se trata de una empresa zombi o no zombi.

Ya en la anterior sección se identificaron con dos criterios diferentes, las pymes zombis y no zombis del grupo de hostelería para el periodo 2019. Sin embargo, tal y como se señaló en la fase de comprensión del negocio, el principal objetivo de la minería de datos es la elaboración de una CNN que prediga con datos contables de periodos anteriores las pymes que pasarán al estado zombi. Por lo tanto, es lógico pensar que para la correcta predicción de las zombis se necesite recopilar partidas contables de periodos anteriores. Es por ello, que se emplean las partidas del año 2017 para predecir las empresas que serán zombis en el año 2019.

Así, con la ayuda de los Excel generados por el *notebook* de la sección anterior, se procede a extraer de *SABI* las partidas contables que figuran en la [Tabla 8](#) junto con el NIF, nombre, forma jurídica, código CNAE y fecha de constitución, para las pymes zombis (Excel **pymes_zombis_variables**) y no zombis (Excel **pymes_no_zombis_variables**). Además, como se puede observar, en dicha tabla también se recogen los números que identifican a estas partidas en *SABI*.

Tabla 8: Partidas contables *SABI*

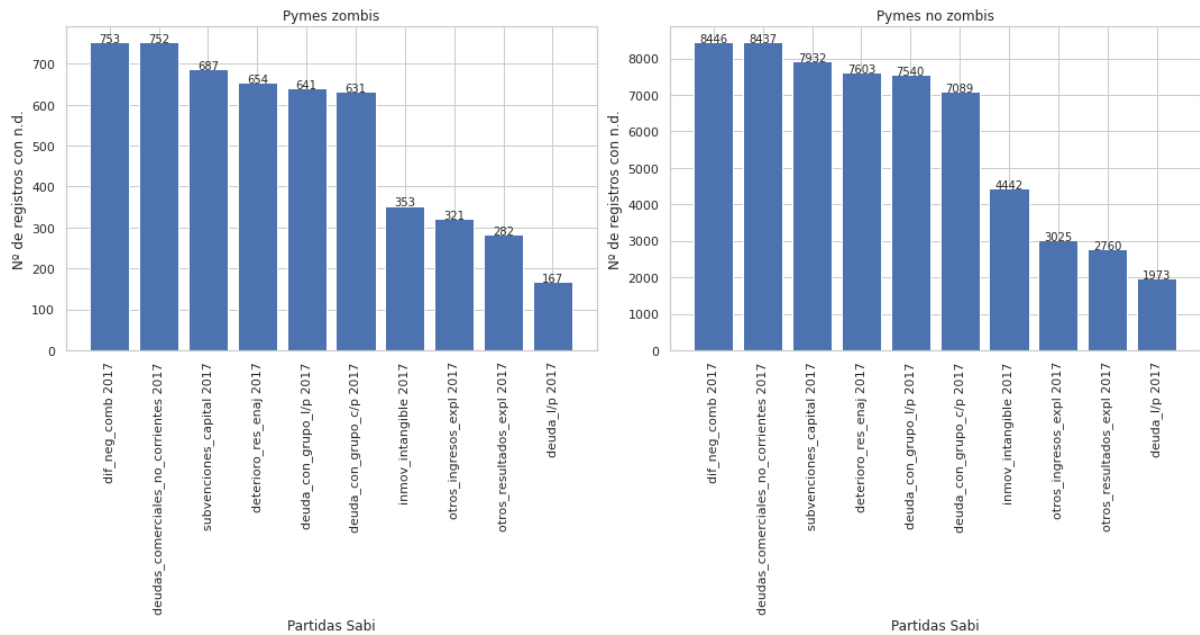
EBIT (732)	EBITDA (733)	Activo no corriente (11000)	Inmovilizado intangible (11100)	Inmovilizado material (11200)	Deudas comerciales no corrientes (11700)
Activo corriente (12000)	Existencias (12200)	Deudores comerciales y otras cuentas a cobrar (12300)	Efectivo y otros activos líquidos equivalentes (12700)	Patrimonio neto (20000)	Subvenciones, donaciones y legados recibidos (23000)
Pasivo no corriente (31000)	Deudas a largo plazo (31200)	Deudas con empresas del grupo y asociadas a largo plazo (31300)	Pasivo corriente (32000)	Deudas a corto plazo (32300)	Deudas con empresas del grupo y asociadas a largo plazo (32400)
Importe neto de la cifra de negocios (40100)	Otros ingresos de explotación (40500)	Deterioro y resultado por enajenaciones del inmovilizado (41100)	Diferencia negativa en combinaciones de negocios (41200)	Otros resultados (41300)	Gastos financieros (41500)
Resultado de explotación (49100)	Resultado del ejercicio (49500)				

Fuente: Elaboración propia del autor

Llegados a este punto, lo que corresponde en la presente fase es la exploración de los datos recogidos en la [Tabla 8](#). Sin embargo, como señalamos antes, dichos datos se emplearán en fases posteriores para derivar distintas variables que se convertirán en imágenes. Es por ello, que se pospone esta exploración hasta la siguiente fase.

Como última tarea de la fase actual es necesario verificar la calidad de los datos extraídos previamente, planteando cuestiones como: ¿los datos son completos? ¿contienen errores? ¿faltan elementos en los datos? De este modo la [Figura 15](#) muestra, tanto para las zombis como para las no zombis, las diez partidas que mayor número de registros *n.d.* presentan.

Figura 15: Top 10 partidas *SABI* con datos *n.d.*



Fuente: Elaboración propia del autor

Así, se concluye, que existen numerosas partidas contables que presentan datos faltantes para las zombis como para las no zombis. Dicha situación revela, que será necesario el tratamiento de estas en la fase de preparación de los datos antes de que sean transformadas en imágenes. Además, los resultados que se muestran en esta sección también se encuentran recogidos en el *notebook* llamado: **Predicción de las pymes zombis españolas**, que a partir de este momento será el *notebook* a seguir para el resto de las fases de la metodología *CRISP-DM*.

6.3. Preparación de los datos

Esta fase de la metodología *CRISP-DM* abarca todas las actividades necesarias para la construcción del *dataset* final a partir de los datos brutos de la fase anterior. Así, el objetivo final de esta sección, consiste en la elaboración de imágenes en escala de grises que se suministrarán como entrada a la CNN. Sin embargo, para lograr dicho objetivo es necesario previamente la selección, limpieza y construcción de los datos que compondrán tales imágenes.

De esta manera, se procede a dividir la actual fase en tres secciones. En la primera de ellas, se procede con el tratamiento de los datos faltantes así como la determinación de los datos a incluir e excluir. Posteriormente, se elaboran las diferentes variables que compondrán las imágenes. Y por último, se procede con la construcción de los *datasets* que se les suministrarán a los modelos de la siguiente fase.

6.3.1. Selección y tratamiento de los datos faltantes

El problema de los datos faltantes es bastante habitual en el ámbito de la ciencia de datos y en otros como economía y empresariales. Así, estos pueden llegar a reducir la potencia estadística de un estudio además de generar estimaciones sesgadas en el mismo, las cuales conducirían hacia conclusiones erróneas (Kang, 2013).

Sin embargo, antes de proceder con el tratamiento de los datos faltantes es imprescindible identificar la causa detrás de estos. Así, se pueden diferenciar entre tres tipos (Swalin, 2018):

- **MCAR** (*missing completely at random*): se producen cuando el valor de un dato faltante no está relacionado con su valor hipotético y con los valores de las variables del *dataset*.
- **MAR** (*missing at random*): suceden cuando la ausencia del dato depende de otras variables observadas en el mismo *dataset*.
- **MNAR** (*missing not at random*): se producen cuando el dato faltante sigue una tendencia, es decir existen razones para que el dato falte.

Teniendo en cuenta dichos tipos, el mecanismo que siguen los datos faltantes del presente estudio es el **MAR**. Esto se debe a que el dato que falta en una partida contable está intrínsecamente relacionado con otras partidas del *dataset*. Un ejemplo de ello es que, en la muestra existen empresas cuyo activo total está formado íntegramente por el activo no corriente. Esto supone que su activo corriente sea nulo y por consiguiente se produzca la ausencia de datos en todas las partidas de este último.

Una vez aclarada la causa detrás de los datos faltantes, se procede con el tratamiento de estos. Así, los dos métodos más utilizados para su tratamiento son: la **eliminación** o **imputación**. El primero de ellos es el más sencillo de implementar. Además, como su propio nombre indica, consiste en la eliminación de distintas muestras del *dataset*. Sin embargo, dicho método no debe llevarse a cabo cuando los datos faltantes sean del tipo MNAR debido a la posibilidad de introducir sesgo en el modelo (Swalin, 2018). Asimismo, en este método es posible diferenciar entre dos tipos de eliminaciones tradicionales:

- **Análisis con datos completos** (*listwise deletion*): consiste en eliminar todas las muestras que presentan algún dato faltante (Kang, 2013; Muñoz Hernández, 2018). Este tipo de eliminación puede suponer que se pierda una gran cantidad de muestras, lo que a su vez puede introducir sesgos en las estimaciones (Muñoz Hernández, 2018).

- **Análisis con datos disponibles (*pairwise deletion*)**: se tienen en cuenta todos los datos observados para cada una de las variables del *dataset* (Muñoz Hernández, 2018). De esta forma, se conserva una mayor cantidad de información en comparación al método anterior (Kang, 2013). No obstante, se puede dar el caso de que exista un número distinto de observaciones para cada variable (Swalin, 2018).

Al margen de las dos técnicas tradicionales comentadas, existe una tercera que consiste en eliminar aquellas variables que no son significativas y para las que el 60% de sus observaciones son datos faltantes (Swalin, 2018).

Por otro lado, los métodos de imputación sustituyen los datos faltantes por valores estimados (Gómez-García *et al.*, 2006). Así, es posible diferenciar entre los métodos de imputación **simples** y **múltiples** (Gómez-García *et al.*, 2006). Los primeros se basan en un algoritmo que estima un único valor que se utilizará para reemplazar al dato faltante, mientras que los segundos realizan múltiples estimaciones que se combinan para sustituir al dato faltante (Sotaquirá, 2021). Además, entre los métodos simples es posible la imputación por la media o mediana, por regresiones o por el vecino más cercano (*NNI: nearest neighbor imputation*), entre otras posibilidades.

En este estudio se emplean ambas metodologías para el tratamiento de los datos faltantes. Así, en primer lugar se procede con la eliminación de las variables: **diferencia negativa en combinaciones de negocios y deudas comerciales no corrientes**. Dichas variables, como puede observarse en la [Figura 15](#), presentan casi la totalidad de sus registros a *n.d.* ya sea para las zombis como para las no zombis.

A continuación, se eliminan todas las pymes que no presentan información sobre su cifra de negocios anual. Estas eliminaciones se deben a que, si una empresa no dispone de información sobre sus ventas, muy difícilmente presentará datos del resto de las partidas de la [Tabla 8](#). Posteriormente, también se eliminan las pymes que presentan datos faltantes en sus gastos financieros y en todas sus deudas remuneradas de la [Tabla 8](#), ya que estas partidas se emplearán para el cálculo de varias ratios financieras. Así, tras este paso la muestra de zombis se reduce de 753 a 647 y la de no zombis de 8.451 a 7.354, lo cual en términos porcentuales representa aproximadamente el 13%.

Llegados a este punto, aún existen partidas que disponen de valor *n.d.* en el *dataset*. Sin embargo, en la mayoría de las empresas investigadas, dichas partidas no presentan valor alguno

para los periodos desde 2019 hasta 2017. Además, no se considera oportuno la estimación de estas ya que se podría distorsionar la estructura económica-financiera de una empresa, generando así inconsistencia entre las partidas extraídas. Es por tal motivo, que se imputa el resto de valores *n.d.* por un cero.

Así, el autor del informe, es consciente que dicha imputación no es la mejor entre las disponibles. Pero, que puede cobrar sentido para partidas como: deudas con empresas del grupo y asociadas a corto y largo plazo, ya que la empresa puede no pertenecer a un grupo; deterioro y resultado por enajenaciones del inmovilizado, otros ingresos de explotación y otros resultados, ya que dichas partidas suelen emplearse por las empresas para recoger resultados ajenos a su actividad. Además, se ha comprobado que cuando la empresa no presenta deuda remunerada a corto y largo plazo es porque posee la totalidad de su deuda con empresas del grupo.

Como último detalle a comentar sobre el tratamiento de los datos faltantes, es que en la muestra existen pymes cuyos activos/pasivos totales están constituidos de manera íntegra por el activo/pasivo corriente o no corriente. Esto supone que las partidas que forman parte de elementos faltantes contengan también el valor *n.d.* en los estados contables de la empresa.

6.3.2. Elaboración de ratios propias

Una vez realizado el tratamiento de los datos faltantes, estos ya están disponibles para el cálculo de las diferentes ratios y variables financieras que formarán la imagen sobre las que la CNN entrenará.

De esta manera, en la [Tabla 9](#) se recogen todas aquellas variables financieras que han resultado ser significativas en los distintos modelos econométricos propuestos en la literatura de estas empresas (*e.g.* Chang *et al.*, 2021; Fang *et al.*, 2020; Goto & Wilbur, 2019; Hoshi, 2006). Dichos modelos en la gran mayoría de los casos, consisten en regresiones que intentan determinar la relación entre estas variables y la probabilidad de ser zombi. Sin embargo, no se ha encontrado evidencia de estudios que empleen las variables anteriores para la predicción de estas empresas, y menos aún con las CNN.

Es por ello, que basándose en las variables de la [Tabla 9](#) (véase [Tabla A4](#) para el cálculo de las diferentes partidas), el estudio intentará predecir todas aquellas pymes que pasarán al estado zombi en el año 2019. Además, como se puede observar en la misma tabla, se han seleccionado variables de distintas índole que representan: la estructura financiera de estas empresas, las

diferentes tipos de rentabilidades que obtienen estas de sus actividades y su capacidad para cubrir los gastos financieros, entre otras.

Un último detalle a aclarar antes de proceder con el cálculo de las variables comentadas es que para estas empresas no es posible calcular con certeza el importe total de las subvenciones recibidas. En este sentido, solo se puede determinar con exactitud el importe de las subvenciones de capital que figuran en la partida “subvenciones, donaciones y legados recibidos” del patrimonio neto. No obstante, las empresas también pueden recibir subvenciones de explotación que en el modelo normal figurarían en la partida de “subvenciones de explotación incorporadas al resultado del ejercicio”, pero que para el caso de estas empresas aparecen recogidas en la partida de “otros ingresos de explotación” de la cuenta de resultados. Es por ello, que el presente estudio basándose en una medida *proxy*: suma de subvenciones, donaciones y legados recibidos, y otros ingresos de explotación, intenta determinar el verdadero valor de las subvenciones recibidas por las pymes españolas.

Tabla 9: Variables para predicción de zombis

Tipo	Nombre	Cálculo
Endeudamiento (apalancamiento)	Total	$r_{end1} = \frac{\text{Pasivo total}}{\text{Activo total}}$
	Remunerado	$r_{end2} = \frac{\text{Deuda remunerada total}}{\text{Activo total}}$
Rentabilidad	Económica	$roa = \frac{\text{EBIT}}{\text{Activo total}}$
	Económica de explotación	$roa_{expl} = \frac{\text{Resultado explotación ajustado}}{\text{Activo funcional}}$
	Negocio	$r_{rent_negc} = \frac{\text{EBITDA}}{\text{Activo total}}$
	Financiera	$roe = \frac{\text{Resultado del ejercicio}}{\text{Patrimonio neto}}$
	Ventas	$ros = \frac{\text{EBIT}}{\text{Importe neto de la cifra de negocios}}$
Proporción de subvenciones	$r_{sub1} = \frac{\text{Subvenciones}}{\text{Importe neto de la cifra de negocios}}$	
	$r_{sub2} = \frac{\text{Subvenciones}}{\text{Activo total}}$	
	$r_{sub3} = \frac{\text{Subvenciones}}{\text{Resultado del ejercicio}}$	
Cobertura de gastos financieros	$icr_1 = \frac{\text{EBIT}}{\text{abs}(\text{Gastos financieros})}$	
	$icr_2 = \frac{\text{EBITDA}}{\text{abs}(\text{Gastos financieros})}$	
Capacidad de servicio de la deuda	$r_{csd} = \frac{\text{EBITDA}}{\text{Deuda remunerada total}}$	
Solvencia	$r_{solv} = \frac{\text{Activo total}}{\text{Pasivo total}}$	
Variables logarítmicas	Activo total	$\log_activos = \log(\text{activo total})$
	Pasivo total	$\log_pasivos = \log(\text{pasivo total})$
	Deuda remunerada total	$\log_drt = \log(\text{deuda remunerada total})$
	Importe neto de la cifra de negocios	$\log_cifra_negocios = \log(\text{importe neto de la cifra de negocios})$
Edad	Año 2017 – Fecha de constitución (en años)	

Fuente: Elaboración propia del autor

Teniendo en cuenta lo mencionado, se procede a calcular las variables anteriores con la ayuda del *notebook* principal del estudio. Además, en la [Tabla 10](#) y [Tabla 11](#) se aportan las medidas estadísticas básicas para las pymes zombis y no zombis, que permitirán realizar la exploración inicial que se había pospuesto en la fase de comprensión de los datos.

Tabla 10: Estadística descriptiva para pymes zombis

Variables	Media	Desviación típica	Min	Max
r_end1	1,06	1,09	0,01	11,93
r_end2	0,82	0,99	0,00	9,88
roa	-0,11	0,26	-4,08	0,19
roa_expl	-0,17	0,42	-4,05	0,27
r_rent_negc	-0,05	0,26	-3,98	0,32
roe	-0,20	3,29	-53,57	28,69
ros	-1,58	34,49	-877,39	0,23
r_sub1	0,04	0,24	0,00	4,14
r_sub2	0,02	0,06	0,00	0,79
r_sub3	1,15	12,41	-30,68	191,14
r_csd	-0,44	5,00	-107,82	5,16
r_solv	1,92	4,23	0,08	85,56
icr_1	6,46E+02	2,09E+04	-3,90E+04	5,31E+05
icr_2	1,40E+03	3,67E+04	-1,11E+04	9,35E+05
activo_total	4,38E+06	1,41E+07	1,83E+04	2,36E+08
pasivo_total	3,48E+06	1,52E+07	1,17E+04	2,68E+08
drt	3,10E+06	1,47E+07	2,40E+02	2,68E+08
cifra_negocios	1,40E+06	2,05E+06	3,00E+02	3,30E+07
edad	14,96	11,34	1,23	97,15

Fuente: Elaboración propia del autor

Tabla 11: Estadística descriptiva para pymes no zombis

Variables	Media	Desviación típica	Min	Max
r_end1	0,70	2,34	0,00	195,94
r_end2	0,43	2,08	0,00	173,65
roa	0,09	0,30	-3,06	21,65
roa_expl	0,14	0,58	-24,08	24,07
r_rent_negc	0,14	0,31	-3,04	21,76
roe	1,70	112,18	-217,48	9.515,47
ros	0,08	1,18	-43,35	68,43
r_sub1	0,12	4,37	0,00	311,36
r_sub2	0,02	0,09	0,00	3,30
r_sub3	27,14	2.327,78	-4.193,20	199.569,40
r_csd	744,99	60.508,59	-4.676,50	5.188.168,00
r_solv	7,69	434,00	0,01	37.217,66
icr_1	2,86E+04	1,88E+06	-2,92E+06	1,60E+08
icr_2	3,04E+04	1,89E+06	-1,78E+06	1,60E+08
activo_total	3,98E+06	1,78E+07	7,90E+02	9,76E+08
pasivo_total	1,92E+06	7,34E+06	1,01E+00	2,81E+08
drt	1,52E+06	6,93E+06	2,00E-02	2,75E+08
cifra_negocios	2,28E+06	3,76E+06	1,70E+03	1,35E+08
edad	16,51	12,54	1,22	129,34

Fuente: Elaboración propia del autor

De este modo, se puede observar que las pymes zombis poseen una mayor cantidad media de deudas, ya sean remuneradas o total, con respecto a las no zombis, característica que se había advertido en la sección 4.1.2. Así, dichas deudas suponen que sus ratios de endeudamiento (1,06; 0,82) también sean superiores a las de no zombis (0,70; 0,43), en valores medios.

Otra característica citada en la sección 4.1.2 y que se cumple en este análisis inicial de las variables, es la baja rentabilidad económica de estas empresas frente a las no zombis. En este

sentido, las zombis presentan un valor medio de ROA (-0,11) inferior a las no zombis (0,09), lo cual aparentemente demuestra la ineficiencia de estas empresas en la gestión de sus recursos económicos. Además, dicha ineficiencia también se refleja en la forma en la que estas gestionan su actividad principal y sus recursos financieros, obteniendo así una rentabilidad económica de explotación (-0,17) y financiera (-0,20) negativas. La misma situación se da en las rentabilidades de negocio (r_{rent_negc}) y de las ventas (ros), donde la última refleja que los beneficios medios de las pymes zombis por cada euro de sus ventas, son negativos (-1,58).

Con respecto a la capacidad de estas para hacer frente a los gastos financieros generados, se observa que las zombis son capaces de soportar tales gastos, sin embargo, sus ratios de coberturas, son significativamente inferiores que las no zombis (646; 1.399). Algo similar ocurre con su capacidad de servicio de la deuda, que refleja que estas empresas no son capaces de cubrir con sus resultados brutos de explotación la deuda remunerada total que poseen.

Por otra parte, si se observa el valor de las diferentes ratios de subvenciones propuestas no parece, al menos en esta exploración inicial, que las pymes zombis dependan exclusivamente de las subvenciones. En este sentido, sus cifras son relativamente inferiores que las no zombis, las cuales si parecen tener una alta participación de subvenciones en su estructura financiera.

Por último, como era de esperar, las pymes no zombis (7,69) son cuatro veces más solventes que las zombis (1,92). Además, destacar que las zombis poseen una edad media (15 años aproximadamente) menor que las no zombis (16,5 años) y que no se han observado diferencias significativas en el valor de los activos ya que se tratan de pymes pertenecientes al mismo sector.

6.3.3. Generación de las imágenes en escala de grises

Tras calcular las variables de la [Tabla 9](#) para todas las pymes de la muestra, es el momento de generar las imágenes en escala de grises que se le suministrarán como entrada a la CNN en la fase del modelado.

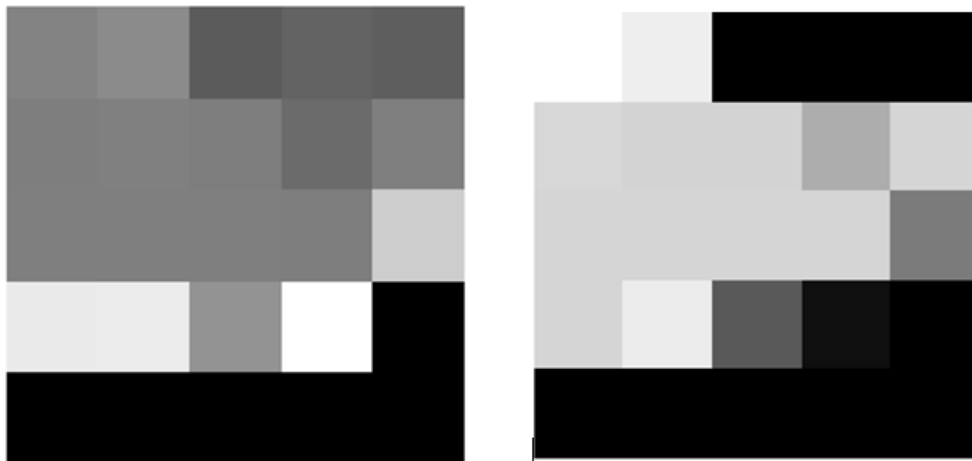
Así, para cada pyme de la muestra se genera una imagen que recoge las distintas variables calculadas previamente. A cada pixel de esta imagen se le asigna el valor de una de las variables de la empresa y cuyo brillo viene determinado por la siguiente ecuación extraída del estudio de Hosaka (2019):

$$\frac{v[R(i)]-m[R(i)]}{\sigma[R(i)]} \times 100 + 128 \quad (10)$$

Donde $v[R(i)]$ representa el valor de la variable, y donde $m[R(i)]$ y $\sigma[R(i)]$ son la media y desviación típica de dicha variable, respectivamente. Además, cuando el valor de la [ecuación 10](#) es mayor que el límite superior del brillo de un pixel: 255 o menor que el límite inferior: 0, se toma como 255 o 0, respectivamente (Hosaka, 2019). De este modo, si una variable financiera es mayor que la media del *dataset*, se le asigna al pixel correspondiente un color más cercano al blanco (255), y si es menor, pues un color más cercano al negro (0).

Sin embargo, con las 19 variables que se han calculado en la [Tabla 9](#) solo se pueden obtener imágenes cuadradas de 5x5 pixeles, algo que es bastante pequeño. Es por ello, que empleando la técnica del vecino más cercano se aumentan las dimensiones de estas imágenes a 224x224 pixeles. De este modo, en la [Figura 16](#) se pueden observar las imágenes en escala de grises para una empresa zombi y no zombi, respectivamente. Además, como se puede apreciar en la misma figura, los últimos 6 pixeles de estas imágenes están a negro. Esto se debe a la no correspondencia de variable alguna a tales pixeles. Por último, dichas imágenes que únicamente representan a las variables de la [Tabla 9](#), se guardan bajo dos directorios diferentes, uno para zombis y otro para no zombis, del *dataset* denominado como *dsa1*.

Figura 16: Imágenes con variables iniciales



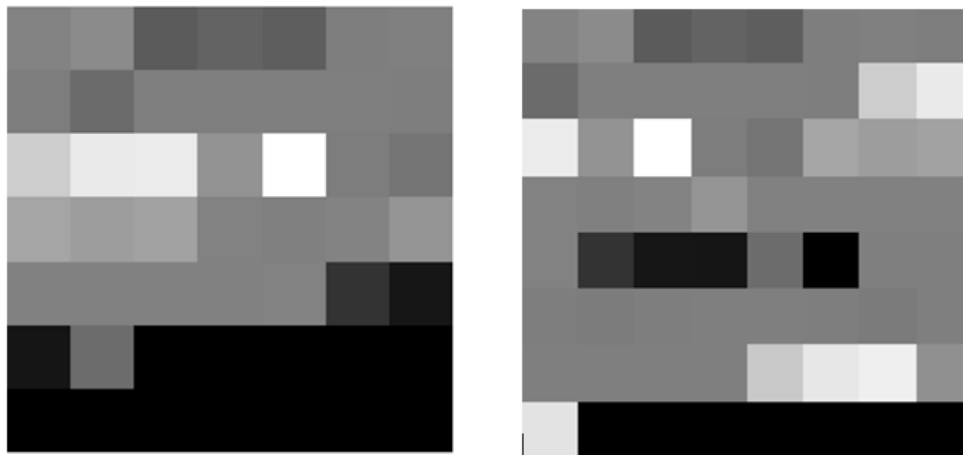
El lado izquierdo representa la imagen para una zombi mientras que el derecho la imagen para una no zombi

Fuente: Elaboración propia del autor

Como se ha visto, las imágenes generadas para el *dataset dsa1* solo recogen las variables inicialmente calculadas por el estudio. Y aunque estas han sido aumentadas en mayores dimensiones, existe cierta incertidumbre de que la CNN pueda discriminar correctamente entre las pymes zombis y no zombis. Por tal motivo, y con el fin de experimentar y obtener unas imágenes con mayores cantidades de información, se generan dos *datasets* adicionales denominados como: *dsa2* y *dsa3*. El primero de ellos contiene imágenes que están compuestas

por las variables iniciales del estudio y sus diferencias con sus respectivas medias. Mientras que el segundo, posee imágenes que representan la misma información que las del *dataset dsa2*, pero a las cuales se han incorporado cada una de las variables iniciales al cuadrado. De este modo, la [Figura 17](#) muestra como serían estas nuevas imágenes para la misma empresa zombi de la [Figura 16](#).

Figura 17: Imágenes con nuevas variables



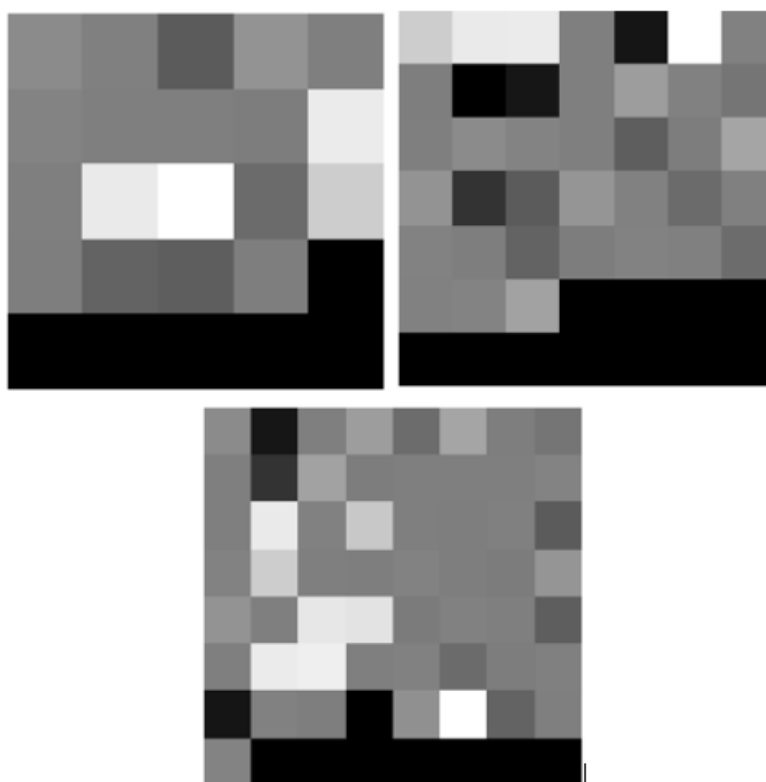
La imagen del lado izquierdo pertenece al *dataset dsa2* mientras que la del lado derecho al *dataset dsa3*

Fuente: Elaboración propia del autor

Además, siguiendo la estrategia anterior, se generan tres *datasets* adicionales: *dsc1*, *dsc2* y *dsc3*, cuyas imágenes se generan con el método correlacionado de Hosaka (2019). En dicho método, se determina la correspondencia entre las variables financieras y las posiciones de los píxeles, de modo que las variables altamente correlacionadas se colocan lo más cerca posible unas de otras (Hosaka, 2019). Así, el fundamento detrás de este método, es que las imágenes naturales para el reconocimiento de los objetos suelen tener correlaciones extremadamente altas entre sus píxeles vecinos (Hosaka, 2019), lo que facilita la detección para la CNN.

De esta manera, en la [Figura 18](#) se recogen las imágenes generadas por el método correlacionado para la misma empresa zombi de la [Figura 16](#) y [Figura 17](#). Sin embargo, es necesario recalcar que tales imágenes no siempre serán idénticas debido a la naturaleza aleatoria del método correlacionado de Hosaka (2019). Esto se debe a que el método se centra en la reducción de una función de energía. Así, si dicha energía puede reducirse intercambiando las posiciones de dos píxeles (que contienen variables) seleccionados al azar, el intercambio se lleva a cabo. Este último paso se repite hasta que no sea pueda reducir la energía en $3N$ veces, donde N es el número de variables en la imagen.

Figura 18: Imágenes con el método correlacionado



Las imágenes de la parte superior pertenecen a los *datasets dsc1* y *dsc2*, respectivamente y la imagen de la parte inferior al *dataset dsc3*.

Fuente: Elaboración propia del autor

Por último, la [Tabla 12](#) resume toda la información necesaria sobre los seis *datasets* generados en esta fase final de preparación de los datos.

Tabla 12: Información sobre *datasets* generados

Nombre	Nº de variables en las imágenes	Dimensiones iniciales (ancho x alto x canales)	Dimensiones finales (ancho x alto x canales)
dsa1, dsc1	19	5x5x1	224x224x3
dsa2, dsc2	38	7x7x1	
dsa3, dsc3	57	8x8x1	

dsa (*dataset normal*); dsc (*dataset correlacionado*)

Fuente: Elaboración propia del autor

6.4. Modelado

En esta fase lo que corresponde es la aplicación de diversas técnicas de modelado y la calibración de sus parámetros hasta alcanzar los valores óptimos. Así, esta sección se divide en cuatro partes bien diferenciadas. En la primera de ellas, se documentan las técnicas de modelado que se emplean en este estudio. A continuación, se especifica el plan a seguir para la división de los *dataset* generados. Posteriormente, se procede con la construcción de los distintos modelos para la predicción de las pymes zombis. Y, por último, se realiza la evaluación de los modelos propuestos con diferentes tipos de métricas.

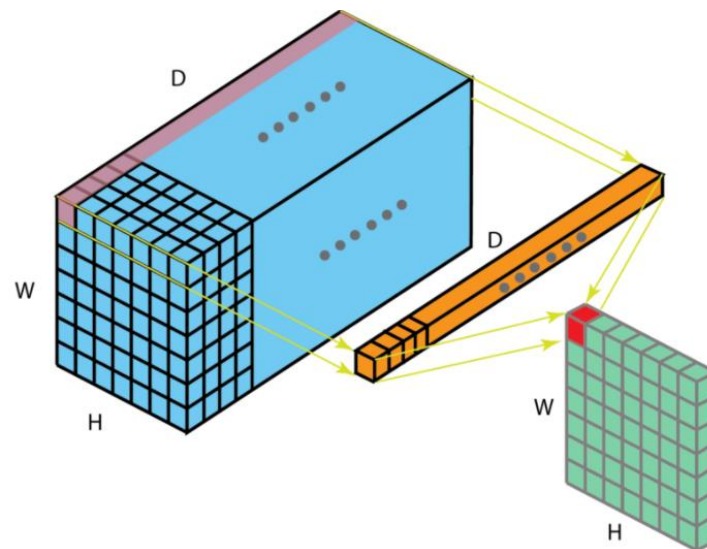
6.4.1. Selección de las técnicas de modelado

Como ya se ha reiterado en repetidas ocasiones, la principal técnica de modelado que se utiliza para este estudio son las CNN. Así, estas redes han demostrado ser muy eficientes sobre entradas que consisten en imágenes, ya sea para la detección de objetos en ellas o para su clasificación (González, 2017; Prabhu, 2018).

Ahora bien, en el estudio no se plantea la construcción de una CNN desde sus raíces, sino que se emplea la misma CNN que el estudio de Hosaka (2019) conocida como *GoogLeNet* (Szegedy *et al.*, 2016). Así, dicha red consta de 27 capas desde la entrada hasta la salida, y cuya configuración ha sido propuesta para ahorrar tiempo de cálculo y mejorar la precisión en la identificación (Hosaka, 2019). Sin embargo, los factores que han hecho que *GoogLeNet* tenga éxito son: los filtros o las operaciones de convolución 1×1 y su módulo llamado *Inception*.

Así, por un lado, dichas operaciones de convolución (Figura 19), permiten la reducción de los canales de entrada de la capa previa, logrando mantener el ancho y largo de la misma (Bai, 2019). Gracias a la alta aplicación de estas operaciones, *GoogLeNet* es capaz de reducir de manera considerable el número de parámetros de la red (Szegedy *et al.*, 2016).

Figura 19: Operación de convolución 1×1



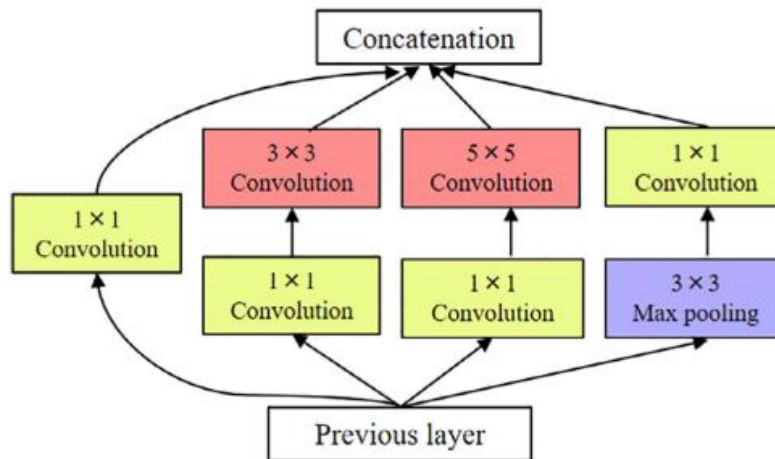
El tamaño del filtro es $1 \times 1 \times D$

Fuente: Extraído de Bai (2019)

Por otro lado, los módulos *Inception* consisten en una unidad muy sencilla, pero a su vez potente (Brownlee, 2019). Así, estos permiten al modelo aprender no solo los filtros paralelos de mismo tamaño, sino también los filtros paralelos de distintos tamaños, los cuales habilitan el aprendizaje a múltiples escalas (Brownlee, 2019). De este modo, las entradas procedentes de la

capa anterior se copian y se pasan por cuatro rutas, en las que el tamaño del filtro es fijo y se le aplican múltiples operaciones de convolución paralelamente (Hosaka, 2019). Finalmente, se concatenan las salidas de todas las anteriores rutas, las cuales avanzan hacia la siguiente capa de la red ([Figura 20](#)).

Figura 20: Módulo *Inception*



Fuente: Extraído de Hosaka (2019)

Con respecto a la arquitectura que sigue *GoogLeNet* ([Figura A1](#)), caben destacar los siguientes aspectos (Szegedy *et al.*, 2016):

- Esta espera una entrada del tamaño 224x224x3, es decir imágenes en formato RGB.
- En ella existen un total de nueve módulos *Inception* ([Figura 20](#)).
- La capa de salida utiliza la función de activación *softmax* para proporcionar 1.000 salidas que se corresponden con el número de categorías del *dataset ImageNet*. Así, dicho *dataset* fue sobre el que originalmente se realizó el entrenamiento de *GoogLeNet*.
- Y por último, existen dos redes adicionales ([Figura A1](#)) que permiten la discriminación en la etapas inferiores del clasificador final, aumentan la señal de gradiente que se obtiene en la retro propagación y proporcionan una regularización adicional.

Sin embargo, no se ha encontrado evidencia del uso de este tipo de redes para la predicción de las empresas zombis, lo cual como mencionamos, es una de las características a destacar del presente estudio. No obstante, un hecho que se repite con bastante frecuencia en los distintos estudios sobre estas empresas, es el empleo de las regresiones logarítmicas binarias para su identificación. Así, son tal el número de estudios que emplean dichas regresiones, que las mismas pueden ser consideradas como una técnica tradicional en la literatura de estas empresas.

Es por dicho motivo, que se seleccionan las regresiones logarítmicas binarias como la segunda técnica de modelado para el presente estudio.

Aunque, la principal razón de escoger una segunda técnica de modelado, se debe al hecho de comparar si de verdad es necesario este enorme esfuerzo cuando con las metodologías tradicionales puede ser posible la predicción de estas empresas de una manera más simple. Así, dichos modelos de regresiones servirán al estudio como una línea base con las que comparar los resultados obtenidos por la CNN y así determinar el éxito o el intento fallido de la prueba de concepto.

6.4.2. Generación del plan de prueba

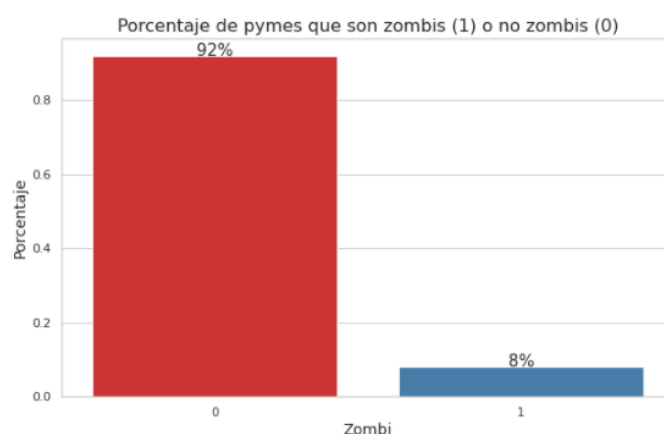
El objetivo principal de esta tarea consiste en determinar la división de los *datasets* generados en subconjuntos de entrenamiento, validación y test. No obstante, antes de proceder con dicha división, es necesario aclarar que *datasets* corresponden a cada técnica de modelado.

Así, la [Tabla 12](#) recoge todos los *datasets* alternativos sobre los que se entrena y evalúa la CNN propuesta en el estudio. En este sentido, se intenta determinar si realmente es significativo incluir una mayor cantidad de variables en las imágenes, además de comparar el rendimiento de la red sobre imágenes generadas con el método correlacionado. Con respecto a esto último, se espera que la CNN muestre una mayor eficacia sobre los *datasets* de las imágenes correlacionadas que las normales.

En cambio, para el modelo de la regresión logística binaria se utilizan, de manera alternativa, los registros de los distintos *dataframes*: *ratios*, *ratios_with_dif* y *ratios_with_sq*, elaborados en el *notebook* principal. Así, estos *dataframes* disponen para cada pyme las diferentes variables comentadas a lo largo del estudio, además de la etiqueta que le corresponde a cada una: zombi (0) o no zombi (1).

Una vez aclarada tal correspondencia, cada *dataset* dispone de 7.354 y 647 imágenes o registros, para las pymes no zombis y zombis, respectivamente. Sin embargo, existe un desbalanceo de datos significativo en cualquiera de los *datasets* generados ([Figura 21](#)). Tal desequilibrio puede afectar negativamente al rendimiento de la CNN (Pulgar-Rubio *et al.*, 2018) o a cualquier otro modelo de aprendizaje automático, perjudicando a las clases minoritarias (zombis). De este modo, si a la CNN se le pasan una gran cantidad de imágenes de pymes no zombis frente a las zombis, la red se limitará a responder que todas las empresas son no zombis, ya que no logra diferenciar una clase de la otra (Ignacio Bagnato, 2019).

Figura 21: Desbalanceo en los datos



Fuente: Elaboración propia del autor

Por lo tanto, con la intención de resolver dicho desequilibrio, en el estudio se emplea lo que se conoce como la **función de pérdida ponderada** (*Weighted Loss Function*). Esta consiste en otorgar diferentes pesos a las clases mayoritarias y minoritarias, los cuales influirán en la clasificación de las estas durante la fase de entrenamiento. Con ello el objetivo es penalizar la clasificación errónea de la clase minoritaria estableciendo un peso más alto y, al mismo tiempo reduciendo el peso para la clase mayoritaria (Singh, 2020).

Teniendo en cuenta lo comentado y con la ayuda de las librerías: *splitfolders* y *scikit-learn*, se procede con la división de los diferentes *datasets* en los subconjuntos de entrenamiento, validación y test. De este modo, para la CNN se selecciona de manera aleatoria el 80% de las muestras para el subconjunto de entrenamiento, mientras que solo el 10% para los subconjuntos de validación (también de forma aleatoria) y test.

Por último, para los modelos de regresiones logísticas se sigue la misma estrategia con respecto al subconjunto de entrenamiento, pero se selecciona el 20% para el test, debido a la no existencia del subconjunto de validación.

6.4.3. Construcción de los modelos

Tal y como se mencionó en la selección de las técnicas de modelado, la CNN que se utiliza en este trabajo es la conocida como *GoogLeNet*. Sin embargo, el principal inconveniente con el que se encuentra el estudio es que, *Keras* no dispone de dicha red entre sus modelos preentrenados y, por consiguiente, no es posible llevar a cabo la famosa técnica emergente en el campo del aprendizaje profundo denominada como aprendizaje por transferencia (*Transfer learning*).

Por dicho motivo, en el *notebook* principal se establecen las funciones necesarias para la construcción de esta CNN desde cero. Así, por un lado la función ***Inception_block()*** devuelve un módulo *Inception* tal como el visto en la [Figura 20](#). Mientras que la función ***GoogLeNet()*** basándose en la anterior, devuelve la CNN principal sobre la que se realizará el entrenamiento.

No obstante, un aspecto a mencionar de esta última función, es la modificación del número original de neuronas desde 1.000 a 1, en las capas de salida del clasificador final y de las dos redes adicionales del modelo. Esto se debe a que el presente estudio se centra en una clasificación binaria (zombi o no zombi) en lugar de una categórica. Además, debido a esta naturaleza binaria, también es necesario reemplazar la función de activación *softmax* por la ***sigmoid***, la cual asegura que la salida de la neurona este comprendida entre 0: no zombi o 1: zombi.

Con respecto a la función de pérdida de la CNN se selecciona la ***BinaryCrossentropy()*** que es la más adecuada para aplicaciones de clasificación binaria. Por otro lado, se utiliza el optimizador ***Adam()***, que consiste en un método de descenso del gradiente estocástico basado en la estimación adaptativa de los momentos de primer y segundo orden (Keras, 2021).

De esta manera, se obtienen un total de **nueve modelos**. Los primeros seis consisten en las CNN entrenadas y validadas sobre los *datasets* mencionados en la [Tabla 12](#). Mientras que los últimos tres consisten en regresiones logísticas binarias entrenadas sobre los *dataframes*: ***ratios***, ***ratios_with_dif*** y ***ratios_with_sq***, respectivamente.

Por último, debido al desequilibrio en los datos identificado en la sección anterior, se les suministra a los nueve modelos anteriores el parámetro: ***class_weight***, que se trata de un diccionario que contiene los distintos pesos para cada una de las clases a predecir. Así la [Figura 22](#) muestra el valor de estos pesos para cada una de las clases.

Figura 22: Pesos de las clases

```
▶ CLASS_WEIGHTS = {}

COUNT_Z = y1_train[y1_train['etiqueta'] == 1].shape[0]
COUNT_NZ = y1_train[y1_train['etiqueta'] == 0].shape[0]
COUNT = COUNT_Z + COUNT_NZ

CLASS_WEIGHTS[0] = (1/COUNT_NZ)*(COUNT)/2.0
CLASS_WEIGHTS[1] = (1/COUNT_Z)*(COUNT)/2.0

CLASS_WEIGHTS

[51]: {0: 0.5438477226376615, 1: 6.2015503875969}
```

Fuente: Elaboración propia de autor

6.4.4. Evaluación de los modelos

En esta última tarea de la fase del modelado, se evalúan los distintos modelos generados previamente. Sin embargo, antes de proceder con dicha tarea, es fundamental mencionar las distintas métricas que nos permitirán realizar una comparación adecuada entre estos modelos.

Así, como primera métrica a emplear, destaca la **matriz de confusión** (*confusion matrix*) la cual permite medir el rendimiento de un modelo de clasificación. Dicha matriz está formada por columnas, que representan el número de predicciones de cada clase y filas, que contienen las instancias de las clases reales ([Figura 23](#)). Además, combinando estas filas y columnas se obtienen las siguientes posibilidades para el presente estudio:

- Verdaderos negativos (*True negatives, TN*): el modelo predijo que la empresa es no zombi y efectivamente era una no zombi.
- Verdaderos positivos (*True positives, TP*): el modelo predijo que la empresa es zombi y efectivamente era una zombi.
- Falsos negativos (*False negatives, FN*): el modelo predijo que la empresa es no zombi cuando realmente era una zombi.
- Falsos positivos (*False positives, FP*): el modelo predijo que la empresa es zombi cuando realmente era una no zombi.

Figura 23: Matriz de confusión binaria

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Fuente: Extraído de Moocarme *et al.* (2019)

A partir de las diferentes combinaciones de las anteriores opciones, es posible obtener el resto de métricas a utilizar en este estudio. Así, una de las métricas más empleada en el campo del aprendizaje automático es la **exactitud** (*accuracy*) y se calcula como:

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (11)$$

Esta medida indica básicamente la cantidad de predicciones, tanto positivas como negativas, que fueron correctas. Sin embargo, la exactitud puede llegar a ser una medida engañosa cuando se trabaja con *datasets* no balanceados, como ocurre en este estudio. Es decir, un modelo puede obtener 90% de exactitud pero no haber identificado ninguna muestra de la clase minoritaria, lo cual genera una falsa sensación de que el modelo funciona correctamente (Ignacio Bagnato, 2019). Es por ello, que se tendrá un especial cuidado con la interpretación de esta medida al evaluar los distintos modelos generados.

Por otro lado, la **precisión** (*precision*) de una clase indica cuanto de confiable es un modelo al predecir que una muestra pertenece a dicha clase. Esta se calcula como:

$$\text{precisión zombi} = \frac{TP}{TP+FP} \quad (12)$$

Otras dos medidas que se pueden derivar de la matriz de confusión son la **sensibilidad** (*recall*) y **especificidad** (*specificity*). Dichas métricas indican la capacidad del modelo para discriminar los casos positivos de los negativos (Barrios Arce, 2019). Así, para este estudio la sensibilidad, o también conocida como tasa de verdaderos positivos (*True positive rates*, TPR), se define como la proporción de zombis que fueron identificadas correctamente por el modelo, mientras que la especificidad indica esa misma proporción pero para las no zombis. Estas se pueden calcular de la siguiente manera:

$$\text{sensibilidad} = \frac{TP}{TP+FN} \quad (13)$$

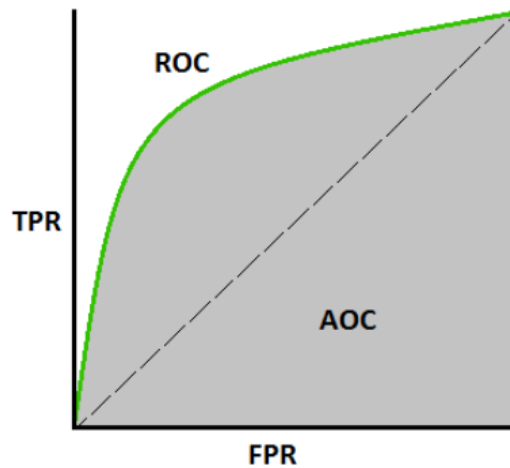
$$\text{especificidad} = \frac{TN}{TN+FP} \quad (14)$$

Asimismo, una métrica que es gran utilidad cuando se poseen datos no balanceados es la conocida como **F1-score**. Dicha métrica es muy empleada ya que resume la sensibilidad y especificidad en una única y se calcula como:

$$\frac{2*\text{sensibilidad}*especificidad}{\text{sensibilidad}+\text{especificidad}} \quad (15)$$

Por último, una de las métricas más importantes para comprobar el rendimiento de cualquier modelo de clasificación son las curvas **AUC-ROC** (Narkhede, 2018). ROC se trata de una curva de probabilidad calculada en base a la TPR y FPR (1-especificidad), mientras que AUC representa el grado o medida de separabilidad del modelo ([Figura 24](#)).

Figura 24: Curva AUC-ROC



Fuente: Extraído de Narkhede (2018)

De este modo, AUC indica la capacidad del modelo para distinguir entre las clases disponibles (Narkhede, 2018). Por lo tanto, cuanto mayor sea AUC, mejor será el modelo para distinguir entre empresas zombis y no zombis. Sin embargo, una AUC de 0.5 indica que el modelo no tiene capacidad de separación alguna (Narkhede, 2018). Asimismo, esta medida es la que se ha empleado para el entrenamiento y validación de la CNN, con el objetivo de lograr una AUC cercana del 1.

Tras explicar las métricas que se utilizan en este estudio, se procede con la evaluación de los distintos modelos generados en la sección anterior. De esta manera, la [Tabla 13](#) recoge todas las métricas enumeradas anteriormente para los distintos modelos del estudio.

Tabla 13: Métricas modelos

Modelo	Exactitud	Precisión		Sensibilidad	Especificidad	F1-score		AUC
		0	1			0	1	
model_dsa1	0,74	0,99	0,23	0,92	0,73	0,84	0,37	0,82
model_dsa2	0,79	1,00	0,27	0,97	0,77	0,87	0,43	0,87
model_dsa3	0,72	0,99	0,21	0,92	0,70	0,82	0,35	0,81
model_dsc1	0,92	0,92	0,00	0,00	1,00	0,96	0,00	0,5
model_dsc2	0,92	0,92	0,00	0,00	1,00	0,96	0,00	0,5
model_dsc3	0,08	0,00	0,00	1,00	0,00	0,00	0,15	0,5
logreg_1	0,77	0,97	0,22	0,72	0,78	0,86	0,34	0,75
logreg_2	0,77	0,97	0,22	0,71	0,78	0,86	0,34	0,74
logreg_3	0,65	0,92	0,08	0,34	0,68	0,78	0,14	0,51

model_dsa1: CNN entrenada sobre *dataset dsa1*; model_dsa2: CNN entrenada sobre *dataset dsa2*; model_dsa3: CNN entrenada sobre *dataset dsa3*; model_dsc1: CNN entrenada sobre *dataset dsc1*; model_dsc2: CNN entrenada sobre *dataset dsc2*; model_dsc3: CNN entrenada sobre *dataset dsc3*; logreg_1: regresión logística binaria sobre el *dataframe ratios*; logreg_2: regresión logística binaria sobre el *dataframe ratios_with_dif*; logreg_3: regresión logística binaria sobre el *dataframe ratios_with_sq*

Fuente: Elaboración propia del autor

La primera característica que destaca en dicha tabla, es que los modelos correlacionados no han logrado obtener un buen rendimiento en la predicción de las pymes zombis. En este sentido, para los tres modelos propuestos el valor de AUC es 0.5, lo cual significa que la CNN no presenta capacidad de separación alguna entre las zombis y no zombis. Las posibles razones que explican el fracaso de estos modelos son:

- La poca cantidad de información original disponible en las imágenes de los *datasets dsc1*, *dsc2* y *dsc3*, en comparación al estudio de Hosaka (2019). Así, el número original de variables para el presente estudio son 19, número considerablemente inferior frente al estudio de Hosaka (2019).
- La condición de parada del método correlacionado. Tal como se comentó en la sección [6.3.3](#), dicho método termina si por 3N veces consecutivas no es capaz de reducir la función de energía. Esto junto a la razón anterior, hacen que el método finalice relativamente pronto y no sea del todo eficiente al posicionar las variables más correlacionadas juntas.

Es por dichas razones, que se procede a descartar todos los modelos correlacionados debido a su baja capacidad de discriminación entre las pymes zombis y no zombis.

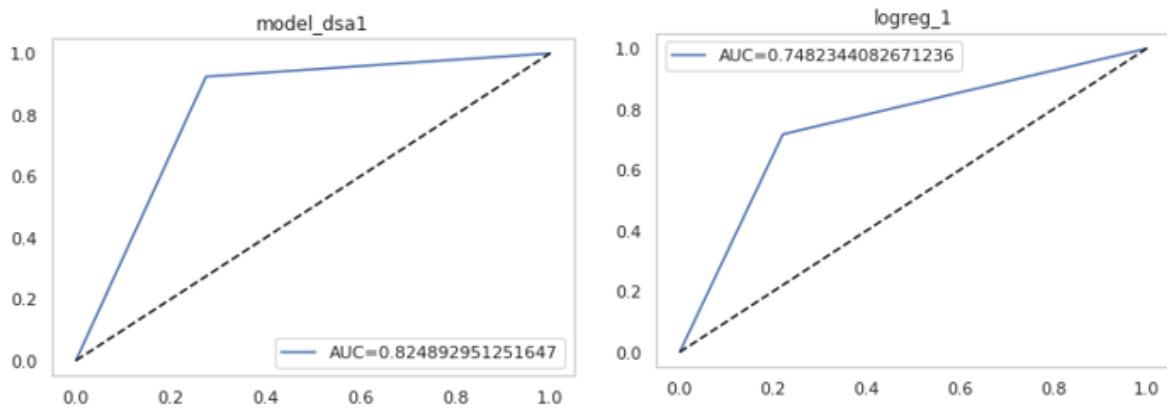
Por contrario, los modelos simples sí que logran obtener unas AUC cercanas a 0,9, lo que indica una buena capacidad de estos para distinguir entre las empresas zombis y no zombis. Además, todos estos modelos presentan unas sensibilidades altas y precisiones bajas para las pymes zombis. Esto último significa que dichos modelos logran detectar bien la clase zombi, aunque incluyen también muestras no zombis. Por otro lado, para las no zombis se da justo la situación

contraria, lo que supone que los modelos de aprendizaje profundo no detectan muy bien la clase, pero cuando lo hacen son altamente confiables.

Otro de los objetivos que se había propuesto, era determinar si de verdad es necesario el uso de las CNN cuando con las metodologías tradicionales es posible su predicción de manera más simple. A tal efecto, las regresiones *logreg_1* y *logreg_2* aun presentando unas AUC de 0.75 y 0.74, respectivamente, no logran superar a las AUC obtenidas por los modelos simples. Asimismo, dichas regresiones también presentan una menor sensibilidad en la identificación de las pymes zombis que la obtenida por las CNN. Todo esto se traduce en que, las CNN poseen una mejor capacidad para distinguir entre las pymes zombis y no zombis que las metodologías tradicionales. No obstante, un aspecto positivo a destacar de estas regresiones, es la obtención de unas tasas de predicción y especificidad similares a las obtenidas por los modelos simples de la [Tabla 13](#).

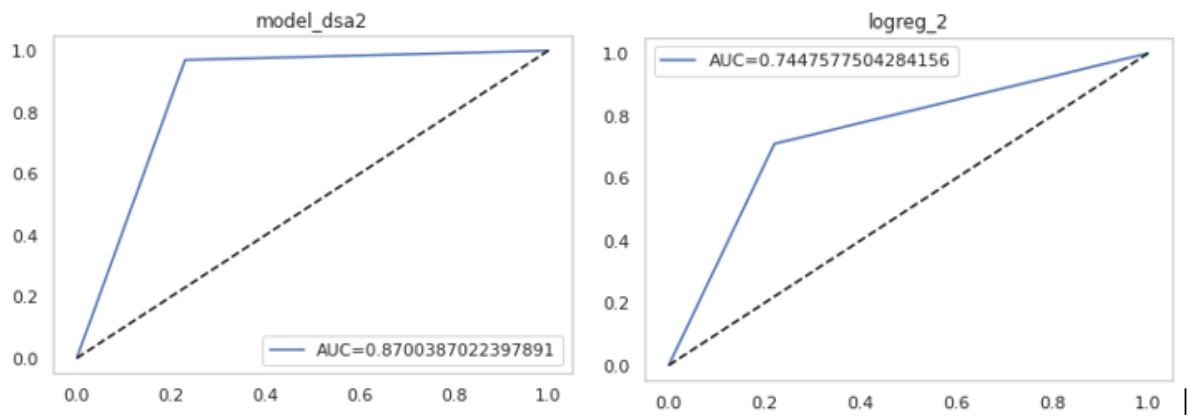
Por último, y con el fin de establecer una comparativa visual entre los modelos simples y las regresiones, se proporcionan las [Figura 25](#), [Figura 26](#) y [Figura 27](#) que muestran la curva AUC-ROC para estos modelos.

Figura 25: Curva AUC-ROC *model_dsa1-logreg_1*



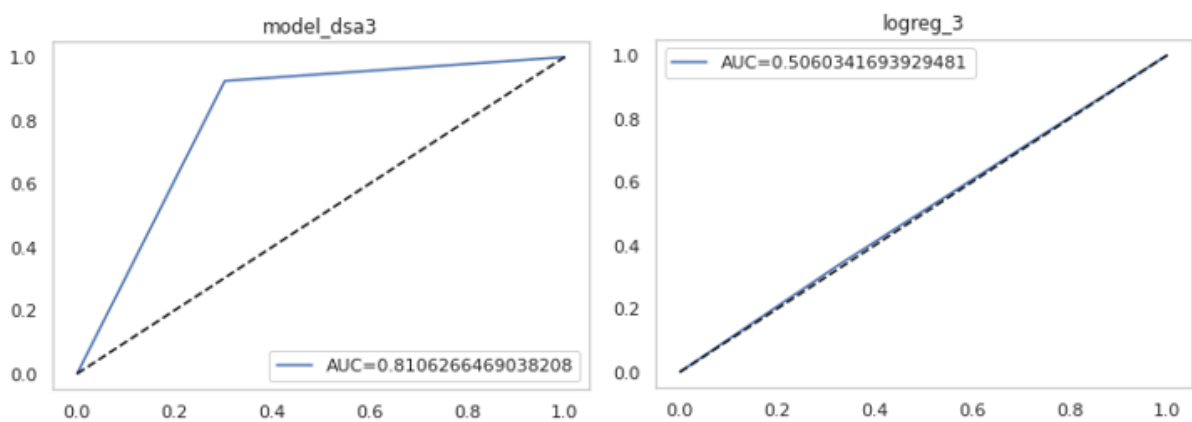
Fuente: Elaboración propia del autor

Figura 26: Curva AUC-ROC *model_dsa2-logreg_2*



Fuente: Elaboración propia del autor

Figura 27: Curva AUC-ROC *model_dsa3-logreg_3*



Fuente: Elaboración propia del autor

Como conclusión de esta última tarea de la fase de modelado, se da por cumplimentado el objetivo de la minería de datos del proyecto. Es decir, se ha demostrado que las CNN poseen la capacidad de predecir con datos contables de periodos anteriores, aquellas pymes españolas que pasarán al estado zombi en un futuro cercano. Además, estas presentan una mejor capacidad de distinción que las regresiones logísticas binarias. Sin embargo, un aspecto negativo a señalar de esta fase es el fracaso del método correlacionado en la predicción de dichas empresas.

6.5. Evaluación

La evaluación realizada en la fase anterior se refiere a los aspectos técnicos del modelo. Sin embargo, en esta fase se mide el grado en que dichos modelos cumplen con los objetivos iniciales del proyecto. Así al final de esta, se aprueban los modelos que mejor cumplan con los criterios de éxito del proyecto.

De este modo, como se puede comprobar en la [Tabla 13](#), el modelo que mejores métricas presenta en comparación a las regresiones logísticas, es el **model_dsa2**. Así, dicho modelo obtiene la AUC más cercana a 1 (0.87) y logra identificar en un 97% y 77%, a las pymes zombis

y no zombies, respectivamente. Es por dicha razón, que se selecciona el *model_dsa2* como el modelo que mejor cumple con el primer criterio de éxito del proyecto y el que continuará hacia la fase del despliegue. Asimismo, en la sección [6.2.1](#) ya había quedado satisfecho el segundo criterio de éxito del proyecto, que consistía en la identificación de las pymes zombies españolas con los dos criterios mencionados en la sección [4.1.4](#).

Así, llegados a este punto, se han cumplido los dos objetivos inicialmente propuestos para este proyecto. Por un lado, la prueba de concepto ha logrado obtener unos resultados prometedores al predecir las pymes zombies en el sector turístico español con la ayuda de las CNN. Mientras que por el otro y a través de la revisión bibliográfica, se ha aportado una mayor luz a la escasa literatura existente sobre estas empresas en España, además de su identificación mediante de dos criterios.

Por lo tanto, a la vista de los resultados obtenidos, no es necesaria otra nueva iteración de la metodología *CRISP-DM*, lo que a su vez significa la finalización del proyecto a falta del despliegue del modelo seleccionado.

6.6. Despliegue

Al tratarse el presente estudio de un trabajo de fin de titulación, la fase de despliegue consiste simplemente en la elaboración del presente informe además de su exposición ante los diferentes tribunales en el mes de febrero 2022.

Sin embargo, una de las tareas de esta fase que si se puede llevar a cabo es la revisión del proyecto. Así, en dicha tarea se realiza una retrospectiva sobre aspectos del proyecto que han salido bien, podrían haber salido mejor y mejoras de cara al futuro.

De esta manera, un aspecto que merece la pena recalcar es el éxito, con resultados prometedores, de la prueba del concepto del actual proyecto. En este sentido, se ha demostrado que las CNN permiten predecir, y además de una manera más efectiva que las metodologías tradicionales, las pymes zombies del sector turístico español.

Finalmente, un aspecto a mejorar del presente estudio es la implementación del método correlacionado, que no es la óptima para la generación de los distintas imágenes de los *datasets* correlacionados.

7. Conclusiones

El presente estudio ofrece una CNN basada en *GoogLeNet*, que tomando como entrada imágenes que representan diferentes variables de una pyme logra predecir si esta se convertirá en zombi en un futuro cercano (3 años).

Con esto, se demuestra que las CNN ofrecen resultados prometedores en la predicción de dichas empresas. Así, el mejor modelo del estudio obtiene una AUC de 0.87 y logra identificar en un 97% a las pymes zombis del sector turístico español. Además, estos resultados superan de manera significativa a los obtenidos por las regresiones logísticas binarias, las cuales consisten en la metodología tradicional en la literatura de estas empresas. Por lo tanto, merece la pena realizar este trabajo tedioso, con el fin de obtener un mayor rendimiento en la predicción de las pymes zombis.

Además, este estudio también logra aportar, a través de la revisión bibliográfica, una mayor luz a la escasa literatura existente sobre dichas empresas en España. Así, se demuestra que el 8,2% de las pymes de hostelería son consideradas zombis bajo el criterio principal del estudio (6,1% bajo el criterio alternativo).

Por último, es de señalar que estas empresas no son errores o equivocaciones en los datos de la contabilidad financiera, sino un grupo especial de empresas que no deberían ser ignoradas por los estudios de la gestión y finanzas. Debido a la creciente amenaza que estas suponen para el desarrollo económico sostenible, consideramos fundamental su predicción exitosa. Así, con la identificación temprana de estas, un sector puede mejorar la forma en la que se asignan sus recursos haciendo que estos se red desplieguen en actividades o empresas más productivas. Asimismo, esto puede posibilitar la entrada de nuevos participantes potenciales más productivos en dicho sector. Finalmente, con la identificación de estas, las empresas no zombis pueden ver mejoradas sus tasas de beneficios, sus oportunidades de crecimiento, un menor riesgo de salida, disminución de los costes de financiación y una menor restricción financiera debido a la no existencia de empresas que atan abundante capital financiero.

8. Futuras investigaciones

El autor del informe, así como los dos tutores académicos, opinan que las empresas familiares pueden presentar una relación positiva con la probabilidad de ser zombis. Así, se considera que por las relaciones que se dan en dichas empresas, estas pueden presentar una mayor resistencia en aceptar su quiebra y por consiguiente convertirse en zombi.

Sin embargo, en *SABI* no es nada fácil determinar que empresas son familiares. Esto se debe al mal formato que se sigue en dicha base de datos para los nombres y apellidos de los distintos grupos de interés de la empresa. Es por ello, que a las futuras investigaciones se le reclama la propuesta de un algoritmo que permita identificar a tales empresas y así poder incluir dicha variable en la predicción de las pymes zombis.

Por último, también se solicita que futuros trabajos incluyan una mayor cantidad de información original en las imágenes, ya que de esta manera se puede aumentar el rendimiento de la CNN. Asimismo, se anima a la utilización de otras arquitecturas de las CNN para la predicción de las pymes zombis.

9. Bibliografía

- Adalet McGowan, M., Andrews, D., & Millot, V. (2018). The walking dead? Zombie firms and productivity performance in OECD countries. *Economic Policy*, 33(96), 685–736. <https://doi.org/10.1093/epolic/eiy012>
- Ahearne, A. G., & Shinada, N. (2005). Zombie firms and economic stagnation in Japan. *International Economics and Economic Policy*, 2(4), 363–381. <https://doi.org/10.1007/s10368-005-0041-1>
- Alam, N., Gao, J., & Jones, S. (2021). Corporate Failure Prediction: An Evaluation of Deep Learning vs Discrete Hazard Models. *Journal of International Financial Markets, Institutions and Money*, 75(266), 101455. <https://doi.org/10.1016/j.intfin.2021.101455>
- Andrade Carrera, H., Sinche Maita, S., & Hidalgo Lascano, P. (2021). Modelo para detectar el uso correcto de mascarillas en tiempo real utilizando redes neuronales convolucionales. *Revista de Investigación En Tecnologías de La Información*, 9(17), 111–120. <https://doi.org/10.36825/riti.09.17.011>
- asesorias. (2021, March 18). *Microempresas. Concepto y características - Asesorías*. <https://asesorias.com/empresas/sociedades/microempresas/>
- Bai, K. (2019, February 11). *A Comprehensive Introduction to Different Types of Convolutions in Deep Learning | by Kunlun Bai | Towards Data Science*. <https://towardsdatascience.com/a-comprehensive-introduction-to-different-types-of-convolutions-in-deep-learning-669281e58215>
- BANCO DE ESPAÑA. (2021, October). *Banco de España - Estadísticas*. https://www.bde.es/webbde/es/estadis/infoest/temas/sb_tiif.html
- Barrios Arce, J. I. (2019, July 26). *La matriz de confusión y sus métricas – Inteligencia Artificial* –. <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>
- Blažková, I., & Dvoutělý, O. (2020). Zombies: Who are they and how do firms become zombies? *Journal of Small Business Management*, 2778. <https://doi.org/10.1080/00472778.2019.1696100>
- Brown L., G. (2020, December 6). *Scratch to SOTA: Build Famous Classification Nets 4 (GoogLeNet) | by Gary L. Brown | Medium*. <https://pasa34987.medium.com/scratch-to->

sota-build-famous-classification-nets-4-googlenet-2e3bac439462

- Brownlee, J. (2019, April 26). *How to Develop VGG, Inception and ResNet Modules from Scratch in Keras*. <https://machinelearningmastery.com/how-to-implement-major-architecture-innovations-for-convolutional-neural-networks/>
- Caballero, R. J., Hoshi, T., & Kashyap, A. K. (2008). Zombie lending and depressed restructuring in Japan. *American Economic Review*, 98(5), 1943–1977. <https://doi.org/10.1257/aer.98.5.1943>
- Calvo, D. (2018, December 8). *Perceptrón Multicapa - Red Neuronal - Diego Calvo*. <https://www.diegocalvo.es/perceptron-multicapa/>
- Caro Mora, C. (2021, August 18). ▷ ¿Qué es el trading? ¿Cómo funciona esta actividad? - Admirals. <https://admiralmarkets.com/es/education/articles/forex-basics/que-es-trading>
- Carreira, C., Teixeira, P., & Nieto-Carrillo, E. (2021). Recovery and exit of zombie firms in Portugal. *Small Business Economics*. <https://doi.org/10.1007/s11187-021-00483-8>
- Chang, Q., Zhou, Y., Liu, G., Wang, D., & Zhang, X. (2021). How does government intervention affect the formation of zombie firms? *Economic Modelling*, 94(December 2019), 768–779. <https://doi.org/10.1016/j.econmod.2020.02.017>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Rudiger, W. (2000). Crisp-Dm 1.0. *CRISP-DM Consortium*, 76.
- Chen, J. H., & Tsai, Y. C. (2020). Encoding candlesticks as images for pattern classification using convolutional neural networks. *Financial Innovation*, 6(1). <https://doi.org/10.1186/s40854-020-00187-0>
- Comisión Europea. (2016). *Guía del usuario sobre la definición del concepto de pyme*. <https://doi.org/10.2873/2552>
- Dai, X., Qiao, X., & Song, L. (2019). Zombie firms in China's coal mining sector: Identification, transition determinants and policy implications. *Resources Policy*, 62(November 2018), 664–673. <https://doi.org/10.1016/j.resourpol.2018.11.016>
- Data Science Process Alliance. (2021). *CRISP-DM - Data Science Process Alliance*. <https://www.datascience-pm.com/crisp-dm-2/>

- De la Garza Cárdenas, M. H., Sánchez Tovar, Y., & Zerón Félix, M. (2019). Impacto de la rentabilidad y la cuota de mercado en las empresas zombis en México. *Revista Mexicana de Economía y Finanzas*, 14(4), 729–743. <https://doi.org/10.21919/remef.v14i4.107>
- Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction. *IJCAI International Joint Conference on Artificial Intelligence, 2015-Janua(Ijcai)*, 2327–2333.
- Dong, X., Duan, H., Wang, T., & Liu, Q. (2020). An Adaptive Weighted Bagging Ensemble Learning Model for Zombie Enterprise Identification. *ICEIEC 2020 - Proceedings of 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication*, 273–276. <https://doi.org/10.1109/ICEIEC49280.2020.9152215>
- Du, W., & Li, M. (2019). Can environmental regulation promote the governance of excess capacity in China's energy sector? The market exit of zombie enterprises. *Journal of Cleaner Production*, 207, 306–316. <https://doi.org/10.1016/j.jclepro.2018.09.267>
- El Ghouli, S., Fu, Z., & Guedhami, O. (2021). Zombie firms: Prevalence, determinants, and corporate policies. *Finance Research Letters*, 41(December 2020), 101876. <https://doi.org/10.1016/j.frl.2020.101876>
- Elsevier. (2021). *Scopus / La mayor base de datos de bibliografía revisada por pares*. <https://www.elsevier.com/es-es/solutions/scopus>
- Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An Introductory Review of Deep Learning for Prediction Models With Big Data. *Frontiers in Artificial Intelligence*, 3(February), 1–23. <https://doi.org/10.3389/frai.2020.00004>
- Fang, J., Gozgor, G., Lau, C. K. M., Wu, W., & Yan, C. (2020). Listed zombie firms and top executive gender: Evidence from an emerging market. *Pacific Basin Finance Journal*, 62(February), 1–17. <https://doi.org/10.1016/j.pacfin.2020.101357>
- FEET. (2021). *FEET / GADE-Objetivos y competencias del título*. <https://feet.ulpgc.es/content/gade-objetivos-y-competencias-del-titulo>
- Fernández Fernández, E. E. (2018, March 26). *¿Qué es el Trading?. Marzo 26. 4 minutos de lectura / by Efraín Enrique Fernández Fernández / Medium*. <https://medium.com/@efrainenriquefernandez/qué-es-el-trading-2d7815efc48a>

- Fu, X., Qi, J., Zou, L., & Hou, Y. (2019). Analysis on evolution model of zombie company under the absence of bank data. *Proceedings of the Annual Hawaii International Conference on System Sciences, 2019-Janua*, 1323–1329.
<https://doi.org/10.24251/hicss.2019.161>
- Fukuda, S. ichi, & Nakamura, J. ichi. (2011). Why did “zombie” firms recover in Japan? *World Economy*, 34(7), 1124–1137. <https://doi.org/10.1111/j.1467-9701.2011.01368.x>
- García-Olalla Olivera, O. (2019, September 16). *Redes Neuronales artificiales: Qué son y cómo se entrenan* / Xeridia. <https://www.xeridia.com/blog/redes-neuronales-artificiales-que-son-y-como-se-entrenan-parte-i>
- Geng, Y., Liu, W., & Wu, Y. (2021). How do zombie firms affect China’s industrial upgrading? *Economic Modelling*, 97(January), 79–94.
<https://doi.org/10.1016/j.econmod.2021.01.010>
- GOBIERNO DE ESPAÑA. (2021a, April 28). *BOE.es - BOE-A-2010-10544 Real Decreto Legislativo 1/2010, de 2 de julio, por el que se aprueba el texto refundido de la Ley de Sociedades de Capital*. <https://www.boe.es/buscar/act.php?id=BOE-A-2010-10544>
- GOBIERNO DE ESPAÑA. (2021b, November 3). *BOE.es - BOE-A-1996-8930 Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia*. <https://www.boe.es/buscar/act.php?id=BOE-A-1996-8930>
- Gómez-García, J., Palarea-Albaladejo, J., & Martín-Fernández, J. A. (2006). Métodos de inferencia estadística con datos faltantes. Estudio de simulación sobre los efectos en las estimaciones. *Estadística Española*, 48, 241–270.
- González, J. C. (2017, November 20). *Use of convolutional neural network for image classification* / Blog. <https://www.apsl.net/blog/2017/11/20/use-convolutional-neural-network-image-classification/>
- Goto, Y., & Wilbur, S. (2019). Unfinished business: Zombie firms among SME in Japan’s lost decades. *Japan and the World Economy*, 49(December 2017), 105–112.
<https://doi.org/10.1016/j.japwor.2018.09.007>
- Guerra Artal, C. (2021). *Jupyter Notebook Viewer*.

<https://nbviewer.org/url/cayetanoguerra.github.io/ia/nbpy/redneuronal6.ipynb>

- Gunduz, H., Yaslan, Y., & Cataltepe, Z. (2017). Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations. *Knowledge-Based Systems*, 137, 138–148. <https://doi.org/10.1016/j.knosys.2017.09.023>
- Han, S., Li, G., Lubrano, M., & Xun, Z. (2020). Lie of the weak: Inconsistent corporate social responsibility activities of Chinese zombie firms. *Journal of Cleaner Production*, 253, 119858. <https://doi.org/10.1016/j.jclepro.2019.119858>
- Han, S., You, W., & Nan, S. (2019). Zombie firms, external support and corporate environmental responsibility: Evidence from China. *Journal of Cleaner Production*, 212, 1499–1517. <https://doi.org/10.1016/j.jclepro.2018.12.136>
- He, Q., Li, X., & Zhu, W. (2020). Political connection and the walking dead: Evidence from China's privately owned firms. *International Review of Economics and Finance*, 69(December 2018), 1056–1070. <https://doi.org/10.1016/j.iref.2018.12.007>
- Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, 117, 287–299. <https://doi.org/10.1016/j.eswa.2018.09.039>
- Hoshi, T. (2006). Economics of the living dead. *Japanese Economic Review*, 57(1), 30–49. <https://doi.org/10.1111/j.1468-5876.2006.00354.x>
- IBERDROLA. (2020). ¿ *Qué es el deep learning y por qué es clave para la inteligencia artificial ?* 1–6. <https://www.iberdrola.com/innovacion/deep-learning>
- Ignacio Bagnato, J. (2018, November 29). *Convolutional Neural Networks: La Teoría explicada en Español | Aprende Machine Learning*. <https://www.aprendemachinelearning.com/como-funcionan-las-convolutional-neural-networks-vision-por-ordenador/>
- Ignacio Bagnato, J. (2019, May 16). *Clasificación con datos desbalanceados | Aprende Machine Learning*. <https://www.aprendemachinelearning.com/clasificacion-con-datos-desbalanceados/>
- Imai, K. (2016). A panel study of zombie SMEs in Japan: Identification, borrowing and investment behavior. *Journal of the Japanese and International Economies*, 39, 91–107.

- <https://doi.org/10.1016/j.jjie.2015.12.001>
- INFORMA. (2021). *SABI: Análisis de Balances de Empresas - Informa D&B*.
<https://www.informa.es/riesgo-empresarial/sabi>
- Jiang, X., Li, S., & Song, X. (2017). The mystery of zombie enterprises – “stiff but deathless.” *China Journal of Accounting Research*, 10(4), 341–357.
<https://doi.org/10.1016/j.cjar.2017.08.001>
- Juan Pérez, J., & Cantero, L. (2021, March 30). *Qué es la Inteligencia Artificial - BLOG - IA SOLVER*. <https://iasolver.es/que-es-la-inteligencia-artificial/>
- jupyter. (2021, December 11). *Project Jupyter / Home*. <https://jupyter.org/>
- kaggle. (2022). *Kaggle: Your Home for Data Science*. <https://www.kaggle.com/>
- KANE, E. J. (1987). DANGERS OF CAPITAL FORBEARANCE: THE CASE OF THE FSLIC AND “ZOMBIE” S&Ls. *Contemporary Economic Policy*, 5(1), 77–83.
<https://doi.org/10.1111/j.1465-7287.1987.tb00247.x>
- Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, 64(5), 402–406. <https://doi.org/10.4097/kjae.2013.64.5.402>
- Keras. (2021). *Adam*. <https://keras.io/api/optimizers/adam/>
- Lam, W. R., Schipke, A., Tan, Y., & Tan, Z. (2017a). Resolving China’s Zombies: Tackling Debt and Raising Productivity. *IMF Working Papers*, 17(266), 1.
<https://doi.org/10.5089/9781484330722.001>
- Lam, W. R., Schipke, A., Tan, Y., & Tan, Z. (2017b). Resolving China ’ s Zombies : Tackling Debt and Raising Productivity. *IMF Working Papers WP/17/266*, 26.
- Liu, G., Zhang, X., Zhang, W., & Wang, D. (2019). The impact of government subsidies on the capacity utilization of zombie firms. *Economic Modelling*, 83(September), 51–64.
<https://doi.org/10.1016/j.econmod.2019.09.034>
- Malik, F. (2019a, March 9). *Everything About Python — Beginner To Advanced / by Farhad Malik / FinTechExplained / Medium*. <https://medium.com/fintechexplained/everything-about-python-from-beginner-to-advance-level-227d52ef32d2>
- Malik, F. (2019b, March 31). *Why Should We Use NumPy?. Understanding NumPy Features*

- Before... | by Farhad Malik | FinTechExplained | Medium.*
<https://medium.com/fintechexplained/why-should-we-use-numpy-c14a4fb03ee9>
- matplotlib. (2021). *Matplotlib documentation — Matplotlib 3.5.0 documentation.*
<https://matplotlib.org/stable/index.html#>
- Moocarme, M., Bhagwat, R., & Abdollahnejad, M. (2019, April). *Confusion Matrix | Applied Deep Learning with Keras.*
https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781838555078/6/ch06lv11sec34/confusion-matrix
- Muñoz Hernández, B. (2018). *Análisis experimental de técnicas de imputación en bases con datos faltantes.*
- Myrianthous, G. (2021, August 28). *Difference Between scikit-learn and sklearn | Towards Data Science.* <https://towardsdatascience.com/scikit-learn-vs-sklearn-6944b9dc1736>
- Narkhede, S. (2018, June 26). *Understanding AUC - ROC Curve | by Sarang Narkhede | Towards Data Science.* <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>
- NumPy. (2021, June 22). *What is NumPy? — NumPy v1.21 Manual.*
<https://numpy.org/doc/stable/user/whatisnumpy.html>
- Nurmi, S., Vanhala, J., & Virén, M. (2020). *Bank of Finland Research Discussion Papers 8 • 2020 The life and death of zombies-evidence from government subsidies to firms Bank of Finland Research* (Issue May).
- pandas. (2021, December 12). *pandas - Python Data Analysis Library.*
<https://pandas.pydata.org/>
- Papworth, T. (2013). *The Trading Dead: The zombie firms plaguing Britain's economy, and what to do about them.*
- Peek, J., & Rosengren, E. S. (2003). ©2003 by Joe Peek and Eric S. Rosengren All rights reserved. Short sections of text not to exceed two paragraphs, may be quoted without explicit permission provided that full credit including ©notice, is given to the source.
- Peña, L. (2020, July 29). *¿Qué es una Microempresa? Definición y Especificaciones - Billin.*
<https://www.billin.net/blog/microempresa/>

- pillow. (2022, January 2). *Pillow — Pillow (PIL Fork) 9.0.0 documentation*.
<https://pillow.readthedocs.io/en/stable/>
- Prabhu. (2018, March 4). *Understanding of Convolutional Neural Network (CNN) — Deep Learning / by Prabhu / Medium*. <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>
- Pulgar-Rubio, F., Rivera-Rivas, A. J., Charte, F., & Díaz, M. J. del J. (2018). Análisis del impacto de datos desbalanceados en el rendimiento predictivo de redes neuronales convolucionales. *XVIII Conferencia de La Asociación Española Para La Inteligencia Artificial (CAEPIA 2018)*, 1213. <http://benchmark.ini.rub.de>
- python. (2021). *What is Python? Executive Summary / Python.org*.
<https://www.python.org/doc/essays/blurb/>
- Rojo Ramírez, A. A., Diéguez Soto, J., & López Delgado, P. (2011). Importancia del concepto de Empresa Familiar en investigación: utilización de la base de datos SABI para su clasificación. *European Journal Of Family Business*, 1(1), 53–67.
<https://doi.org/10.24310/ejfbefb.v1i1.5034>
- Schivardi, F., Sette, E., & Tabellini, G. (2017). Credit Misallocation During the European Financial Crisis. *SSRN Electronic Journal*, 66. <https://doi.org/10.2139/ssrn.3051189>
- scikit-learn. (2022). *scikit-learn: machine learning in Python — scikit-learn 1.0.2 documentation*. <https://scikit-learn.org/stable/>
- Sezer, O. B., & Ozbayoglu, A. M. (2018). Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Applied Soft Computing Journal*, 70, 525–538. <https://doi.org/10.1016/j.asoc.2018.04.024>
- Shen, G., & Chen, B. (2017). Zombie firms and over-capacity in Chinese manufacturing. *China Economic Review*, 44(October 2016), 327–342.
<https://doi.org/10.1016/j.chieco.2017.05.008>
- Singh, K. (2020, October 6). *How To Dealing With Imbalanced Classes in Machine Learning*. <https://www.analyticsvidhya.com/blog/2020/10/improve-class-imbalance-class-weights/>
- Sotaquirá, M. (2021, June 15). *Guía completa para el Manejo de Datos Faltantes / Codificando Bits*. <https://www.codificandobits.com/blog/manejo-datos-faltantes/#la->

- Storz, M., Koetter, M., & Setzer, R. (2003). Do we want these two to tango? On zombie firms and stressed banks in Europe. In *Review* (Vol. 85, Issue 6).
<https://doi.org/10.20955/r.85.67>
- Swalin, A. (2018, January 31). *How to Handle Missing Data*. “The idea of imputation is both... | by Alvira Swalin | Towards Data Science. <https://towardsdatascience.com/how-to-handle-missing-data-8646b18db0d4>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2016). Going Deeper with Convolutions. *Journal of Chemical Technology and Biotechnology*, 91(8), 2322–2330. <https://doi.org/10.1002/jctb.4820>
- Tan, Y., Huang, Y., & Woo, W. T. (2016). Zombie Firms and the Crowding-Out of Private Investment in China. *Asian Economic Papers*, 15(3), 32–55.
https://doi.org/10.1162/ASEP_a_00474
- Terra, J. (2021, May 25). *Keras vs Tensorflow vs Pytorch: Popular Deep Learning Frameworks*. <https://www.simplilearn.com/keras-vs-tensorflow-vs-pytorch-article>
- Tsantekidis, A., Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M., & Iosifidis, A. (2017). Forecasting stock prices from the limit order book using convolutional neural networks. *Proceedings - 2017 IEEE 19th Conference on Business Informatics, CBI 2017*, 1, 7–12.
<https://doi.org/10.1109/CBI.2017.23>
- ULPGC. (2017). *Reglamento para la realización y evaluación de Trabajos de Fin de Título. Facultad de Economía, Empresa y Turismo*. https://feet.ulpgc.es/sites/default/files/wp-content/uploads/2017/11/nuevo_reglamento_tft_feet_2017_con_cambios_serv._juridicos_para_web.pdf
- ULPGC. (2020). *especificidades de los Trabajos de Fin de Título E SCUELA DE INGENIERÍA INFORMÁTICA*.
- ULPGC. (2021a). *Reglamento General De Los Trabajos De Fin De Grado Y De Fin De Máster De Las Titulaciones Oficiales De La Universidad De Las Palmas De Gran Canaria*. https://www.ulpgc.es/sites/default/files/ArchivosULPGC/Normativa_y_reglamentos/Reglamentos/reglamento_general_de_los_trabajos_de_fin_de_grado_y_de_fin_de_master_de_las_titulaciones_oficiales_de_la_ulpgc_boulpgc_5_diciembre_2019_t

ransitorio.pdf

ULPGC. (2021b). *ULPGC - Titulaciones ULPGC*.

https://www2.ulpgc.es/index.php?pagina=plan_estudio&ver=pantalla&numPantalla=01&codTitulacion=4008&codPlan=40&tipotitulacion=G

Urionabarrenetxea, S., Garcia-Merino, J. D., San-Jose, L., & Retolaza, J. L. (2018). Living with zombie companies: Do we know where the threat lies? *European Management Journal*, 36(3), 408–420. <https://doi.org/10.1016/j.emj.2017.05.005>

Velotio Technologies. (2019, June 25). *The Ultimate Beginner's Guide to Jupyter Notebooks* | by Velotio Technologies | Velotio Perspectives | Medium. <https://medium.com/velotio-perspectives/the-ultimate-beginners-guide-to-jupyter-notebooks-6b00846ed2af>

Verdence, C. (2020, October 29). *What is an Activation Function in a Neural Network?* | by Chris Verdence | Writers' Blokke | Medium. <https://medium.com/writers-blokke/what-is-an-activation-function-in-a-neural-network-ae0eb2bd2f82>

Vida Barea, J. (2021). *CNAE 2009 grupo I*.

<https://www.cnae.com.es/actividades.php?grupo=I>

Wang, Y., & Zhu, Y. (2021). The Financing and Investment Crowding-out Effect of Zombie Firms on Non-zombie Firms: Evidence from China. *Emerging Markets Finance and Trade*, 57(7), 1959–1985. <https://doi.org/10.1080/1540496X.2019.1711370>

Wu, H., Yang, M., & Gu, J. (2021). Are zombie firms more incentivized to financialize? *China Journal of Accounting Research*, 100199. <https://doi.org/10.1016/j.cjar.2021.05.003>

Yu, M., Guo, Y. M., Wang, D., & Gao, X. (2021). How do zombie firms affect debt financing costs of others: From spillover effects views. *Pacific Basin Finance Journal*, 65(September 2020), 101471. <https://doi.org/10.1016/j.pacfin.2020.101471>

Zhu, H., & Fan, H. E. (2016). Quantity and Characteristics of Zombie Firms in China. *Journal of Beijing Technology & Business University*, 31(2016), 2021.

Anexo I

Tabla A1: Abreviaturas empleadas en *el* anexo

Comunes	
profit	Beneficio de la empresa
rl	Ídem ecuación 1
rs	Ídem ecuación 1
B	Préstamos bancarios (a corto y largo plazo)
BL	Ídem ecuación 1
BL1	Pasivos a largo plazo
BS	Ídem ecuación 1
BS1	Pasivos a corto plazo menos las cuentas, los impuestos y otras partidas por pagar. Mide la deuda bancaria a corto plazo.
EBIT	Beneficio antes de intereses e impuestos
EBITDA	Resultados antes de intereses, impuestos, depreciaciones y amortizaciones
ICR	Ratio de cobertura de intereses
R	Pago de intereses reales de la empresa
R*	Pago mínimo de intereses exigidos a la empresa
ROA	Rentabilidad económica
Zombie lending and depressed restructuring in Japan (2008)	
x	Diferencia entre pago de intereses, ídem a la ecuación 2
Zombie Firms and the Crowding-Out of Private Investment in China (2016)	
gap	Diferencia entre pago de intereses
rd	Tipo de depósito bancario a un año
AI	Inventario de la empresa
AR	Cuentas por cobrar de la empresa
AT	Activos líquidos de la empresa
B1	BS1+BL1
RA	Pago mínimo de intereses exigidos a la empresa
RB	Ingresos por intereses de la empresa procedentes de los depósitos bancarios
RC	Pago de intereses reales neto
Credit Misallocation During the European Financial Crisis (2017)	
PRIME	Tipo de interés medio aplicado a las líneas de crédito de la empresas más seguras
Living with zombie companies: Do we know where the threat lies? (2018)	
W	Signos de recuperación. Representa si la actividad continuada de la empresa puede generar un rendimiento positivo en términos de capital y, por lo tanto, la desaparición de su estado problemático.
X	Extensión o alcance del problema. Pérdidas causadas a la economía por parte de la empresa
Y	Efecto contagio. Nivel de riesgo asumido por los acreedores
Z	Inmediatez del problema
Zombie firms in China's coal mining sector: Identification, transition determinants and policy implications (2019)	
BL2	Deudas a largo plazo
BS2	Deudas a corto plazo
Can environmental regulation promote the governance of excess capacity in China's energy sector? The market exit of zombie enterprises (2019)	
cs	Subvención de créditos
gs	Subvenciones del gobierno
lil	Tipo de interés mínimo del préstamo a largo plazo
minIP	Gasto mínimo por intereses
sil	Tipo de interés mínimo del préstamo a corto plazo
IP	Gastos de intereses de la empresa
LL	BL
SL	BS

Political connection and the walking dead: Evidence from China's privately owned firms (2020)	
rlb_{min} over last 5 years	Tipo de cupón mínimo observado de cualquier bono corporativo a largo plazo emitido en los cinco años anteriores
rsb	Tipo de cupón mínimo observado de cualquier bono corporativos a corto plazo y de los instrumentos del mercado monetario emitidos
LB	Bonos a largo plazo
SB	Bonos a corto plazo (con un vencimiento inferior a un año)
An Adaptive Weighted Bagging Ensemble Learning Model for Zombie Enterprise Identification (2020)	
cnkj	Volumen de negocio total de los activos
cwkj	Flujo de caja neto de las actividades de explotación/deudas pendientes
cxzz()	Crecimiento sostenido
dn	Beneficios antes de intereses e impuestos del año
fjcsy	Pérdidas y ganancias anuales no recurrentes
fz	Pasivos anuales
hyzws	Mediana del sector
jjkj	(EBIT + depreciación)/activos totales
jlz	Beneficio neto anual
mqsc	Estimación del tipo de interés más bajo disponible actualmente para la mayoría de las empresas del mercado
qydk	Aumento de los préstamos externos
qyz	Relación entre activos y pasivos
qysjzf	Tipo de interés real pagado por la empresa
rzgm	Escala de financiación anual
wbcwyz	(Subvenciones del gobierno + nuevos préstamos bancarios)/salidas de efectivo de las actividades empresariales
zy	Tipo de interés óptimo calculado
How do zombie firms affect debt financing costs of others: From spillover effects views (2021)	
How does government intervention affect the formation of zombie firms? (2021)	
rbond	Tipos de interés de los bonos en circulación
Bonds1	Bonos convertibles, bonos de capital privado, bonos corporativos, herramientas direccionales, etc.
The Financing and Investment Crowding-out Effect of Zombie Firms on Non-zombie Firms: Evidence from China (2021)	
r_{cb}_{min} over 5,t	Tipo mínimo del cupón del bono durante los últimos cinco años en el mercado
rl1	Tipo de interés mínimo de los préstamos a largo plazo en el mercado
rs1	Tipo de interés mínimo de los préstamos a corto plazo en el mercado
Bonds2	Saldo de los bonos en el mercado
BL3	Saldo del préstamo a largo plazo de la empresa en el mercado
BS3	Saldo del préstamo a corto plazo de la empresa en el mercado

Fuente: Elaboración propia del autor

Tabla A2: Primeros estudios sobre empresas zombis

Nombre Artículo (Año)	Criterios empleados para identificar a zombis
Do we want these two to tango? On zombie firms and stressed banks in Europe (2003)	Criterio del estudio , empresa zombi si cumple con los siguientes criterios: <ol style="list-style-type: none"> 1. $ROA \left(\frac{\text{Ingresos netos}}{\text{Activos totales}} \right) < 0$ 2. Inversiones netas (variación activos fijos respecto año anterior) < 0 3. Capacidad servicio de deuda $\left(\frac{EBITDA}{\text{Deuda financiera total}} \right) < 5\%$ 4. Condiciones anteriores para dos años consecutivos Criterios alternativos basados en Adalet McGowan y Gebauer: <ol style="list-style-type: none"> 1. $ICR = \left(\frac{EBITDA}{\text{Intereses pagados}} \right) < 2$: Empresa Zombi 2. $\frac{\text{Nivel de endeudamiento}}{\text{Activos totales}} > 85\%$: Empresa Zombi
Economics of the living dead (2006)	Criterio del estudio basado en CHK y sigue los siguientes pasos: <ol style="list-style-type: none"> 1. Cálculo de $R_{i,t}^*$ ídem a la ecuación 1 2. Si $(R_{i,t} - R_{i,t}^*) < 0$: Empresa Zombi
Zombie lending and depressed restructuring in Japan (2008)	Criterio del estudio 1: mencionado en la sección 4.1.3 Criterio del estudio 2: $z(x, d_1, d_2) = \begin{cases} \text{Zombi si } x < d_1 \\ \frac{d_2 - x}{d_2 - d_1} \text{ si } d_1 \leq x \leq d_2 \text{ donde } d_1 \leq 0 \leq d_2 \\ \text{Cierta grado de zombi} \\ \text{No Zombi si } x > d_2 \end{cases}$
Zombie Firms and the Crowding-Out of Private Investment in China (2016)	Criterio del estudio basado en CHK y FN y sigue los siguientes pasos: <ol style="list-style-type: none"> 1. Cálculo de $RA_{i,t}$ ($R_{i,t}^*$ nombrado de otra forma) de manera simplificada: $RA_{i,t} = rs_{t-1} BS1_{i,t-1} + \left(\frac{1}{5} \sum_{j=1}^5 rl_{t-j} \right) BL1_{i,t-1}$ 2. Cálculo de $RB_{i,t}$: $RB_{i,t} = (AT_{i,t-1} - AR_{i,t-1} - AI_{i,t-1}) * rd_t$ 3. Cálculo del $gap_{i,t}$ y si $gap_{i,t} < 0$: Empresa Zombi $gap_{i,t} = \frac{RC_{i,t} - (RA_{i,t} - RB_{i,t})}{B1_{i,t-1}}$ 4. Además, si $profit_{i,t} + RC_{i,t} > RA_{i,t} - RB_{i,t}$: empresa reclasificada como No Zombi Criterios alternativos , tras aplicar los cuatro pasos anteriores: <ol style="list-style-type: none"> 1. Si apalancamiento $\left(\frac{\text{Deuda total}}{\text{Activo total}} \right) > 50\%$ y deuda continua en aumento en el año t: Empresa Zombi 2. Uso de media móvil (dos años) de la deuda de la empresa para calcular $RA_{i,t}$ 3. Cálculo de $profit_{i,t}$ extrayendo las subvenciones y devoluciones de impuestos por parte del gobierno
The mystery of zombie enterprises—stiff but deathless (2017)	Criterio del estudio: ídem al criterio CHK de la sección 4.1.3 Criterio alternativo: <ul style="list-style-type: none"> ➤ Si $(\text{Beneficio por acción} - \text{pérdidas y ganancias anormales}) < 0$, durante tres años consecutivos: Empresa Zombi
Zombie firms and over-capacity in Chinese manufacturing (2017)	Criterio del estudio 1 basado en FN, empresa zombi si cumple con los siguientes criterios: <ol style="list-style-type: none"> 1. Apalancamiento $> 50\%$ 2. Beneficio anual real $(\text{Beneficio} - \text{subvenciones} - \text{ingresos no operativos}) < 0$ 3. Pasivos_t $>$ Pasivos_{t-1} Criterio del estudio 2 basado en Zhu y Fan: <ul style="list-style-type: none"> ➤ Beneficio real (ídem criterio estudio 1) < 0, durante tres años consecutivos: Empresa Zombi

Resolving China's Zombies: Tackling Debt and Raising Productivity (2017)	<p>Criterio del estudio basado en FN: empresas que se ajustan al FN modificado (criterio alternativo 2) pero considerando sólo una cuarta parte de la deuda a corto plazo al aplicar FN modificado.</p> <p>Criterios alternativos:</p> <ol style="list-style-type: none"> 1. Pérdidas tres años consecutivos: Empresa Zombi 2. FN modificado, empresa zombi cuando cumple con los siguientes criterios: <ol style="list-style-type: none"> a. $R_{i,t} < R_{i,t}^*$ o b. Apalancamiento $\left(\frac{\text{Deuda total}}{\text{Activo total}}\right) > 50\%$ y en aumento a lo largo del año, y $\text{EBIT} < R_{i,t}^*$ 3. Empresas ajustables a la definición de FN modificado (punto 2) durante dos años sucesivos: Empresa Zombi
Credit Misallocation During the European Financial Crisis (2017)	<p>Criterio del estudio:</p> <ol style="list-style-type: none"> 1. $\text{ROA} \left(\frac{\text{media móvil tres años EBITDA}}{\text{Activos totales}} \right) < \text{PRIME (media móvil tres años)}$ 2. Apalancamiento $\left(\frac{\text{Deuda financiera total (excluye deuda con accionistas)}}{\text{Activos totales}} \right) > \text{la mediana de apalancamiento del grupo de empresas que salieron y tuvieron ROA} < \text{PRIME, al menos una vez en los últimos dos años}$

Fuente: Elaboración propia del autor

Tabla A3: Estudios recientes sobre empresas zombis

Nombre Artículo (Año)	Criterios empleados para identificar a zombis
The walking dead? Zombie firms and productivity performance in OECD countries (2018)	<p>Criterio del estudio: mencionado en la sección 4.1.3</p> <p>Criterio alternativo basado en CHK y sigue los siguientes pasos:</p> <ol style="list-style-type: none"> Cálculo de $R_{i,t}^*$ de manera simplificada: $R_{i,t}^* = rs_{t-1} BS_{i,t-1} + \left(\frac{1}{5} \sum_{j=1}^5 rl_{t-j}\right) BL_{i,t-1}$ Si $(R_{i,t} - R_{i,t}^*) < 0$: Empresa Zombi
Living with zombie companies: Do we know where the threat lies? (2018)	<p>Criterio del estudio, plantean índice EZ que recoge cuatro dimensiones de los zombis extremos:</p> $X = \frac{\text{Patrimonio neto negativo}}{\sum_{i=1}^N \text{Total Patrimonio neto negativo}_i}$ $Y = -\frac{\text{Patrimonio neto negativo}}{\text{Activos}} \quad \text{o} \quad Y = 1 - \frac{\text{Activos}}{\text{Deudas}}$ $W = 1 - \frac{CF}{\text{Deudas}} \quad \text{o} \quad W = 1 - \frac{EBIT}{\text{Activos}}$ $Z = 1 - \frac{\text{Activos}}{\text{Deudas corto plazo}}$ <p>Índice EZ = $k1 (X) + k2 (Y) + k3 (Z) + k4 (W)$</p>
The impact of government subsidies on the capacity utilization of zombie firms (2019)	<p>Criterio del estudio, basado en CHK y FN y sigue los siguientes pasos:</p> <ol style="list-style-type: none"> Cálculo de $R_{i,t}^*$ ídem a la ecuación 1. Identifican como zombis a las siguientes empresas: <ol style="list-style-type: none"> Si $(R_{i,t} - R_{i,t}^*) < 0$: Empresa Zombi. Asimismo, si $(\text{Beneficio neto} - \text{ganancias y pérdidas no recurrentes}) > R_{i,t}^*$ y $\frac{\text{Activos}}{\text{Pasivos}} \geq 50\%$: Empresa No Zombi Si $R_{i,t}^* < (\text{Beneficio neto} - \text{ganancias y pérdidas no recurrentes})$, $(\text{Beneficio neto} - \text{ganancias y pérdidas no recurrentes}) < 0$ y $\frac{\text{Activos}_{t-1}}{\text{Pasivos}_{t-1}} > 50\%$ pero la empresa continua solicitando préstamos: Empresa Zombi $(\text{Beneficio neto} - \text{ganancias y pérdidas no recurrentes}) < 0$, para tres años consecutivos: Empresa Zombi <p>Criterio alternativo: empresa cumple criterios zombis por dos años consecutivos</p>
Zombie firms in China's coal mining sector: Identification, transition determinants and policy implications (2019)	<p>Criterio del estudio, basado en CHK y FN y sigue los siguientes pasos:</p> <ol style="list-style-type: none"> Aplicación del criterio CHK de la sección 4.1.3 para identificar a zombis pero, para el cálculo de $R_{i,t}^*$, se emplea la siguiente fórmula: $R_{i,t}^* = rs_{t-1} BS_{i,t-1} + \left(\frac{1}{5} \sum_{j=1}^5 rl_{t-j}\right) BL_{i,t-1}$ Además en la ecuación 2, no se utiliza CP y Bonds. FN ligeramente modificado: <ol style="list-style-type: none"> Si $EBIT > 0$: Empresa No Zombi Las empresas que fueron clasificadas como no zombis en el paso 1, se examinan de nuevo y se marcan como zombis si: Apalancamiento $\left(\frac{\text{Deuda total}_{t-1}}{\text{Activos totales}_{t-1}}\right) > 50\%$, $(\text{Beneficio de explotación} - \text{subvenciones del gobierno}) < 0$ en el año t y $\text{Deuda total}_t > \text{Deuda total}_{t-1}$ <p>Criterios alternativos:</p> <ol style="list-style-type: none"> Si Beneficios reales $(\text{Beneficio} - \text{subvenciones} - \text{otros ingresos no operativos}) < 0$ en el año t y $t-1$: Empresa Zombi Si Beneficios reales $(\text{Beneficio} - \text{subvenciones} - \text{otros ingresos no operativos}) < 0$, apalancamiento $\left(\frac{\text{Deudas}}{\text{Activos}}\right) > 50\%$ y el endeudamiento aumentó al final del año t: Empresa Zombi
Can environmental regulation promote the governance of excess capacity in China's energy sector? The	<p>Criterio del estudio, basado en Imai y FN:</p> <p>➤ Si $\sum_{t=1}^T (\text{profit}_{i,t} - \text{gs}_{i,t} - \text{cs}_{i,t}) < 0$: Empresa Zombi</p> $\text{cs}_{i,t} = \min IP_{i,t} - IP_{i,t} \text{ y } \min IP_{i,t} = SL_{i,t} * \text{sil}_{i,t} - LL_{i,t} * \text{lil}_{i,t}$

market exit of zombie enterprises (2019)	
Zombie firms, external support, and corporate environmental responsibility: Evidence from China (2019)	<p>Criterio del estudio, basado en CHK y FN y sigue los siguientes pasos:</p> <ol style="list-style-type: none"> 1. Aplicación del criterio CHK de la sección 4.1.3, con la única diferencia de que en la ecuación 2 no se utiliza CP. 2. Aplicación de los dos criterios de FN, tal y como se mencionan en la sección 4.1.3. <p>Criterio alternativos:</p> <ol style="list-style-type: none"> 1. Modificación del criterio de rentabilidad de FN: EBIT – ingresos por subvenciones del gobierno. 2. Empleo del segundo criterio del estudio de Caballero.
Unfinished business: Zombie firms among SME in Japan's lost decades (2019)	<p>Criterios del estudio, basados en CHK, FN y Imai:</p> <ol style="list-style-type: none"> 1. CHK: cálculo de $R_{i,t}^*$ ídem a la ecuación 1 y si $(R_{i,t} - R_{i,t}^*) < 0$: Empresa Zombi 2. FN ligeramente modificado: si $EBIT < R_{i,t}^*$ y $(B_{i,t} > B_{i,t-1}$ o $R_{i,t} < R_{i,t}^*)$: Empresa Zombi 3. Imai: ídem al mencionado en la sección 4.1.3
Impacto de la rentabilidad y la cuota de mercado en las empresas zombis en México (2019)	<p>Criterio del estudio, basado en CHK y sigue estos pasos:</p> <ol style="list-style-type: none"> 1. Calculan $R_{i,t}^*$ de la siguiente manera: $R_{i,t}^* = rs_{t-1} BS_{i,t-1} + \left(\frac{1}{5} \sum_{j=1}^5 rl_{t-j}\right) BL_{i,t-1}$ 2. Si $\frac{R_{i,t} - R_{i,t}^*}{B_{i,t-1}} < 0$: Empresa Zombi
Political connection and the walking dead: Evidence from China's privately owned firms (2020)	<p>Criterio del estudio, basado en CHK y sigue los siguientes pasos:</p> <ol style="list-style-type: none"> 1. Cálculo de $R_{i,t}^*$ de la siguiente manera: $R_{i,t}^* = rs_{t-1} BS_{i,t-1} + \left(\frac{1}{5} \sum_{j=1}^5 rl_{t-j}\right) BL_{i,t-1} + rsb_{t-1} SB_{i,t-1} + rlb_{\min \text{ over last 5 years}} LB_{i,t-1}$ y si $(R_{i,t} - R_{i,t}^*) < 0$: Empresa Zombi 2. Además, debe cumplir que su ROA < mediana del sector. <p>Criterios alternativos: empleo de la misma medida CHK del paso 1 pero de manera aislada y para dos y tres años consecutivos.</p>
Listed zombie firms and top executive gender: Evidence from an emerging market (2020)	<p>Criterio del estudio: ídem al CHK mencionado en la sección 4.1.3</p>
Lie of the weak: Inconsistent corporate social responsibility activities of Chinese zombie firms (2020)	<p>Criterio del estudio, basado en CHK y FN y sigue los siguientes pasos:</p> <ol style="list-style-type: none"> 1. Aplicación del criterio CHK de la sección 4.1.3, con la única diferencia de que en la ecuación 2 no se utiliza CP. 2. Aplicación de los dos criterios de FN, tal y como se mencionan en la sección 4.1.3. <p>Criterio alternativo: empleo del segundo criterio del estudio de Caballero.</p>
Bank of Finland Research Discussion Papers 8: The life and death of zombies-evidence from government subsidies to firms Bank of Finland Research (2020)	<p>Criterio del estudio, basado en Adalet McGowan y sigue los siguientes pasos:</p> <ol style="list-style-type: none"> 1. $ICR = \frac{EBIT}{\text{Intereses} + \text{Gastos financieros}} < 1$, para tres años consecutivos 2. Además, la tasa de crecimiento anual de la empresa en términos de trabajadores empleados no es positiva en promedio durante un período de dos años dentro del período zombi.
An Adaptive Weighted Bagging Ensemble Learning Model for Zombie Enterprise Identification (2020)	<p>Criterios del estudio:</p> <ol style="list-style-type: none"> 1. CHK: $qysjzf < mqsc$ 2. FN-CHK: $qy fz > 0.5$ y $qy dk > 0$ y $dn < zy$ 3. Pérdidas consecutivas: 3 años($jlr - fjcsy < 0$) 4. Excesivos prestamos: 3 años($jlr < 0$) y 3 años($fz > 0$) y 3 años($cxzz(rzgm)$) 5. Reconocimiento cuatro dimensiones: 2 años($cw kj < 1$) y 2 años($jj kj < hyzws$) y 2 años($cnkj < hyzws$) y 2 años($wbcwyz < hyzws$)

How do zombie firms affect China's industrial upgrading? (2021)	Criterio del estudio , basado en FN y Imai y deben cumplir los siguientes criterios: <ol style="list-style-type: none"> 1. Apalancamiento > 50% 2. Suma de beneficios reales (Beneficios – subvención gobierno) < 0, para tres años consecutivos 3. $Deuda_t > Deuda_{t-1}$
How do zombie firms affect debt financing costs of others: From spillover effects views (2021) How does government intervention affect the formation of zombie firms? (2021)	Criterio del estudio , basado en Zhu y Fan: <ul style="list-style-type: none"> ➤ Beneficios netos negativos (excepto las ganancias y pérdidas no recurrentes), por tres años consecutivos: Empresas Zombis Criterios alternativos: <ol style="list-style-type: none"> 1. Empleo del criterio de Zhu y Fan, como se menciona en la sección 4.1.3 2. Aplicación del criterio CHK ídem a la sección 4.1.3 pero calculan $R_{i,t}^*$ con la siguiente formula: $R_{i,t}^* = rs_{t-1} BS_{i,t-1} + \left(\frac{1}{5} \sum_{j=1}^5 rl_{t-j} \right) BL_{i,t-1} + \sum_j r_{bond_{i,j,t-1}} * Bonds1_{i,j,t-1}$
Are zombie firms more incentivized to financialize? (2021)	Criterio del estudio , basado en CHK, FN y Imai y sigue los siguientes pasos: <ol style="list-style-type: none"> 1. Cálculo de $R_{i,t}^*$ ídem a la ecuación 1 2. Si la media móvil de tres años de: (Ingresos netos – subvenciones gobierno – $(R_{i,t} - R_{i,t}^*)$) < 0: Empresa Zombi Criterio alternativo: <ul style="list-style-type: none"> ➤ Pérdidas por tres años consecutivos: Empresa Zombi
Zombie firms: Prevalence, determinants, and corporate policies (2021)	Criterio del estudio , basado en Adalet McGowan y siguen estos pasos: <ol style="list-style-type: none"> 1. Edad > 10 años 2. $ICR = \frac{EBITDA}{\text{Gastos por intereses}} < 1$, para tres años consecutivos
Recovery and exit of zombie firms in Portugal (2021)	Criterio del estudio , basado en Schivardi y sigue los siguientes pasos: <ol style="list-style-type: none"> 1. $ROA \left(\frac{EBITDA}{\text{Activos totales}} \right) < \text{tipo de interés medio anual Euribor a 12 meses}$, por tres años consecutivos 2. Apalancamiento $\left(\frac{\text{Deuda del pasivo corriente y largo plazo}}{\text{Activos totales}} \right) > \text{valor de la mediana de apalancamiento del grupo de empresas que salieron}$ 3. Edad > 5 años 4. Además, incluyen aquellas empresas zombis que se convierten en no zombis en t+1 y zombis de nuevo en t+2 Criterios alternativos: <ol style="list-style-type: none"> 1. Adalet McGowan: mencionado en la sección 4.1.3 2. Schivardi: mencionado en la tabla A2 3. Shen y Chen: emplean su criterio 1 mencionado en la tabla A2 4. Storz: mencionado en la tabla A2 5. Criterio del propio estudio modificado ligeramente, siguiendo estos pasos: <ol style="list-style-type: none"> a. Comparación de ROA (misma definición) con la media anual de tipo de interés de las operaciones de préstamo a las empresas b. Apalancamiento (misma definición) > media del sector (o del percentil 75) de las empresas que salieron c. Edad > 10 años
The Financing and Investment Crowding-out Effect of Zombie Firms on Non-zombie Firms: Evidence from China (2021)	Criterio del estudio , basado en CHK y FN y sigue los siguientes pasos: <ol style="list-style-type: none"> 1. Cálculo de $R_{i,t}^*$ de la siguiente manera: $R_{i,t}^* = rs_{t-1} BS_{i,t-1} + \left(\frac{1}{5} \sum_{j=1}^5 rl_{t-j} \right) BL_{i,t-1} + r_{cb_{\min \text{ over } 5,t}} Bonds2_{i,t-1}$ 2. Aplicación de los dos criterios de FN como mencionados en la sección 4.1.3 3. Además se aplican los siguientes criterios alternativos a los anteriores dos: <ol style="list-style-type: none"> a. Si beneficio neto > 0 antes de deducir una subvención del gobierno y < 0 tras deducirla, y apalancamiento $\left(\frac{\text{Pasivos}}{\text{Activos totales}} \right) > 50\%$: Empresa Zombi b. Si beneficio neto < 0, $\frac{\text{Subvención gobierno}}{\text{Beneficio neto}} < -0.5$ y apalancamiento $\left(\frac{\text{Pasivos}}{\text{Activos totales}} \right) > 50\%$: Empresa Zombi

	<p>c. Si $\frac{\text{beneficio neto}}{\text{Subvención gobierno}} < 0$, $\text{Subvención gobierno} > 0$, $\frac{\text{Beneficio neto}}{\text{Subvención gobierno}} > -0.5$, pérdidas del año aumentan en 100% con respecto al año anterior, $\text{Subvención gobierno}_t > \text{Subvención gobierno}_{t-1}$ y $\text{Apalancamiento}_t > \text{Apalancamiento}_{t-1}$: Empresa Zombi</p> <p>4. Además, debe cumplir los criterios para dos años consecutivos</p>
--	--

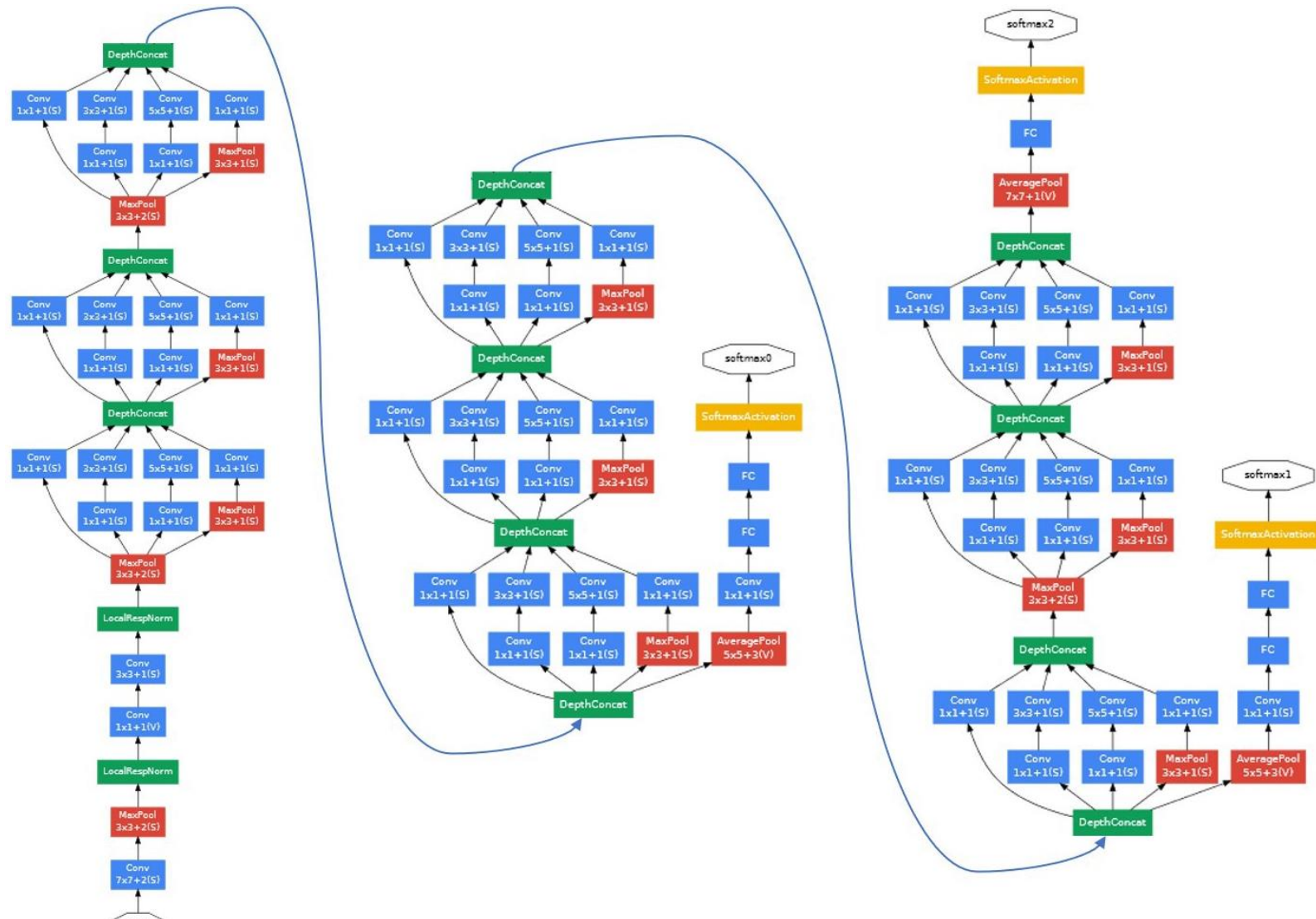
Fuente: Elaboración propia del autor

Tabla A4: Partidas de las variables de predicción

Variable	Cálculo
Activo funcional	Suma de inmovilizado intangible, inmovilizado material, existencias, deudores comerciales y otras cuentas a cobrar, y efectivo y otros activos líquidos equivalentes
Activo total	Suma de activo corriente y no corriente
Deuda remunerada total	Suma de deudas a corto y largo plazo y de deudas con empresas del grupo y asociadas a corto y largo plazo
Pasivo total	Suma de pasivo corriente y no corriente
Resultado de explotación ajustado	Resultado de explotación tras eliminar el deterioro y resultado por enajenaciones del inmovilizado y otros resultados
Subvención	Suma de subvenciones, donaciones y legados recibidos, y otros ingresos de explotación

Fuente: Elaboración propia del autor

Figura A1: Arquitectura GoogLeNet



Fuente 1: Extraído de Brown L. (2020)