



**UNIVERSIDAD
DE BURGOS**

DEPARTAMENTO DE DIGITALIZACIÓN

PROGRAMA DE DOCTORADO EN TECNOLOGÍAS INDUSTRIALES E
INGENIERÍA CIVIL

TESIS DOCTORAL

**APORTACIONES A LA TOMA
DE DECISIONES DE GESTIÓN
EMPRESARIAL UTILIZANDO
METODOLOGÍAS DE APREN-
DIZAJE MÁQUINA**

Autor

Luis Santiago García Pineda

Directores:

Dr. Carlos Alonso de Armiño Pérez

Dr. Daniel Urda Muñoz

Burgos, julio de 2023

Agradecimientos

Mi agradecimiento en primer lugar al entorno familiar en el que me he criado. Mis padres y hermanos me han enseñado desde siempre que el valor del esfuerzo y de la ayuda a los demás, siempre tiene recompensa.

A mis directores de tesis y compañeros de la Universidad sin cuya ayuda y apoyo hubiera sido imposible realizar este trabajo.

Y como no, a mi familia: a Sira por tener paciencia y aguante en los momentos difíciles. Y a mis hijos, para que sigan la estela.

ÍNDICE GENERAL

AGRADECIMIENTOS	III
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	9
INTRODUCCIÓN	11
CAPÍTULO I: LOS DATOS Y LA TOMA DE DECISIONES	15
1.1. Los datos en la toma de decisiones.	16
1.1.1.- Importancia de los datos en la toma de decisiones.	16
1.2.- Ámbitos de toma de decisiones en gestión.	17
1.3.- Metodologías de recogida de datos.	19
1.3.1.- Encuestas.	20
1.3.2.- Entrevistas.	21
1.3.3.- Observación.	23
1.3.4.- Análisis de documentos.	24
1.3.5.- Experimentos.	24
1.3.6.- Recopilación de datos en línea.	25
1.3.7.- Grupos focales.	25
1.4.- Métodos de análisis de datos.	26
CAPÍTULO II: METODOLOGÍAS DE TRATAMIENTO DE DATOS APLICADOS A LA INVESTIGACIÓN REALIZADA	31
2.1 Métodos Proyeccionistas.	32
2.1.1.- Análisis de componentes principales (PCA).	33
2.1.2.- Cooperative Maximum Likelihood Hebbian Learning (CMLHL).	35
2.2 Métodos de agrupamiento.	36
2.2.1.- Métodos de clusterización.	36
2.2.2.- Método K-means.	40
2.2.3.- Método de clustering aglomerativo.	42
2.3 Redes Neuronales.	43
2.3.1.- Métodos neuronales para predicción de series temporales.	43
2.3.2.- Nonlinear AutoResgresive (NAR).	44
2.3.3.- Nonlinear Input-Outup (NIO).	45
2.3.4.- Nonlinear AutoResgresive with eXogenous input (NARX).	46

2.4 Metodología Altman Z-score.	47
2.4.1.- Estimación de la fórmula.	48
2.4.2.- Precedentes.	49
2.4.3.- Precisión y eficacia.	50
2.4.4.- Definición de variables del indicador Z de Altman original.	50
2.4.5.- Estimación del índice Z' de Altman para empresas privadas.	51
2.4.6.- Valor de predicción Z'' de Altman para empresas no manufactureras y mercados emergentes.	52
CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS	53
3.1 Una herramienta de visualización inteligente para analizar la sostenibilidad del transporte por carretera.	54
3.1.1.- Introducción y trabajo previo.	55
3.1.2.- Materiales y método.	60
3.1.3.- Resultados	63
3.2 Modelos neuronales no lineales para predecir el precio del acero HRC en España.	69
3.2.1.- Introducción y trabajo previo.	70
3.2.2.- Materiales y Métodos.	72
3.2.3.- Experimentos y Resultados.	74
3.3 Análisis de la sostenibilidad económica en la cadena de suministro aplicando el predictor Altman Z-score.	78
3.3.1.- Introducción.	78
3.3.2.- Marco teórico.	80
3.3.3.- Metodología.	87
3.3.4.- Estudio empírico del octavo modelo de la regresión logística Altman Z-score.	88
3.3.5.- Discusión.	89
CAPÍTULO IV: CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURO	93
4.1.- Conclusiones individuales (por estudio)	95
4.1.1.- Conclusiones y futuras líneas de investigación referentes al estudio “Una herramienta de visualización inteligente para analizar la sostenibilidad del transporte por carretera”.	95
4.1.2.- Conclusiones y futuras líneas de investigación referentes al estudio “Modelos neuronales no lineales para predecir el precio del acero HRC en España”.	96
4.1.3.- Conclusiones y futuras líneas de investigación referentes al estudio “Análisis de la sostenibilidad económica en la cadena de suministro aplicando el predictor Altman Z-score”.	97
4.2.- Conclusiones generales	98
BIBLIOGRAFÍA	100
ANEXOS	119
A.- Artículo 1	121
B.- Artículo 2	137
C.- Artículo 3	147

ÍNDICE DE FIGURAS

Figura 1.- Dendograma ejemplo con 20 datos. Fuente: Elaboración propia.....	37
Figura 2.- Representación de MATLAB para modelo NAR con 10 neuronas en capa oculta y una salida con un retardo de 6. Fuente: The Mathworks Inc.	45
Figura 3.- Representación de MATLAB para modelo NIO con 5 neuronas en capa oculta y una entrada exógena con 5 valores y un retardo de 8. Fuente: The Mathworks Inc. ...	46
Figura 4.- Representación de MATLAB para modelo NARX con 30 neuronas en la capa oculta y una neurona con la capa de salida, una entrada exógena con 5 valores y un retardo de 4 y una única salida con un retardo de 10. Fuente: The Mathworks Inc.	47
Figura 5.- Serie de datos de edad de la flota de transporte, dividida en cuartiles y relacionada con las fases determinadas para la Gran Depresión. Fuente: Elaboración propia	59
Figura 6.- HUEP obtenidos mediante la aplicación de agrupamiento aglomerativo ($k = 3$, distancia = euclidiana, enlace = promedio) al conjunto de datos analizado, variando la técnica EPP: a) CMLHL, b) KPCA, c) SM.	64
Figura 7.- HUEP obtenido al aplicar KPCA y agrupamiento aglomerativo ($k=3$, distancia = euclidiana, enlace = promedio) al conjunto de datos analizado, utilizando la función de sostenibilidad en la metáfora del glifo.....	65
Figura 8.- HUEP obtenidos con KPCA y agrupación aglomerativa (distancia = euclidiana, enlace = promedio), que muestra la característica de sostenibilidad en la metáfora del glifo y un número variable de agrupaciones: a) $k=2$, b) $k=4$ y c) $k=8$	66
Figura 9.- HUEP obtenidos mediante la aplicación de agrupaciones de KPCA y k-means (distancia = euclidiana, enlace = promedio), que muestran la característica de sostenibilidad en la metáfora del glifo y un número variable de agrupaciones: a) $k=2$, b) $k=3$, y c) $k=4$	68
Figura 10.- Proyección obtenida al aplicar KPCA de forma aislada al conjunto de datos analizado, que muestra la característica de sostenibilidad en la metáfora del glifo. a) 2D y b) 3D.	69

ÍNDICE DE TABLAS

Tabla 1.- Comparativa de las características de los distintos tipos de encuestas. Fuente: Martínez Candil (2015).	20
Tabla 2.- Plataformas de videoconferencia para la realización de entrevistas online. Río Grande del Sur. Fuente: Schmidt, Palazzi y Piccinini (2020)	22
Tabla 3.- Clasificación de los distintos tipos de análisis de datos. Fuente: Elaboración propia con datos de los autores referenciados.....	27
Tabla 4.- Clasificación de los distintos tipos de aprendizaje automático. Fuente: Elaboración propia.....	28
Tabla 5.- Algunas de las medidas de distancia conocida y ampliamente empleadas en los métodos de agrupación. Fuente: Alonso de Armiño (2019).	39
Tabla 6.- Medidas de distancia para k-means empleadas en este estudio. Fuente: Elaboración propia.....	41
Tabla 7.- Métodos de enlace (linkage) empleados para el agrupamiento aglomerativo en este estudio. Fuente: Elaboración propia.....	42
Tabla 8.- Métodos neuronales de aprendizaje supervisado para predicción de series temporales. Fuente: Elaboración propia	44
Tabla 9.- Emission standards adopted by the EU for diesel cars and heavy goods vehicles. (Comisión_Europea, 2005)	57
Tabla 10.- Leyenda de los gráficos utilizando la metáfora del glifo según los valores del rasgo relacionado con la sustentabilidad (Edad promedio de la flota de vehículos).	65
Tabla 11.- MSE de los resultados obtenidos por los modelos neuronales NAR, NIO y NARX. Los resultados promediados para las 3 series de datos se muestran por el número de retrasos de entrada. N retrasos de entrada.	75
Tabla 12.- MSE (Error Cuadrático Medio) de los resultados obtenidos por los modelos neuronales NAR, NIO y NARX. Los resultados promedio para las 3 series de datos se muestran por el número de neuronas ocultas.	76
Tabla 13.- MSE de los resultados obtenidos por los modelos neuronales NAR, NIO y NARX. Los resultados promediados para las 3 series de datos se muestran según el algoritmo de entrenamiento.	76
Tabla 14.- MSE of the results obtained by the NIO, and NARX neural models. Averaged results are shown per the data series.	77
Tabla 15.- MSE de los resultados obtenidos por los modelos neuronales NAR, NIO y NARX. Se muestran los mejores resultados individuales (promediados solo para las 10 ejecuciones).	77
Tabla 16.- Key figures of the initial Z-Altman model. Altman, E.I. Financial ratios, Discriminant analysis and the prediction of Corporate Bankruptcy. J. Financ.1968, 23 589-609	83

Tabla 17.- The seven hypotheses considered by Altman in the updated model. Barboza, F.; Kimura, H.; Altman, E. Machine Learning Models and Bankruptcy Prediction. Expert Syst. Appl. 2017, 83, 405–417	85
Tabla 18.- Table of equivalences of economic activities between CNAE and Z-score Altman 2014. Source: Own elaboration	87
Tabla 19.- Table of descriptive values of the eighth model of the Altman Z-Score Logistic Regression of the research sample based on the bankruptcy financial classification of Z"-score model. Source: Own elaboration	88
Tabla 20.- Equivalence of the financial classification between the Z"-Score model and the renewed measurement of the eighth model of the Altman Z-Score. Source: Own elaboration	88
Tabla 21.- Measurement of the eighth model of the Altman Z-Score for different industries. Source: Own elaboration	89

INTRODUCCIÓN

La tesis doctoral que se refleja en este documento, se ha planteado mediante el trabajo de investigación realizado en los últimos años, dentro de un grupo de investigación conformado por investigadores de las áreas de Organización de Empresas y de Ciencia de la Computación e Inteligencia Artificial de la Universidad de Burgos.

Este grupo viene trabajando desde hace más una década en el análisis de datos, mediante la aplicación de distintas técnicas relacionadas con la inteligencia artificial, que ayuden a la toma de decisiones aplicadas a la gestión eficiente en la empresa industrial.

En este sentido son numerosos los proyectos de investigación competitiva realizados por los miembros del grupo, cuya orientación principal de los resultados obtenidos ha sido su aplicabilidad al entorno industrial más cercado, y siempre contando con la participación de tejido empresarial más próximo.

Los diferentes resultados obtenidos de estas investigaciones han sido publicados en revistas científicas indexadas en los más importantes índices académicos, tanto en las ramas relacionadas con la inteligencia artificial, como en las relacionadas con la organización de empresas.

Este doctorando se ha incorporado reciente a este grupo de investigación y, aunque ya poseía resultados publicados anteriormente, para el objetivo de obtener el grado de doctor, se aportan solamente las tres últimas investigaciones realizadas dentro del grupo, del que forman parte los dos directores de la tesis. Dos de estas investigaciones están publicadas en revistas con importante impacto dentro del *Journal Citation Reports* (JCR). El tercer trabajo ya finalizado, se ha presentado como comunicación a un congreso y, en estos momentos, se está profundizando en el análisis de los datos para intentar su posible publicación en una revista indexada en el JCR.

El documento soporte de esta investigación, se ha estructurado en cinco partes. En primer lugar, se ha realizado un sencillo estudio del estado del arte referido a los datos y las diferentes metodologías que se han ido desarrollando a lo largo del tiempo, intentando obtener una clasificación resumida de todas ellas. Este análisis ha permitido conocer las distintas herramientas desarrolladas y, así, poder seleccionar aquellas que se consideran más adecuadas, para aplicarlas a la problemática de gestión en la que se pretende aportar conocimiento.

En segundo lugar, se profundiza en los fundamentos de las metodologías que se consideran más adecuadas para el análisis de los datos obtenidos en los campos de la logística y la financiación empresarial, materias en las que he desarrollado mi actividad docente, durante los últimos años en la Universidad de Burgos.

Estas técnicas, que podríamos encuadrar bajo el paraguas de las denominadas Machine Learning, se han descrito, examinado y analizado en este apartado, para poder aplicarlas con el máximo rigor científico en los tres trabajos que se presentan como resultados de la investigación.

En el tercer apartado se incluyen los tres trabajos de investigación realizados. El primero se refiere a una herramienta de visualización inteligente para analizar la sostenibilidad del transporte por carretera, donde se analiza la sostenibilidad de la actividad del transporte, proponiendo la aplicación de novedosas técnicas de Inteligencia Artificial (más concretamente, Machine Learning), ampliando el uso de Parcelas Exploratorias Híbridas No Supervisadas con nuevas técnicas de Persecución de Proyección Exploratoria. La propuesta se prueba con un gran conjunto de datos de la encuesta oficial de transporte por carretera en España, que se llevó a cabo durante un período de 7 años.

El segundo trabajo de investigación trata sobre el análisis de la sostenibilidad económica del Sector Supply Chain aplicando el predictor Altman Z-Score. Estudiando, desde la perspectiva financiera, la medición del desempeño de la cadena de suministro relacionado con la falta de un indicador de probabilidad de quiebra, y proponiendo un predictor, que es el octavo modelo de la regresión logística Altman Z-Score. Se obtiene un ranking de probabilidad de quiebra para las cadenas de suministro de las empresas, según la industria a la que pertenecen, estableciéndose tres categorías de empresas según el predictor. Finalmente se observa que la industria de la construcción es la que tiene la mayor probabilidad de quiebra.

El último trabajo propuesto para esta tesis doctoral es el realizado más recientemente, y se avanza en el análisis de series temporales para pronosticar el precio del acero en España, dado su alto impacto en los costes de sus productos. Por ello, muchas industrias de todo el mundo tratan de pronosticar su precio, aplicándose por primera vez modelos neuronales no lineales a diferentes conjuntos de datos. Los datos han sido tomados durante los últimos siete años, con el objetivo de validar su idoneidad para predecir el precio del acero laminado en caliente en España. Los resultados obtenidos, pretenden aportar información a las empresas para una mejor toma de decisiones en la programación de las operaciones de abastecimiento, de acuerdo con la previsión de precios.

Se ha considerado oportuno aportar una copia de los dos trabajos ya publicados en la revista *Sustainability*. También se aporta el tercer trabajo presentado en

CAPÍTULO I: MACROECONOMÍA, CRISIS Y CAMBIOS

el congreso SOCO 2022, celebrado en Salamanca. Por ello, una copia original de los tres trabajos se ha incluido en el anexo de este documento.

En todo trabajo de investigación se deben llegar a establecer unas conclusiones, resaltado aquellas aportaciones que se realizan, bien a la comunidad científica y/o a la sociedad.

Así, aunque en cada uno de los trabajos de investigación, como no puede ser de otra manera, ya se establecen sus conclusiones particulares, se ha considerado conveniente ordenar y homogeneizar las mismas en un único apartado, con el objetivo de entrelazar y resaltar aquellos resultados que, por su relevancia, puedan ser más interesantes para la comunidad científica.

Finalmente, en el apartado de bibliografía, se reflejan todos los documentos científicos que se han utilizado para llevar a cabo esta investigación y que están referenciados a lo largo de todo este documento y en las tres investigaciones publicadas.

CAPÍTULO I: LOS DATOS Y LA TOMA DE DECISIONES

1.1. Los datos en la toma de decisiones.

En este primer capítulo se pretenden estudiar los fundamentos básicos que van a soportar todo el trabajo de investigación. Al realizarse la investigación dentro de un grupo mixto, conformado por investigadores de inteligencia computacional e ingeniería de organización, se ha considerado oportuno realizar un breve análisis del estado del arte referido a los datos, como soporte fundamental en la toma de decisiones.

Así se comienza resaltando la importancia de los datos en la toma de decisiones, identificando las razones fundamentales de esta necesidad para cualquier organización o empresa.

A continuación, se realiza un breve repaso de las aportaciones realizadas por algunos investigadores en los distintos ámbitos de gestión empresarial, determinándonos en los tres que consideramos más cercanos a los trabajos presentados posteriormente, como son la logística, la dirección de operaciones y el financiero.

En el tercer apartado, se analizan las metodologías de recogida de datos, revisando en mayor profundidad las siete metodologías clásicas referenciadas por la mayoría de los autores.

Finalmente, en el último apartado, se revisan y clasifican brevemente, los métodos de análisis de datos, como elemento de partida para profundizar en la explicación de los métodos aplicados a las investigaciones realizadas.

1.1.1.- Importancia de los datos en la toma de decisiones.

Los datos son fundamentales para la toma de decisiones adecuadas en cualquier tipo de organización o empresa. La importancia de los datos radica en su capacidad para proporcionar información objetiva y cuantificable además de utilizarse para identificar problemas, tomar decisiones estratégicas y medir el éxito o no de las iniciativas empresariales.

Dentro de las razones por la que se considera que los datos son importantes para la toma de decisión podemos destacar:

- Identificación de problemas: Los datos pueden ayudar a identificar problemas específicos dentro de cualquier organización o empresa ya que, al recopilarlos y analizarlos, los líderes son capaces de identificar áreas susceptibles de mejora y crear oportunidades para implementar cambios que pueden mejorar el rendimiento de la organización.

- Toma de decisiones estratégicas: Los datos a menudo proporcionan información valiosa a los líderes que les ayuda a tomar decisiones estratégicas importantes. Al analizar los datos estos pueden evaluar diferentes escenarios y tomar las decisiones adecuadas basadas en hechos concretos
- Medición del éxito: Para medir el éxito de las medidas tomadas en una organización es necesario también llevar a cabo una recopilación y análisis de datos. En base a los resultados se llevan a cabo las iniciativas y ajustes que se crean oportunos encaminados a mejorar el rendimiento de la organización
- Identificación de tendencias: Los gráficos obtenidos con los datos adecuados, pueden ayudar a identificar tendencias importantes de los mercados o de las preferencias de los consumidores y tomar las medidas necesarias.

Como resumen, podemos decir que los datos son fundamentales para la toma de decisiones en cualquier organización o empresa. Al recopilar y analizar datos adecuadamente, los líderes son capaces de tomar decisiones estratégicas, medir el éxito y anticipar tendencias importantes.

1.2.- Ámbitos de toma de decisiones en gestión.

Son muy numerosos los artículos que recogen afirmaciones sobre la importancia de los datos para la toma de decisiones en diversos campos empresariales. Por su importancia destacamos tres, en el ámbito de la logística, el entorno financiero y el del marketing, aunque su aplicación abarca un amplio espectro de disciplinas:

En el ámbito de la logística podemos tomar como referencia a (Parada Gutiérrez, 2009) que, en su trabajo realizado sobre el enfoque multicriterio para la toma de decisiones en la gestión de inventarios, pone de manifiesto que, sobre esta temática, ha sido reconocido ampliamente esta necesidad, destacando los aportes de (John F. Magee et al., 1973); (Ballou & de Lemus, 1991); (Arbones Malisani E., 1992); (Prida & Gutiérrez, 1996); (Levy, 1997); (Pau Cos & De Navascués, 2001); (Torres et al., 2019).

Más en concreto, (Prida & Gutiérrez, 1996), en su trabajo sobre la logística de aprovisionamiento, analizando el cambio en las relaciones proveedor-clientes, como un nuevo desafío para la empresa del siglo XXI, se describe la evolución de la gestión de suministro desde las compras tradicionales a lo que los autores llaman ahora “gestión de fábrica externa”, poniendo de relieve la nueva relación entre proveedor y cliente.

Otros autores como (Pau Cos & De Navascués, 2001), que han trabajado sobre la logística integral, indican que la logística es una función clave dentro de las organizaciones ya que tiene una función determinante dentro del organigrama

y funcionamiento de la empresa. Analizan los sistemas actuales de la organización logística, para centrarse más adelante en las compras, el aprovisionamiento y el transporte, su recepción y el control del mismo. Se analizan los distintos tipos de embalajes y aparatos de manutención, para acabar profundizando en el almacenamiento, transporte y distribución y su relación con los operadores logísticos, para así poder calcular los costes logísticos. Es en este punto donde inciden en la importancia de los datos para que estos cálculos sean correctos.

No sólo en el ámbito de la logística, sino también en el campo de la dirección de operaciones, cuando se tratan los aspectos tácticos y operativos en la producción y los servicios, ha de tenerse en cuenta la importancia de la toma de datos correctos, como ponen de manifiesto (Domínguez Machuca & Álvarez Gil, 1995) en sus análisis sobre la dirección de operaciones, tanto en los aspectos estratégicos como en los tácticos y operativos.

Par tanto, con este breve análisis se observa la importancia de la toma de datos en los procesos de gestión de la logística en la empresa, siendo fundamentales para poder realizar su posterior análisis y poder tomar las adecuadas decisiones de gestión.

En el ámbito financiero, distintos autores como (Calderón et al., 2021) ponen de relieve la importancia de los datos para el análisis de los estados financieros en la toma de decisiones. El objetivo de su investigación es demostrar la importancia de los datos de los estados financieros para la toma de decisiones dentro de una organización, así como las consecuencias y efectos que se pueden producir si estas decisiones se toman con datos reales y veraces.

Esta veracidad y exactitud de los datos en los estados financieros son muy importantes tanto en las grandes compañías como en las PYMES tal y como indican (Sandoval et al., 2008) en su trabajo sobre los estados financieros básicos, su uso e interpretación para la toma de decisiones en las pequeñas y medianas empresas.

Y también por (Carvajal et al., 2020) en su investigación sobre los estados financieros y la toma de decisiones en las pymes, indica que en numerosas pequeñas empresas no se cuenta con la información contable básica, o se desconoce cómo realizar el análisis financiero que puede ocasionar toma de decisiones erróneas con los inconvenientes que ello acarrea

No solo en estas áreas, sino también en otros tales como el marketing, en la que (Villegas, 2019) analiza la importancia de las aplicaciones estadísticas a las decisiones de marketing, poniendo de manifiesto la importancia de la toma de datos y su tratamiento, mediante las adecuadas herramientas estadísticas para la toma de decisiones correctas en los departamentos comerciales.

1.3.- Metodologías de recogida de datos.

Justificada la importancia de los datos en toma de decisiones, es necesario crear una estructura adecuada para realizar una clara, rigurosa y adecuada obtención de solo aquellos datos que se consideren necesarios.

Por ello debemos analizar las metodologías existentes. En la actualidad existen distintos tipos de recogida de datos para su posterior análisis. Entre ellos podemos destacar varios como:

1. Encuestas. Se trata de un método muy común para recopilar datos en los cuales se hacen una serie de preguntas a un grupo de personas para obtener información sobre sus opiniones, actitudes o comportamientos.
2. Entrevistas. Esta técnica de recopilación de datos requiere hablar con personas a través del teléfono, de manera presencial o en línea con el fin de obtener información detallada de un tema en particular.
3. Observación. En esta técnica de recopilación de datos que se caracteriza porque el investigador observa y registra el comportamiento o los eventos que ocurren en un ambiente natural.
4. Análisis de documentos. En este caso el investigador recopila y analiza documentos tales como informes, actas, registros y otros documentos de los cuales extrae la información relevante. Los documentos que se analizan pueden ser tanto documentos de investigación como actas de reuniones y cualquier otro tipo de documento.
5. Experimentos. En este tipo de recopilación de datos el investigador tiene en cuenta una o más variables independientes, para conocer el posible efecto que tienen en la variable dependiente. Se pueden llevar a cabo tanto en ambientes controlados como en la vida real.
6. Recopilación de datos en línea. Hablamos de este tipo de obtención de datos cuando se realiza a través de aspectos que están en la red. Internet es una fuente inagotable de datos. Puede incluir encuestas realizadas en línea, análisis de redes sociales, y otro tipo de recopilación de datos basados en la red.
7. Grupos focales. En este caso se reúne a un grupo de personas para discutir un tema específico en profundidad. Suelen utilizarse para obtener información sobre experiencias, actitudes, creencias y comportamientos.

Pasamos a continuación a explicar con una mayor profundidad cada una de las metodologías referenciadas anteriormente.

1.3.1.- Encuestas.

Autores como (Martínez Candil, 2015) han trabajado sobre el diseño de encuestas y cuestionarios de investigación. En su trabajo se explica las distintas técnicas de recogida de información en la investigación de mercados y se indican distintos diseños de cuestionarios (sus elementos y estructura) para la adecuada toma de datos, así como el desarrollo de diversas aplicaciones informáticas de diseño y ejecución de encuestas. En la tabla 1 se refleja un análisis comparativo de las características de los diferentes tipos de encuestas, bajo cuatro aspectos diferentes.

Propiedades de las encuestas personal, telefónica y postal e Internet			
Propiedades	Personal	Telefónica	Postal e Internet
Libertad de expresión	Muy alto	Poco	Nulo
Influencia del entrevistador	Muy alto	Alto	Nulo
Complejidad	Muy alto	Alto	Poco
Claridad	Poco	Alto	Muy alto
Coste	Muy alto	Alto	Poco
Tiempo	Muy alto	Alto	Poco
Participación	Muy alto	Alto	Poco

Tabla 1.- Comparativa de las características de los distintos tipos de encuestas. Fuente: Martínez Candil (2015).

Las encuestas son el método más utilizado para recogida de datos en multitud de investigaciones, destacando por su relevancia las de las EPA (Encuesta de Población Activa). Esta metodología se ha tenido que adaptar a las nuevas técnicas y circunstancias sociales para cumplir con sus objetivos de calidad, básicamente, sobre las características de la no respuesta en la encuesta de población activa (Ballano, 2005) y (Ballano et al., 2001).

La aplicación de estos nuevos métodos ha conllevado el desarrollo de los cuestionarios y recogida mediante CATI (Computer Assisted Telephone Interviewing). A través de esta metodología se realizan encuestas telefónicas guiadas por un cuestionario que aparece en la pantalla del ordenador, en las que el entrevistador toma nota de las respuestas, utilizando el teclado y el ratón según las respuestas que han sido previamente codificadas.

Un ejemplo interesante de la utilización de esta metodología es el descrito por (Amro, 2023), recogiendo a través de entrevistas telefónicas datos referidos a la reciente pandemia del COVID 19, en el trabajo sobre validación de cuestionarios de diagnóstico de migraña, adaptado a una entrevista telefónica asistida por ordenador (Porta-Etessam et al., 2010). En este, se realizan entrevistas telefónicas a 50 individuos elegidos al azar, llevado a cabo una investigación en el campo médico (estudio de migrañas), validando las entrevistas de forma telefónica (CATI).

Una variante de CATI es el método CAPI (Computer Assisted Personal Interviewing), que es similar a CATI, salvo que en este caso la encuesta se lleva a cabo de forma presencial en lugar de realizarse a través del teléfono.

1.3.2.- Entrevistas.

Distintos autores han estudiado los procesos de la realización de entrevistas, analizando cuales son las cuestiones fundamentales para llevarlas a cabo. Para ello, generalmente, han utilizado distintos niveles: teórico, epistemológico, ético y práctico. Concretamente, autores como (Flick U., 2007) detallan con mayor profundidad la planificación y el diseño de la investigación cualitativa y la metodología sobre cómo realizar las entrevistas.

Entre otros documentos, en la colección de investigación cualitativa, se resumen formas distintas o adicionales de recogida de datos. Así, (Angrosino, 2012), en su estudio sobre la etnografía y observación participante en la investigación cualitativa, presenta la observación participante y hace hincapié en la etnografía y en cómo representar este tipo de datos. Además, se hace referencia al contexto cambiante de la investigación para el siglo XXI analizando la tecnología, la globalización y los mundos virtuales, así como el uso de las entrevistas en este contexto.

La pandemia del COVID 19 ha impulsado en gran medida el uso de las nuevas tecnologías, avanzando en desarrollo de nuevas herramientas para la recolección de datos. Así, se ha extendido el uso de plataformas de videoconferencias tales como como: Google Meet, GoToMeeting, Skype, Webex, WhatsApp y Zoom.

Por otra parte, (Schmidt B. et al., 2020) analizan en el cuadro reflejado en la tabla 2, los distintos tipos de plataformas que son más ampliamente utilizados

para la realización de entrevistas on line y, sobre todo, durante y tras la pandemia del COVID-19.

	General features ^a	Type of recording	Type of access
Google Meet^b	Free for audio/video interviews, both individual and in group, for an unlimited time until March 31 st , 2021. Website: https://apps.google.com/intl/pt-BR/meet/how-it-works	Audio/video recording on Google Drive for paid plans. Allows instant closed captions during interviews.	Participants can download the platform or use it in a web browser. Interviewer and participants need to create a Google account.
GoToMeeting	Paid plan for audio/video interviews, both individual and in group, for an unlimited time. Website: https://www.gotomeeting.com/pt-br	Audio/video recording saved in the platform cloud and sent by email to the interviewer after the interview.	Participants need to click on an invitation link to download the platform and enter the interview, after identifying themselves. Participants do not need to create an account.
Skype	Free for audio/video interviews, both individual and group, for an unlimited time. Website: https://www.skype.com/pt-br/	Audio/video recording available for 30 days in the cloud. Before recording, the platform alerts the interviewer to obtain the participants' consent.	Participants can download the platform or use it in a web browser. Interviewer and participants need to create an account before using the program.
Webex	Free for audio/video interviews, both individual and in group, for an unlimited time. Website: https://www.webex.com/	Audio/video recording available only in paid versions; restricted access to the interviewer.	Participants must download the platform. Participants do not need to create an account.
WhatsApp^c	Free for audio/video interviews, for up to eight participants, for an unlimited time. Website: https://www.whatsapp.com	Audio/video recording not available.	Participants must download the platform on their cell phone. On the computer, WhatsApp Web does not allow video calls, only exchanging messages.
Zoom	Free for audio/video interviews, unlimited for up to two participants (in addition to the interviewer), and for up to 40 minutes with more participants. Website: https://zoom.us/	Audio/video recording, controlled and accessed by interviewer, who may or may not share it with participants. Paid version allows simultaneous transcription of the recording.	Participants can download the platform or use it in a web browser. Only the interviewer needs to download the platform, create an account, and send the link with an ID and password

Tabla 2.- Plataformas de videoconferencia para la realización de entrevistas online. Río Grande del Sur. Fuente: Schmidt, Palazzi y Piccinini (2020)

La realización de entrevistas para reclutamiento de personal es un método habitualmente utilizado en todo tipo de empresas. Además, se observa que cada vez se están utilizando con más frecuencia las entrevistas con esta finalidad a través de redes sociales (Eugenia et al., 2016).

Los avances que se están llevando a cabo con la Inteligencia Artificial, hacen que, para determinadas funciones, que hasta hace muy poco se venían realizando con humanos, empiecen a llevarse a cabo con ayuda de máquinas de inteligencia artificial. El primer software conocido que se desarrolló para realización de entrevistas mediante esta tecnología fue el denominado SAPIA de entrevistas, aplicación que apareció en 2018 con base en Australia.

En la actualidad hay varios programas tales como el propio SAPIA, MOCKMATE, Engadget, y otros, que ofrecen a los candidatos a distintos puestos de trabajo realizar entrenamientos mediante la inteligencia artificial, para poder realizar mejor las posteriores entrevistas de trabajo presenciales.

También existe la posibilidad para los reclutadores de las empresas utilizar dispositivos de última generación, como el seguimiento de movimiento de ojos "eye tracking" (Adiani et al., 2022) para mejorar la fiabilidad de los resultados de las entrevistas, ya que los comportamientos no verbales, tales como la mirada fija al entrevistador, se consideran rasgos importantes para la empleabilidad del candidato.

1.3.3.- Observación.

Autores como (Torres et al., 2019), describen los métodos de recolección de datos para una investigación, estableciendo una serie de criterios a considerar para diseñar la herramienta de recolección de información, así como los métodos de recogida, para lograr en una investigación con resultados fiables. Una investigación es científicamente válida al estar sustentada en información verificable, que responda a lo que se pretenda demostrar con la hipótesis formulada. Para ello es imprescindible realizar un proceso de recolección de datos en forma planificada y teniendo claros los objetivos sobre el nivel y profundidad de la información a recoger.

También (Kawulich, 2005), desarrolló el concepto de la observación participante como método de recolección de datos. Este concepto de observación participante o participativa es una herramienta muy interesante para recoger datos en estudios de investigación cualitativa.

Por otra parte, (Marshall C. & Rossman G.B., 2016), definen la observación como "*la descripción sistemática de eventos, comportamientos y artefactos en el escenario social elegido para ser estudiado*".

Mientras que (de Munck V.C. & Sobo E, 1988), describen la observación participante, en su trabajo sobre el uso de métodos de campo, como el primer método usado por los antropólogos al hacer trabajo de campo. El trabajo de campo involucra "mirada activa, una memoria cada vez mejor, entrevistas informales, escribir notas de campo detalladas, y, tal vez lo más importante, paciencia". Así mismo, provee el contexto para desarrollar directrices de muestreo y guías de entrevistas.

Para (Dewalt K. M. & DeWalt B. R., 2002), siempre es mejor intentar completar cada sesión con notas de campo escritas lo antes posible después de las observaciones y, desde luego, no más de tres días después de llevar a cabo la observación. Además, estos autores definen la observación participante como "el proceso de aprendizaje a través de la exposición y el involucrarse en el día a día o las actividades de rutina de los participantes en el escenario del investigador"

Por último, Schensul et al.(1999) en su estudio sobre los métodos etnográficos esenciales, en el que analizan las observaciones, entrevistas y cuestionarios, indican como transformar los hechos en propuestas teóricas o modelos que sirven para formar a los investigadores actuales y futuros. Introducen el concepto de etnografía como un enfoque científico para descubrir e investigar patrones y su significado en distintas comunidades, instituciones y otros grupos sociales

1.3.4.- Análisis de documentos.

Como nos indican (Caffarel-Serra et al., 2017), son muy numerosos los trabajos científicos que inciden en la importancia de la utilización de documentos para su posterior análisis y para la toma de decisiones.

Pero cuando es necesario tomar de decisiones estratégicas en la empresa, se emplean habitualmente documentos de tipo financiero, tanto para empresas del sector comercial como en el sector industrial (Sociedad et al., 2020). También el trabajo de (Lavalle Burguete, 2017) explica cómo realizarlo, indicando que con un estudio detallado de los mismos se pueden llegar a distinguir las causas reales de los problemas y así poder tomar acciones correctivas.

Existen muchos documentos cuantitativos disponibles para su interpretación y, entre ellos, se incluyen informes usados para la toma de decisiones, tales como los informes de desempeño, y otros (Kendall & Kendall, 2005)

Respecto al proceso de investigación científica (León, 2012), indica que para la toma de decisiones debemos apoyarnos en el análisis de los documentos , ya que “se trata de una fuente importante, objetiva y barata para la realización de una investigación cualitativa siendo necesario en primer lugar seleccionar el contenido, aplicar un sistema de clasificación y medidas, transformar o codificar los datos en unidades que permitan la descripción del contenido y definir categorías para su agrupación”.

1.3.5.- Experimentos.

Esta metodología se utiliza con mucha frecuencia en campos como el de la enseñanza, para la realización de experimentos (Molina et al., 2011). La misma aporta una descripción del nuevo paradigma metodológico conocido como la investigación de diseño (desing research), ya que resulta de mucha utilidad su aplicación a la didáctica de las ciencias, integrando la descripción que aporta la literatura con la experiencia de la puesta en práctica.

En lo referente a la preparación de aquellos aspirantes a doctor en la rama de la pedagogía, hay una continua preocupación por los métodos utilizados para el análisis tanto en lo cualitativo como en lo cuantitativo (Ramírez et al., 2010). Este método de trabajo se puede utilizar también para poder crear un elemento de reflexión de estos enfoques con la idea de aplicar, ofrecer criterios y recomendaciones a los potenciales doctores.

Otro campo de aplicación importante donde se utiliza este tipo de técnicas es la psicología, siendo este método considerado como una de las nueve categorías definidas por (Montero & León, 2005) en el sistema de clasificación del método en los informes de investigación en esta disciplina. Esta clasificación se está utilizando actualmente como modelo para los informes de investigación en algunas de las más importantes revistas del sector.

1.3.6.- Recopilación de datos en línea.

La recopilación de datos en línea ya se lleva haciendo muchos años, pero con el confinamiento de la población durante la pandemia del COVID -19 se ha reforzado, incluyendo a las personas investigadoras, y ha puesto de manifiesto la necesidad de adaptar las metodologías, sus técnicas y herramientas a un nuevo contexto. Así, en la investigación de (Hernán-García et al., 2021), sobre cómo adaptar una investigación cualitativa a contextos de confinamiento, se indica cómo poderlo llevara cabo a través de diez sencillas reglas para el diseño de una e-investigación.

Tanto Internet como los medios de comunicación o las redes sociales, posibilitan la recogida de datos textuales, secuencias e imágenes o narrativas sobre una realidad limitada para el acceso a ella de forma presencial.

Este tipo de recopilación de datos se utiliza, no sólo en los ámbitos de la medicina y la salud (Giabicani et al., 2023); (Robin et al., 2023), sino también en la empresa (Lyublinskaya & Du, 2023), en la educación (Wao et al., 2023) y en el deporte (Dufour W., 1983) sobre los métodos de objetivación del comportamiento motor de la recogida de datos en fútbol.

También se está llevando a cabo en temas, que últimamente están siendo ampliamente investigados, como son la influencia de diversos factores que afectan a los estudiantes de bachillerato, en su decisión para la elección de las disciplinas STEM en sus posteriores estudios universitarios (Dabdoub et al., 2023).

1.3.7.- Grupos focales.

La correcta formación de los grupos focales se hace imprescindible para recolectar información de tipo cualitativo. En su reseña metodológica sobre los grupos focales, (Rubén & Reyes, 2012) indican que la profundidad del problema de estudio, el número de participantes, el perfil de los participantes y el rol del moderador, hacen del grupo focal una técnica con valor y utilidad para recolectar la información.

También otros autores como (Arredondo Dominguez E R. et al., 2019) utilizan esta técnica como método idóneo en la investigación de mercados.

Al igual que otras técnicas de recolección de datos, existe abundante literatura de su uso en los ámbitos de la salud y de la educación, o incluso en ambos (Rudrakumar et al., 2023).

1.4.- Métodos de análisis de datos.

Una vez recogidos los datos y utilizados diversos métodos de recopilación nos encontramos con la fase más importante de todo investigador que es el tratamiento y análisis de estos.

Hasta ahora hemos considerado brevemente los métodos tradicionales de obtención de datos. En la presente tesis, y para el tratamiento de los datos, nos vamos a centrar en métodos de análisis que podemos calificar más sofisticados, sin perder de vista los métodos de análisis tradicionales, que no por sencillos son menos importantes.

En la tabla 3 se ha desarrollado una clasificación de los distintos tipos de análisis de datos para las investigaciones científicas.

CAPÍTULO I: LOS DATOS Y LA TOMA DE DECISIONES

AUTOR (AÑO)	COMENTARIO	TIPO DE ANALISIS
Jordi Olle (2016)	Clasifica los distintos tipos de análisis para la mentorización de tesis doctorales https://www.linkedin.com/in/jordiolllesanchez/?originalSubdomain=es	Descripción Exploración Asociación Correlación Comparativas Modelos estadísticos o causales Segmentación de datos Reducción dimensional Algoritmos y modelos Predictivos
Power data	Empresa dedicada a la gestión de datos, data warehousing, enmascaramiento de datos, master data management, big data, creación de centro de competencia, etc...	Media aritmética Desviación estándar Determinación del tamaño de la muestra Regresión Prueba de hipótesis Aprendizaje automático Redes neuronales Aprendizaje de reglas de asociación Algoritmos genéticos Análisis de series temporales
Marytere Narvaez	Ha generado dentro de Question Pro una herramienta que permite crear y distribuir cuestionarios	Descriptivo Exploratorio De diagnóstico Inferencial Predictivo Prescriptivo Mecanicista De clústeres De cohortes De textos Factorial Conjoint
Jörg Hecher	Es el CEO de Atlas.ti que es una empresa asociada a Open AI que a su vez es la principal organización de investigación sobre AI	Descriptivos Exploratorios Predictivos De diagnóstico Prescriptivo Técnicas de análisis de datos
Manuel Orois García (2020)	Distintas técnicas de machine learning y comparativa de cada una de ellas	Random forest K-NN Redes neuronales
Andrés Bueno Crespo (2021)	Utilización de técnicas de aprendizaje automático para realizar predicciones en distintos campos (medicina, agricultura, etc...)	Predictivos
Daniel Peña 2002	Libro Peña, D. (2002). Análisis de datos multivariantes (Vol. 24). Madrid: McGraw-hill	Análisis de datos multivariantes
Jaime Andreu Abeñila (2002)	Propuesta de utilización de distintas técnicas de análisis de contenidos	Técnicas cuantitativas y cualitativas
H Pedroza, L Dicovskyi (2007)	Análisis estadístico de datos provenientes de investigación experimental y no experimental utilizando el software SPSS	Técnica experimental y no experimental

Tabla 3.- Clasificación de los distintos tipos de análisis de datos. Fuente: Elaboración propia con datos de los autores referenciados.

Por otra parte, en la en la tabla 4 se ha desarrollado una clasificación de los distintos tipos de análisis de datos para las investigaciones científicas, más concretamente referidas al aprendizaje automático.

APRENDIZAJE AUTOMATICO	
SUPERVISADO	NO SUPERVISADO
En el aprendizaje supervisado se proporciona al algoritmo de aprendizaje un conjunto de datos de entrenamiento etiquetados, es decir, cada ejemplo de entrenamiento tiene una etiqueta o clase asignada. El objetivo del algoritmo es aprender a mapear las entradas a las salidas deseadas.	En el aprendizaje no supervisado, no se proporcionan etiquetas en los datos de entrenamiento. El algoritmo debe descubrir patrones, estructuras o relaciones ocultas en los datos por sí mismo.
TIPOS	TIPOS
1. Clasificación: El algoritmo aprende a asignar instancias de entrada a categorías predefinidas. Por ejemplo, clasificar correos electrónicos como spam o no spam.	1. Agrupamiento (Clustering): El algoritmo agrupa los datos en conjuntos basados en similitudes o características comunes. No se le proporciona información sobre las clases o categorías esperadas de antemano. Por ejemplo, agrupar clientes en segmentos con características similares para realizar análisis de mercado
2. Regresión: El algoritmo aprende a predecir valores continuos basados en las características de entrada. Por ejemplo, predecir el precio de una casa en función de su tamaño y ubicación.	2. Asociación: El algoritmo descubre patrones de asociación o correlación entre diferentes elementos en un conjunto de datos. Por ejemplo, encontrar la relación entre productos comprados por los clientes en un supermercado.
	3. Reducción de dimensionalidad: El algoritmo reduce la cantidad de variables o características en un conjunto de datos, manteniendo la información más relevante. Ayuda a simplificar el modelo y mejorar la eficiencia computacional. Por ejemplo, reducir la dimensionalidad de un conjunto de imágenes conservando las características más distintivas.

Tabla 4.- Clasificación de los distintos tipos de aprendizaje automático. Fuente: Elaboración propia.

Como puede extraerse de este breve análisis, el número y tipo de técnicas a utilizar para los trabajos de investigación es muy variado y depende fundamentalmente del propósito o finalidad del mismo.

CAPÍTULO I: LOS DATOS Y LA TOMA DE DECISIONES

Esta ordenación personal reflejada en la tabla pretende aportar una nueva clasificación de las diferentes técnicas de análisis de datos, con el objetivo de conocer los diferentes investigadores que han realizado investigaciones en este campo, así como conocer los distintos criterios de clasificación y sus características más relevantes.

Esta tarea de recopilación ordenada del estado del arte ha resultado imprescindible para poder elegir aquellas técnicas más adecuadas para realizar la investigación de los tres trabajos que conforman esta tesis doctoral.

Para la elaboración de la presente tesis se han utilizado y analizado en profundidad técnicas de aprendizaje automático o "*Machine Learning*", que tienen una mayor complejidad que las descritas hasta el momento en esta introducción y que pasaremos a definir y estudiar en profundidad en los capítulos siguientes.

Los distintos trabajos que forman parte de esta tesis doctoral se centran en varias de estas técnicas elegidas en función de los objetivos que se han perseguido en cada uno de ellos, abordando temas diferentes ligados a aspectos logísticos en la empresa, a predicción de la supervivencia o quiebra de empresas según distintos indicadores y a predicciones de precios del acero futuros en España.

CAPÍTULO II: METODOLOGÍAS DE TRATAMIENTO DE DATOS APLICADOS A LA INVESTIGACIÓN REALIZADA

A lo largo de este capítulo, se pretende profundizar en las metodologías de tratamiento de datos que se han utilizado el análisis de las variables que conforman el cuerpo principal de esta tesis doctoral, con el objetivo de comprender mejor los resultados obtenidos.

Así, en primer lugar, se realiza una breve explicación de los métodos proyectivistas, deteniéndonos especialmente en el análisis de componentes principales y en el Cooperative Maximum Likelihood Hebbian Learning (CMLHL).

Dada su relación con estos, a continuación, nos adentramos en el análisis de los métodos de agrupamiento, profundizando en el estudio de la clusterización, el método K-means y el método de clustering aglomerativo.

En el tercer epígrafe se estudian las redes neuronales a través de los métodos neuronales para la predicción de series temporales y los métodos NAR, NIO y NARX.

Por último, se profundiza en la metodología de Altman Z-score, teniendo en cuenta los precedentes y la estimación de la fórmula primera, su valoración y eficacia. Todo ello nos ha permitido conocer y definir las variables del índice Z-Altman original, así como las variables e indicadores utilizados en los índices Z' para empresas privadas y Z'' para empresas no manufactureras y mercados emergentes, que dan lugar a los modelos correspondientes aplicados en la investigación.

2.1 Métodos Proyeccionistas.

La comunidad científica viene preocupándose desde hace mucho tiempo, y más aún en el último período del análisis de datos multidimensionales. No siempre es fácil extraer conclusiones y por tanto conocimiento de estos datos debido a la gran cantidad de datos empleados y por la complejidad de su análisis.

Los métodos proyectivistas, conocidos por sus siglas como EPP, acrónimo en inglés de “*Exploratory Projection Pursuit*”, tienen como finalidad buscar proyecciones de baja dimensionalidad (2D ó 3D) partiendo de datos de una mayor dimensionalidad.

Estos métodos realizan una exploración de los datos con el fin de identificar una posible estructura que puede ir asociada a todos ellos y que, de antemano, es desconocida. Este es el principal cometido de los EPP: identificar la estructura de los datos hasta el mínimo detalle.

Se trata de métodos de aprendizaje no supervisado, ya que los datos que analizan no están etiquetados o clasificados *a priori*. Pretenden conseguir, gracias

a la proyección de los mismos, una visualización que de forma intuitiva nos permita definir la estructura de esos datos. Una vez obtenida esa proyección y visualización, se requiere la intervención de un experto que será el responsable de interpretarlos y extraer las conclusiones que se deriven del análisis de esas proyecciones.

Teniendo en mente el significado de los métodos proyeccionistas se han definido para el primer artículo que forma parte de la presente tesis doctoral los métodos proyeccionistas de análisis de componentes principales, PCA y el CMLHL.

2.1.1.- Análisis de componentes principales (PCA).

El método PCA (Análisis de Componentes Principales), (Pearson K., 1901) es un método estadístico que tiene por objetivo llevar a cabo el análisis de la información de una manera más sencilla, reajustando la dimensionalidad de los datos. Intenta reducir el número de variables para poder interpretar de forma adecuada la información que los datos nos ofrecen. Cuando los datos se comprimen desaparecen algunos campos, que llamamos características, de los datos originales. Esto conlleva una pérdida de información, pero se hace de tal forma que la pérdida de información sea lo menor posible. De todo el conjunto de datos, habrá múltiples variables que desaparecerán en favor de otras pocas, utilizando para ello métodos cuya finalidad es minimizar, en la medida de lo posible, la pérdida de información, maximizando la varianza.

En la proyección se nos revela una estructura de los datos que deberemos interpretar adecuadamente estudiando la relación entre los factores y las variables origen. Posteriormente incidiremos tanto en las relaciones identificadas, como en su signo y en su magnitud. De esta manera, obtendremos la proyección y las variables o componentes principales. Nuestro cometido será la lectura e interpretación adecuada de la proyección y su vinculación con los datos y variables originales. Estos aspectos son fundamentales en la aplicación de método PCA.

(Pearson, 1901) ideó este método a principios del siglo pasado, aunque unos años más tarde (Hotelling, 1933) lo estudió en profundidad y lo desarrolló. En cualquier caso, fue a partir de la aparición de los ordenadores a mediados de siglo cuando su aplicación y difusión tuvo su esplendor aprovechando los sistemas de información y el desarrollo de los programas que facilitaron su implementación.

Al aplicar la metodología PCA se consiguen dos funciones: por una parte, se comprimen las observaciones de un espacio N-dimensional a un espacio bi o tridimensional prefijado, a través de sus componentes principales y, por otra

parte, las variables originales, habitualmente correlacionadas entre sí, pasan a ser nuevas variables ortogonales entre ellas, que son precisamente sus componentes principales.

La definición de (Brown et al., 2009) es muy acertada cuando indica que *“el PCA es un método estadístico para la contracción de la dimensionalidad en grupos de datos de alta dimensionalidad. O lo que es lo mismo, si partimos de un conjunto de datos multivariable, la finalidad es reajustarlo a un menor número de variables, en orden decreciente de importancia, con la mínima pérdida de información, de manera que las variables resultantes sean combinación lineal de las variables origen, e independientes entre sí”*.

Aunque en ocasiones la utilización de este modelo se emplea para comprender mejor la información que nos dan los datos, también sirve para determinar e identificar patrones de comportamiento de los datos objeto de análisis. Cuando se realiza un cambio de base es más fácil observar los patrones que siguen los datos, aunque para hacerlo de forma correcta es necesario tener un conocimiento elevado y realizar un análisis en profundidad de los resultados obtenidos sobre patrones que de antemano son desconocidos. La opción que emplea PCA con la idea de garantizar la independencia entre los componentes principales, es encontrar la base ortogonal en función de la varianza de los datos. Para ello define primero la dirección que cuenta con la máxima varianza. Esta será la que se corresponde con el primer vector de la base que será el primer componente principal. A continuación, se determina la dirección que contiene la mayor varianza restante. De esta forma habremos definido el segundo vector de la base que será la segunda componente principal. Continuaríamos siguiendo el mismo criterio hasta llegar al máximo de la varianza de la dimensionalidad que habíamos definido previamente. Al final de este proceso, proyectamos los datos sobre las primeras componentes principales, de modo que se habrá reducido la dimensionalidad de estos, y tendrán tanta variabilidad como sea posible. Cuando tengamos los datos proyectados en una dimensionalidad reducida, podremos visualizarlos fácilmente, manteniendo sobre ella el máximo nivel de información.

(Bishop C, 1995) describe de manera formal el PCA como un mapeo de vectores X^d en un espacio de entrada N-dimensional en vectores Y^d en un espacio M-dimensional de salida, siendo $M \leq N$. X puede ser representado como una combinación lineal de un conjunto de N vectores ortonormales W_i

$$x = \sum_{i=1}^N y_i W_i$$

Los vectores W_i satisfacen la siguiente relación de ortonormalidad:

$$W_i^t W_j = \delta_{ij}$$

Donde δ_{ij} es la delta de Kronecker.

En lo que se refiere a la aplicación del método, este es adecuado cuando se trata de analizar conjuntos de datos de elevada dimensionalidad.

Para el caso de análisis de datos de transporte recogidos de la EPTMC (Encuesta Permanente del Transporte de Mercancías por Carretera) parece que la utilización del método PCA puede ser una buena opción

El método PCA ha sido utilizado en múltiples ocasiones por distintos investigadores y en distintos ámbitos. ; (S. Huang et al., 2015) lo han utilizado con el fin de evaluarlos sobre distintas bases de datos científicas de empleo común. El procesamiento de imagen por sistemas inteligentes (Kountcheva & Kountchev, 2017) es otro de los campos donde ha sido ampliamente utilizado, además de reconocimiento de rostros (Puyati & Walairacht, 2008). También ha sido utilizado en estudios médicos y forenses (Lee et al., 2016), en procesos de fabricación (Z. Liu et al., 2015), para la detección de déficits de conocimiento (Herrero et al., 2010), clasificación de señales en detectores multinivel (T. Wang et al., 2015) y en muchos otros.

2.1.2.- Cooperative Maximum Likelihood Hebbian Learning (CMLHL).

El Cooperative Maximum Likelihood Hebbian Learning (CMLHL) es un algoritmo de aprendizaje utilizado en el campo de las redes neuronales no supervisadas (MacDonald et al., 2004) y en el aprendizaje automático. Combina los principios de estimación de máxima verisimilitud y el aprendizaje "hebbiano" para entrenar redes neuronales de manera cooperativa.

El aprendizaje hebbiano es una regla de aprendizaje biológico que establece que "las células que disparan juntas, se conectan juntas". Sugiere que la conexión entre dos neuronas debe fortalecerse si se activan simultáneamente. La estimación de máxima verosimilitud, por otro lado, es un método estadístico utilizado para estimar los parámetros de un modelo maximizando la función de verosimilitud.

En cuanto a su comportamiento el CMLHL es similar al PCA, ya que permite proyectar datos en una dimensionalidad inferior y de esta manera ser visualizados con mayor facilidad. La diferencia fundamental subyace en el hecho de que CMLHL no tiene en cuenta la varianza sino momentos estadísticos de orden superior.

En el artículo "Una herramienta de visualización inteligente para analizar la sostenibilidad del transporte por carretera" (*An intelligent visualization tool to analyze the sustainability of Road Transportation*) que forma parte del cuerpo de la presente tesis, el método CMLHL es de ayuda para identificar los distintos trimestres estudiados en la proyección obtenida. Estos trimestres se rigen por la similitud entre ellos y se basan en las variables de transporte que han sido

extraídas de la EPTMC (Encuesta Permanente del Transporte de Mercancías Por Carretera) y otras fuentes similares.

Este modelo neuronal ha sido utilizado por muchos expertos en sus investigaciones (Alonso et al., 2010) y (Zurutuza et al., 2011), que lo han aplicado en ciberseguridad. (Herrero et al., 2011) en gestión del conocimiento y en análisis de riesgo país. Además (Arroyo et al., 2017) lo han utilizado en investigaciones sobre datos ambientales y otros autores lo han empleado en muy variados ámbitos. No hay constancia, al menos en nuestro conocimiento, de que esta técnica haya sido aplicada con anterioridad para estudiar los datos de transportes de mercancías por carretera, bien por sí mismos o en relación con datos macroeconómicos.

2.2 Métodos de agrupamiento.

Los métodos de agrupamiento o “clustering” se pueden equiparar, en cierto modo, a los métodos proyeccionistas: EPP (Exploratory Projection Pursuit), aunque existen diferencias significativas entre ellos.

Para nuestros trabajos de investigación, se ha considerado oportuno utilizar los métodos básicos de clusterización, el método K-means y el método de clustering aglomerativo. Por ello, en los siguientes epígrafes, se profundiza en el conocimiento de estas metodologías.

2.2.1.- Métodos de clusterización.

El objetivo de este método de “clusterización” es agrupar los datos de entrada en diferentes grupos tipológicos que llamamos clústeres o grupos, aplicando un criterio determinado, en vez de interpretar con la ayuda gráfica de las proyecciones los datos que se utilizan para su estudio tal y como proponen los métodos proyeccionistas. Los métodos de agrupamiento son también de aprendizaje no supervisado, que intentan extraer conocimiento de los datos, aún sin conocer su estructura. Se tienen en cuenta los datos en sí mismos, pero no se tiene en cuenta otra posible información adicional para definir las agrupaciones. En lo que se refiere a su salida, estos métodos no nos van a proporcionar una visualización intuitiva, como hacen los métodos proyeccionistas. Por el contrario, en este caso los datos se agrupan y permiten que se profundice en la naturaleza de cada uno de los grupos llevando a cabo un análisis posterior.

(Xu R. & Wunsch D.C., 2009) definen el análisis de clústeres, como la búsqueda de una organización para una colección de registros con múltiples datos o patrones, que se representan, generalmente, como un vector de

medidas, o un punto en el espacio multidimensional. Las agrupaciones se identifican basándose en la similitud y son las que se consideran como clusteres propiamente dichos. Se toman funciones de distancia entre pares de patrones y se considera la proximidad entre ellos. Se cumple que los patrones dentro de un cluster válido son más similares a otros del mismo cluster que a los patrones que pertenecen a otros clusteres. Para conformar las distintas agrupaciones se utilizan diferentes medidas de distancia (Andreopoulos et al., 2009), (Zhuang et al., 2012), tal y como se explica mas adelante. La salida del agrupamiento puede ser estricta (hard) o difusa. Si es estricta se le asigna un patrón a un solo grupo, mientras que si es difusa cada patrón tiene un grado variable de pertenencia a cada uno de los distintos grupos de salida. Las agrupaciones difusas se pueden convertir en estricta, si se asigna cada patrón a la agrupación con el grado de pertenencia más alto.

Existen distintos criterios de distancia y también distintos enfoques para definir los grupos de datos que se pueden clasificar a su vez en jerárquicos y particionales. Los métodos jerárquicos se ilustran habitualmente mediante un dendrograma (diagrama de árbol) y producen una serie anidada de particiones, similar al mostrado en la figura 1.

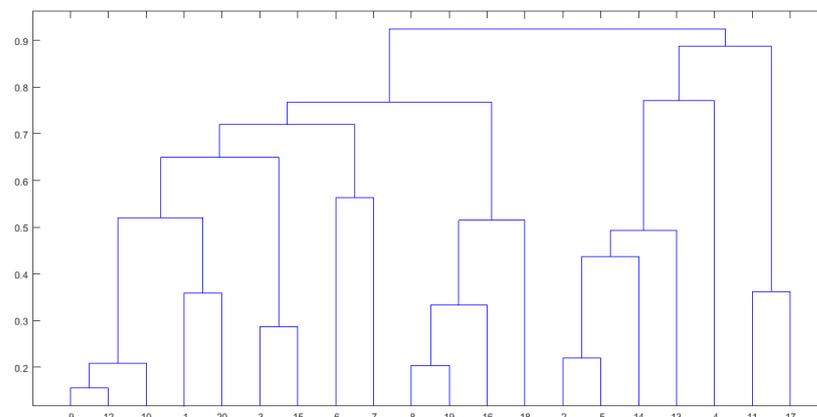


Figura 1.- Dendrograma ejemplo con 20 datos. Fuente: Elaboración propia.

Por su parte en los métodos particionales las series que anidan el método están basadas en un criterio de similitud para fusionar o dividir agrupaciones. Estos métodos particionales son capaces de identificar aquella partición que optimiza un criterio de agrupación, generalmente a nivel local. La jerarquización de los grupos proporciona generalmente una mayor flexibilidad que otros métodos. Si cortamos el árbol de grupos a un determinado nivel podemos obtener una partición de los datos.

Adentrándonos en los métodos jerárquicos podemos subdividir estos en dos tipos:

1. Jerárquicos aglomerativos: Se parte de la base de que hay tantos grupos como datos, que se van uniendo sucesivamente a otros grupos, hasta que llegado el caso se forma un grupo único o hasta el momento en que se cumple un criterio de detención.
2. Jerárquicos divisivos: En este caso el método empieza con todos los patrones en un solo grupo y estos se van dividiendo sucesivamente hasta que se cumple el criterio de detención, o bien hasta que cada patrón esté en un grupo diferente. Este método no se ha utilizado en la presente investigación.

El objetivo del agrupamiento particional no es obtener una estructura de agrupamiento, como puede ser el diagrama de árbol o dendrograma, sino obtener directamente una sola partición de los datos. La mayor parte de estos métodos tienen como base la optimización iterativa de una función de criterio que viene a reflejar la similitud entre un nuevo dato y cada uno de los patrones que han sido seleccionados para una iteración específica.

Entre las ventajas de los métodos de partición se puede destacar que son útiles en aplicaciones que involucran un gran número de conjuntos de datos, y en ese caso la construcción de un dendrograma es prácticamente inviable. Como desventaja se puede indicar que es necesario definir a priori un número determinado de grupos de salida o clústeres.

Si se desea un número de agrupaciones óptimo, se requiere una búsqueda exhaustiva en todo el conjunto inicial de datos que, computacionalmente hablando, lo hace prohibitivo. Este es el motivo por el que en la práctica el algoritmo se ejecuta varias veces con diferentes estados de inicio. La mejor configuración que se obtenga después de todas las ejecuciones será la que se tome como agrupamiento de salida.

Existen otras técnicas adicionales para la agrupación apoyadas en métodos basados en densidad probabilística, gráfica-teórica (Tu et al., 2012), (Brailovsky, 1991), (Argyrou, 2009) y resolución de mezclas, aunque no han sido utilizadas en este trabajo.

Teniendo en cuenta que la similitud es la relación más importante para la definición de los clústeres, es esencial medir esta similitud en la mayor parte de los métodos de clusterización, la cual tendrá que ser elegida con mucho cuidado. En la presente tesis se aplican criterios de distancia que son bien conocidos y se utilizan para datos con características continuas y se ven reflejados en la tabla 5.

Métrica	Descripción
Euclidean	<p>Distancia Euclídea. Se calcula de la siguiente manera:</p> $D_{ab} = \sqrt{\sum_{j=1}^p (x_{aj} - x_{bj})^2}$ <p>Dónde: x_{aj}, x_{bj} son los valores adoptados por la variable jésima para los objetos a y b, respectivamente en el espacio multivariable p-dimensional.</p>
sEuclidean	<p>Distancia Euclídea Estandarizada. Para su cálculo cada diferencia de coordenadas entre filas en X se escala dividiendo por el elemento correspondiente de la desviación estándar.</p>
Cityblock	<p>También conocida como distancia Manhattan. Se calcula de la siguiente manera:</p> $D_{ab} = \sum_{j=1}^p x_{aj} - x_{bj} $ <p>Dónde: x_{aj}, x_{bj} son los valores tomados por la jésima variable para los objetos a y b, respectivamente en el espacio multivariable p-dimensional.</p>
Minkowski	<p>Distancia Minkowski, que es en realidad una expresión compleja de las dos métricas anteriores:</p> $D_{ab} = \sqrt[\lambda]{\sum_{j=1}^p x_{aj} - x_{bj} ^\lambda}$ <p>Donde: x_{aj}, x_{bj} son los valores tomados por la jésima variable para los objetos a y b, respectivamente en el espacio multivariable p-dimensional. Para valor $\lambda=1$ obtenemos la distancia Cityblock. Para valor $\lambda=2$ obtenemos la distancia Euclídea.</p>
Chebychev	<p>Distancia Chebychev: aplica el valor máximo de diferencia entre las diferentes coordenadas de los valores.</p>
Mahalanobis	<p>Distancia Mahalanobis que emplea la máxima covarianza de los vectores X:</p> $D_{ab} = \sqrt{(x_a - x_b)^T S^{-1} (x_a - x_b)}$ <p>Dónde: x_a, x_b son los valores tomados por los objetos a y b, respectivamente en el espacio multivariable. S es la matriz de covarianza.</p>
Cosine	<p>Se calcula como 1 menos el coseno del ángulo trazado entre los puntos, siendo estos tratados como vectores.</p>
Correlation	<p>Se calcula como 1 menos la correlación muestral entre puntos, tratados estos como secuencias de valores.</p>

Tabla 5.- Algunas de las medidas de distancia conocida y ampliamente empleadas en los métodos de agrupación. Fuente: Alonso de Armiño (2019).

La distancia que más se utiliza es la euclídea. La métrica euclídea se trata de un caso específico de la métrica de Minkowski, para el caso $\lambda=2$. Cuando el conjunto de datos está formado por grupos compactos o aislados, se obtienen buenos resultados (Mao & Jain, 1996). Existe un problema si se utilizan directamente las métricas de Minkowski y consiste en que las características de mayor escala tienden a dominar al resto. Para resolver este problema se opta por normalizar las características continuas haciendo uso de la distancia sEuclídea.

Cuando existe una correlación lineal entre las características, las medidas de distancia pueden verse alteradas. En este caso el problema se atenúa haciendo uso de la distancia de Mahalanobis al cuadrado, asignando pesos diferentes a las distintas características en función de variaciones y correlaciones lineales por pares.

2.2.2.- Método K-means.

El método k-means es probablemente el más conocido y, en el mismo, se parte de un número de clústeres que se pretende obtener como resultado. Para ello se aplica una de las medidas de distancia que están reflejadas en la tabla 6, de tal forma que se parte de la media de distancia que vamos a aplicar y del número de clústeres que deseamos obtener con nuestros datos. Se prueban distintas combinaciones de valores haciendo uso de los dos parámetros de entrada (distancia y número de grupos) y los resultados obtenidos se analizan con el fin de seleccionar el mejor de ellos.

DISTINTAS METRICAS UTILIZADAS PARA UTILIZACIÓN DEL METODO K-MEANS	
SQ EUCLIDEAN	
Toma como base la raíz cuadrada de la distancia euclídea. Cada centroide es la media de los puntos que componen el clúster	
CITYBLOCK	
Tiene en cuenta la suma de las diferencias absolutas. Cada centroide es la mediana en cuanto a los componentes de los puntos de ese grupo	
COSINE	
Considera el valor 1 menos el coseno del ángulo incluido entre los puntos (tratados como vectores). Cada centroide es la media de los puntos de ese grupo, una vez que se han normalizado esos puntos a la unidad de longitud euclídea	
CORRELATION	
Considera el valor 1 menos la correlación muestral entre los puntos tratando estos como secuencias de valores. Cada centroide es la media en cuanto a componentes de los puntos en cada uno de los grupos, una vez centrados y normalizados los puntos en la media de desviación estándar cero.	

Tabla 6.- Medidas de distancia para k-means empleadas en este estudio. Fuente: Elaboración propia.

En este método se consideran los centroides como puntos medios de cada grupo y es importante referirse a ellos para obtener las medias y, así, son considerados como representativos de cada uno de los clústeres. Si incluyéramos un nuevo punto a un grupo, será a uno cuyo centroide sea el más próximo, y una vez incluido recalcularemos el nuevo centroide considerando también este nuevo punto dentro del clúster. Este proceso se repetirá tantas veces como sea necesario hasta que se obtenga un centroide estable.

El criterio que se sigue para medir la calidad del agrupamiento obtenido es el SSE del error (Suma de los Errores Cuadráticos).

Este error tratará de minimizarlo el propio algoritmo teniendo para ello en cuenta los centroides de cada grupo.

$$SSE = \sum_{j=1}^k \sum_{x \in G_j} \frac{p(x_i, c_j)}{n}$$

Siendo k el número de grupos, p la función de proximidad, c_j el centroide del grupo j y n el número de muestras de datos incluido

2.2.3.- Método de clustering aglomerativo.

Teniendo en cuenta cómo se actualiza la matriz de proximidad entre puntos, y para fusionar los distintos grupos se han definido distintos métodos de enlace (linkage).

En alguno de los artículos del cuerpo de la presente tesis se han utilizado los distintos métodos de enlace reflejados en la tabla 7.

DISTINTOS METODOS DE ENLACE UTILIZADOS EN EL CLUSTERING AGLOMERATIVO	
SIMPLE (SINGLE)	
Tiene en cuenta la distancia más corta:	
$d'(\bar{k}, \{i, j\}) = \min\{d(\bar{k}, i), d(\bar{k}, j)\}$	
COMPLETO (COMPLETE)	
Tiene en cuenta la distancia más larga:	
$d'(\bar{k}, \{i, j\}) = \max\{d(\bar{k}, i), d(\bar{k}, j)\}$	
MINIMA VARIANZA (WARD)	
Utiliza el algoritmo de varianza mínima que tiene en cuenta la distancia al cuadrado interior. Se trata de un método apropiado únicamente para distancias euclídeas.	
MEDIANA (MEDIAN)	
Utiliza el método WPGMC (Weighted Pair Group Method with Centroid Averaging) ¹¹ o de centros de distancia de masa ponderada. Se trata de un método apropiado únicamente para distancias euclídeas.	
MEDIA (AVERAGE)	
Utiliza el método UPGMA: (Unweighted Pair Group Method with Arithmetic Averaging) o de distancia promedio no ponderada.	
CENTROIDE (CENTROID)	
Utiliza el método UPGMC: (Unweighted Pair Group Method with Centroid Averaging) considera las distancia entre centroides y se trata de un método apropiado únicamente para distancias euclídeas.	
PONDERADA (WEIGHTED)	
Utiliza el método WPGMA: (Weighted Pair Group Method with Arithmetic Averaging) considernado distancias medias ponderadas.	

Tabla 7.- Métodos de enlace (linkage) empleados para el agrupamiento aglomerativo en este estudio. Fuente: Elaboración propia.

Los algoritmos de enlace único y los de enlace completo son los más ampliamente utilizados por los algoritmos de agrupación jerárquica. Estas dos tipologías de algoritmos se diferencian fundamentalmente en la manera de caracterizar la similitud entre un par de grupos:

1. En los algoritmos de enlace único (single), se considera la distancia entre dos grupos, como el valor mínimo de las distancias entre los pares de patrones que están incluidos en cada uno de los clústeres.

2. En los algoritmos de enlace completo, por el contrario, se considera la distancia entre dos grupos, como el valor máximo de todas las distancias entre los pares de patrones que están incluidos en cada uno de los clústeres.

En ambos casos la fusión de dos grupos para formar uno mayor se produce teniendo en cuenta los criterios de mínima distancia. Por un lado, el algoritmo de enlace completo produce enlaces compactos (Baeza-Yates R, 1992), mientras que por otro el algoritmo de enlace único se ve expuesto al llamado efecto en-cadenamiento, consistente en que existe una tendencia a producir agrupamientos alargados, que son desiguales. En cuanto a ventajas y desventajas de cada uno de los algoritmos se puede destacar que, el algoritmo de enlace completo produce generalmente jerarquías más útiles en multitud de aplicaciones. Por su parte el algoritmo de enlace único ofrece una mayor versatilidad. Este es el motivo por el que se ha experimentado con ambas alternativas en la presente investigación.

2.3 Redes Neuronales.

Los modelos de aprendizaje máquina conocidos como Redes Neuronales Artificiales (RNA) están inspirados en el funcionamiento de las neuronas biológicas. Se lleva a cabo mediante una simulación con un conjunto de ellas conectadas entre sí. Este conjunto de unidades se transmiten señales a través de las conexiones que las unen. Pueden utilizarse para casos muy diferentes ya que su naturaleza es muy diferente.

Se explican a continuación las tres metodologías necesarias para realizar la investigación desarrollada. Así, en primer lugar, abordaremos los métodos neuronales para la predicción de series temporales para, posteriormente, conocer las metodologías NAR, NIO y NARX.

2.3.1.- Métodos neuronales para predicción de series temporales.

Estos métodos se han utilizado, en la presente investigación, para obtener proyecciones, tal y como se ha explicado con anterioridad. También se pueden utilizar para realizar análisis y predicciones de series temporales. Para casos de aprendizaje supervisado como es este último, se utilizan con mucha frecuencia. Se presentan a continuación.

Matlab incluye varios algoritmos de entrenamiento del modelo neuronal y se han recogido en la siguiente tabla 8.

METODOS NEURONALES DE APRENDIZAJE SUPERVISADO PARA PREDICCIÓN DE SERIES TEMPORALES	
NOMBRE	BREVE DESCRIPCIÓN
1.- LEVENBERG - MARQUARDT	LMA o también LM. Es el método de los mínimos cuadrados amortiguados (DLS). Se utiliza para resolver problemas de mínimos cuadrados no lineales
2.- BATCH GRADIENT DESCENT	Algoritmo de optimización iterativa de primer orden para encontrar el mínimo de una función
3.- GRADIENT DESCENT WITH MOMENTUM	Mejora el algoritmo anterior al agregar un término de impulso
4.- ADAPTIVE LEARNING RATE BACK PROPAGATION	Realiza una búsqueda progresiva de mejora en base a resultados previos
5.- GRADIENT DESCENT WITH MOMENTUM AND ADAPTIVE LEARNING RATE	Es un método que combina propiedades de los dos anteriores
6.- SCALED CONJUGATE GRADIENT	Mejora del método de descenso de gradiente en base a magnitudes
7.- BROYDEN - FLETCHER - GOLDFARB SHANNO BACKPROPAGATION (QUASI-NEWTON)	Método iterativo para optimización de problemas no lineales, dentro de la familia de métodos cuasi newtonianos

Tabla 8.- Métodos neuronales de aprendizaje supervisado para predicción de series temporales. Fuente: Elaboración propia

Estos algoritmos se han utilizado para realizar distintas pruebas, lo que nos ha permitido llevar a cabo distintos experimentos con numerosas combinaciones de valores.

El resultado de cada experimento surge al combinar distintos valores de los parámetros (número de neuronas en capa oculta y número de retardos en las distintas fuentes de datos)

En este caso se han utilizado con la finalidad primera de obtener proyecciones, aunque también se ha pretendido usarlos para analizarlos y obtener predicciones de series temporales.

Por tratarse de un aprendizaje supervisado, se dispone de un valor de salida real y se puede medir el error cometido por el modelo en cada uno de los experimentos. Se ha realizado utilizando la métrica MSE (*Mean Squared Error*) que viene definida por:

$$MSE = \frac{1}{N} \sum_{i=1}^N (t - a)^2$$

Siendo t el valor a predecir de la serie y a el valor realmente obtenido por el modelo.

2.3.2.- Nonlinear AutoRegresive (NAR).

En el modelo NAR (*Nonlinear Autoregressive*) se realiza una predicción en el instante de tiempo t de una serie temporal (y), a partir exclusivamente de valores de la propia serie temporal en instantes anteriores de tiempo. De esta forma se cumple:

$y(t) = f(y(t-1), \dots, y(t-n_y))$ Siendo n_y el número de retardos a considerar en el modelo. Es decir, la cantidad de valores anteriores de la serie

Al utilizar MATLAB para su implantación, se crea una red neuronal con propagación hacia adelante que incorpora una capa oculta. Esta capa oculta dispone de una función de activación tangente hiperbólica y una única neurona en la capa de salida, con función de activación lineal. Además, se utilizan distintos algoritmos para el entrenamiento de este modelo. El modelo puede verse en la figura 2.

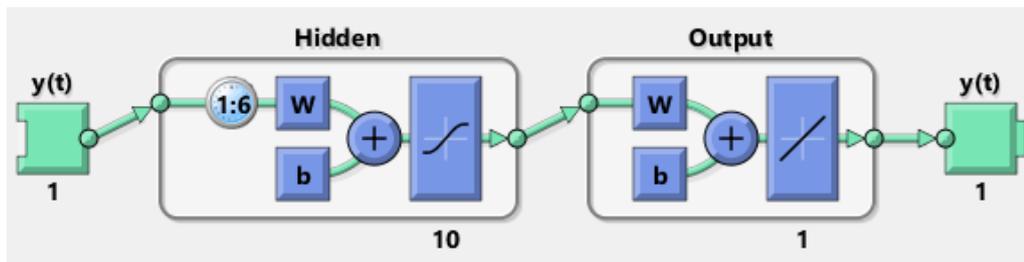


Figura 2.- Representación de MATLAB para modelo NAR con 10 neuronas en capa oculta y una salida con un retardo de 6. Fuente: The Mathworks Inc.

En este modelo probamos distintos valores para los parámetros ajustables (número de neuronas y el número de retardos (n_y)) que se corresponden a la serie de entrada de los datos previos de la serie, realizando distintos experimentos junto con el algoritmo de entrenamiento.

2.3.3.- Nonlinear Input-Output (NIO).

En el modelo NIO (Nonlinear Input - Output) también se realiza una predicción de futuros valores de una serie temporal $y(t)$. La predicción también se lleva a cabo a partir exclusivamente de valores de una serie temporal en instantes anteriores de tiempo. No obstante, y a diferencia del modelo NAR, esta serie no es la propia entrada sino una entrada distinta o exógena. El modelo queda detallado a través de la siguiente ecuación:

$y(t) = f(x(t-1), \dots, x(t-n_y))$ Siendo n_y el número de retardos a considerar en el modelo. Es decir, la cantidad de valores anteriores de la serie exógena a considerar para calcular un nuevo valor de la serie a predecir.

En la implementación de MATLAB empleada, se crea una red neuronal con propagación hacia adelante que incorpora una capa oculta que incluye una función de activación tangente hiperbólica, y una única neurona en la capa de salida cuya función de activación es lineal. tal y como se indica en la figura 3. Además,

se emplean los distintos algoritmos descritos para el entrenamiento de este modelo.

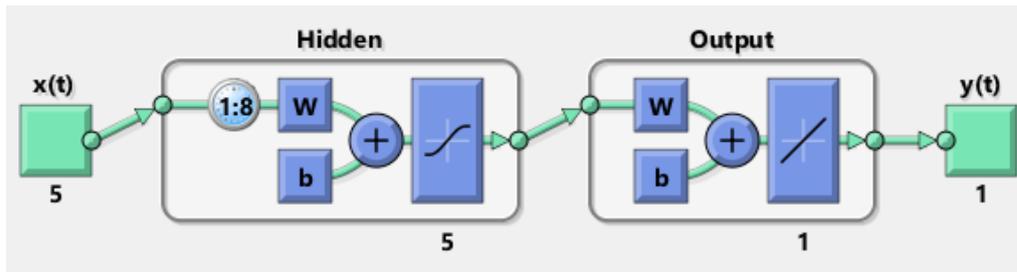


Figura 3.- Representación de MATLAB para modelo NIO con 5 neuronas en capa oculta y una entrada exógena con 5 valores y un retardo de 8. Fuente: The Mathworks Inc.

Al igual que en el caso de NAR, en este modelo se realizan varios experimentos y se prueba con distintos valores para los parámetros ajustables junto con el algoritmo de entrenamiento. Estos parámetros son el número de neuronas y el número de retardos (n_x).

2.3.4.- Nonlinear AutoRegressive with eXogenous input (NARX).

El modelo neuronal conocido como NARX (Nonlinear Autoregressive with Exogenous Input) (Leontaritis and Billings 1985), también se ha utilizado en esta investigación. El modelo predice valores de una serie temporal $y(t)$, a partir tanto de valores anteriores de la propia, como de otra serie temporal, $x(t)$. Su formulación general sigue la siguiente expresión:

$$y(t) = f(y(t-1), \dots, y(t-n_y), x(t-1), \dots, x(t-n_x)) + \xi(t)$$

Siendo n_y el número de retardos en la salida (feedback), n_x el número de retardos en la entrada y $\xi(t)$ un término asociado a ruido (asumido como gaussiano y blanco generalmente).

Cuando se aplica este modelo se trata de hallar la función no lineal que lleva a cabo el mapeo entre los datos de entrada (valores previos x e y) y la salida (valor a predecir de la serie y). De tal forma que NARX se puede considerar como una generalización no lineal de los modelos ARX, que son modelos dinámicos lineales de auto regresión, utilizando variables exógenas, que se aplican muy frecuentemente.

En estas investigaciones también se ha utilizado una implementación neuronal de este modelo en MATLAB. Se ha creado una red neuronal con tan sólo una capa oculta y una neurona en la capa de salida. La capa oculta dispone de una función de activación tangente hiperbólica y la capa de salida una función de activación lineal. Como principal característica de esta red es que permite como

entrada tanto a los valores de la propia serie a predecir, $y(t)$, como a los valores de la serie externa o exógena, $x(t)$. Su representación es la que se indica en la siguiente figura 4.

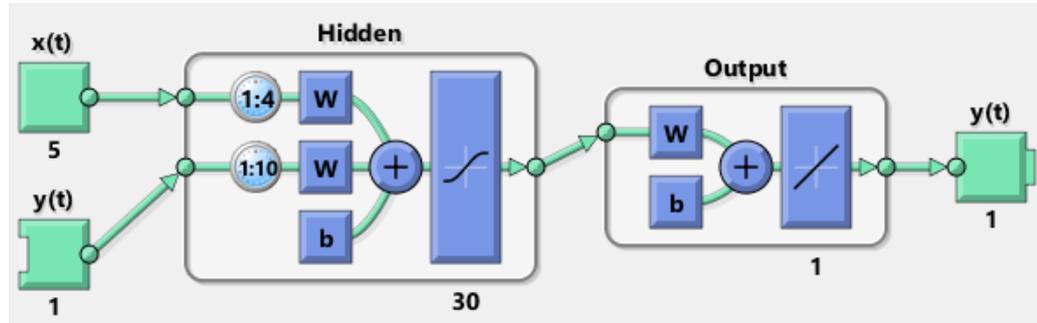


Figura 4.- Representación de MATLAB para modelo NARX con 30 neuronas en la capa oculta y una neurona con la capa de salida, una entrada exógena con 5 valores y un retardo de 4 y una única salida con un retardo de 10. Fuente: The Mathworks Inc.

En el caso de predicción de series cronológicas con largas dependencias temporales se recomienda el empleo de modelos NARX con descenso de gradiente frente al uso de modelos basados en Perceptrones Multicapa (T. Lin et al., 1998). El rendimiento de este modelo para la predicción de series temporales es mejor que el de otras implementaciones neuronales (Menezes & Barreto, 2008). Este ha sido uno de los motivos para seleccionar este modelo en el trabajo de investigación que nos ocupa.

2.4 Metodología Altman Z-score.

La economía financiera de la empresa es la rama de la ciencia económica que se ocupa de los problemas económicos financieros de la empresa. Por una parte, trata de interpretar la realidad económico - financiera de la empresa y por otro trata de explicar su comportamiento. Alcanza tanto al conocimiento de “lo que es” la empresa en sí, como a “lo que puede llegar a ser” mediante la formulación de predicciones científicas y una vez incorporados juicios de valor, puede llegar a dar modelos de “lo que debe ser” (Partal Ureña A. et al., 2016).

Ya en la primera mitad de los años treinta del pasado siglo XX, algunos autores como (Fitzpatrick, 1932), además de (Smith, 1935), aplicaron métodos de análisis univariantes muy básicos con la idea de predecir de una manera científica la probabilidad de quiebra de las empresas, teniendo como base las distintas ratios financieras.

Años más tarde, a mediados de los años sesenta algunos otros investigadores como (Beaver, 1966) decidieron incluir en sus trabajos un análisis univariante

discriminatorio evaluando separadamente cada una de las ratios financieras. Esto supuso un importante avance en la capacidad predictiva de quiebra, no obstante, fueron desechados en los años subsiguientes y reemplazados por los análisis multivariantes, ya que estos son más precisos.

En el año 1968 el entonces profesor de finanzas de la Universidad de Nueva York Edward I. Altman publicó una fórmula de un indicador que llamó Z para predecir la bancarrota de las empresas. La fórmula se podía utilizar para predecir si una empresa, en función de la información que se desprendía de sus cuentas iba a estar en bancarrota en un período de dos años. Este y los posteriores indicadores Z se pueden utilizar para predecir problemas financieros, tratándose de una medida de control fácil de calcular.

El indicador Z utiliza múltiple ratios y valores que se obtienen tanto del balance como de la cuenta de pérdidas y ganancias de las empresas

2.4.1.- Estimación de la fórmula.

El estimador Z combina de forma lineal varios índices financieros comunes, generalmente cuatro o cinco, ponderados por coeficientes. Para estimar los coeficientes se identificaron un grupo de empresas que se habían declarado en quiebra y se compararon con otras de tamaño similar y del mismo sector que no había sufrido la bancarrota

En un principio, Altman recogió los datos de empresas que cotizaban en bolsa y aplicó para su análisis métodos estadísticos de análisis discriminante lineal.

Más adelante volvió a analizar datos de otras empresas, pero no sólo empresas manufactureras o de producción, si no también, para empresas de servicios.

La muestra inicial constaba de datos pertenecientes a 66 empresas, de las cuales la mitad de ellas se declararon en bancarrota. En este caso todas ellas eran manufactureras y se descartaron las que se consideraron pequeñas por tener activos inferiores a 1 millón de dólares (USD)

La fórmula original para el valor de Z era:

$$Z = 1.2X1 + 1.4X2 + 3.3X3 + 0.6X4 + 1.0X5.$$

Siendo:

CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS

X1 = Fondo de maniobra / Activos Totales. Mide los activos líquidos en relación con el tamaño de la empresa.

X2 = Reservas / Activos Totales. Mide la rentabilidad y refleja la edad y el poder adquisitivo de la empresa.

X3 = Beneficios Antes de Intereses e Impuestos (EBIT) / Activos Totales. Mide la eficiencia operativa excluyendo los factores fiscales y de apalancamiento. Reconoce que los beneficios de la actividad son importantes para la viabilidad a largo plazo.

X4 = Valor de Mercado de la empresa (capitalización bursátil) / Valor en Libros del Pasivo Total. Agrega una dimensión de mercado que puede mostrar la fluctuación del precio de las acciones como una posible señal de alerta.

X5 = Ventas / Activos Totales. Medida estándar de la rotación total de activos (varía mucho de una industria a otra).

Altman descubrió que para el grupo de las empresas en quiebra se obtenía un valor medio de -0,25 de media, mientras que para el grupo que no quebró el valor medio se situaba en torno al +4,48.

2.4.2.- Precedentes.

El trabajo de Altman se basó en el trabajo desarrollado por del investigador de aspectos contables William Beaver (Beaver, 1966) y otros. A partir de la década de 1930, (Mervin, 1930) y otros recolectaron un número mayor de muestras y apreciaron que varios índices contables parecían ser valiosos para predecir la bancarrota. El indicador Z de Altman es una versión personalizada de la técnica de análisis discriminante de (FISHER, 1936).

El trabajo de William Beaver, publicado en 1966 y 1968, fue el primero en el que se aplicaba un método estadístico (método t-test) para predecir la quiebra de una muestra de empresas emparejadas. Beaver aplicó este método para evaluar la importancia de cada uno de varios índices contables basándose en análisis univariantes, usando cada ratio contable de uno en uno. La principal mejora de Altman fue aplicar otro método estadístico: el análisis discriminante, que podía tener en cuenta múltiples variables simultáneamente, y no sólo una variable cada vez.

2.4.3.- Precisión y eficacia.

En su prueba inicial, se encontró que el Z-Score de Altman tenía una precisión del 72 % en la predicción de la quiebra dos años antes del evento, con un error de tipo II (falsos negativos) del 6 % (Altman, 1968). En una serie de pruebas posteriores en las que se cubrieron tres períodos durante los siguientes 31 años (hasta 1999), se encontró que el modelo tenía una precisión de aproximadamente 80% a 90% considerando predicción de la quiebra un año antes del evento, con un error tipo II (clasificando a la empresa como “quiebra” cuando en realidad fue “no quiebra”) de aproximadamente 15%–20% (Altman, 2000).

Aproximadamente desde 1985 en adelante, los predictores Z obtuvieron una amplia aceptación por parte de auditores, jefes de contabilidad, tribunales y sistemas de bases de datos utilizados para la evaluación de préstamos (Eidleman 1995). El enfoque de la fórmula se ha utilizado en una gran variedad de contextos y países, aunque fue diseñado originalmente para empresas manufactureras públicas que cotizaban en bolsa con activos de más de 1 millón de dólares. Las variaciones posteriores de Altman fueron diseñadas para poder aplicarlas también en empresas privadas (Indicador Altman Z') y en empresas no manufactureras (Indicador Altman Z'').

No se recomienda el uso de los modelos de Altman, ni otros tipos de modelos basados en las cuentas anuales de las empresas, para organizaciones financieras.

Esto se debe a la opacidad de los balances de las empresas financieras y al uso frecuente de partidas fuera del balance. Existen fórmulas basadas en el mercado que se utilizan para predecir el incumplimiento de las empresas financieras tal y como indican Merton y Thackor (2022) en el modelo de Merton, pero tienen un valor predictivo limitado porque se basan en datos de mercado (fluctuaciones de precios de acciones y precio de las opciones que implican fluctuaciones en los valores de los activos) para predecir un evento de mercado como por ejemplo incumplimiento, es decir, la disminución de los valores de los activos por debajo del valor de los pasivos de una empresa.

2.4.4.- Definición de variables del indicador Z de Altman original.

En su definición original se consideraron las siguientes variables e indicadores, descritas a continuación, para poder obtener el modelo.

CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS

X1 = Fondo de Maniobra / Activos totales

X2 = Reservas / Activos totales

X3 = Beneficios antes de intereses e impuestos / Activos totales

X4 = Valor de Mercado del Patrimonio Neto / Pasivo Total

X5 = Ventas / Activos Totales = rotación de activos

Modelo de quiebra del indicador Z original:

$$Z = 1.2X1 + 1.4X2 + 3.3X3 + 0.6X4 + .999X5$$

Zonas de Discriminación:

$Z > 2.99$ -Zona "segura"

$1.81 < Z < 2.99$ -Zona "gris"

$Z < 1.81$ -Zona "peligro"

2.4.5.- Estimación del índice Z' de Altman para empresas privadas.

En este caso, para empresas privadas, fueron consideradas las siguientes ratios financieras

X1 = Capital Circulante / Activo Total

X2 = Beneficios acumulados / Activos totales

X3 = Beneficios antes de intereses e impuestos / Activos totales

X4 = Valor en Libros del Patrimonio Neto / Pasivo Total

X5 = Rotación de activos = (Ventas/ Activos Totales)

Y se obtuvo el siguiente modelo de predicción de bancarrota de Altman Z' para empresas privadas:

$$Z' = 0.717X1 + 0.847X2 + 3.107X3 + 0.420X4 + 0.998X5$$

Zonas de Discriminación:

$Z' > 2.9$ -Zona "segura"

$1.23 < Z' < 2.9$ -Zona "gris"

$Z' < 1.23$ -Zona "peligro"

2.4.6.- Valor de predicción Z'' de Altman para empresas no manufactureras y mercados emergentes.

También se estudió la situación particular de las empresas no manufactureras y mercados emergentes, utilizando las cuatro ratios que se indican a continuación.

$$X1 = (\text{Activo Corriente} - \text{Pasivo Corriente}) / \text{Activo Total}$$

$$X2 = \text{Beneficios acumulados} / \text{Activos totales}$$

$$X3 = \text{Beneficios antes de intereses e impuestos} / \text{Activos totales}$$

$$X4 = \text{Valor en Libros del Patrimonio} / \text{Pasivo Total}$$

Los resultados de la investigación se plasmaron en los dos modelos Z'' que a continuación se describen.

Modelo de quiebra Z-Score:

$$Z'' = 3,25 + 6,56X1 + 3,26X2 + 6,72X3 + 1,05X4$$

Modelo de quiebra Z-Score (para mercados emergentes):

$$Z'' = 6,56X1 + 3,26X2 + 6,72X3 + 1,05X4$$

Zonas de discriminación:

$Z > 2.6$ -Zona "segura"

$1.1 < Z < 2.6$ -Zona "gris"

$Z < 1.1$ - Zona "peligro"

CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS

En el tercer capítulo, se refleja y justifica toda la investigación realizada aplicando las metodologías expuestas en el capítulo segundo, a distintos problemas de gestión empresarial.

En primer lugar,

3.1 Una herramienta de visualización inteligente para analizar la sostenibilidad del transporte por carretera.

El transporte por carretera es parte integrante de la actividad económica y, por tanto, fundamental para su desarrollo. Como contrapartida, supone el 30% de las emisiones mundiales de GEI (Gases de Efecto Invernadero), de las que casi un tercio corresponden al transporte de mercancías en vehículos pesados por carretera. Además, los medios de transporte siguen evolucionando técnicamente y están sujetos a normativas cada vez más exigentes que pretenden reducir sus emisiones. Para analizar la sostenibilidad de esta actividad, este estudio propone la aplicación de novedosas técnicas de Inteligencia Artificial (más concretamente, Machine Learning). En esta investigación se amplía el uso de Parcelas Exploratorias Híbridas No Supervisadas (HUEP: Hybrid Unsupervised Exploratory Plots) con nuevas técnicas de Persecución de Proyección Exploratoria (EPP: Exploratory Projection Pursuit). Estos, junto con las técnicas de agrupamiento, forman una herramienta de visualización inteligente que permite obtener conocimiento a partir de un conjunto de datos previamente desconocido. La propuesta se prueba con un gran conjunto de datos de la encuesta oficial de transporte por carretera en España (EPTMC), que se llevó a cabo durante un período de 7 años. Los resultados obtenidos son interesantes y proporcionan evidencia alentadora para el uso de esta herramienta como un medio de análisis inteligente en el tema de la evolución de la sostenibilidad del transporte por carretera.

3.1.1.- Introducción y trabajo previo.

Los 17 Objetivos de Desarrollo Sostenible del plan director de la ONU incluyen una línea de acción determinada para frenar el calentamiento global, con una relación directa con la reducción de las emisiones de CO₂. La Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC) y el Protocolo de Kyoto, ambos firmados en 2002, apuntan en la misma dirección.

Además, el transporte por carretera es consustancial a la actividad económica y, por tanto, esencial para el futuro de nuestra civilización. Pero como contraposición, representó el 30% (OECD/IFT, 2016) de las emisiones globales de CO₂ en 2015, y el 29% de las 314.529 kilotoneladas de GEI (Gases de Efecto Invernadero) emitidas en 2019 en España, con un total de 91.372 kilotoneladas de CO₂ equivalente emitidas. A su vez, se estima a nivel mundial y de la UE que casi una cuarta parte de estas emisiones corresponden al transporte de mercancías en vehículos pesados y autobuses (Ministerio para la Transición Ecológica, 2021), siendo esta una de las áreas que ha registrado un crecimiento sostenido en años recientes. Mientras tanto, otros modos de transporte reducen progresivamente su cuota, estimándose un aumento de cuatro veces sus emisiones actuales para 2050 si no se adoptan medidas al respecto. A la vista de estos datos y previsiones, parte de los estudios científicos y líneas de actuación para la gobernanza han puesto el foco en este ámbito, avanzando en el desarrollo de propuestas de regulación de emisiones de los equipos que realizan esta actividad, que supone el 6% de las emisiones totales de GEI de la Unión Europea (Consejo-Europeo, 1998).

3.1.1.1.- Sostenibilidad en el transporte.

La revisión sistemática de los sistemas de intervención para la sostenibilidad del transporte de mercancías por carretera reveló que la línea de actuación más evidente es la aplicación de tecnologías emergentes combinadas con políticas energéticas decisivas (Tob-Ogu et al., 2018). Otras líneas de actuación complementarias se centran en la utilización y adecuación de las infraestructuras al desarrollo efectivo de esta actividad y el fomento de la intermodalidad en la gestión eficiente de las cadenas de suministro (Shankar et al., 2019). En muchos casos, estos se centran en una apuesta firme por la sincro-modalidad (Agbo A.A. et al., 2017) e incluso se inclinan por medir la sostenibilidad del transporte en función de su intermodalidad (Kelle et al., 2019). También se distingue la eficacia de distintas medidas a favor de la sostenibilidad en el transporte de recogida y entrega, con una mayor aplicabilidad de las técnicas colaborativas apoyadas en sistemas de información, frente al transporte de larga distancia, donde los resultados de la optimización tecnológica en materia de transporte, el consumo

de equipos y la adecuación de las infraestructuras son más convincentes (Meyer, 2020). Algunos modelos, que reflejan la estrecha conexión entre la economía y el transporte, se basan en el análisis de un enfoque de sostenibilidad económica de los modelos logísticos (Rossi et al., 2020) y, en la misma línea, algunos de ellos se centran en el análisis de determinados bienes a partir de su estructura productiva y distribución de mercancías por carretera, analizando su actividad completa (De Campos et al., 2019).

En este sentido, podríamos decir que estamos siguiendo una perspectiva clásica en la búsqueda de la sostenibilidad, a la que apunta (Corlu et al., 2020) en su revisión del estado del arte sobre la optimización del consumo energético en los procesos de transporte, como combinación de tres principales líneas de actuación para el transporte por carretera: i) la mejora del factor de carga, entendido como un reductor de emisiones en función de la ocupación máxima del medio de transporte, ii) el uso de técnicas colaborativas para optimizar la asignación de los medios de transporte y iii) la definición y el seguimiento de los objetivos de sostenibilidad en el desarrollo de las operaciones de transporte. No obstante, lo más destacable del conjunto es que todos los modelos y estudios apuntan, en definitiva, a un objetivo subyacente común; la optimización de la energía consumida en el proceso de las operaciones de transporte de mercancías.

Una línea de investigación directamente asociada al objetivo central de mejora energética es el estudio de las emisiones de los equipos de transporte en función de su antigüedad. (Hassani et al., 2021) determinan que las emisiones en vehículos ligeros pueden ser hasta 5 veces mayor, dependiendo de su antigüedad. Debido a este hallazgo, se le da especial relevancia a esta variable en este estudio.

Estos resultados no son arbitrarios; más bien son el resultado de la implementación de políticas adoptadas por el sector de fabricación de vehículos. Ya en 2002, (Ang-Olson & Schroeder, 2002) determinaron que la implementación adecuada de soluciones técnicas en la producción de vehículos pesados de transporte por carretera podría llevar a una reducción de más de 11 mil millones de litros de consumo anual de combustible en un período de 10 años, asociado con una disminución de 8,3 millones de toneladas de emisiones de gases de efecto invernadero solo en los Estados Unidos (EE. UU.). Desde entonces, los mecanismos de gobernanza de EE. UU. y la UE han seguido imponiendo mandatos a la producción de vehículos de transporte, con el objetivo de reducir el consumo y las emisiones (Comisión Europea, 2002). Como consecuencia de estos mandatos, se han ido desarrollando progresivamente las denominadas Euro-normas de emisión, con requisitos cada vez más exigentes sobre consumos y emisiones de los elementos de transporte, tal y como se muestra en la

Tabla 9, sobre Monóxido de Carbono (CO), Óxidos de Nitrógeno (NOX) y material en partículas (PM).

Emission Standards for Diesel Cars				
Standard	Date	CO g/Km	NOX g/Km	PM g/Km
Euro 4	2005	0.50	0.30	0.025
Euro 5	2010	0.50	0.23	0.005
Euro 6	2015	0.50	0.17	0.005
Emission Standards for Heavy Goods Vehicles				
Standard	Date	CO g/KWh	NOX g/KWh	PM g/KWh
Euro IV	2005	1.50	3.50	0.020
Euro V	2008	1.50	2.00	0.020
Euro VI	2013	1.50	0.40	0.010

Tabla 9.- Emission standards adopted by the EU for diesel cars and heavy goods vehicles. (Comisión Europea, 2005)

Si nos centramos específicamente en los vehículos pesados, (Haugen & Bishop, 2018) establecen dos muestreos completos de emisiones de vehículos en puntos de carga y descarga con importantes movimientos de mercancías, determinando finalmente que una reducción de la edad media de estos vehículos de 7,8 a 6 años se traduce en una reducción significativa de las emisiones de hasta un 87% de material en forma de partículas en suspensión. Un estudio posterior (Yu et al., 2021) también determinaron que la reducción de las emisiones nocivas de NOX disminuyó de 38 g a 9 g por kg de combustible diésel consumido por vehículos pesados de 2005 a 2020, lo que representa una reducción del 76,3 % en los vehículos producidos entre estas fechas. es decir, una reducción promedio de 5% en las emisiones por cada año de producción de vehículos durante este período.

Con las posibles diferencias en la cuantificación de las mejoras de emisiones, hay una conclusión clara; los vehículos pesados contribuyen de manera importante a la emisión de gases nocivos, y la antigüedad de dichos vehículos es también un factor determinante en su eficiencia y sostenibilidad.

3.1.1.2.- Trabajo previo de digitalización.

Surgen nuevas perspectivas al incorporar el enfoque de digitalización al transporte por carretera. Un estudio realizado por un panel de 52 expertos (Perneštal et al., 2020), pone de manifiesto el valor de la automatización de procesos, la recopilación completa de datos sobre información digital y la base para

la aplicación de inteligencia artificial para un adecuado desempeño y planificación. Dando un paso más en esta dirección, y con un claro enfoque en la mejora de la gestión de las cadenas de suministro, se realizó un estudio sobre los datos recogidos por la UE en las Encuestas Permanentes de Transporte de Mercancías por Carretera (Mangina et al., 2020). Con este fin, se realizó una modelización de esta actividad, que fue recopilada por EUROSTAT para sus países miembros entre 2011 y 2014. Se llevó a cabo bajo un modelo de Colaboración Horizontal (Soysal et al., 2018) y se analizaron las mejoras que hubieran resultado de la aplicación de técnicas de Pooling, similares a los agrupamientos de carga optimizados, y un modelo de implementación del llamado Internet Físico (Montreuil, 2011). Esto se hizo mediante el desarrollo de modelos informáticos que simulan cada opción, y que se utilizan para estimar un cálculo de emisiones. El resultado final es una clara ventaja para el modelo de Internet físico actualmente puramente teórico.

Más específicamente, se han realizado contribuciones anteriores sobre el uso de la Inteligencia Artificial (IA) en general y el Aprendizaje Automático (ML) en particular, para abordar los problemas de sostenibilidad en el transporte de mercancías por carretera. En (Kumar Dadsena et al., 2019), se propuso un enfoque integrado de análisis de efectos y modo de falla difusa para la selección de estrategias de mitigación de riesgos en tareas de seguimiento y localización en la industria, cuyo objetivo era ayudar a los directivos a elegir una estrategia considerando la criticidad de los riesgos cuando el presupuesto es bajo. Además, en (Wong et al., 2020) se presenta un nuevo modelo de optimización del transporte de contenedores que incorpora la red de carreteras junto con métricas de conectividad, con el objetivo de minimizar la distancia total del viaje, el costo del combustible del camión, el costo del alquiler de contenedores y los movimientos de contenedores entre múltiples consignatarios y transportistas. Más recientemente, (Budak & Sarvari, 2021) proponen un enfoque novedoso para predecir el margen de beneficio, en función del cliente, en el sector del transporte sostenible de mercancías por carretera, mediante la combinación de diferentes métodos de aprendizaje automático. Esto ayuda a los directivos a obtener información útil sobre perspectivas estratégicas y de desarrollo sostenible.

A diferencia de este estudio anterior, los autores de este artículo han aplicado diferentes técnicas de ML (Alonso de Armiño et al., 2020a) a la misma familia de datos (ver apartado 3.1.2). Se han obtenido resultados significativos en cuanto a su vinculación con los ciclos de actividad económica. También se han identificado los puntos de inflexión económicos que se produjeron durante la Gran Recesión en España, basándose exclusivamente en diversas técnicas de agrupación del conjunto de datos de transporte de mercancías por carretera. Se observó una fuerte recesión hasta el segundo trimestre de 2012, seguida de

una severa depresión durante el período siguiente hasta el segundo trimestre de 2013, seguida de una recuperación gradual hasta el segundo trimestre de 2015, donde se produjo una clara fase de recuperación económica.

El conjunto de datos que muestra la edad media de la flota de vehículos pesados de carretera nunca dejó de aumentar en ningún momento durante el período en cuestión (aunque aumentó a diferentes niveles). Coincidió casi exactamente en los puntos de inflexión económica con la distribución por cuartiles de los datos, como puede verse en la Figura 5. Este crecimiento acelerado o sostenido no corresponde más que a un patrón de sobre amortización de los medios de producción en tiempos de recesión económica, como medida para proteger la rentabilidad de su actividad económica.

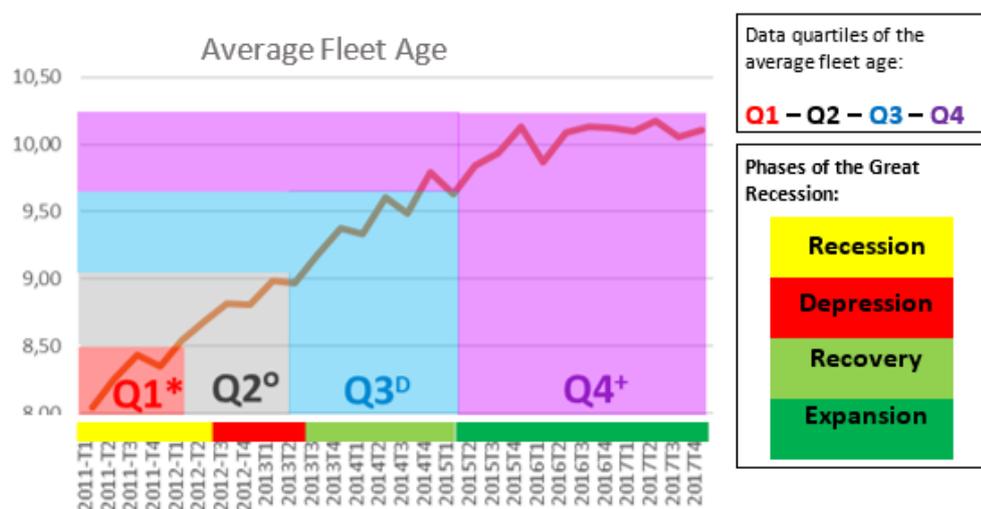


Figura 5.- Serie de datos de edad de la flota de transporte, dividida en cuartiles y relacionada con las fases determinadas para la Gran Depresión. Fuente: Elaboración propia

Hemos decidido centrar nuestros estudios en esta serie de datos agrupados en cuartiles por las siguientes razones: en primer lugar, está específicamente ligada a la sostenibilidad de la actividad del transporte; y, en segundo lugar, su distribución de datos por cuartiles coincide con las fases económicas. Además, el estudio de esta serie, clave para la interpretación de la sostenibilidad del parque de vehículos de transporte de mercancías, no tuvo en cuenta la evolución y estudio del resto de series que reflejaban aumentos o descensos puntuales en la fase de depresión, que hace aconsejable aplicarle técnicas de análisis complementarias.

Para abordar este problema, en este estudio se presentan herramientas de visualización pioneras basadas en ML. Más específicamente, las técnicas de agrupamiento y Búsqueda de Proyección Exploratoria (EPP) se han combinado por primera vez, bajo el marco de Parcelas Exploratorias No Supervisadas Híbridas (HUEP), para apoyar el análisis visual de los datos de sostenibilidad relacionados con el transporte por carretera. Asimismo, se amplía la formulación original de los HUEPs, y también se proponen y validan nuevas técnicas de proyección.

El resto de los apartados de este estudio se organizan de la siguiente manera: en el apartado 3.1.2 se describen los métodos empleados junto con los datos sobre los que se aplican. Los resultados obtenidos en el estudio experimental se presentan en el apartado 3.1.3, y en la sección 3.1.4 se presentan las principales conclusiones en relación con estos, así como algunas propuestas de trabajo futuro,

3.1.2.- Materiales y método.

Como se indicó anteriormente, en este estudio se investigó la actividad del transporte por carretera, con un enfoque específico en su sostenibilidad. Esto se hizo analizando un conjunto de datos descrito en la subsección 3.1.2.1 con las técnicas novedosas que se presentan en la subsección 3.1.2.2

3.1.2.1.- Datos.

Los datos se obtuvieron de dos fuentes diferentes:

- El Ministerio de Transportes, Movilidad y Agenda Urbana de España, a través de su Subdirección General de Estudios Económicos y Estadísticas.
- La encuesta sobre el transporte europeo de mercancías por carretera (*ERFT: European Road Freight Transport*). Esta encuesta se refiere a la actividad de los vehículos pesados matriculados en España para el transporte de mercancías. Tiene un nivel de muestreo lo suficientemente alto como para tener representatividad estadística para cada Comunidad Autónoma, con el fin de medir sus operaciones de transporte. Con este objetivo, la encuesta registra el movimiento de una sola clase de mercancías, desde un punto de partida hasta un destino. La investigación se realizó de acuerdo con la normativa correspondiente (Consejo-Europeo, 1998) y su posterior revisión (Consejo-Europeo, 2012). El total de registros incluidos en dicha base de datos fue de 1.932.671 que tiene una representatividad muestral de 1.259.938.252 operaciones de transporte.

Se utilizaron datos de entre 2011 y 2017. Todos los datos representaban niveles de agregación trimestrales, por lo que incluían cada variable en estudio, un total de 28 valores. Las variables consideradas fueron:

- Costes de transporte (B): basado en un incremento del 100 por ciento sobre los precios medios anuales en 2000, según lo determinado por los estudios de investigación trimestrales del Ministerio de Fomento.
- Gastos de combustible en España: valores medios trimestrales ponderados en céntimos de euro, según indican los datos recabados por el Ministerio de Fomento.

CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS

- Gastos de combustible en la UE: valores medios trimestrales ponderados en céntimos de euro, según indican los datos recabados por el Ministerio de Fomento.
- Número de toneladas transportadas (A, B, C): peso de la mercancía transportada.
- Viajes completados (A, B, C): número de operaciones de transporte y distancia en vacío.
- Distancia en vacío (A, B): kilómetros recorridos sin mercancías.
- Carga máxima para operaciones de transporte (A, B): límite superior de peso para viajes realizados en toneladas.
- Carga máxima para distancia en vacío (A, B): límite superior de peso para la distancia en vacío recorrida en toneladas.
- Distancia de transporte (A, B): kilómetros recorridos.
- Distancia de transporte en vacío (A, B): kilómetros recorridos sin mercancías.
- Cantidad de vehículos representados (A): número de vehículos representados.
- Capacidad de carga representada (A): límite superior de carga de los vehículos representados.
- Toneladas-km (A, B, C): total de toneladas transportadas y distancia recorrida en cada operación de acarreo.
- Edad media de la flota (A, B): promedio de años transcurridos desde la matriculación de los vehículos. Como se indicó anteriormente en la sección 1.1, este es un hallazgo de datos importante con respecto a la sostenibilidad. Debido a esto, también se usa en la metáfora del glifo.
- Antigüedad media de la flota por recorrido en vacío (A, B): promedio de años transcurridos desde la matriculación de los vehículos que circulan sin mercancías.

Las series de datos con letras asignadas se subdividieron según:

(A) Tipo de transporte:

A1) Todos los transportes; A2) Transporte mercancías propias; A3) Servicio público.

(B) Rango de distancia:

B1) Todas las distancias; B2) < 50 km; B3) 51-100 km; B4) 101-200 km; B5) 201-300 km; B6) >300 km.

(C) Ámbito geográfico:

C1) Todos los ámbitos; C2) Municipal; C3) Regional; C4) Nacional; C5) Importación; C6) Exportación; C7) Cabotaje.

Se calcularon 113 series de datos de transporte, con los valores de los 28 trimestres antes indicados en cada una.

Como resultado, se requiere analizar un conjunto de datos con alta dimensionalidad para investigar la sostenibilidad del transporte por carretera.

3.1.2.2.- HUEP.

Las parcelas exploratorias híbridas no supervisadas (HUEP: Hybrid Unsupervised Exploratory Plots) (Herrero et al., 2019) se han propuesto recientemente como una nueva herramienta de visualización para combinar los resultados de la búsqueda de proyección exploratoria (EPP: Exploratory Projection Pursuit) y los métodos de agrupamiento de una manera novedosa e informativa. Para abordar el conocido desafío de la "maldición de la dimensionalidad" y avanzar en el análisis de datos descriptivos, los métodos EPP y Clustering se han aplicado de forma independiente y sus resultados se han combinado de una nueva manera. En particular, 3 métodos EPP (principalmente basados en redes neuronales artificiales), comúnmente conocidos como Análisis de Componentes Principales (PCA: Principal Component Analysis), Aprendizaje Hebbiano de Máxima Verosimilitud (MLHL: Maximum Likelihood Hebbian Learning) y MLHL Cooperativo (CMLHL: Cooperative Maximum Likelihood Hebbian Learning). Además, se incluye una extensión a este estudio para mejorar la capacidad de visualización de los HUEP.

Para generar los gráficos, cada vector x original (desde el espacio de entrada) se procesa de la siguiente manera:

1. Se obtiene una proyección 2D del vector mediante el método EPP aplicado.
2. Se calcula la salida del método de agrupación (es decir, el número de agrupación asignado).
3. Las dos salidas anteriores se combinan en un vector 3D que se ubica en el espacio de salida.
4. Opcionalmente, se agrega más información (datos de sostenibilidad en el presente estudio) a la visualización mediante el uso de la metáfora del glifo.

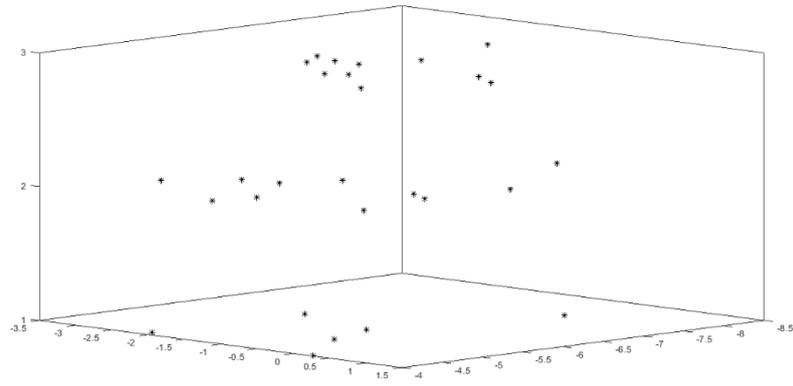
Originalmente, los HUEP se concibieron como una nueva forma de mostrar datos de manera intuitiva mediante la aplicación de un método de agrupación particional (k -medias) o uno jerárquico (aglomerativo) junto con un método EPP. Como evolución de esta propuesta inicial, este estudio valida la incorporación de métodos de visualización complementarios y conocidos, como Kernel-PCA (KPCA) (Schölkopf et al., 1998) y Sammon Mapping (SM) (Sammon, 1969).

KPCA es una extensión no lineal de PCA convencional que toma la mayoría de las funciones del Kernel para obtener proyecciones de datos más interesantes al extraer componentes principales no lineales y mantener el costo de computación a un nivel razonable. Por otro lado, el método Sammon Mapping se propuso como un caso especial de la familia de escalamiento multidimensional métrico basado en la distancia, siendo en sí mismo una de las primeras propuestas de aprendizaje múltiple. Además, SM puede considerarse como el primer método de aprendizaje múltiple no lineal propuesto. Estos métodos EPP no lineales se proponen por primera vez en el marco de HUEP como uno de los principales métodos EPP no lineales. Se analizan en este estudio y se validan con los datos descritos anteriormente.

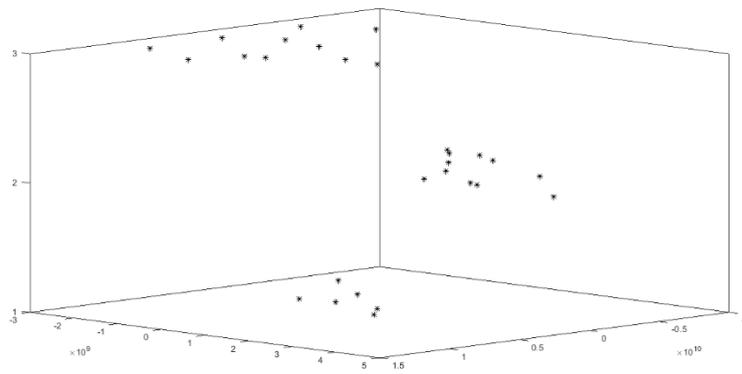
3.1.3.- Resultados.

En esta sección se muestran los gráficos HUEP obtenidos. En primer lugar, la figura. 6 muestra el gráfico HUEP obtenido al combinar el agrupamiento aglomerativo con diferentes técnicas de EPP. Como resultado, se puede comparar la influencia de las diferentes técnicas de EPP en los resultados obtenidos. Debido a esto, no se muestra información adicional a través de la metáfora del glifo en esta figura, para permitir la comparación única de las proyecciones. En aras de la brevedad, se muestran las visualizaciones gráficas más interesantes, pero no se incluyen en la figura. 6 las obtenidas por algunos de los métodos EPP

a)



b)



c)

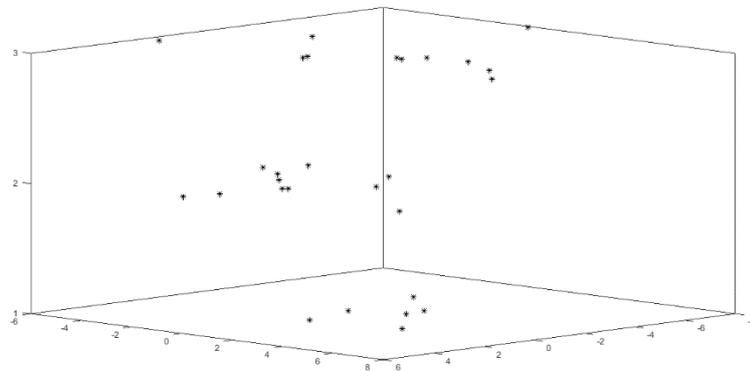


Figura 6.- HUEP obtenidos mediante la aplicación de agrupamiento aglomerativo ($k = 3$, distancia = euclidiana, enlace = promedio) al conjunto de datos analizado, variando la técnica EPP: a) CMLHL, b) KPCA, c) SM.

La visualización obtenida por b) (KPCA: Kernel Principal Component Analysis) puede considerarse la más reveladora. Permite observar con mayor claridad la estructura de los datos, ya que representa los datos de forma más compacta y, por lo tanto, permite analizar las tendencias con mayor claridad. Dado que no

es posible incluir todos los resultados obtenidos en este estudio, en el resto de esta sección solo se muestran los resultados obtenidos con KPCA.

3.1.3.1.- Resultados incluyendo la metáfora del glifo.

Después de haber seleccionado KPCA como el modelo EPP que ofrece mejores proyecciones para los datos analizados, en esta sección se presentan los resultados utilizando la metáfora del glifo. Para ello, en los siguientes gráficos se incorpora información adicional sobre la variable Edad de la flota. Los símbolos para cada dato se diferencian según el cuartil al que pertenecen, de acuerdo con el valor tomado para esa variable, y de acuerdo también con los símbolos que se muestran en la Tabla 10.

Q	Glyph
1	*
2	o
3	▷
4	+

Tabla 10.- Leyenda de los gráficos utilizando la metáfora del glifo según los valores del rasgo relacionado con la sustentabilidad (Edad promedio de la flota de vehículos).

De acuerdo con lo anterior, se muestra el gráfico HUEP previamente seleccionado (figura 6.b), aunque ahora con la característica de sostenibilidad incorporada. Adicionalmente, la figura está realizada con una línea discontinua para guiar al lector en la interpretación de los resultados.

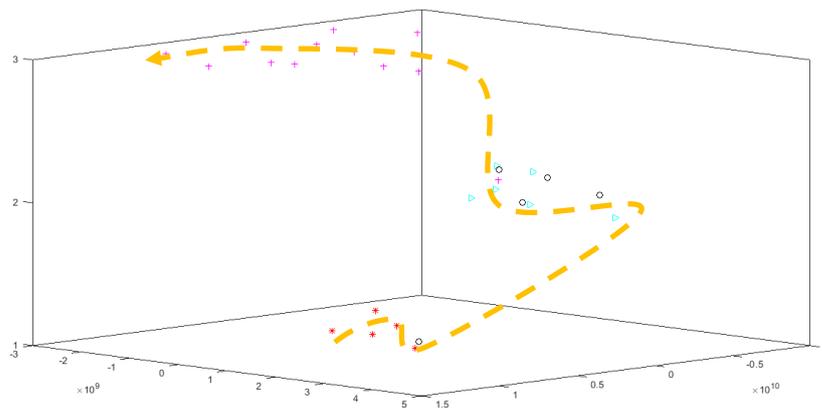


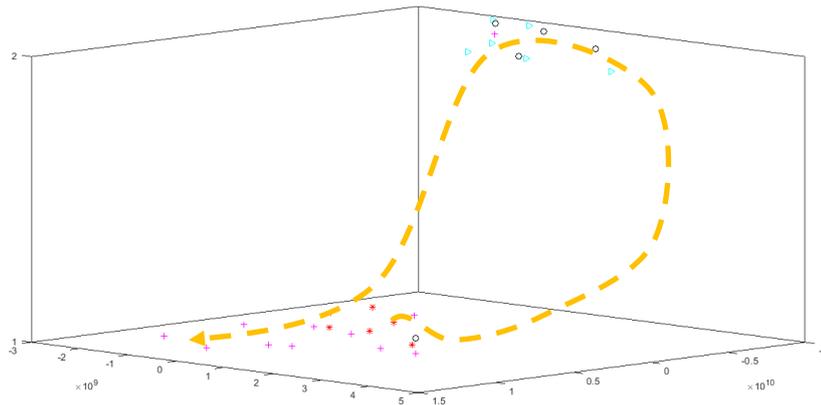
Figura 7.- HUEP obtenido al aplicar KPCA y agrupamiento aglomerativo (k=3, distancia = euclidiana, enlace = promedio) al conjunto de datos analizado, utilizando la función de sostenibilidad en la metáfora del glifo.

En la figura 7 se puede observar que se obtiene una clara diferenciación de los cuartiles de datos de la serie, pudiendo marcarse una clara línea de progresión sobre el resultado (línea discontinua amarilla). Desde un punto de vista práctico, el gráfico muestra que la visualización obtenida es útil a la hora de determinar las fases de progresión de la serie de edad; 1. la edad inicial, 2. la fase de

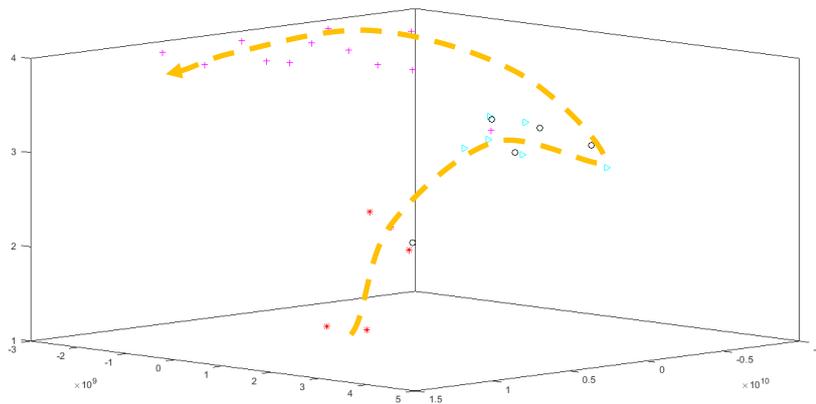
sobre amortización de los medios de transporte y 3. la fase de nueva edad media de la flota.

A continuación, se presentan en el gráfico diferentes visualizaciones para apreciar el impacto del parámetro k de la técnica de agrupamiento (figura 8), con los mismos modelos EPP y de agrupamiento, pero con un número variable de agrupamientos.

a)



b)



c)

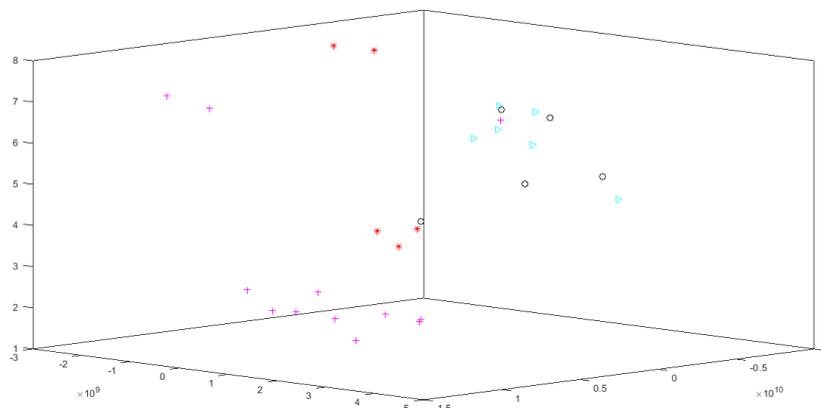
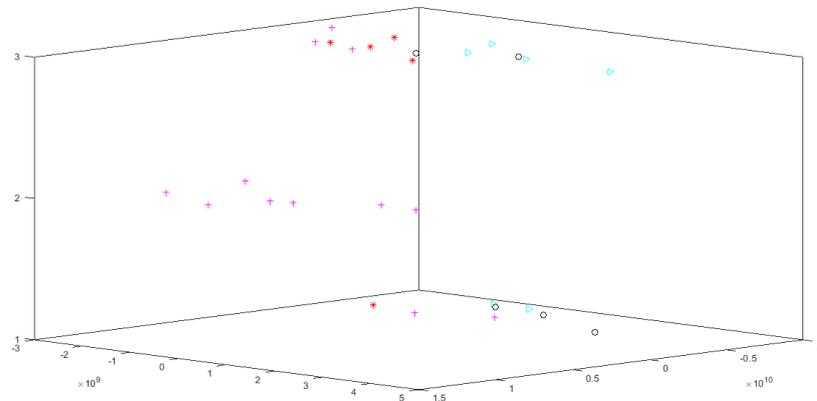


Figura 8.- HUEP obtenidos con KPCA y agrupación aglomerativa (distancia = euclidiana, enlace = promedio), que muestra la característica de sostenibilidad en la metáfora del glifo y un número variable de agrupaciones: a) $k=2$, b) $k=4$ y c) $k=8$.

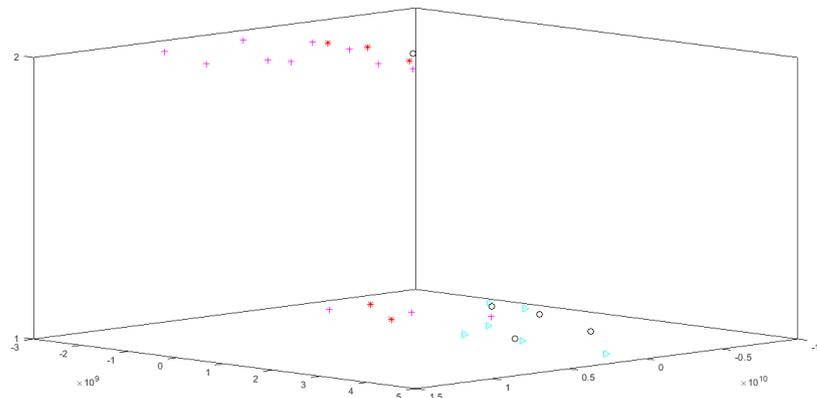
Tras analizar los resultados mostrados en la figura 8, podemos deducir que para $k=2$, la visualización se limita a diferenciar entre las fases de amortización sostenida y las de amortización creciente. Para $k=4$, podemos ver una progresión conforme aumenta la edad, pero en comparación con el resultado significativo para $k=3$ mostrado en la figura 7, se limita a diferenciar un nuevo grupo en los períodos iniciales del estudio. El resultado para $k=8$ no parece significativo. Por lo tanto, se puede concluir que 3 es un número apropiado de conglomerados para el conjunto de datos en estudio.

A continuación, se presentan las visualizaciones obtenidas por agrupamiento parcial, con el fin de observar la influencia de la técnica de agrupamiento utilizada, comparándola con los resultados anteriores (agrupamiento aglomerativo). Dado que 3 se ha revelado como un número apropiado de conglomerados, en la figura 9 se muestran diferentes HUEP obtenidos con k-means para un número de conglomerados en torno a 3.

a)



b)



c)

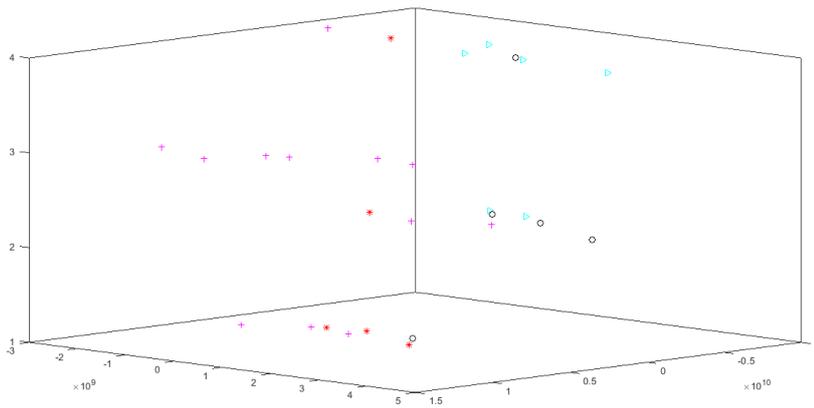
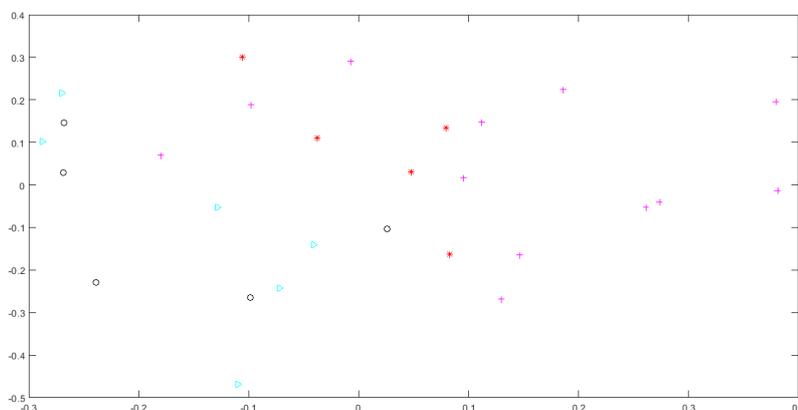


Figura 9.- HUEP obtenidos mediante la aplicación de agrupaciones de KPCA y k-means (distancia = seuclediana, enlace = promedio), que muestran la característica de sostenibilidad en la metáfora del glifo y un número variable de agrupaciones: a) k=2, b) k=3, y c) k=4.

Los resultados que se muestran en la Fig. 9 no proporcionan una visualización clara o convincente. No se puede inferir un agrupamiento significativo con respecto a la progresión de los datos en términos de antigüedad de la flota de transporte en ninguna de las configuraciones probadas. Por lo tanto, podemos concluir que la aplicación de k-medias no proporciona una mejor visualización de los datos en estudio en comparación con el método aglomerativo.

Finalmente, la visualización mejorada obtenida por los HUEP extendidos se compara con la obtenida por el método KPCA recientemente aplicado para identificar qué pueden proporcionar los HUEP a los usuarios. Los resultados obtenidos por KPCA (2D y 3D) se muestran en la figura 9, que comprende la metáfora del glifo para respaldar la comparación directa.

a)



b)

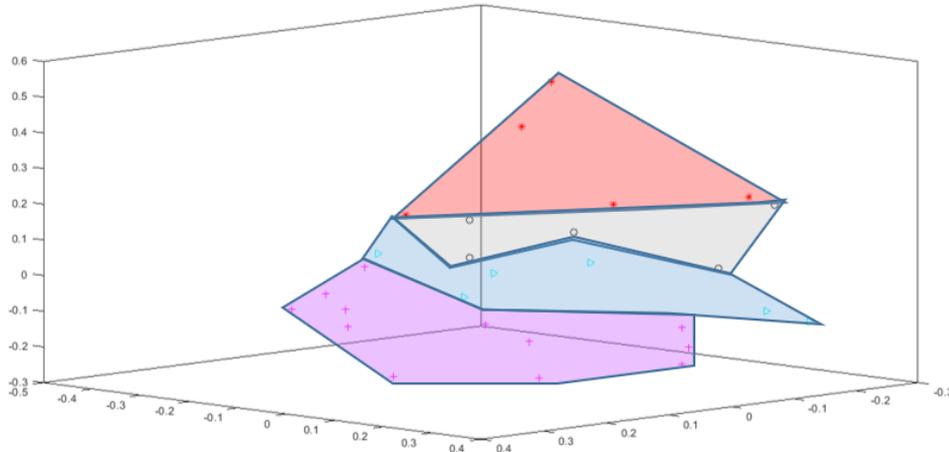


Figura 10.- Proyección obtenida al aplicar KPCA de forma aislada al conjunto de datos analizado, que muestra la característica de sostenibilidad en la metáfora del glifo. a) 2D y b) 3D.

En la figura 10, podemos ver que después de aplicar KPCA de forma aislada, la visualización ofrece menos información que la obtenida a través de los HUEP, ya que no permite una identificación clara de la estructura de los datos. Por otro lado, se obtiene cierta mejora en su visualización 3D frente a la visualización 2D. En la vista 3D, los grupos significativos de cuartiles de edad de la flota se agrupan en áreas relacionadas del espacio tridimensional, como se indica en la figura por las áreas coloreadas correspondientes a los cuartiles.

3.2 Modelos neuronales no lineales para predecir el precio del acero HRC en España.

El acero es una materia prima muy utilizada en la industria debido a sus ventajas frente a otras alternativas, tales como su coste, reciclaje rápido y respetuoso con el medio ambiente, facilidad de uso, alta resistencia y diferentes acabados y calidades. Pronosticar los precios del acero ha sido una tarea importante y desafiante que tradicionalmente se ha abordado con técnicas econométricas, estocásticas-gaussianas y de series temporales. Avanzando desde trabajos previos y sobre este nuevo desafío, en el presente trabajo se aplican por primera vez algunas Redes Neuronales Artificiales para pronosticar el precio del acero laminado en caliente en España. Más precisamente, algunas redes neuronales no lineales se aplican a varias series de tiempo de entrada diferentes. El objetivo de esta investigación es doble; por un lado, identificar cuál de los modelos neuronales supera a los otros a la hora de predecir los precios del acero y, por otro lado, validar diferentes series de datos para dicha predicción. Los principales resultados de esta investigación, tras validar los modelos neuronales sobre datos reales de los últimos 7 años, contribuyen en gran medida a este campo ya que se obtienen conclusiones novedosas y relevantes.

3.2.1.- Introducción y trabajo previo.

El acero es una aleación del mineral de hierro con otros elementos, principalmente carbono, pero también con otros minerales para modificar sus propiedades. Además de utilizar estos minerales, el acero también se puede producir a partir de chatarra de acero, lo que lo convierte en un material altamente reciclable que representa una verdadera economía circular en la actualidad. El descubrimiento del acero y su posterior mejora en la tecnología de producción fueron parte de la primera y la segunda revolución industrial, por lo que el acero ha sido esencial para la modernización del mundo y continúa siéndolo hoy.

La producción mundial de acero superó los 1.875 millones de toneladas en 2019, según la Asociación Internacional del Acero, y es el sector industrial con mayor consumo energético del mundo (Conejo et al., 2020).

Por ello, el acero es de gran importancia para la industria mundial, ya que se utiliza en diferentes formatos para un gran número de sectores industriales (construcción, electrodomésticos, vehículos, etc.), y es un material con una excelente relación entre sus propiedades físicas (elasticidad, fuerza, resistencia) y su costo. También es un material versátil que se usa de formas nuevas y variadas todos los días gracias al desarrollo de nuevas aleaciones y recubrimientos con propiedades mejoradas (Gutierrez & Vianna, 2020).

Inicialmente, el acero se obtenía en un horno, al que se le añadía mineral de hierro y una capa de coque para fundir estos minerales al fuego. Se utilizó soplado de aire para impulsar el fuego del horno para forzar el paso de los gases de combustión y mezclar las materias primas para obtener arrabio, que es una aleación que se funde a una temperatura más baja que el acero. Luego, el arrabio se tritura y muele, y se mezcla con carbón en un crisol, que se sella para evitar la oxidación y se enfría lentamente para obtener la pieza de acero, que luego se utiliza para forjar. Posteriormente se realizaron diversos avances tecnológicos en el proceso (proceso Bessemer, proceso Gilchrist-Thomas, Siemens-Martin), para producir grandes cantidades a bajo costo. Hoy en día, los mini hornos eléctricos también se utilizan para producir acero a partir de chatarra (chatarra ferrosa y finos de mineral de hierro Fe 62%). Existen dos rutas diferentes en el proceso de fabricación del acero, la primaria o BOS (Basic Oxygen Furnace), que utiliza principalmente mineral de hierro y carbón, y la ruta secundaria o EAF (Electric Arc Furnace), que utiliza chatarra de acero (European Commission, 2021).

CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS

Se puede distinguir entre diferentes tipos de acero en función de su composición química (acero al carbono, inoxidable, eléctrico y especial), su forma física (bobinas, aceros planos y largos), su etapa de procesamiento (laminado en caliente, acero laminado, galvanizado), así como su acabado, lo que hace que el acero no sea un producto homogéneo (European Commission, 2021).

El precio de los diferentes tipos de acero depende de muchos factores, como las materias primas (mineral de hierro, carbón), el proceso de producción, la capacidad de producción, las inversiones, los costos de energía, la demanda, la oferta y los factores políticos (Malanichev & Vorobyev, 2011). Además, parece que puede haber cierta correlación entre el precio del acero y el precio de las acciones de las empresas siderúrgicas (Manu, 2019). Por otro lado, hay estudios que sugieren una correlación unidireccional entre algunas variables, como el precio del petróleo que influye en el precio del acero, pero no al revés (Chou & Yang, 2012). Y los precios de los metales han sido considerados como indicadores relevantes para los mercados debido a su capacidad para ajustarse a las condiciones macroeconómicas y especulativas (Omura et al., 2016).

Debido al uso generalizado del acero en muchos sectores, el precio del acero es un factor importante que influye en la competitividad de muchas empresas. Por lo tanto, predecir las tendencias de los precios puede ayudar a las empresas a tomar decisiones de compra y venta de acero para aumentar su rentabilidad.

En general, los precios de los productos de acero todos siguen el mismo patrón, pero la bobina de chapa laminada en caliente (HRC: Hot Rolled Coil) suele considerarse como el punto de referencia porque representa el mayor volumen de exportaciones de acero (Rodríguez Liboreiro & Rodríguez Liboreiro, 2019).

Si bien el mercado del acero es global, el precio del acero puede diferir de un país a otro, debido a diferentes razones, como aranceles, demanda del mercado e impuestos, entre otros. No obstante, la evolución de los precios del acero es paralela en cada país (Gutierrez & Vianna, 2020b), (Pauliuk et al., 2013), (Panasiyk et al., 2016). Además, el precio de las materias primas, incluido el acero, parece experimentar fluctuaciones correlacionadas relativamente diferentes para cada país, lo que puede deberse a términos comerciales internacionales similares (Adewuyi et al., 2020).

Para la predicción del precio del acero se han utilizado múltiples modelos, como el modelo Auto-Regressive Integrated Moving Average (ARIMA) (Zola P & Carpita M, 2016), (Adli, 2020), así como el Long Short-Term Memory LSTM (Cetin

et al., 2019). Por otro lado, las redes neuronales no lineales aplicadas lo han sido previamente a algunos otros problemas como la logística (Alonso de Armiño et al., 2020) o la agricultura de precisión (Yartu et al., 2022) entre otros. A diferencia de ese trabajo previo, en el presente trabajo se aplican por primera vez algunas redes neuronales no lineales a este problema. Además, las series de datos novedosos se utilizan como datos de entrada para tales modelos de predicción.

Este estudio se centra en la predicción del precio de la BLC (Bobina Laminada en Caliente) para España e incluye una comparación de varios modelos de redes neuronales (descritos en la sección 3.2.2), a saber: Entrada-Salida no lineal (NIO), Autorregresivo no lineal (NAR) y autorregresivo no lineal con entrada exógena (NARX). Estos modelos de aprendizaje supervisado se aplican para la previsión de series temporales de precios del acero. Además, también se evalúa el impacto del ajuste de parámetros para estos modelos. También, se estudian y comparan algunas series de datos diferentes (descritas en la subsección 3.2.3.1) para identificar cuál contribuye a la predicción más precisa.

3.2.2.- Materiales y Métodos.

En la presente sección se introducen tanto los datos utilizados (Subsección 3.2.2.1) como los modelos aplicados (Subsección 3.2.2.2).

3.2.2.1.- Conjunto de datos.

Como se mencionó anteriormente, la presente investigación se enfoca en predecir el precio de la bobina laminada en caliente (HRC) para el mercado español. La fuente de estos datos es CRU-Commodities Research Unit, que es un proveedor independiente líder de información sobre el mercado del acero. Para predecir el precio se estudian tres grupos de datos:

- Grupo 1: Cotizaciones bursátiles de las tres principales empresas siderúrgicas mundiales, que son Posco, ArcelorMittal y NipponSteel, cuya fuente de datos es NYSE - Nasdaq Real Time Price en USD.
- Grupo 2: Situación económica en España, incluyendo datos relativos a tres temas: a) Precio (Índice 2010 =100): Índice de Precios de Consumo Armonizado, Índice de Precios al Consumidor Todos los artículos, Índice de Precios al Productor Índice de Todos los Productos Básicos; b) Actividad Económica- Producción Industrial (Índice 2010 =100): Índice Total, Índice Manufacturero, Índice Minero; c) Mercancías de Comercio Exterior (Dólares Estadounidenses): Valor de las Importaciones (Costo, Seguro, Flete CIF), Valor de las Exportaciones (Franco a Bordo FOB). (Fuente de datos del Fondo Monetario Internacional.

- Grupo 3: Precio de otros *commodities*: precio del Aluminio; crudo Brent; Cobalto; Cobre; crudo de Dubái; Oro; Plomo; Molibdeno; Plata; Zinc (Fuente de datos del Fondo Monetario Internacional).

La serie de datos para todas las características corresponde a una periodicidad mensual durante el período 2013 - 2019. Para la fuente de datos se utiliza una periodicidad mensual ya que esta es la unidad de tiempo generalmente establecida en la indexación del precio del acero y es suficiente para reflejar la volatilidad del precio del acero (Guo et al., 2019). Como muestra de los datos objetivo a pronosticar, se identificó el precio máximo de la BLC (Bobina Laminada en Caliente) en la serie bajo análisis.

3.2.2.2.- Modelos neuronales no lineales.

Para predecir el precio del acero HRC, se han utilizado los datos descritos en la subsección anterior como datos de entrada para varios modelos neuronales que utilizan series temporales no lineales (Leontaritis & Billings, 2007), los siguientes: Entrada-Salida no lineal (NIO), autorregresivo no lineal (NAR) y autorregresivo no lineal con entrada exógena (NARX). Los tres se pueden definir como redes neuronales (NN: Neural Networks) que incluyen un retraso de derivación asociado al peso de entrada. Los modelos tienen una respuesta dinámica finita a los datos de entrada de series temporales gracias a dicho retraso.

Aunque estos tres modelos tienen características similares, también existen algunas diferencias, siendo principalmente los datos que se proporcionan al modelo. En primer lugar, están los modelos NIO que intentan predecir una serie de datos solo teniendo en cuenta los valores anteriores de otras series de datos, pero no la de destino. Es decir, la serie de datos a predecir no se utiliza como entrada. Por el contrario, el modelo NAR, como su nombre lo indica, simplemente modela una relación entre los valores anteriores de la serie de datos cuyos valores futuros se intenta predecir. Por lo tanto, este modelo solo utiliza la serie que se va a predecir. Estos dos enfoques se combinan en el modelo NARX, ya que predice los valores de una serie de datos utilizando los valores anteriores (entrada endógena) y considera también otras series de datos (entradas exógenas).

De una manera más formal, estos modelos se pueden definir matemáticamente de la siguiente manera, considerando $y(t)$ la característica que se va a predecir en el instante de tiempo t y $f()$ la función que se va a aproximar mediante el modelo de predicción de series temporales. El modelo NAR consiste principalmente en un modelo que predice una serie (y) utilizando únicamente n valores previos de la misma, se puede expresar como:

$$Y(t)=f(y(t-1),\dots,y(t-ny))$$

Desde una perspectiva opuesta, el modelo NIO intenta predecir una serie (y) utilizando n valores previos de otra diferente (x):

$$y(t)=f(x(t-1),\dots,x(t-nx))$$

Finalmente, el NARX es un modelo que combina valores anteriores tanto de la serie a predecir (y) como de otra diferente (x) que se considera exógena:

$$y(t)=f(y(t-1),\dots,y(t-ny),x(t-1),\dots,x(t-nx))$$

3.2.3.- Experimentos y Resultados.

Esta sección presenta información sobre los experimentos realizados, así como los resultados obtenidos. Al definir los experimentos y en base a resultados anteriores, los parámetros de estos modelos se han ajustado con combinaciones de los siguientes valores:

- Algoritmo de entrenamiento: {1 - Levenberg-Marquardt, 2 - Descenso de gradiente por lotes, 3 - Descenso de gradiente con impulso, 4 - Propagación inversa de tasa de aprendizaje adaptable, 5 - Descenso de gradiente con impulso y tasa de aprendizaje adaptativo, 6 - Gradiente conjugado escalado, 7 - Broyden–Fletcher–Goldfarb–Shanno Backpropagation}

- Número de neuronas ocultas: {1, 5, 10, 15, 20}
- Número de retrasos de entrada: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
- Número de retardos de salida: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}

Para obtener resultados más significativos estadísticamente, se han realizado 10 ejecuciones para la misma configuración de los parámetros. Debido a la combinación de todos estos valores, se han realizado 350 ejecuciones para los modelos NIO y NAR por cada una de las series de datos. En el caso de NARX, al combinar un número variable de retardos tanto de entrada como de salida, se han realizado 3.500 ejecuciones por cada una de las series de datos.

CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS

En esta sección se proporciona el error cuadrático medio (MSE) promedio para las 10 ejecuciones. En cada una de las tablas, se destaca en negrita el menor valor de error por columna.

Inicialmente, los resultados promedio obtenidos por los modelos de series temporales se presentan por el número de retrasos de entrada en la Tabla 11.

Nº retrasos de entrada	NAR	NIO	NARX
1	44.777,62	34.768,56	30.731,95
2	36.811,43	33.635,41	1.259.633,09
3	35.092,46	36.092,17	31.739,07
4	37.317,69	30.805,89	31.725,67
5	35.941,02	34.324,44	34.231,69
6	32.765,64	34.953,35	34.211,34
7	29.586,15	35.951,07	34.099,02
8	35.067,41	42.498,11	36.024,10
9	35.714,85	41.546,65	37.190,99
10	27.439,20	39.195,92	39.736,90

Tabla 11.- MSE de los resultados obtenidos por los modelos neuronales NAR, NIO y NARX. Los resultados promediados para las 3 series de datos se muestran por el número de retrasos de entrada. N retrasos de entrada.

De los resultados de la Tabla 11, se puede decir que se han obtenido resultados opuestos para algunos de los modelos: en el caso de NAR, el menor error se obtiene con el mayor número de retardos de entrada (10) mientras que en el caso de NARX se ha obtenido con el más bajo (1). En el caso de NIO, un valor intermedio (4) para retardos de entrada es el que tiene el error más bajo.

Los resultados obtenidos por los modelos neuronales también se presentan por el número de neuronas ocultas en la Tabla 12.

Al considerar el número de neuronas en la capa oculta de los modelos neuronales, se encuentran resultados opuestos: cuando solo se utiliza la serie de datos de precios (modelo NAR), el error más bajo se obtiene con la capa oculta más grande (20 neuronas). Sin embargo, cuando se utilizan las series de datos

endógenos (modelos NIO y NARX), el error más bajo se obtiene con la capa oculta más pequeña (1 neurona).

Neurons	NAR	NIO	NARX
1	35.092,46	10.662,66	10.453,78
5	36.205,07	21.073,80	20.342,18
10	34.353,33	33.933,62	646.951,91
15	32.326,78	52.494,22	47.046,96
20	31.577,02	63.721,49	59.867,09

Tabla 12.- MSE (Error Cuadrático Medio) de los resultados obtenidos por los modelos neuronales NAR, NIO y NARX. Los resultados promedio para las 3 series de datos se muestran por el número de neuronas ocultas.

En la Tabla 13, se presentan los resultados obtenidos por el algoritmo de entrenamiento.

Training algorithm	NAR	NIO	NARX
1	2.029,26	12.615,28	11.009,04
2	115.329,59	94.843,10	947.446,69
3	117.558,44	95.041,44	89.026,89
4	3.321,75	15.365,87	12.867,00
5	3.845,27	13.214,13	16.612,29
6	1.448,84	11.720,95	10.919,50
7	1.826,26	11.839,34	10.645,26

Tabla 13.- MSE de los resultados obtenidos por los modelos neuronales NAR, NIO y NARX. Los resultados promediados para las 3 series de datos se muestran según el algoritmo de entrenamiento.

Como se puede observar en la Tabla 13, existen fuertes diferencias en cuanto a los errores obtenidos al entrenar los modelos con diferentes algoritmos, identificándose claramente los mejores y peores resultados. Los dos algoritmos asociados a los mejores resultados (menor error) son "Gradiente conjugado escalado" (6) y "Broyden-Fletcher-Goldfarb-Shanno Back-propagation" (7) para los tres modelos. Por otro lado, los peores resultados (tasas de error más altas) están asociados con el "Batch Gradient Descent" (2) y el Gradient Descent with Momentum (3) para los tres modelos.

CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS

Como uno de los objetivos de la presente investigación es identificar la mejor serie temporal para predecir el precio, en la Tabla 14 también se presentan los resultados promediados por serie de datos utilizada (solo los modelos NIO y NARX, ya que el modelo NAR solo se proporciona con el precio en sí).

Al utilizar únicamente la serie de datos exógenos para realizar la predicción (NIO), la serie de datos 3 (Precio de otras materias primas) es la que se asocia a mejores resultados y la serie de datos 1 (Precios bursátiles) la segunda mejor. Al combinar la serie de datos endógenos con el propio precio (NARX), ocurre lo contrario: la serie de datos 1 es la que se asocia a mejores resultados y la serie de datos 3 es la segunda mejor. Para ambos modelos, la serie de datos 2 (Situación económica de España) es la que se asocia a peores resultados. En cuanto al modelo neuronal que se aplica, se puede decir que cada una de las series de datos, el mejor resultado lo obtiene siempre el modelo NARX, siendo NIO en todos los casos el segundo mejor.

Data series	NIO	NARX
1	29.775,95	26.434,71
2	52.250,05	417.830,09
3	27.105,46	26.532,35

Tabla 14.- MSE of the results obtained by the NIO, and NARX neural models. Averaged results are shown per the data series.

Finalmente, para identificar claramente los mejores resultados individuales, la Tabla 15 presenta las tasas de error más bajas para cada combinación de parámetros. En este caso, el error se ha promediado solo para las 10 ejecuciones ejecutadas con exactamente los mismos valores para los diferentes parámetros.

Input data series	NAR	NIO	NARX
Price itself	566,63	-	
1	-	789,75	460,41
2		1.150,34	796,86
3		664,81	582,63

Tabla 15.- MSE de los resultados obtenidos por los modelos neuronales NAR, NIO y NARX. Se muestran los mejores resultados individuales (promediados solo para las 10 ejecuciones).

Para estos resultados individuales (no se promedia el error para todas las combinaciones de parámetros), se pueden observar resultados similares a los de la Tabla 14: El mejor resultado se obtiene con la serie de datos 3 en el caso del modelo NIO mientras que es la serie de datos 3 en el caso de los modelos NARX. En ambos casos, los peores resultados se obtienen con la serie de datos 2. En cuanto a los modelos neuronales, para todas las series de datos endógenos, en todos los casos NARX supera claramente a NIO.

Con todo, el mejor resultado individual (10 ejecuciones) se obtiene al combinar el modelo NARX y la serie de datos 1. El segundo mejor lo obtiene el modelo NAR, por lo tanto, utilizando solo la serie de datos de precios.

3.3 Análisis de la sostenibilidad económica en la cadena de suministro aplicando el predictor Altman Z-score.

A modo de resumen, este trabajo pretende también rellenar un vacío desde la perspectiva financiera de la medición del desempeño de la cadena de suministro, relacionado con la falta de un indicador de probabilidad de quiebra, y propone un predictor que es el octavo modelo de la regresión logística Altman Z-Score. Además, se establece para las cadenas de suministro de las empresas un ranking de probabilidad de quiebra, según el sector al que pertenecen. Además, los valores se fijan para establecer tres categorías de empresas según el predictor. Se analiza y estudia la probabilidad de quiebra para la cadena de suministro de diferentes industrias. Se revela que la industria de la construcción tiene la mayor probabilidad de quiebra.

3.3.1.- Introducción.

La definición de “cadena de suministro” ha sido ampliamente establecida en el estado del arte como una red de materiales, información y servicios vinculados en tres estados: compras, producción y distribución. La cadena de suministro más compleja es la denominada cadena de suministro final, que incluye todo el sector de empresas de diferentes industrias que atañen a todos los flujos de productos, servicios, finanzas e información desde los proveedores hasta los clientes finales (Mentzer et al., 2001) (Min et al., 2019). El sector de la cadena de suministro denominado cadena de suministro final representa la alineación de múltiples empresas que llevan productos o servicios a los consumidores finales en el mercado. Las industrias a las que pertenecen las empresas en la

última cadena de suministro son: fabricantes, proveedores, transportistas, almacenes, mayoristas, minoristas y otros intermediarios (Mentzer et al., 2001).

En los últimos años, se han publicado varios sistemas de medición del desempeño de la cadena de suministro (SCPM: Supply Chain Performance Measurement), para cuantificar la eficiencia y eficacia de los procesos de la cadena de suministro, sus relaciones y las múltiples funciones de las organizaciones y empresas para permitir orquestar la cadena de suministro (Arzu Akyuz & Erman Erkan, 2009), (Khan et al., 2021). Las empresas pueden compararse con otras mediante el uso de modelos de madurez de la cadena de suministro. Tanto en los modelos de madurez como en SCPM, se incluyen aspectos relacionados con la regulación ambiental para entender que los recursos son limitados, por lo que las empresas adoptan prácticas sostenibles de recursos naturales (Santos et al., 2020). Además de esta perspectiva medioambiental, estudios recientes incluyen métricas de empresa desde otras perspectivas que pueden influir en su cadena de suministro, tales como: proceso interno, aprendizaje y crecimiento, cliente, social y financiero (Santos et al., 2021). Desde una perspectiva financiera, tampoco los modelos de madurez ni los modelos de medición de mejora de la cadena de suministro incluyen un indicador de la sostenibilidad económica de la empresa, como la probabilidad de quiebra de la empresa, para medir la continuidad de la empresa y su influencia en la cadena de suministro.

Es importante perseguir la sostenibilidad económica, ya que las dificultades financieras y los fracasos de los proveedores críticos desencadenan anuncios de interrupciones en la cadena de suministro que penalizan severamente operaciones comerciales (Jahani et al., 2019), (Yang et al., 2015). Por lo tanto, las empresas en una cadena de suministro deben mantener situaciones financieras saludables para lograr la integración exitosa de sus actividades y la adecuada coordinación de todos sus procesos clave. De lo contrario, la incertidumbre y los riesgos de quiebra de una empresa en la cadena de suministro puede amenazar las operaciones y actividades del resto de empresas en la cadena de suministro (Craighead et al., 2009).

Predecir el riesgo de quiebra es importante para la detección temprana de signos de deterioro de las condiciones financieras y puede permitir que se tomen medidas correctivas en el funcionamiento de la cadena de suministro (Zej & Id, 2018), (Karan et al., 2013). En consecuencia, las empresas con posiciones financieras más fuertes pueden asumir la liquidez necesaria para mantener su actividad en la cadena de suministro, al aceptar condiciones de pago temporal más altas o pagar por adelantado para aumentar el crédito comercial a empresas con una posición financiera más débil y, así complementar la salud financiera de estas empresas (Karan et al., 2013), (Zhang et al., 2019).

En este contexto, las preguntas de investigación (Research Questions, RQ) consideradas en esta investigación son dos.

RQ1: ¿Cuáles son los sectores de las cadenas de suministro de las empresas españolas con mayor probabilidad de quiebra?

RQ2: ¿Qué valores permiten medir la probabilidad de quiebra, delimitada en tres zonas para la clasificación financiera de una empresa?

Además de la introducción, este trabajo se divide en cuatro secciones. Inicialmente, se hace una revisión del concepto teórico de la cadena de suministro y la probabilidad de bancarrota, así como el predictor Z de bancarrota de Altman. A continuación, se presenta la metodología del estudio. Más adelante, se realiza el estudio empírico para responder a las preguntas de investigación. Finalmente, se presenta la discusión y conclusiones del estudio, destacando los sectores de la cadena de suministro con mayor probabilidad de riesgo

3.3.2.- Marco teórico.

El estudio del marco teórico de esta investigación, como se acaba de indicar, se basa en un análisis breve de la cadena de suministro y de la revisión conceptual del predictor de quiebra del modelo Z-score de Altman, que desarrollamos a continuación.

3.3.2.1.- La cadena de suministro.

La gestión de la cadena de suministro facilita el flujo de bienes, información y dinero. De estos tres elementos, el menos estudiado en el estado del arte es el flujo de dinero en la cadena de suministro, que se denomina financiación de la cadena de suministro (Supply Chain Finance SCF) (Zhang et al., 2019).

Debido a la interdependencia cada vez mayor de las empresas de la cadena de suministro final, la quiebra de una empresa puede provocar que otras compañías de la cadena caigan en dificultades (Xu et al., 2010).

La provisión de liquidez a la cadena de suministro ayuda a asegurar su estabilidad financiera. La reducción del ciclo de conversión de efectivo aumenta la liquidez de las empresas en la cadena de suministro, y por lo tanto reduce la probabilidad de quiebra de estas empresas (Zhang et al., 2019). El ciclo de conversión de efectivo se puede reducir enfocándose en la optimización del capital circulante (working capital) en términos de cuentas a pagar, cuentas a cobrar, inventarios y, a veces, incluso financiación de activos fijos (Gelsomino et al., 2016). La financiación mediante créditos comerciales es otra opción para aumentar la liquidez, aunque en ocasiones aumenta el nivel de riesgo y quiebra, ya que puede estimular el aumento de pedidos (Yan & He, 2020), (R. Huang & Ren, 2016).

La cadena de suministro final, que incluye cadenas de suministro de servicio únicamente (SOSC: Service Only Supply Chain) y las cadenas de suministro de servicios de productos (PSSC: Product Service Supply Chain) (Y. Wang et al., 2015), es diferente para cada industria, por lo que la causa de quiebra puede variar de unas industrias a otras (W. Liu et al., 2017).

A nivel industrial, la demanda del mercado, el crecimiento de las ventas y el crecimiento sostenible de cada empresa en la cadena de suministro define el comportamiento estratégico de la empresa en varias categorías (expansivo, dominante, en retroceso, en reestructuración, en declive y desenfocado) que se puede utilizar para predecir la rentabilidad de las empresas (Alcalde-Delgado et al., 2020).

La visibilidad de una cadena de suministro y la integración de proveedores para compartir información y así poder detectar la quiebra de un proveedor de la cadena, permite la gestión de riesgos interrumpiendo intencionadamente la cadena de suministro (DuHadway et al., 2019).

Por lo tanto, la medición de la probabilidad de quiebra es importante para la sostenibilidad económica de la cadena de suministro de cada sector. Dado que no existe un indicador de esa naturaleza en sistemas SCPM (Supply Chain Performance Measurements), en esta investigación se aborda una propuesta para introducir un predictor de análisis de quiebra para cada sector industrial.

3.3.2.2.- Predictor de quiebra: el modelo Z-score de Altman.

En la década de los 1930, se llevaron a cabo trabajos pioneros de investigación científica sobre la predicción de fechas de quiebra empresarial, cuando (Fitzpatrick, 1932) y (Smith, 1935) aplicaron análisis univariantes muy básicos basados en el estudio de la evolución de las ratios financieras (Elwell, 1936).

Las contribuciones de (Beaver, 1966) incorporaron en su investigación un análisis discriminante univariante con una evaluación separada de la capacidad predictiva de cada uno de las ratios consideradas, lo que supuso un salto cualitativo, aunque pronto fue abandonado por técnicas multivariantes (Beaver, 1966).

(Altman, 1968) fue el pionero en aplicar técnicas de análisis discriminante multivariante (MDA: Multivariate Discriminant Analysis) al estudio de la predicción de la quiebra empresarial, sustituyendo a la contribución previa de Beaver (Altman, 1968). Con esta investigación y otras posteriores, se alcanzaron buenos resultados con errores menores de clasificación (Altman, 2018).

Desde entonces, esta línea de investigación se ha enriquecido con numerosas aportaciones de diferentes autores con nuevos enfoques y la incorporación de

otras técnicas, como la inteligencia artificial, redes neuronales, mapas autoorganizados, escalas multidimensionales y la técnica Logit (que permite estimar la probabilidad de fallo de un conjunto de restricciones o atributos) (Pozuelo Campillo et al., 2013), (Mai et al., 2019).

A pesar de estas otras contribuciones, la herramienta de Altman es la más ampliamente aceptada y utilizada para anticipar quiebras financieras y riesgo de bancarrota en varios sectores industriales o empresas dentro de la misma cadena de suministro (Ellinger et al., 2011), (Ashraf et al., 2019).

3.3.2.2.1- Modelo de puntuación Z de Altman.

El modelo inicial se obtuvo en 1968 con una muestra de 66 empresas, de las cuales 33 se declararon en concurso de acreedores y otras 33 no. Todas las empresas pertenecían a los Estados Unidos, sector manufacturero, eran de tamaño medio y cotizaban en bolsa desde hacía 20 años (1946-1965). Los 66 estados financieros se obtuvieron del manual industrial de Moody, y de ellos Altman extrajo 22 índices financieros relacionados con: rentabilidad, endeudamiento, actividad, liquidez y solvencia. Agrupó estos indicadores en cinco variables, obtenidas a partir de siete datos de contabilidad y resultados financieros (Altman, 1968), como se puede observar en la Tabla 16.

Como resultado de aplicar la técnica de análisis de discriminación múltiple a la muestra de empresas, se obtuvo la siguiente ecuación:

$$Z = 1.2 X_1 + 1.4 X_2 + 3.3 X_3 + 0.6 X_4 + 1.0 X_5$$

Las empresas de la muestra de Edward Altman que obtuvieron un valor de Z por debajo de 1,81 terminaron en bancarrota. Mientras que las empresas que obtuvieron una puntuación superior a 2,99 se encontraban en una zona financiera saludable. Como consecuencia, aquellas empresas con puntuaciones entre 1,81 y 2,99 se consideraron en zona de precaución.

Este modelo fue criticado por ser aplicable únicamente a las empresas manufactureras que cotizan en el mercado de capitales y no a las empresas comerciales y de servicios, cotizadas o no. Por este motivo, Altman accedió a revisar el modelo, si bien es cierto que inicialmente se resistió, argumentando que no era necesario. Finalmente, publicó dos nuevas versiones del modelo, el Z' y la Z'' (Altman & Hotchkiss, 1993), (Siddiqui, 2012).

A pesar de esto, se pueden encontrar artículos académicos recientes que aún utilizan este modelo en lugar de la última actualización (Panigrahi, 2019), (Nandi et al., 2019).

CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS

3.3.2.2- Modelo de puntuación Altman Z'.

En 1983 se hizo un ajuste al modelo original para que pudiera ser aplicado a empresas manufactureras no cotizadas. En particular, en el cálculo de la ratio X4, se reemplazó el numerador con el valor en libros del capital en lugar del valor de mercado y dividido por el valor en libros de los pasivos totales (Altman et al., 2014). En consecuencia, se obtuvo la nueva ecuación:

$$Z' = 0.717 X_1 + 0.847 X_2 + 3.107 X_3 + 0.42 X_4 + 0.998 X_5$$

Variable	Name	Formula	Explanation
X ₁	Liquidity ratio	WCTA = working capital/ total assets	The part of the current assets that is financed using permanent resources is called the working capital. The relationship between the working capital and total assets is related to the capacity of a company to comply with its short-term financial obligations, that is to say, liquidity. Altman distinguished X ₁ as one of the most important ratios, since it is proportional to the company's liquidity and therefore to its financial health.
X ₂	Accumulated return ratio	RETA = accumulated reserves/total assets	The age of the company is quite relevant to predicting bankruptcy and insolvency, since a younger company is more likely not to survive compared to a company with an established operation or economic activity. Normally, older operating companies have a higher level of accumulated reserves. When the business accumulates losses, reserves may become negative. However, when it accumulates profits, the reserves will be positive, evidencing a capacity for growth and reinvestment to self-finance its projects.
X ₃	Profitability ratio	EBITTA = profitability before tax and interest (EBIT)/total assets	If a company fails to consolidate its operation to generate sufficient resources, it will subsequently disappear from one year to the next because it is unable to meet its payment obligations. The higher the value of this variable, the lower the probability of bankruptcy.
X ₄	Financial structure ratio	BVETD = market value/ total liabilities	For quoted companies, the market value can be calculated using the stock market price. Its division into total liabilities allows it to be related to the book value. The result is a variable related to the financial structure. The higher the value of this variable, the lower the probability of bankruptcy of the company, since the market is valuing the company above its book value.
X ₅	Asset turnover rate	SALTA = net sales/ total assets	This indicator is associated with the speed of movement of assets, i.e., the administrative and commercial capacity of the company in relation to competition from other companies in the sector in which it operates. The result is interpreted as the number of times sales contain the asset.

Tabla 16.- Key figures of the initial Z-Altman model. Altman, E.I. *Financial ratios, Discriminant analysis and the prediction of Corporate Bankruptcy. J. Financ.* 1968, 23 589-609

En este modelo, los resultados por debajo de 1,23 se consideran propensos a fallar, mientras que los que están por encima de 2,99 se consideran saludables. Cuando se encuentran entre 2.99 y 1.23 la empresa está en la zona de precaución.

Altman no validó este modelo con una segunda muestra debido a la falta de empresas privadas en la base de datos. Consideró que la ratio X5 (ventas netas/activo total) estaba influenciada por el sector de la empresa o industria, por lo que decidió hacer un nuevo modelo válido para todas las industrias o sectores, a lo que llamó Z''-score (Altman E & Hotchkiss E, 2011). Sin embargo, en

estudios académicos recientes o artículos, es habitual observar que se sigue utilizando el modelo Z' en lugar de la última actualización (Jahani et al., 2019), (González-Martín et al., 2018).

3.3.2.2.3- Modelo de puntuación Z'' de Altman.

Con el fin de obtener un modelo adecuado para su uso con todas las categorías de empresas, la variable X_5 se excluye de la ecuación anterior, ya que es un valor que varía significativamente de una empresa a otra en diferentes sectores o industrias, lo que provoca una distorsión que puede influir en el modelo. Entonces, se obtiene la siguiente expresión:

$$Z'' = 3.25 + 6.56 X_1 + 3.26 X_2 + 6.72 X_3 + 1.05 X_4$$

En este otro modelo Z'' , una empresa con una puntuación inferior a 1,1 se considera una alta probabilidad de quiebra, mientras que si es superior a 2,6 se considera saludable. Por otra parte, las empresas con un valor entre 2,6 y 1,1 se consideran en una zona de precaución financiera (Altman & Hotchkiss, 1993), (Altman E & Hotchkiss E, 2011).

Al igual que en el caso anterior, se pueden encontrar artículos académicos recientes usando modelos previos en vez de la última actualización (Hosaka, 2019), (Alfaro et al., 2019).

3.3.2.2.4- Modelo Z-score de Altman Z 2014.

Desde el año 2000 hasta el 2014 se han publicado numerosos artículos científicos mencionando el modelo Z-score Altman. De una selección de 33 artículos, 16 de ellos utilizaron el modelo para medir la solidez financiera de empresas o la quiebra; otros 14 estudios verificaron y modificaron el modelo, y en 3 casos se utilizó para validar su robustez. A pesar de que este modelo ha sido publicado durante muchos años y es ampliamente aceptado, parece lógico creer que requiere una revisión, en la que se reestimen sus parámetros con nuevos datos y haciendo uso de nuevas técnicas de estimación.

Por esta razón, Altman y otros han realizado una nueva versión del modelo Z-score (Altman et al., 2014), con una nueva muestra de datos de 2.640.778 empresas (2.602.563 no en régimen concursal y 38.215 en quiebra) de USA, China, Colombia y 31 países europeos. También se han incluido nuevas variables (país, industria, tamaño y edad) y se han agrupado en base a siete hipótesis (H1 a H7) para mejorar el rendimiento del modelo (Barboza et al., 2017), que se muestran en la Tabla 17

CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS

Hypothesis	Description
H1	Obsolescence of the coefficients. In the previous model, the same sample of companies was used as those in the initial Altman model carried out many years earlier.
H2	Estimation method. In the previous model, the method used for estimation is the MDA, which is based on the least squares method and requires the data to comply with the principles of multinormality, homoeasticity and linearity, which is not always possible.
H3	Bankruptcy year. Year in which the bankruptcy occurs. In contrast to the previous model, whose sample corresponds to a long period of time, specifically 1946–1965, the year is considered.
H4	Size of the company. The previous model did not include data from very small or very large companies.
H5	Age of the company. This is an additional contribution to previous model, since it does not explicitly consider the age of the company, even though it does influence the bankruptcy. Bankruptcy is more likely for young companies.
H6	Company sector or industry. The previous model was made for the productive industrial sector. The sector or industry influences the financial analysis and bankruptcy of companies [49].
H7	Country of origin. It allows the model to be adjusted to the country, since samples of companies from various countries have been used (USA, China, Colombia and 31 European countries)

Tabla 17.- The seven hypotheses considered by Altman in the updated model. Barboza, F.; Kimura, H.; Altman, E. *Machine Learning Models and Bankruptcy Prediction. Expert Syst. Appl.* 2017, 83, 405–417

Para crear este nuevo modelo de puntuación Z se utilizó la expresión de regresión logística, ya que no requiere que los datos cumplan con homocedasticidad, normalidad y colinealidad, lo cual, si es un requisito requerido en modelos anteriores de puntaje Z, ya que son obtenidas por MDA. La puntuación del nuevo modelo Z-score se obtiene con la ecuación siguiente, cuyo valor representa la probabilidad de quiebra, donde el valor 1 es el máximo y 0 el mínimo.

$$Z = \frac{1}{1 + e^{-L}}$$

El valor de la variable L en la ecuación, se obtiene de una expresión lineal que depende de las hipótesis consideradas en el modelo. Hay siete hipótesis, una para cada modelo. La ecuación posterior representa la expresión para el conjunto de todas las hipótesis a la vez, el cual ha sido llamado el octavo modelo de la regresión logística de Altman Z-score.

$$\begin{aligned} L = & -13.302 - 0.459 X1 - 1.160 X2 - 1.682 X3 - 0.013 X4 \\ & - 0.034 D1 - 0.150D2 - 0.631D3 \\ & + 1.837 S1 - 0.061 S2 \\ & + 0.186 A1 - 0.099 A2 \\ & - 0.628 I1 + 0.365 I2 - 0.157 I3 - 0.176 I4 + 0.095 I5 - 0.472 I6 - 0.915 I7 \\ & - 0.014 C1 \end{aligned}$$

Las variables X_i en esta ecuación corresponden a la Tabla 16. La variable ficticia D_i se refiere al último año de datos disponibles en el estado contable de la empresa, de modo que los valores de D_i son, si el año es igual o anterior a 2008 ($D_1 = 1, D_2 = 0, D_3 = 0$), si es 2009 ($D_1 = 0, D_2 = 1, D_3 = 0$), y si es igual o mayor a 2010 ($D_1 = 0, D_2 = 0, D_3 = 1$). Las variables ficticias I_i son cero a menos que la empresa corresponda a una de estas industrias: restaurantes y hoteles ($I_1 = 1$), construcción ($I_2 = 1$); venta al por mayor y al por menor ($I_3 = 1$); agricultura ($I_4 = 1$); manufactura ($I_5 = 1$); producción de energía y agua ($I_6 = 1$); o tecnología de la información ($I_7 = 1$). La variable ficticia C_1 se refiere al riesgo país de el país de origen de la empresa y se basa en la calificación de riesgo país de Standard & Poor's. Esta variable tiene en cuenta el riesgo de pago de un país e implícitamente evalúa la situación económica del mismo. Con base en el valor de riesgo país de S&P, el valor de C_1 es: AAA = 1, AA+ = 2, AA = 3, AA- = 4, A+ = 5, A = 6, A- = 7, BBB+ = 8, BBB = 9, BBB- = 10, BB+ = 11, BB = 12, BB- = 13, B+ = 14, B = 15, B- = 16, CCC+ = 17, CCC = 18, CCC- = 19, CC = 20, C = 21, D = 22.

Han pasado 50 años desde la publicación del primer modelo Z-score, y el modelo Z-score se ha convertido en un estándar para estimar la probabilidad de quiebra y permite realizar análisis con fines de concesión de créditos, inversiones y fusiones y adquisiciones (Altman, 2018).

Una revisión de la literatura académica indica que el modelo original fue desarrollado para empresas que cotizaban en bolsa, mientras que el modelo Z' estaba enfocado a empresas privadas de manufactura, y ambos modelos (Z'' y el Z-score 2014) se adaptaron a diferentes empresas (privadas y cotizadas) y diferentes propósitos comerciales (fabricantes o no manufactureros) (Altman E et al., 2019).

El modelo de Altman se ha utilizado en muchos artículos académicos. Sin embargo, una revisión a los últimos artículos científicos que tratan este tema revela que se utilizan algunos de los modelos de puntuación Z más antiguos, en lugar de la última versión que es más completa y actualizada. No está claro por qué la comunidad de científicos sigue utilizando un modelo anterior, cuando se ha demostrado que el modelo renovado es más preciso (H. Lin, 2015).

Una posible explicación podría ser la tipología de la industria de las empresas en estudio, pero se debiera utilizar el último modelo Z-Score 2014, ya que es válido para todas las tipologías de empresa. Otra posible causa podría ser la falta de datos que permitan la categorización del resultado de la probabilidad de riesgo de quiebra en algunas zonas, como sí lo incluyen los modelos anteriores. Esta es una de las contribuciones de este trabajo.

3.3.3.- Metodología.

La metodología consiste en un estudio estadístico de la última versión del predictor Altman de quiebra en un conjunto de datos de empresas españolas de diferentes industrias durante un período de tiempo.

La muestra del conjunto de datos se basa en 1379 empresas españolas (el 73% son sociedades anónimas y el 27% son sociedades de responsabilidad limitada) que cuentan con datos contables para el periodo comprendido entre 2010 y 2013. Los datos de esta muestra se han obtenido de la base de datos SABI (Sistema de Análisis de Balances Ibéricos).

La base de datos SABI ha clasificado la actividad principal de las empresas según la Clasificación Nacional de Actividades Económicas de España (CNAE 2009). Para aplicar el modelo Z-Score (Altman et al., 2014), se debe utilizar la clasificación de este modelo, por lo que se ha llevado a cabo una equivalencia entre ambas codificaciones. La tabla 18 muestra esta equivalencia entre la clasificación de 2009 que propone el CNAE y la clasificación del Z-score. Esta equivalencia es sencilla, ya que hay casi equivalencia unívoca entre ambas codificaciones. Así, las empresas del estudio se agrupan en siete tipos de industrias principales, y las empresas restantes se agrupan en la categoría “otros”

CNAE 2009		Z-Score Altman 2014	
I	Restaurants and hotels	1	Restaurants and hotels
F	Building	2	Building
G	Wholesale and retailing	3	Wholesale and retailing
A	Agriculture, farming, forestry and fishing	4	Agriculture
C	Manufacturing industry	5	Manufacturer
D, E	Supply of electricity, gas, steam and air conditioning; water supply, sanitation, waste management and decontamination activities	6	Energy and water
J	Information and communications	7	Information Technology
B, H, K, L, M, N, O, P, Q, R, S, T, U	Extractive industries; financial and insurance activities; real estate activities; professional, scientific and technical activities; administrative and support service activities; public administration and defence; compulsory social security; education; human health and social work activities; arts, entertainment and recreation; service activities; activities of households as employers of domestic personnel; activities of households as producers of goods and services for own use; activities of extraterritorial organisations and bodies	8	Others

Tabla 18.- Table of equivalences of economic activities between CNAE and Z-score Altman 2014. Source: Own elaboration

3.3.4.- Estudio empírico del octavo modelo de la regresión logística Altman Z-score.

Existe una amplia muestra de datos, compuesta por 5516 estados financieros de empresas españolas pertenecientes a diferentes tipos de industrias sobre las que se han realizado los cálculos para obtener los valores para el modelo Z''-score [36,37] y el octavo modelo del Z-score de Altman (Altman et al., 2014).

Se ha utilizado la clasificación financiera de bancarrota de las empresas según la versión Z'', y ha servido de base para calcular los valores del octavo modelo de la regresión logística Altman Z-Score. Los resultados de estos cálculos se encuentran en la tabla 19.

Financial Classification of the Company Z''-Score Model	N	Average	Standard		The 95% Confidence Interval for the Mean		Min.	Max.
			Deviation	Error	Lower	Upper		
Likely bankruptcy	100	0.315033	0.1340710	0.0134071	0.288431	0.341636	0.0833	0.5865
Caution	419	0.274204	0.1140356	0.0055710	0.263254	0.285155	0.0583	0.5665
Healthy	4997	0.222384	0.1051514	0.0014875	0.219468	0.225300	0.0020	0.5958
Total	5516	0.228000	0.1079477	0.0014535	0.225151	0.230849	0.0020	0.5958

Tabla 19.- Table of descriptive values of the eighth model of the Altman Z-Score Logistic Regression of the research sample based on the bankruptcy financial classification of Z''-score model. Source: Own elaboration

Los valores medios de la probabilidad de quiebra, según la última puntuación Z de cada agrupación de estados financieros de las empresas, han servido para calcular los límites de cada una de las tres categorías (quiebra probable, cautela y saludable) que han permitido dar respuesta a la pregunta de investigación RQ2.

En la tabla 20 se muestra el resumen de datos obtenidos, que establece la equivalencia de las agrupaciones para ambos modelos de quiebra. Los rangos de probabilidad varían de 0 a 1. Se obtiene que una empresa entra en la categoría de probabilidad de quiebra cuando este valor de Z según el octavo modelo es mayor que 0.315.

Financial Classification of the Company	Z''-Score	The Eighth Model of the Altman Z-Score
Healthy	$2.6 < Z''\text{-score}$	$0 \leq Z\text{-score} < 0.2223$
Caution	$1.1 \leq Z''\text{-score} \leq 2.6$	$0.2223 \leq Z\text{-score} \leq 0.3150$
Likely bankruptcy	$Z''\text{-score} < 1.1$	$0.3150 < Z\text{-score} < 1$

Tabla 20.- Equivalence of the financial classification between the Z''-Score model and the renewed measurement of the eighth model of the Altman Z-Score. Source: Own elaboration

Los resultados estadísticos del octavo modelo de la regresión logística Altman Z-Score se muestran en la Tabla 6. Esta tabla está ordenada de menor a mayor probabilidad de quiebra para cada sector (industries) con el fin de responder a la pregunta de investigación RQ1.

Industries	N	Average	Standard		95% of the Confidence Interval for the Mean		Min.	Max.
			Deviation	Error	Lower	Upper		
Information technology	236	0.14	0.06	0.00	0.13	0.14	0.02	0.33
Restaurants and hotels	124	0.18	0.07	0.01	0.17	0.20	0.08	0.42
Agriculture	56	0.20	0.06	0.01	0.18	0.21	0.11	0.39
Wholesale and retailing	1344	0.21	0.09	0.00	0.21	0.22	0.05	0.47
Energy and water	176	0.25	0.07	0.00	0.24	0.26	0.10	0.39
Manufacturer	2376	0.27	0.09	0.00	0.27	0.27	0.06	0.59
Building	316	0.37	0.10	0.01	0.36	0.38	0.10	0.60
Others	888	0.12	0.05	0.00	0.11	0.12	0.00	0.31
Total	5516	0.23	0.11	0.00	0.23	0.23	0.00	0.60

Tabla 21.- Measurement of the eighth model of the Altman Z-Score for different industries. Source: Own elaboration

Las empresas en la cadena de suministro del sector de la construcción tienen la mayor probabilidad de quiebra, que se encuentra en la categoría de quiebra probable. A continuación, las industrias manufactureras y de energía y agua están en la categoría de precaución financiera. Finalmente, en la categoría saludable, de acuerdo con la probabilidad de riesgo, se encuentran los siguientes sectores: mayorista y minorista, agricultura, restaurantes y hoteles e industrias de tecnología de la información, en ese orden.

La agrupación de industrias consideradas como otras tiene la menor probabilidad de quiebra.

Si se comparan los valores máximos de cada sector se observa que la clasificación continúa siendo la misma, aunque también los datos indican una desviación típica cercana a 0,1, por lo que se puede concluir que dentro de cada industria hay situaciones financieras muy diferentes entre unas empresas y otras del mismo sector

3.3.5.- Discusión.

Este estudio introduce la conveniencia de incluir un nuevo indicador para medir el desempeño de la cadena de suministro (SCPM: Supply Chain Performance Measurement) para evaluar la probabilidad de quiebra de la empresa. Estos indicadores SCPM se pueden compartir entre las empresas de la misma cadena de suministro para mejorar las operaciones entre ellas y permitir una cadena de suministro más eficiente. Además, se han encontrado diferencias significativas en la probabilidad de quiebra para cada uno de los sectores.

La cadena de suministro de la industria de la tecnología de la información se compone de servicios tecnológicos que están experimentando un rápido progreso en la computación en la nube para reducir costos, así como productos de tecnología que permiten a los clientes finales obtener eficiencias de costos en sus flujos de información (Ferguson & Hadar, 2011), (Reyes et al., 2014). Esta mayor eficiencia para el cliente final podría conducir a una mayor liquidez y

crédito financiero para las empresas de este sector industrial, lo que reduce la probabilidad de la quiebra de la empresa y la interrupción de la cadena de suministro de las empresas que lo componen.

La cadena de suministro de la industria de restauración y hostelería está relacionada con el poder económico de los consumidores finales y la actividad económica de otros sectores (Chen et al., 2018), (Alberca & Parte, 2018). Debido al hecho de los datos corresponden a un momento de crecimiento económico, posterior a la crisis de 2008, que potenció la eficiencia de las empresas del resto de sectores, puede explicarse la situación saneada de las empresas de este sector y que se muestra en la Tabla 21.

En el sector agrícola, se ha encontrado que las relaciones en el tiempo de las empresas que conforman la cadena de suministro son más estables, influyendo positivamente en el valor de probabilidad de quiebra. Por lo que se obtiene una posición saludable de este sector (van Bergen et al., 2019), (Sun et al., 2018).

La cadena de suministro mayorista y minorista requiere niveles altos de efectivo y, por lo tanto, riesgo financiero, que suele estar protegido por seguros y otros mecanismos para mejorar las decisiones con créditos comerciales y bancarios. Por otro lado, un aumento en las tasas de los préstamos bancarios es un indicador de la reducción de los beneficios de estas empresas. No obstante, durante el período de muestreo en las empresas se observan tasas bajas de interés, lo cual mejora los beneficios de este sector y le proporciona una situación financiera sana (Xing Y.U. & Zhongkai WAN M.D., 2020), (Q. Lin & Xiao, 2018).

La cadena de suministro de energía y agua implica una alta inversión en infraestructura regulada por el gobierno del país, lo que requiere constantes adaptaciones y cambios en las regulaciones que requieren una inversión de capital adicional. El aumento de la inversión y la falta de precios regulados por el gobierno actualizados para estas empresas provoca que los márgenes de empresas en esta industria se reduzcan drásticamente. Por lo tanto, la probabilidad de quiebra aumentó, lo que hizo que este sector estuviera en la zona de precaución (Ameyaw et al., 2017), (Ye et al., 2019).

La industria manufacturera está compuesta por cadenas de suministro que deben estar particularmente atentos a los riesgos en las operaciones, ya que esto influye en la eficiencia de la cadena de suministro y plantea riesgos potenciales. Por tanto, además del producto, se debe mejorar la calidad de la información y el servicio para reducir el riesgo de suministro y evitar la probabilidad de quiebra. Esta necesidad de gestionar el riesgo de suministro provocó que el sector manufacturero se encontrara en una zona de precaución (Truong & Hara, 2018), (Benedettini et al., 2017).

CAPÍTULO III: RESULTADOS E INVESTIGACIONES REALIZADAS Y PUBLICADAS

La cadena de suministro del sector de la construcción tiene numerosos riesgos que, si no se gestionan adecuadamente puede contribuir a una mayor probabilidad de quiebra. Además, los directivos de este sector tratan de utilizar su experiencia en lugar de herramientas analíticas para reducir el riesgo. Las características especiales del sector de la construcción, así como la cultura y la fragmentación de sus cadenas de suministro, entre otras, han llevado a la renuncia a mejorar y reducir la complejidad de la gestión de la cadena de suministro. Como resultado, el sector de la construcción estaba en la zona de probable bancarrota (Shojaei & Haeri, 2019), y (Aloini et al., 2012).

En consecuencia, el nuevo indicador de SCPM (Supply Chain Performance Measurement) recomendado como predictor es el Altman Z-score, que, a pesar de ser sencillo, ha demostrado ser capaz de predecir de forma fiable el fracaso o la quiebra de las empresas y actualmente está siendo utilizado por los inversores (Giroux & Cassell, 2011), (Srebro et al., 2021).

Los modelos anteriores de Altman (Z , Z' y Z'') se basaban en el método de análisis discriminante múltiple (MDA) de múltiples parámetros. Este método requiere que las variables independientes de los parámetros cumplan unos requisitos (normalidad, linealidad, homocedasticidad), que no se cumplen para las variables financieras utilizadas.

Se ha aplicado la técnica de regresión logística o logit al renovado modelo de Altman (Altman et al., 2014), eliminando la limitación de los modelos MDA, que requieren que los datos satisfagan las condiciones de homocedasticidad, normalidad y colinealidad. Además, se han incorporado nuevas variables que proporcionan un modelo más preciso (Altman et al., 2014), (Laitinen & Suvas, 2013). A pesar de esto, es más probable que las publicaciones académicas usen las versiones anteriores de Altman, ya que son más familiares y más fácil de usar.

Se considera que la nueva revisión del modelo Z-score de Altman tiene algunos parámetros que ya están obsoletos. Por ejemplo, una variable ficticia se refiere al último año con datos disponible del estado contable de la empresa, y sólo se referencia para los años 2007 a 2010, por lo que si se hace un estudio en la actualidad esta variable será cero para todas las empresas, ya que muchos de ellos tendrán estados contables del 2010 a no ser que fueran de reciente creación. De la misma forma, la variable para considerar el país de origen de la empresa es respaldada por la calificación de riesgo país de Standard & Poor's, pero no se ha establecido qué hacer si esta entidad decide introducir nuevas calificaciones, como sucedió en la última crisis de 2008.

La introducción en el modelo Z-score de 2014 de Altman de una variable que refleja la edad de la empresa y otra variable que refleja el riesgo país se considera positiva, ya que supera la miopía de los métodos discriminatorios en los ejes temporal y macroeconómico.

Esta investigación es una de las primeras en aplicar el modelo 8 de puntuación Z de 2014 de (Altman et al., 2014), para realizar una clasificación de la probabilidad de quiebra para las empresas de la cadena de suministro de diferentes sectores industriales. Además, este estudio proporciona por primera vez un ranking de la probabilidad de quiebra de las empresas para este nuevo modelo z-score.

Las principales limitaciones de este estudio son que se ha realizado para empresas españolas durante un período de tiempo posterior a la crisis de 2008.

CAPÍTULO IV: CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURO

4.1.- Conclusiones individuales (por estudio)

En este capítulo se recogen las conclusiones de cada una de las investigaciones llevadas a cabo para la elaboración de la presente tesis y se indican por separado las futuras líneas de investigación propuestas para cada uno de los estudios

4.1.1.- Conclusiones y futuras líneas de investigación referentes al estudio “Una herramienta de visualización inteligente para analizar la sostenibilidad del transporte por carretera”.

Por un lado, es evidente que el desarrollo de la actividad del transporte de mercancías por carretera es intrínsecamente necesario para la actividad económica y el creciente intercambio de bienes materiales entre diferentes zonas del planeta, mientras que, por otro lado, su despliegue conlleva una importante tasa de emisiones de GEI (Gases de Efecto Invernadero). Además, también es evidente que la positiva progresión en el desarrollo de los medios de transporte ha supuesto importantes reducciones en la emisión de gases y partículas a la atmósfera con los equipos más modernos. Como consecuencia de lo anterior, las herramientas para el estudio, control y predicción de la evolución de la (edad) antigüedad de los medios de transporte pueden resultar de gran valor para el gobierno corporativo y la gestión de esta actividad. Más concretamente, en sus líneas de compromiso con la sostenibilidad, que les permitan poner en marcha planes de estímulo y políticas de responsabilidad social encaminadas a la renovación de la flota de vehículos de transporte.

En este sentido, los HUEPs (Hybrid Unsupervised Exploratory Plots) constituyen una herramienta visual válida, rápida e intuitiva para estudiar la progresión de la edad de la flota a partir de un conjunto de datos relacionados con la actividad del transporte de mercancías por carretera.

En cuanto a las técnicas empleadas para el conjunto de datos en estudio, vale la pena señalar que una de las nuevas técnicas EPP (Exploratory Projection Pursuit) probadas (KPCA: Kernel Principal Component Analysis) es la que proporciona mejores visualizaciones, lo que confirma aún más la extensión propuesta de la formulación original de los HUEP. También se ha demostrado que la combinación de KPCA con un método de agrupamiento proporciona mejores resultados que las pantallas 3D generadas solo con la técnica EPP. En cuanto a las técnicas de clustering, se ha confirmado la idoneidad del método aglomerativo frente a k-means, aunque es necesario ajustar el número de clusters (parámetro k) a los datos.

Para trabajos futuros, es importante destacar que existen también otros factores relacionados con el estudio de la sostenibilidad de la actividad de transporte de mercancías por carretera, que podrían ser objeto de un estudio de aplicación de estas y otras técnicas vinculadas al ML, lo que permitiría probablemente arrojar valiosas conclusiones. Un claro ejemplo sería el estudio de la ocupación de los medios de transporte en el desarrollo de su actividad, además de escalar la dimensión de los datos, considerando el conjunto de Europa. En cuanto a las técnicas de ML, se propone la aplicación de HUEPs en este campo, añadiendo otras técnicas de clustering como las basadas en densidad.

4.1.2- Conclusiones y futuras líneas de investigación referentes al estudio “Modelos neuronales no lineales para predecir el precio del acero HRC en España”.

Este estudio ha presentado los resultados empíricos obtenidos con tres conjuntos de datos (mercado de valores, económico y otros productos básicos o *commodities*) sobre el rendimiento de varios modelos no lineales (NIO, NAR, NARX) para la previsión del precio del acero. Se ha comparado el rendimiento de los diferentes modelos neuronales y sus parámetros. Los modelos utilizados no se han aplicado previamente para la previsión del precio del acero.

Después de analizar los resultados individuales, se puede concluir sin ninguna duda que las mejores predicciones del precio del acero HRC se obtienen con el modelo NARX y la serie de datos 1, mostrando un rendimiento superior a los otros modelos y series de datos. Por lo tanto, se puede concluir que existe una fuerte relación entre el precio de las acciones de las empresas siderúrgicas y el precio del acero.

Desde la perspectiva del ajuste de parámetros, se puede concluir que para algunos de los parámetros no existe un consenso que afecte a todos los modelos. Los mejores valores para el número de retrasos de entrada y neuronas ocultas varían de un modelo a otro, por lo que debe ajustarse caso por caso. Sin embargo, en el caso del algoritmo utilizado para entrenar el modelo, se puede concluir que el “Gradiente Conjugado Escalado” y “Broyden—Fletcher—Goldfarb—Shanno Back-propagation” son los más recomendables para los tres modelos.

CAPÍTULO IV: CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURO

Al considerar la serie de datos de entrada en términos generales, el “Precio de otras materias primas” y los “Precios bursátiles” son los que se asocian a mejores resultados. Los peores resultados se han obtenido siempre al utilizar la serie de datos “Situación económica en España”.

Desde una perspectiva empresarial, el resultado obtenido muestra que la evolución bursátil de las acciones de las empresas siderúrgicas está correlacionada con la evolución del precio del acero. Esta conexión es consistente con el efecto de que aumentos en el precio de venta del material de una empresa, con igual demanda, tienen una influencia positiva en la rentabilidad de la empresa, y por lo tanto el valor de la acción se incrementa para los potenciales inversores.

En futuros estudios se podrán utilizar nuevas técnicas de computación y otras series de datos para mejorar los modelos predictivos existentes.

4.1.3.- Conclusiones y futuras líneas de investigación referentes al estudio “Análisis de la sostenibilidad económica en la cadena de suministro aplicando el predictor Altman Z-score”.

Este artículo llena el vacío en la perspectiva financiera de la mejora de la medición del desempeño de la cadena de suministro. (SCPM: Supply Chain Performance Measurement), relacionado con la falta de un predictor de quiebra, que permite a la empresa resumir la probabilidad de quiebra y compararse con otras empresas en su sector y éste con otros sectores. Para SCPM y para predecir la probabilidad de quiebra de la empresa, se recomienda la utilización de un indicador tal como el indicado en el octavo modelo de Altman Z-score.

Además, se ha establecido un ranking de probabilidad de quiebra por sector, a partir de un estudio de una muestra amplia de estados financieros de empresas españolas, lo que permite responder a la pregunta de investigación RQ1. Este ranking se puede utilizar para comparar hacer benchmarking entre empresas en cadena de suministro de un sector industrial, ya que se puede comparar con el valor promedio obtenido para la probabilidad de quiebra para cada caso. Se ha utilizado una muestra empírica muy amplia de empresas españolas, lo que ha permitido una mayor precisión en los resultados obtenidos.

Además, este trabajo es pionero en la clasificación de las empresas en tres categorías (saludable, cautelosa y probable bancarrota) para la predicción de dificultades financieras utilizando el nuevo modelo de puntuación Z de Altman, que ha respondido a la pregunta de investigación RQ2.

Se pueden sugerir futuros estudios, como el análisis de la evolución de las empresas estudiadas y cuáles quebraron después de 2014. De igual manera, se puede revisar el comportamiento estratégico de cada empresa para determinar la influencia de la quiebra de la empresa en la cadena de suministro del sector al que pertenece.

4.2.- Conclusiones generales

A través de los diferentes estudios que componen el cuerpo fundamental de la presente tesis se ha podido observar y concluir que las distintas técnicas utilizadas de aprendizaje máquina han resultado de gran ayuda para la toma de decisiones para una gestión eficiente en las empresas industriales y de otros ámbitos en los que se ha experimentado.

Se ha demostrado mediante los estudios que componen la tesis el valor añadido que supone para las empresas y sectores analizados del entorno industrial la utilización de las técnicas de aprendizaje máquina destacando la aplicabilidad de cada una de ellas como la más adecuada a cada uno de los casos estudiados (logística, análisis de financiación empresarial y su posibilidad de quiebra en base a su salud financiera, predicción de precios a futuro en el sector del acero, etc.)

La visualización inteligente utilizando las Parcelas Exploratorias Híbridas No supervisadas (HUEP,s) combinada con métodos de agrupamiento han resultado ser de gran utilidad para evaluar de forma rápida e intuitiva la progresión de la edad de la flota en actividades de transporte de mercancías por carretera.

Por su parte, la evolución del indicador de predicción de quiebra de una empresa “Altman Z-score” ha ido mejorando a lo largo del tiempo con sus nuevas adaptaciones y mejoras sucesivas. El octavo modelo utilizado en nuestro estudio arroja datos muy interesantes atendiendo a la clasificación de empresas en las tres categorías (saludable, con riesgo moderado y probable bancarrota) sobre el tipo de empresa dentro de un determinado sector, además de demostrar ser útil para elaborar un ranking por sectores industriales. Más aún, ha resultado ser muy interesante para determinar la influencia de cada empresa dentro

CAPÍTULO IV: CONCLUSIONES Y LÍNEAS DE TRABAJO FUTURO

de su cadena de suministro dentro de un determinado sector de actividad industrial.

Por tanto, podemos concluir que las herramientas utilizadas de aprendizaje máquina aportan un importante valor en la toma de decisiones para la empresa como se ha demostrado en los tres trabajos realizados y que conforman el soporte de la presente tesis doctoral.

BIBLIOGRAFÍA

- Adewuyi, A. O., Wahab, B. A., & Adeboye, O. S. (2020). Stationarity of prices of precious and industrial metals using recent unit root methods: Implications for markets' efficiency. *Resources Policy*, 65, 101560. <https://doi.org/10.1016/J.RESOURPOL.2019.101560>
- Adiani, D., Qu, C., Gass, T., Gurram, S., LeMay, D., Bhusal, A., Sarkar, M., & Sarkar, N. (2022). Evaluation of Webcam-Based Eye Tracking for a Job Interview Training Platform: Preliminary Results. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*: Vol. 13336 LNAI. https://doi.org/10.1007/978-3-031-05643-7_22
- Adli. (2020). *Forecasting Steel Prices Using ARIMAX Model: A Case Study of Turkey by Kaveh A. Adli* :: SSRN. S SNR. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3728117
- Agbo A.A., Li W., Zheng L., Zhang Y., & Atombo C. (2017). Optimisation of inter-modal freight transport network. *Eur. Transp. - Trasp. Eur.*, No. 63.
- Alberca, P., & Parte, L. (2018). Operational efficiency evaluation of restaurant firms. *International Journal of Contemporary Hospitality Management*, 30(3), 1959–1977. <https://doi.org/10.1108/IJCHM-09-2016-0547/FULL/PDF>
- Alcalde-Delgado, R., Sáiz-Barcena, L., Olmo, R., & Alonso de Armiño, C. (2020). Empirical study of the business growth strategy related to the added value by intellectual capital. *International Journal of Production Management and Engineering*, 8(1), 1–12. <https://doi.org/10.4995/IJPME.2020.10817>
- Alfaro, L., Asis, G., Chari, A., & Panizza, U. (2019). Corporate debt, firm size and financial fragility in emerging markets. *Journal of International Economics*, 118, 1–19. <https://doi.org/10.1016/J.JINTECO.2019.01.002>
- Aloini, D., Dulmin, R., Mininno, V., & Ponticelli, S. (2012). Supply chain management: A review of implementation risks in the construction industry. *Business Process Management Journal*, 18(5), 735–761. <https://doi.org/10.1108/14637151211270135/FULL/PDF>
- Alonso, Á., Porras, S., Garitano, I., Arenaza, I., Uribeetxeberria, R., Zurutuza, U., Herrero, Á., & Corchado, E. (2010). *On the Visualization of Honeypot Data through Projection Techniques*. 27–30.

BIBLIOGRAFÍA

- Alonso de Armiño, C., Manzanedo, M. Á., & Herrero, Á. (2020a). Analysing the intermeshed patterns of road transportation and macroeconomic indicators through neural and clustering techniques. *Pattern Analysis and Applications*, 23(3), 1059–1070. <https://doi.org/10.1007/S10044-020-00872-X/TABLES/4>
- Alonso de Armiño, C., Manzanedo, M. Á., & Herrero, Á. (2020b). Analysing the intermeshed patterns of road transportation and macroeconomic indicators through neural and clustering techniques. *Pattern Analysis and Applications*, 23(3), 1059–1070. <https://doi.org/10.1007/S10044-020-00872-X/TABLES/4>
- Altman E, & Hotchkiss E. (2011). *Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy ... - Edward I. Altman, Edith Hotchkiss - Google Libros*. John Wiley & Sons, Inc.: New York, NY, USA. https://books.google.es/books?hl=es&lr=&id=_5n0-J37Zz4C&oi=fnd&pg=PT26&dq=Altman,+E.I.%3B+Hotchkiss,+E.+Corporate+Financial+Distress+and+Bankruptcy:+Predict+and+Avoid+Bankruptcy,+Analyze+and+Invest+in+Distressed+Debt,+3rd+ed.%3B+John+Wiley+%26+Sons+Inc.:+New+York,+NY,+USA,+2011%3B+ISBN+978-04-7169-189-1&ots=E5Blf4lz9M&sig=l8oXAm-1CycLDBdd9jplfS5ZNOk#v=onepage&q&f=false
- Altman E, Hotchkiss E, & Wang W. (2019). *Corporate Financial Distress, Restructuring, and Bankruptcy: Analyze ... - Edward I. Altman, Edith Hotchkiss, Wei Wang - Google Libros*. John Wiley & Sons, Inc.: New York, NY, USA. <https://books.google.es/books?hl=es&lr=&id=V1-HDwAAQBAJ&oi=fnd&pg=PR9&dq=Altman,+E.I.%3B+Hotchkiss,+E.%3B+Wang,+W.+Corporate+Financial+Distress,+Restructuring,+and+Bankruptcy,+4th+ed.%3B+John+Wiley+%26+Sons,+Inc.:+New+York,+NY,+USA,+2019%3B+ISBN+978-11-1954-192-9.&ots=Lppt6VtFYH&sig=FNMO6NI9X-tWCs3FBvO849E-RPU#v=onepage&q&f=false>
- Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589. <https://doi.org/10.2307/2978933>
- Altman, E. I. (2018). Applications of Distress Prediction Models: What Have We Learned After 50 Years from the Z-Score Models? *International Journal of Financial Studies* 2018, Vol. 6, Page 70, 6(3), 70. <https://doi.org/10.3390/IJFS6030070>
- Altman, E. I., & Hotchkiss, E. (1993). *Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt Third Edition Corporate Financial Distress and Bankruptcy*.

- Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2014). Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.2536340>
- Ameyaw, E. E., Chan, A. P. C., Owusu-Manu, D.-G., Edwards, D. J., & Dartey, F. (2017). A Fuzzy-Based Evaluation of Financial Risks in Build–Own–Operate–Transfer Water Supply Projects. *Journal of Infrastructure Systems*, 23(4), 04017033. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000390](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000390)
- Amro, A. (2023). Evaluation of Using Computer Assisted Telephone Interview Mode of Data Collection in the Time of COVID–19 Pandemic—Palestine Case. In *Studies in Systems, Decision and Control* (Vol. 216). https://doi.org/10.1007/978-3-031-10212-7_18
- Andreopoulos, B., An, A., Wang, X., & Schroeder, M. (2009). A roadmap of clustering algorithms: finding a match for a biomedical application. *Briefings in Bioinformatics*, 10(3), 297–314. <https://doi.org/10.1093/BIB/BBN058>
- Ang-Olson, J., & Schroeder, W. (2002). Energy Efficiency Strategies for Freight Trucking: Potential Impact on Fuel Use and Greenhouse Gas Emissions. *https://doi.org/10.3141/1815-02*, 1815, 11–18. <https://doi.org/10.3141/1815-02>
- Angrosino, M. (2012). Etnografía y observación participante en investigación cualitativa. *Ed Morata. Madrid*, 138.
- Arbones Malisani E. (1992). *LOGÍSTICA EMPRESARIAL. de ARBONES MALISANI, EDUARDO A.: Bueno (1992) | Librería Vobiscum*. Librería Vobiscum. https://www-origin.iberlibro.com/servlet/BookDetailsPL?bi=6380035372&cm_sp=snippet_-_srp1_-_tile8&searchurl=an%3Deduardo%2Barbones%2Bmalisani%26ds%3D10%26n%3D100121503%26sortby%3D17
- Argyrou, A. (2009). Clustering hierarchical data using self-organizing map: A graph-theoretical approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5629 LNCS, 19–27. https://doi.org/10.1007/978-3-642-02397-2_3/COVER
- Arredondo Dominguez E R., Villalta Jadan B. E., Martinez Patiño E.A., Aldaz bombón O. R., & Gómez Cardenas R.E. (2019). *The impact of Focus Groups as a market research tool - ProQuest*. Asesorías y Tutorías Para La Investigación Científica En La Educación Puig-Salabarría S.C. <https://www.proquest.com/docview/2245651347/abstract/A4FC3D6EEFE14128PQ/1?accountid=17195>

BIBLIOGRAFÍA

- Arroyo, Á., Herrero, Á., Tricio, V., & Corchado, E. (2017). Analysis of meteorological conditions in Spain by means of clustering techniques. *Journal of Applied Logic*, 24, 76–89. <https://doi.org/10.1016/J.JAL.2016.11.026>
- Arzu Akyuz, G., & Erman Erkan, T. (2009). Supply chain performance measurement: a literature review. *Http://Dx.Doi.Org/10.1080/00207540903089536*, 48(17), 5137–5155. <https://doi.org/10.1080/00207540903089536>
- Ashraf, S., Félix, E. G. S., & Serrasqueiro, Z. (2019). Do Traditional Financial Distress Prediction Models Predict the Early Warning Signs of Financial Distress? *Journal of Risk and Financial Management* 2019, Vol. 12, Page 55, 12(2), 55. <https://doi.org/10.3390/JRFM12020055>
- Baeza-Yates R. (1992). *Information Retrieval: CHAPTER 2: INTRODUCTION TO DATA STRUCTURES AND*. <http://icodex.com/vc/10book/books/book5/chap02.htm>
- Ballano, C. (2005). *La recogida de datos de la EPA 2005*.
- Ballano, C., Saralegui Gil, J., Álvarez, F., Pérez, C., Sánchez, A., Borrás, F., & Esteban Lefler, M. D. (2001). Generación de un universo artificial representativo de la población española. *XXVI Congreso Nacional de Estadística e Investigación Operativa: Úbeda, 6-9 de Noviembre de 2001, 2001, ISBN 84-8439-080-2*, 64. <https://documat.unirioja.es/servlet/articulo?codigo=3121965&info=resumen&idioma=SPA>
- Ballou, R. H. , & de Lemus, P. R. (1991). *Logística empresarial: control y planificación*. Díaz de Santos.
- Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405–417. <https://doi.org/10.1016/J.ESWA.2017.04.006>
- Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, 4, 71. <https://doi.org/10.2307/2490171>
- Benedettini, O., Swink, M., & Neely, A. (2017). Examining the influence of service additions on manufacturing firms' bankruptcy likelihood. *Industrial Marketing Management*, 60, 112–125. <https://doi.org/10.1016/J.INDMARMAN.2016.04.011>
- Bishop C. (1995). *Neural Networks for Pattern Recognition - Christopher M. Bishop* - Google Libros. [https://books.google.es/books?hl=es&lr=&id=TOS0BgAAQBAJ&oi=fnd&pg=PP1&dq=Bishop,+C.+M.+\(1995\).+Neural+networks+for+pattern+recognition,+Clarendon+Press.&ots=jNa-xF8usg&sig=dsKV2rAi8fLy-WgHkzTYuGewhO_c&redir_esc=y#v=onep-](https://books.google.es/books?hl=es&lr=&id=TOS0BgAAQBAJ&oi=fnd&pg=PP1&dq=Bishop,+C.+M.+(1995).+Neural+networks+for+pattern+recognition,+Clarendon+Press.&ots=jNa-xF8usg&sig=dsKV2rAi8fLy-WgHkzTYuGewhO_c&redir_esc=y#v=onep-)

age&q=Bishop%2C%20C.%20M.%20(1995).%20Neural%20net-works%20for%20pattern%20recognition%2C%20Clarendon%20Press.&f=false

- Brailovsky, V. L. (1991). A probabilistic approach to clustering. *Pattern Recognition Letters*, 12(4), 193–198. [https://doi.org/10.1016/0167-8655\(91\)90031-G](https://doi.org/10.1016/0167-8655(91)90031-G)
- Brown, S. D. (Steven D.), Sarabia, L. A., & Trygg, Johan. (2009). *Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice*. A211–A226. <https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-75771>
- Budak, A., & Sarvari, P. A. (2021). Profit margin prediction in sustainable road freight transportation using machine learning. *Journal of Cleaner Production*, 314, 127990. <https://doi.org/10.1016/J.JCLEPRO.2021.127990>
- Caffarel-Serra, C., Ortega-Mohedano, F., & Gaitán-Moya, J. A. (2017). Investigación en Comunicación en la universidad española en el período 2007-2014. *Profesional de La Información*, 26(2), 218–227. <https://doi.org/10.3145/EPI.2017.MAR.08>
- Calderón, X. C., Yohana, G., Castillo, C., & Calderón, R. E. (2021). LA IMPORTANCIA DE LOS ESTADOS FINANCIEROS EN LA TOMA DE DECISIONES FINANCIERA-CONTABLES. *Revista FAECO Sapiens*, 4(2), 82–96. https://revistas.up.ac.pa/index.php/faeco_sapiens/article/view/2179
- Carvajal, V. M. P., Sanabria, M. E. A., Flores, Á. S. V., & Arbeláez, T. Z. B. (2020). Los estados financieros y la toma de decisiones en las pymes. *Journal of Science and Research*, 5(CICACI), 2528–8083. <https://doi.org/10.5281/zenodo.4726284>
- Cetin, K., Aksoy, S., & Iseri, I. (2019). Steel Price Forecasting Using Long Short-Term Memory Network Model. *UBMK 2019 - Proceedings, 4th International Conference on Computer Science and Engineering*, 612–617. <https://doi.org/10.1109/UBMK.2019.8907015>
- Chen, Y. J., Tsai, H., & Liu, Y. F. (2018). Supply chain finance risk management: Payment default in tourism channels. *Tourism Economics*, 24(5), 593–614. https://doi.org/10.1177/1354816618762187/ASSET/IMAGES/LARGE/10.1177_1354816618762187-FIG4.JPEG
- Chou, M. T. , C. S. , & Yang, Y. L. (2012). A Study of the Dynamic Relationship between Crude Oil Price. *Better Advances Press, Canada in Its Journal Review of Economics & Finance*,.
- Comisión Europea. (2005). *Emisiones de los vehículos pesados (Euro VI): normas de certificación*. Diario Oficial. <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=LEGISSUM:mi0029&from=ES>

BIBLIOGRAFÍA

- Conejo, A. N., Birat, J. P., & Dutta, A. (2020). A review of the current environmental challenges of the steel industry and its value chain. *Journal of Environmental Management*, 259, 109782. <https://doi.org/10.1016/J.JENVMAN.2019.109782>
- Consejo-Europeo. (1998). *REGLAMENTO (CE) N° 1172/98 DEL CONSEJO de 25 de mayo de 1998 sobre la relación estadística de los transportes de mercancías por carretera. OPOCE.*
- Consejo-Europeo. (2012). *REGLAMENTO (UE) No 70/2012 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 18 de enero de 2012 sobre la relación estadística de los transportes de mercancías por carretera.*
- Corlu, C. G., De La Torre, R., Serrano-Hernandez, A., Juan, A. A., & Faulin, J. (2020). Optimizing Energy Consumption in Transportation: Literature Review, Insights, and Research Opportunities. *Energies 2020, Vol. 13, Page 1115, 13(5)*, 1115. <https://doi.org/10.3390/EN13051115>
- Craighead, C. W., Hult, G. T. M., & Ketchen, D. J. (2009). The effects of innovation–cost strategy, knowledge, and action in the supply chain on firm performance. *Journal of Operations Management*, 27(5), 405–421. <https://doi.org/10.1016/J.JOM.2009.01.002>
- Dabdoub, A., Snyder, L. A., Finken, A., & Lin, L. (2023). The importance of personal and contextual background factors on declaration of STEM majors: A focus on Native American undergraduates. *Science Education*, 107(2), 459–509. <https://doi.org/10.1002/sce.21777>
- De Campos, R. S., Simon, A. T., & De Campos Martins, F. (2019). Assessing the impacts of road freight transport on sustainability: A case study in the sugar-energy sector. *Journal of Cleaner Production*, 220, 995–1004. <https://doi.org/10.1016/J.JCLEPRO.2019.02.171>
- de Munck V.C., & Sobo E, J. (1988). *Using Methods in the Field: A Practical Introduction and Casebook - Google Libros*. Altamira Press. Walnut Creek CA. [https://books.google.es/books?hl=es&lr=&id=7u5yz2l4hn4C&oi=fnd&pg=PA7&dq=Victor+C.+de+Munck+%26+Elisa+J.+Sobo+\(1988\).+Using+Methods+in+the+Field.+Ed:+Rowman+Altamira+&ots=8x_x0Fi_7B&sig=sG_ESO_mH9wDa8zQasDSw6cMSWY&redirect_esc=y#v=onepage&q=Victor%20C.%20de%20Munck%20%26%20Elisa%20J.%20Sobo%20\(1988\).%20Using%20Methods%20in%20the%20Field.%20Ed%3A%20Rowman%20Altamira&f=false](https://books.google.es/books?hl=es&lr=&id=7u5yz2l4hn4C&oi=fnd&pg=PA7&dq=Victor+C.+de+Munck+%26+Elisa+J.+Sobo+(1988).+Using+Methods+in+the+Field.+Ed:+Rowman+Altamira+&ots=8x_x0Fi_7B&sig=sG_ESO_mH9wDa8zQasDSw6cMSWY&redirect_esc=y#v=onepage&q=Victor%20C.%20de%20Munck%20%26%20Elisa%20J.%20Sobo%20(1988).%20Using%20Methods%20in%20the%20Field.%20Ed%3A%20Rowman%20Altamira&f=false)
- Dewalt K. M., & DeWalt B. R. (2002). Participant observation. *A Guide for Fieldworkers*. Walnut Creek, CA.

- Domínguez Machuca, J. Antonio., & Álvarez Gil, M. J. (1995). *Dirección de operaciones : aspectos estratégicos en la producción y los servicios*. McGraw-Hill.
- Dufour W. (1983). Los métodos de objetivación del comportamiento motor de la recogida de datos en fútbol. *Apuntes Medicina Deportiva*. <https://www.apuntes.org/index.php?p=revista&tipo=pdf-simple&pii=X021337178304803X&r=276>
- DuHadway, S., Carnovale, S., & Hazen, B. (2019). Understanding risk management for intentional supply chain disruptions: risk detection, risk mitigation, and risk recovery. *Annals of Operations Research*, 283(1–2), 179–198. <https://doi.org/10.1007/S10479-017-2452-0/FIGURES/2>
- Ellinger, A. E., Natarajarathinam, M., Adams, F. G., Brian Gray, J., Hofman, D., & O'Marah, K. (2011). Supply Chain Management Competency and Firm Financial Success. *Journal of Business Logistics*, 32(3), 214–226. <https://doi.org/10.1111/J.2158-1592.2011.01018.X>
- Elwell, F. H. (1936). *Changes in the Financial Structure of Unsuccessful Industrial Corporations*.
- Essential Ethnographic Methods: Observations, Interviews, and Questionnaires - Stephen L. Schensul, Jean J. Schensul, Margaret Diane LeCompte - Google Libros*. (n.d.). Retrieved April 25, 2023, from [https://books.google.es/books?hl=es&lr=&id=QNKTORJO8kC&oi=fnd&pg=PR8&dq=SCHESENSUL,+SCHESENSUL+and+LeCOMPTE+\(1999&ots=DUowWVd1DN&sig=JHtw3Qysq2iw-LGWQsXg2HKTUms#v=onepage&q=SCHESENSUL%2C%20SCHESENSUL%20and%20LeCOMPTE%20\(1999&f=false](https://books.google.es/books?hl=es&lr=&id=QNKTORJO8kC&oi=fnd&pg=PR8&dq=SCHESENSUL,+SCHESENSUL+and+LeCOMPTE+(1999&ots=DUowWVd1DN&sig=JHtw3Qysq2iw-LGWQsXg2HKTUms#v=onepage&q=SCHESENSUL%2C%20SCHESENSUL%20and%20LeCOMPTE%20(1999&f=false)
- Eugenia, P., Castro, G., Laura, M., Barrientos, G., Rosa, E., Sosa, C., Gatica, K. L., Del, R., Vargas Hernández, R., García, J. H., Ramos Velázquez, V. A., María, D., & Díaz, M. (2016). Procesos de reclutamiento y las redes sociales. *RIDE Revista Iberoamericana Para La Investigación y El Desarrollo Educativo*, 6(12), 60–76. <http://ride.org.mx/index.php/RIDE/article/view/226/1028>
- European Commission. (2021). *Towards competitive and clean European steel* .
- Ferguson, D. F., & Hadar, E. (2011). Optimizing the IT business supply chain utilizing cloud computing. *2011 8th International Conference and Expo on Emerging Technologies for a Smarter World, CEWIT 2011*. <https://doi.org/10.1109/CEWIT.2011.6135869>
- FISHER, R. A. (1936). THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. *Annals of Eugenics*, 7(2), 179–188. <https://doi.org/10.1111/J.1469-1809.1936.TB02137.X>

BIBLIOGRAFÍA

- Fitzpatrick, P. J. (1932). *Fitzpatrick, P. J. (1932). A comparison of the ratios of successful industrial enterprises with those of failed companies.*
- Flick U. (2007). *El diseño de la Investigación Cualitativa - Uwe Flick - Google Libros.* Ed Morata. Madrid. [https://books.google.es/books?hl=es&lr=&id=b5ojEAAAQBAJ&oi=fnd&pg=PT30&dq=Flick,+U.+\(2014\).+La+gesti%C3%B3n+de+la+calidad+en+Investigaci%C3%B3n+Cualitativa.+Espa%C3%B1a:+Ediciones+Morata&ots=fBKzgcxyFI&sig=rcRhB8zKgoNqAEX26bIUzcnvSWQ&redir_esc=y#v=onepage&q&f=false](https://books.google.es/books?hl=es&lr=&id=b5ojEAAAQBAJ&oi=fnd&pg=PT30&dq=Flick,+U.+(2014).+La+gesti%C3%B3n+de+la+calidad+en+Investigaci%C3%B3n+Cualitativa.+Espa%C3%B1a:+Ediciones+Morata&ots=fBKzgcxyFI&sig=rcRhB8zKgoNqAEX26bIUzcnvSWQ&redir_esc=y#v=onepage&q&f=false)
- F.R.S., K. P. (2010). LIII. On lines and planes of closest fit to systems of points in space. *Https://Doi.Org/10.1080/14786440109462720*, 2(11), 559–572. <https://doi.org/10.1080/14786440109462720>
- Gelsomino, L. M., Mangiaracina, R., Perego, A., & Tumino, A. (2016). Supply chain finance: a literature review. *International Journal of Physical Distribution and Logistics Management*, 46(4), 348–366. <https://doi.org/10.1108/IJPDLM-08-2014-0173/FULL/PDF>
- Giabicani, M., Le Terrier, C., Poncet, A., Guidet, B., Rigaud, J.-P., Quenot, J.-P., Mamzer, M.-F., Pugin, J., Weiss, E., Bourcier, S., Héroult, M.-C., & Payen, J.-F. (2023). Limitation of life-sustaining therapies in critically ill patients with COVID-19: a descriptive epidemiological investigation from the COVID-ICU study. *Critical Care*, 27(1). <https://doi.org/10.1186/s13054-023-04349-1>
- Giroux, G., & Cassell, C. (2011). Changing audit risk characteristics in the public client market. *Research in Accounting Regulation*, 23(2), 177–183. <https://doi.org/10.1016/J.RACREG.2011.06.009>
- González-Martín, J. M., Sánchez-Medina, A. J., & Alonso, J. B. (2018). [Optimization of the prediction of financial problems in Spanish private health companies using genetic algorithms]. *Gaceta Sanitaria*, 33(5), 462–467. <https://doi.org/10.1016/J.GACETA.2018.01.001>
- Guo, S., Li, H., An, H., Sun, Q., Hao, X., & Liu, Y. (2019). Steel product prices transmission activities in the midstream industrial chain and global markets. *Resources Policy*, 60, 56–71. <https://doi.org/10.1016/J.RESOURPOL.2018.11.014>
- Gutierrez, J. P., & Vianna, A. C. (2020a). Price effects of steel commodities on worldwide stock market returns. *The North American Journal of Economics and Finance*, 51, 100849. <https://doi.org/10.1016/J.NAJEF.2018.09.007>
- Gutierrez, J. P., & Vianna, A. C. (2020b). Price effects of steel commodities on worldwide stock market returns. *The North American Journal of Economics and Finance*, 51, 100849. <https://doi.org/10.1016/J.NAJEF.2018.09.007>

- Hassani, A., Safavi, S. R., & Hosseini, V. (2021). A comparison of light-duty vehicles' high emitters fractions obtained from an emission remote sensing campaign and emission inspection program for policy recommendation. *Environmental Pollution*, 286, 117396. <https://doi.org/10.1016/J.ENVPOL.2021.117396>
- Haugen, M. J., & Bishop, G. A. (2018). Long-Term Fuel-Specific NO_x and Particle Emission Trends for In-Use Heavy-Duty Vehicles in California. *Environmental Science and Technology*, 52(10), 6070–6076. https://doi.org/10.1021/ACS.EST.8B00621/ASSET/IMAGES/LARGE/ES-2018-00621U_0005.JPEG
- Hernán-García, M., Lineros-González, C., & Ruiz-Azarola, A. (2021). How to adapt qualitative research to confinement contexts | Cómo adaptar una investigación cualitativa a contextos de confinamiento. *Gaceta Sanitaria*, 35(3), 298–301. <https://doi.org/10.1016/j.gaceta.2020.06.007>
- Herrero, Á., Corchado, E., & Jiménez, A. (2011). Unsupervised neural models for country and political risk analysis. *Expert Systems with Applications*, 38(11), 13641–13661. <https://doi.org/10.1016/J.ESWA.2011.04.136>
- Herrero, Á., Corchado, E., Sáiz, L., & Abraham, A. (2010). DIPKIP: A CONNECTIONIST KNOWLEDGE MANAGEMENT SYSTEM TO IDENTIFY KNOWLEDGE DEFICITS IN PRACTICAL CASES. *Computational Intelligence*, 26(1), 26–56. <https://doi.org/10.1111/J.1467-8640.2009.00351.X>
- Herrero, Á., Jiménez, A., & Bayraktar, S. (2019). Hybrid unsupervised exploratory plots: A case study of analysing foreign direct investment. *Complexity*, 2019. <https://doi.org/10.1155/2019/6271017>
- 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>
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6), 417–441. <https://doi.org/10.1037/H0071325>
- Huang, R., & Ren, S. (2016). Do Credit Ratings Incorporate Business Linkages Along the Supply Chain? *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.2823323>
- Huang, S., Yang, D., Yongxin, G., & Zhang, X. (2015). Combined supervised information with PCA via discriminative component selection. *Information Processing Letters*, 115(11), 812–816. <https://doi.org/10.1016/J.IPL.2015.06.010>
- Jahani, H., Alavifard, F., Ivanov, D., & Ghasemishabankareh, B. (2019). Managing the risk of supply chain bankruptcy in supply chain network redesign.

BIBLIOGRAFÍA

- IFAC-PapersOnLine*, 52(13), 2431–2436.
<https://doi.org/10.1016/J.IFACOL.2019.11.571>
- John F. Magee, William C. Copacino, & Donald B. Rosenfield. (1973). *Modern Logistics Management: Integrating Marketing, Manufacturing and ...* - John F. Magee, William C. Copacino, Donald B. Rosenfield - Google Libros. John Wiley & Sons. New York.
https://books.google.es/books?hl=es&lr=&id=9y1owOW4d-4C&oi=fnd&pg=PA1&dq=+Modern+logistics+management&ots=KEuf-LXE60&sig=g27NlzwwdPdOELM5evReqXyYalA&redir_esc=y#v=onepage&q=Modern%20logistics%20management&f=false
- Karan, M. B., Ulucan, A., & Kaya, M. (2013). Credit risk estimation using payment history data: A comparative study of Turkish retail stores. *Central European Journal of Operations Research*, 21(2), 479–494.
<https://doi.org/10.1007/S10100-012-0242-Y/METRICS>
- Kawulich, B. B. (2005). *La observación participante como método de recolección de datos 1*. <http://www.qualitative-research.net/fqs/>
- Kelle, P., Song, J., Jin, M., Schneider, H., & Claypool, C. (2019). Evaluation of operational and environmental sustainability tradeoffs in multimodal freight transportation planning. *International Journal of Production Economics*, 209, 411–420. <https://doi.org/10.1016/J.IJPE.2018.08.011>
- Kendall, K. E., & Kendall, J. E. (2005). *Análisis y diseño de sistemas*.
- Khan, S. A. R., Yu, Z., Golpira, H., Sharif, A., & Mardani, A. (2021). A state-of-the-art review and meta-analysis on sustainable supply chain management: Future research directions. *Journal of Cleaner Production*, 278, 123357. <https://doi.org/10.1016/J.JCLEPRO.2020.123357>
- Kountcheva, R., & Kountchev, R. (2017). Sliding Recursive Hierarchical Adaptive PCA for 3D image processing. *ICIT 2017 - 8th International Conference on Information Technology, Proceedings*, 174–181. <https://doi.org/10.1109/ICITECH.2017.8079996>
- Kumar Dadsena, K., Sarmah, S. P., & Naikan, V. N. A. (2019). Risk evaluation and mitigation of sustainable road freight transport operation: a case of trucking industry. <https://doi.org/10.1080/00207543.2019.1578429>, 57(19), 6223–6245. <https://doi.org/10.1080/00207543.2019.1578429>
- Laitinen, E. K., & Suvas, A. (2013). International Applicability of Corporate Failure Risk Models Based on Financial Statement Information: Comparisons across European Countries. *Journal of Finance & Economics*, 1(3), 1–26. www.todayscience.org/jfe

- Lavalle Burguete. (2017). *Análisis financiero - Ana Consuelo Lavalle Burguete, Editorial Digital UNID - Google Libros*. <https://books.google.es/books?hl=es&lr=&id=RuE2DAAAQBAJ&oi=fnd&pg=PT2&dq=lavalle+burguete+&ots=o1sIRqDRNj&sig=OOmUHRymJMnL8tZ4hV3K0tfy7lo#v=onepage&q=lavalle%20burguete&f=false>
- Lee, L. C., Liong, C. Y., Osman, K., & Jemain, A. A. (2016). Comparison of several variants of principal component analysis (PCA) on forensic analysis of paper based on IR spectrum. *AIP Conference Proceedings*, 1750(1), 60012. <https://doi.org/10.1063/1.4954617/586637>
- León, R. A. H. (2012). El proceso de investigación científica (2a. ed.). *Editorial Universitario*, 1–132. <https://bv.unir.net:2769/es/ereader/unir/71501>
- Leontaritis, I. J., & Billings, S. A. (2007). Input-output parametric models for non-linear systems Part I: deterministic non-linear systems. *Http://Dx.Doi.Org/10.1080/0020718508961129*, 41(2), 303–328. <https://doi.org/10.1080/0020718508961129>
- Levy, D. L. (1997). Lean Production in an International Supply Chain. *MIT Sloan Management Review*. <https://sloanreview.mit.edu/article/lean-production-in-an-international-supply-chain/>
- Lin, H. (2015). Default Prediction Model for SME's: Evidence from UK Market Using Financial Ratios. *International Journal of Business and Management*, 10(2), p81. <https://doi.org/10.5539/IJBM.V10N2P81>
- Lin, Q., & Xiao, Y. (2018). Retailer credit guarantee in a supply chain with capital constraint under push & pull contract. *Computers & Industrial Engineering*, 125, 245–257. <https://doi.org/10.1016/J.CIE.2018.08.029>
- Lin, T., Horne, B. G., & Giles, C. L. (1998). How embedded memory in recurrent neural network architectures helps learning long-term temporal dependencies. *Neural Networks*, 11(5), 861–868. [https://doi.org/10.1016/S0893-6080\(98\)00018-5](https://doi.org/10.1016/S0893-6080(98)00018-5)
- Liu, W., Bai, E., Liu, L., Wei, W., Hua, G., Cheng, T. C. E., Chen, F., & Wang, S. (2017). A Framework of Sustainable Service Supply Chain Management: A Literature Review and Research Agenda. *Sustainability 2017, Vol. 9, Page 421*, 9(3), 421. <https://doi.org/10.3390/SU9030421>
- Liu, Z., Chai, T., Yu, W., & Tang, J. (2015). Multi-frequency signal modeling using empirical mode decomposition and PCA with application to mill load estimation. *Neurocomputing*, 169, 392–402. <https://doi.org/10.1016/J.NEU-COM.2014.08.087>

BIBLIOGRAFÍA

- Lyublinskaya, I., & Du, X. (2023). Annotated digital timelining: Interactive visual display for data analysis in mixed methods research. *Methods in Psychology*, 8. <https://doi.org/10.1016/j.metip.2022.100108>
- MacDonald, D., Corchado, E., & Fyfe, C. (2004). Analysing spectroscopic data using hierarchical cooperative maximum likelihood hebbian learning. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 2972, 282–291. https://doi.org/10.1007/978-3-540-24694-7_29/COVER
- Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2), 743–758. <https://doi.org/10.1016/J.EJOR.2018.10.024>
- Malanichev, A. G., & Vorobyev, P. V. (2011). Forecast of global steel prices. *Studies on Russian Economic Development*, 22(3), 304–311. <https://doi.org/10.1134/S1075700711030105/METRICS>
- Mangina, E., Narasimhan, P. K., Saffari, M., & Vlachos, I. (2020). Data analytics for sustainable global supply chains. *Journal of Cleaner Production*, 255, 120300. <https://doi.org/10.1016/J.JCLEPRO.2020.120300>
- Manu, M. V. (2019). Challenging the Status Quo: Steel Producer Case Study on the Enterprise Value for M&A. *Management and Economics Review*, 4(1), 41–56.
- Mao, J., & Jain, A. K. (1996). A self-organizing network for hyperellipsoidal clustering (HEC). *IEEE Transactions on Neural Networks*, 7(1), 16–29. <https://doi.org/10.1109/72.478389>
- Marshall C., & Rossman G.B. (2016). *Designing Qualitative Research - Catherine Marshall, Gretchen B. Rossman - Google Libros*. Ed SAGE. Los Angeles CA. [https://books.google.es/books?hl=es&lr=&id=zncBQAAQBAJ&oi=fnd&pg=PP1&dq=Marshall,+%26+Rossman,+GB+\(1989\).+Designing+qualitative+research.USA+Los+Angeles.+Ed:+SAGE+6th+edition&ots=Lgm-boIVy5U&sig=u87f1KTbrjG3Wt5wxiUzzQcjE-8&redir_esc=y#v=onepage&q&f=false](https://books.google.es/books?hl=es&lr=&id=zncBQAAQBAJ&oi=fnd&pg=PP1&dq=Marshall,+%26+Rossman,+GB+(1989).+Designing+qualitative+research.USA+Los+Angeles.+Ed:+SAGE+6th+edition&ots=Lgm-boIVy5U&sig=u87f1KTbrjG3Wt5wxiUzzQcjE-8&redir_esc=y#v=onepage&q&f=false)
- Martínez Candil, Ignacio. (2015). *UF2120 : Diseño de encuestas y cuestionarios de investigación*. https://books.google.com/books/about/UF2120_Dise%C3%B1o_de_encuestas_y_cuestionar.html?hl=es&id=LGNWDwAAQBAJ
- Menezes, J. M. P., & Barreto, G. A. (2008). Long-term time series prediction with the NARX network: An empirical evaluation. *Neurocomputing*, 71(16–18), 3335–3343. <https://doi.org/10.1016/J.NEUCOM.2008.01.030>

- Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). DEFINING SUPPLY CHAIN MANAGEMENT. *Journal of Business Logistics*, 22(2), 1–25. <https://doi.org/10.1002/J.2158-1592.2001.TB00001.X>
- Mervin, R. F. (1930). *The Social Secretary of Detroit 1930*.
- Meyer, T. (2020). Decarbonizing road freight transportation – A bibliometric and network analysis. *Transportation Research Part D: Transport and Environment*, 89, 102619. <https://doi.org/10.1016/J.TRD.2020.102619>
- Min, S., Zacharia, Z. G., & Smith, C. D. (2019). Defining Supply Chain Management: In the Past, Present, and Future. *Journal of Business Logistics*, 40(1), 44–55. <https://doi.org/10.1111/JBL.12201>
- Ministerio para la Transición Ecologica. (2021). *NVENTARIO NACIONAL DE EMISIONES A LA ATMÓSFERA EMISIONES DE GASES DE EFECTO INVERNADERO*.
- Molina, M., Castro, E., Molina, J. L., & Castro, E. (2011). Un acercamiento a la investigación de diseño a través de los experimentos de enseñanza. *Enseñanza de Las Ciencias: Revista de Investigación y Experiencias Didácticas*, 29(1), 75–88. <https://raco.cat/index.php/Ensenanza/article/view/243824>
- Montero, I., & León, O. G. (2005). Sistema de clasificación del método en los informes de investigación en Psicología. *International Journal of Clinical and Health Psychology*, 5(1), 115–127. <https://www.redalyc.org/articulo.oa?id=33701007>
- Montreuil, B. (2011). Toward a Physical Internet: meeting the global logistics sustainability grand challenge. *Logistics Research 2011 3:2*, 3(2), 71–87. <https://doi.org/10.1007/S12159-011-0045-X>
- Nandi, A., Sengupta, P. P., & Dutta, A. (2019). Diagnosing the Financial Distress in Oil Drilling and Exploration Sector of India through Discriminant Analysis. <https://doi.org/10.1177/0972262919862920>, 23(4), 364–373. <https://doi.org/10.1177/0972262919862920>
- Negocios, D. V. Z.-R. I. y, & 2019, undefined. (n.d.). La importancia de la estadística aplicada para la toma de decisiones en Marketing. *Scielo.Org.Bo*. Retrieved June 1, 2023, from http://www.scielo.org/bo/scielo.php?pid=S2521-27372019000200004&script=sci_arttext
- OECD/IFT. (2016). *International Transport Forum: Global dialogue for better transport The Carbon Footprint of Global Trade Tackling Emissions from International Freight Transport*.
- Omura, A., Todorova, N., Li, B., & Chung, R. (2016). Steel scrap and equity market in Japan. *Resources Policy*, 47, 115–124. <https://doi.org/10.1016/J.RESOURPOL.2016.01.001>

BIBLIOGRAFÍA

- Panasiyk, D., Laratte, B., & Remy, S. (2016). Steel Stock Analysis in Europe from 1945 to 2013. *Procedia CIRP*, 48, 348–351. <https://doi.org/10.1016/J.PROCIR.2016.04.084>
- Panigrahi, C.) A. (2019). *Validity of Altman's 'Z' Score Model in Predicting Financial Distress of Pharmaceutical Companies*. <https://papers.ssrn.com/abstract=3326312>
- Parada Gutiérrez, O. (2009). A multi-criterion focus for decision-making in inventory management | Un enfoque multicriterio para la toma de decisiones en la gestión de inventarios. *Cuadernos de Administracion*, 22(38), 169–187.
- Partal Ureña A., Moreno Bonilla F., Cano Rodriguez M., & Gomez Fernandez-Aguado C. (2016). Introducción a las finanzas empresariales. *Ed Piramide*.
- Pau Cos, J., & De Navascués, R. (2001). *Pau Cos-Manual-de-Logistica-Integral libro.pdf*. https://books.google.com/books/about/Manual_de_log%C3%ADstica_integral.html?hl=es&id=dxTImJ4ipCMC
- Pauliuk, S., Wang, T., & Müller, D. B. (2013). Steel all over the world: Estimating in-use stocks of iron for 200 countries. *Resources, Conservation and Recycling*, 71, 22–30. <https://doi.org/10.1016/J.RESCONREC.2012.11.008>
- Pernestål, A., Engholm, A., Bemler, M., & Gidofalvi, G. (2020). How Will Digitalization Change Road Freight Transport? Scenarios Tested in Sweden. *Sustainability* 2021, Vol. 13, Page 304, 13(1), 304. <https://doi.org/10.3390/SU13010304>
- Porta-Etessam, J., López-Gil, A., Matías-Guiu, J., & Fernández, C. (2010). Validation of a diagnostic questionnaire for migraine adapted for a computer assisted telephone interviewing | Validación de un cuestionario diagnóstico para migraña adaptado para encuesta telefónica asistida por ordenador. *Neurologia*, 25(3), 163–167. [https://doi.org/10.1016/S0213-4853\(10\)70004-9](https://doi.org/10.1016/S0213-4853(10)70004-9)
- Pozuelo Campillo, J., Labatut Serer, G., & Veres Ferrer, E. (2013). Validez de la información financiera en los procesos de insolvencia. Un estudio de la pequeña empresa española. *Cuadernos de Economía y Dirección de La Empresa*, 16(1), 29–40. <https://doi.org/10.1016/J.CEDE.2012.05.001>
- Prida, B., & Gutiérrez, G. (1996). Supply management: From purchasing to external factory management. *Production and Inventory Management Journal*, 37(4), 38–43.
- Puyati, W., & Walairacht, A. (2008). Efficiency improvement for unconstrained face recognition by weighting probability values of modular PCA and Wavelet PCA. *International Conference on Advanced Communication*

Technology, ICACT, 2, 1449–1453.
<https://doi.org/10.1109/ICACT.2008.4494037>

Ramírez, N. de A., Verde, A. R. M., & Fernández, N. L. (2010). Dos formas de orientar la investigación en la educación de postgrado: lo cuantitativo y lo cualitativo. *Pedagogía Universitaria, 15*(5), 13–29.
<https://go.gale.com/ps/i.do?p=IFME&sw=w&issn=16094808&v=2.1&it=r&id=GALE%7CA466617343&sid=googleScholar&linkaccess=fulltext>

Reyes, P., Raisinghani, M. S., & Singh, M. (2014). Global Supply Chain Management in the Telecommunications Industry: The Role of Information Technology in Integration of Supply Chain Entities. *Http://Dx.Doi.Org/10.1080/1097198X.2002.10856325, 5*(2), 48–67.
<https://doi.org/10.1080/1097198X.2002.10856325>

Robin, Valen, F. S., Nomleni, A., Turnip, G., Luhulima, M. Y., & Insani, L. (2023). Presence of non-native freshwater fish in Indonesia: A review - Risk and ecological impacts. *AAJ Bioflux, 16*(1), 3384–3397.

Rodríguez Liboreiro, P., & Rodríguez Liboreiro, P. (2019). COMPETENCIA, RENDIMIENTOS CRECIENTES Y EXCESO DE CAPACIDAD: LA INDUSTRIA SIDÉRGICA MUNDIAL (2000-2014). *Cuadernos de Economía, 38*(76), 137–171. <https://doi.org/10.15446/CUAD.ECON.V37N76.61257>

Rossi, T., Pozzi, R., Pirovano, G., Cigolini, R., & Pero, M. (2020). A new logistics model for increasing economic sustainability of perishable food supply chains through intermodal transportation. *Https://Doi.Org/10.1080/13675567.2020.1758047, 24*(4), 346–363.
<https://doi.org/10.1080/13675567.2020.1758047>

Rubén, N., & Reyes, M. (2012). Reseña metodológica sobre los grupos focales. *Diá-Logos, 9*, 47–53. <https://doi.org/10.5377/DIALOGOS.V1I9.15651>

Rudrakumar, S., Varshney, N., & Taylor, R. D. (2023). Medical student perspectives on substance misuse education in the medical undergraduate programme: a grounded theory approach. *BMC Medical Education, 23*(1).
<https://doi.org/10.1186/s12909-023-04145-z>

Sammon, J. W. (1969). A Nonlinear Mapping for Data Structure Analysis. *IEEE Transactions on Computers, C-18*(5), 401–409. <https://doi.org/10.1109/T-C.1969.222678>

Sandoval, Loenel Luis, & Abreu. (2008). *sandoval (Author) AND SME companies (Topic) - 1 - All Databases.* <https://www.webof-science.com/wos/alldb/summary/247ed57c-cb40-49e1-a3ce-4ff65d7d30e5-81f3c2f1/relevance/1>

BIBLIOGRAFÍA

- Santos, D. de A., Quelhas, O. L. G., Gomes, C. F. S., & Filho, J. R. de F. (2021). Theoretical Proposal for an Integrated Sustainability Performance Measurement System in the Supply Chain. *Frontiers in Sustainability*, 2, 720763. <https://doi.org/10.3389/FRSUS.2021.720763/BIBTEX>
- Santos, D. de A., Quelhas, O. L. G., Gomes, C. F. S., Zotes, L. P., França, S. L. B., de Souza, G. V. P., de Araújo, R. A., & Santos, S. da S. C. (2020). Proposal for a Maturity Model in Sustainability in the Supply Chain. *Sustainability* 2020, Vol. 12, Page 9655, 12(22), 9655. <https://doi.org/10.3390/SU12229655>
- Schmidt B., Palazzi A., & Piccinini C.A. (2020). Potencialidades y desafíos para la recogida de datos en el contexto de la pandemia de COVID-19*. *Revista Família, Ciclos de Vida e Saúde No Contexto Social*. <https://doi.org/10.18554/refacs.v8i4.4877>
- Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10(5), 1299–1319. <https://doi.org/10.1162/089976698300017467>
- Shankar, R., Pathak, D. K., & Choudhary, D. (2019). Decarbonizing freight transportation: An integrated EFA-TISM approach to model enablers of dedicated freight corridors. *Technological Forecasting and Social Change*, 143, 85–100. <https://doi.org/10.1016/J.TECHFORE.2019.03.010>
- Shojaei, P., & Haeri, S. A. S. (2019). Development of supply chain risk management approaches for construction projects: A grounded theory approach. *Computers & Industrial Engineering*, 128, 837–850. <https://doi.org/10.1016/J.CIE.2018.11.045>
- Siddiqui, S. A. (2012). Business Bankruptcy Prediction Models: A Significant Study of the Altman's Z-Score Model. *SSRN Electronic Journal*. <https://doi.org/10.2139/SSRN.2128475>
- Smith, R. F. , W. A. H. (1935). Changes in the financial structure of unsuccessful industrial corporations. *Illinois University*. <https://cir.nii.ac.jp/crid/1130282273231139712>
- Sociedad, U. Y., Belén, N., & Granda, B. (2020). Análisis financiero: factor sustancial para la toma de decisiones en una empresa del sector comercial. *Revista Universidad y Sociedad*, 12(3), 129–134. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2218-36202020000300129&lng=es&nrm=iso&tlng=es
- Soysal, M., Bloemhof-Ruwaard, J. M., Haijema, R., & van der Vorst, J. G. A. J. (2018). Modeling a green inventory routing problem for perishable products with horizontal collaboration. *Computers & Operations Research*, 89, 168–182. <https://doi.org/10.1016/J.COR.2016.02.003>

- Srebro, B., Mavrenski, B., Arsić, V. B., Knežević, S., Milašinović, M., & Travica, J. (2021). Bankruptcy Risk Prediction in Ensuring the Sustainable Operation of Agriculture Companies. *Sustainability 2021, Vol. 13, Page 7712, 13(14)*, 7712. <https://doi.org/10.3390/SU13147712>
- Sun, Y., Liu, Z., & Yang, H. (2018). How Does Suppliers' Fairness Affect the Relationship Quality of Agricultural Product Supply Chains? *Journal of Food Quality, 2018*. <https://doi.org/10.1155/2018/9313068>
- Tob-Ogu, A., Kumar, N., Cullen, J., & Ballantyne, E. E. F. (2018). Sustainability Intervention Mechanisms for Managing Road Freight Transport Externalities: A Systematic Literature Review. *Sustainability 2018, Vol. 10, Page 1923, 10(6)*, 1923. <https://doi.org/10.3390/SU10061923>
- Torres, M., Paz, K., & Salazar, F. G. (2019). Métodos de recolección de datos para una investigación. *UDG Virtual*. <http://148.202.167.116:8080/xmlui/handle/123456789/2817>
- Truong, H. Q., & Hara, Y. (2018). Supply chain risk management: Manufacturing- and service-oriented firms. *Journal of Manufacturing Technology Management, 29(2)*, 218–239. <https://doi.org/10.1108/JMTM-07-2017-0145/FULL/PDF>
- Tu, Q., Lu, J. F., Yuan, B., Tang, J. B., & Yang, J. Y. (2012). Density-based hierarchical clustering for streaming data. *Pattern Recognition Letters, 33(5)*, 641–645. <https://doi.org/10.1016/J.PATREC.2011.11.022>
- van Bergen, M., Steeman, M., Reindorp, M., & Gelsomino, L. (2019). Supply chain finance schemes in the procurement of agricultural products. *Journal of Purchasing and Supply Management, 25(2)*, 172–184. <https://doi.org/10.1016/J.PURSUP.2018.08.003>
- Wang, T., Xu, H., Han, J., Elbouchikhi, E., & Benbouzid, M. E. H. (2015). Cascaded H-Bridge Multilevel Inverter System Fault Diagnosis Using a PCA and Multiclass Relevance Vector Machine Approach. *IEEE Transactions on Power Electronics, 30(12)*, 7006–7018. <https://doi.org/10.1109/TPEL.2015.2393373>
- Wang, Y., Wallace, S. W., Shen, B., & Choi, T. M. (2015). Service supply chain management: A review of operational models. *European Journal of Operational Research, 247(3)*, 685–698. <https://doi.org/10.1016/J.EJOR.2015.05.053>
- Wao, H., Kersaint, G., Smith, C. A. S., Campbell-Montalvo, R., Puccia, E., Skvoretz, J., Martin, J. P., Lee, R., & MacDonald, G. (2023). Examining how social networks influence women and under-represented minority stu-

BIBLIOGRAFÍA

- dents' pursuit of engineering in university: when, who, and how? *International Journal of STEM Education*, 10(1). <https://doi.org/10.1186/s40594-023-00415-w>
- Wong, E. Y. C., Tai, A. H., & So, S. (2020). Container drayage modelling with graph theory-based road connectivity assessment for sustainable freight transportation in new development area. *Computers & Industrial Engineering*, 149, 106810. <https://doi.org/10.1016/J.CIE.2020.106810>
- Xing Y.U., & Zhongkai WAN M..D. (2020). *SUPPLY CHAIN FINANCING MODEL UNDER A NEW MECHANISM OF BANKRUPTCY GUARANTEE*.
- Xu R., & Wunsch D.C. (2009). Clustering. hoboken. *IEEE Press*.
- Xu, X., Sun, Y., & Hua, Z. (2010). Reducing the probability of bankruptcy through supply chain coordination. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 40(2), 201–215. <https://doi.org/10.1109/TSMCC.2009.2031092>
- Yan, N., & He, X. (2020). Optimal trade credit with deferred payment and multiple decision attributes in supply chain finance. *Computers & Industrial Engineering*, 147, 106627. <https://doi.org/10.1016/J.CIE.2020.106627>
- Yang, S. A., Birge, J. R., & Parker, R. P. (2015). The Supply Chain Effects of Bankruptcy. <https://doi.org/10.1287/Mnsc.2014.2079>, 61(10), 2320–2338. <https://doi.org/10.1287/MNSC.2014.2079>
- Yartu, M., Cambra, C., Navarro, M., Rad, C., Arroyo, Á., & Herrero, Á. (2022). Humidity forecasting in a potato plantation using time-series neural models. *Journal of Computational Science*, 59, 101547. <https://doi.org/10.1016/J.JOCS.2021.101547>
- Ye, H., Hu, X., & Xie, M. (2019). *Research on Supply Chain Finance Model based on Agricultural Logistics Park Information Platform*. 226–230. <https://doi.org/10.2991/SSMI-18.2019.40>
- Yu, K. A., McDonald, B. C., & Harley, R. A. (2021). Evaluation of Nitrogen Oxide Emission Inventories and Trends for On-Road Gasoline and Diesel Vehicles. *Environmental Science and Technology*, 55(10), 6655–6664. <https://doi.org/10.1021/ACS.EST.1C00586>/ASSET/IMAGES/LARGE/ES1C00586_0006.JPEG
- Zeji, B., & Id, P. (2018). Review of Research into Enterprise Bankruptcy Prediction in Selected Central and Eastern European Countries. *International Journal of Financial Studies* 2018, Vol. 6, Page 60, 6(3), 60. <https://doi.org/10.3390/IJFS6030060>

- Zhang, T., Zhang, C. Y., & Pei, Q. (2019). Misconception of providing supply chain finance: Its stabilising role. *International Journal of Production Economics*, 213, 175–184. <https://doi.org/10.1016/J.IJPE.2019.03.008>
- Zhuang, W., Ye, Y., Chen, Y., & Li, T. (2012). Ensemble clustering for internet security applications. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 42(6), 1784–1796. <https://doi.org/10.1109/TSMCC.2012.2222025>
- Zola P, & Carpita M. (2016). Statistica & Applicazioni : XIV, 1, 2016. *Statistica & Applicazioni*, XIV(1), 1–103.
- Zurutuza, U., Ezpeleta, E., Herrero, Á., & Corchado, E. (2011). Visualization of misuse-based intrusion detection: Application to honeynet data. *Advances in Intelligent and Soft Computing*, 87, 561–570. https://doi.org/10.1007/978-3-642-19644-7_59/COVER

ANEXOS

Article

An Intelligent Visualisation Tool to Analyse the Sustainability of Road Transportation

Carlos Alonso de Armiño ¹, Daniel Urda ^{2,*}, Roberto Alcalde ³, Santiago García ¹ and Álvaro Herrero ²

¹ Departamento de Ingeniería de Organización, Escuela Politécnica Superior, Universidad de Burgos, Av. Cantabria S/N, 09006 Burgos, Spain; caap@ubu.es (C.A.d.A.); lgpineda@ubu.es (S.G.)

² Grupo de Inteligencia Computacional Aplicada-GICAP, Departamento de Ingeniería Informática, Escuela Politécnica Superior, Universidad de Burgos, Av. Cantabria S/N, 09006 Burgos, Spain; ahcosio@ubu.es

³ Departamento de Economía y Administración de Empresas, Facultad de Ciencias Económicas y Empresariales, Universidad de Burgos, Pza. de la Infanta D^a. Elena, S/N, 09001 Burgos, Spain; radelgado@ubu.es

* Correspondence: durda@ubu.es

Abstract: Road transport is an integral part of economic activity and is therefore essential for its development. On the downside, it accounts for 30% of the world's GHG emissions, almost a third of which correspond to the transport of freight in heavy goods vehicles by road. Additionally, means of transport are still evolving technically and are subject to ever more demanding regulations, which aim to reduce their emissions. In order to analyse the sustainability of this activity, this study proposes the application of novel Artificial Intelligence techniques (more specifically, Machine Learning). In this research, the use of Hybrid Unsupervised Exploratory Plots is broadened with new Exploratory Projection Pursuit techniques. These, together with clustering techniques, form an intelligent visualisation tool that allows knowledge to be obtained from a previously unknown dataset. The proposal is tested with a large dataset from the official survey for road transport in Spain, which was conducted over a period of 7 years. The results obtained are interesting and provide encouraging evidence for the use of this tool as a means of intelligent analysis on the subject of developments in the sustainability of road transportation.

Keywords: artificial intelligence; unsupervised machine learning; exploratory projection pursuit; clustering; road transportation; transport sustainability; age of transport means



Citation: Alonso de Armiño, C.; Urda, D.; Alcalde, R.; García, S.; Herrero, Á. An Intelligent Visualisation Tool to Analyse the Sustainability of Road Transportation. *Sustainability* **2022**, *14*, 777. <https://doi.org/10.3390/su14020777>

Academic Editor: Marc A. Rosen

Received: 7 October 2021

Accepted: 7 January 2022

Published: 11 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction & Previous Work

The 17 Sustainable Development Goals of the UN's master plan include a determined line of action to halt global warming, with a direct link to reducing CO₂ emissions. The United Nations Framework Convention on Climate Change (UNFCCC) and the Kyoto Protocol, both signed in 2002, point in the same direction.

Moreover, road transport is consubstantial to economic activity and, thus, essential to the future of our civilisation. Its counterpart accounted for 30% [1] of global CO₂ emissions in 2015, and 29% of the 314,529 kilo-tonnes of GHG emitted in 2019 in Spain, with a total of 91,372 kilo-tonnes of CO₂-equivalent emissions. In turn, it is estimated at a global and an EU level that nearly a quarter of these emissions correspond to freight transportation in heavy goods vehicles and buses [2,3], with this being one of the areas that has recorded sustained growth in recent years. Meanwhile, other modes of transport are progressively reducing their share, with an estimated increase of four times their current emissions by 2050 if no measures are adopted on the matter. In view of these data and forecasts, part of the focus of scientific studies and lines of action for governance have centred on this area, with progress being made on the development of proposals to regulate emissions for the equipment that performs this activity, which clearly accounts for 6% of the European Union's total GHG emissions [4].

1.1. Sustainability in Transportation

The systematic review of intervention systems for the sustainability of road freight transport revealed that the most obvious line of action is the application of emerging technologies combined with decisive energy policies [5]. Other complementary lines of action focus on the dedication and adaptation of infrastructures to the effective development of this activity and the promotion of intermodality in the efficient management of supply chains [6]. In many cases, these are centred on a firm commitment to synchro-modality [7] and are even inclined to measure the sustainability of transport on the basis of its intermodality [8]. A distinction is also made between the effectiveness of different measures in favour of sustainability in collection and delivery transport, with improved applicability of collaborative techniques supported by information systems, as opposed to long-distance transport, where the results of technological optimisation in terms of transport equipment consumption and the suitability of infrastructures are more convincing [9]. Some models, reflecting the close connection between the economy and transportation, are based on the analysis of an economic sustainability approach to logistics models [10] and, along the same lines, some of them focus on the analysis of certain goods based on their production structure and the distribution of goods by road, analysing their entire activity [11]. There are also studies aimed at optimizing the transport process based on streamlining some of its sub-processes such as loading [12,13].

To this extent, we could say that we are following a classic perspective in the search for sustainability, which Corlu [14] points to in his review of the state of the art on optimising energy consumption in transport processes, as a combination of three main lines of action for road transport: (i) improving the load factor, understood as a reducer of emissions based on maximum occupancy of the means of transport, (ii) the use of collaborative techniques to optimise the allocation of means of transport and (iii) defining and monitoring sustainability objectives in the development of transport operations. Nevertheless, the most remarkable aspect of the ensemble is that all the models and studies point, in short, to a common underlying objective; the optimisation of energy consumed in the process of freight transport operations.

A line of research directly associated with the central objective of energy improvement is the study of emissions from transport equipment based on their age. Hassani et al. [15] determined that emissions in light vehicles can be up to five times higher depending on their age. Owing to this finding, this idea (sustainability based on the vehicles age) is given special relevance in this study.

These results are not arbitrary; instead, they are the result of the implementation of policies adopted by the vehicle manufacturing sector. Back in 2002, Ang-Olson and Schroer [16] determined that the proper implementation of technical solutions in the production of heavy goods road transport vehicles could lead to a reduction of more than 11 billion litres of annual fuel consumption within a period of 10 years, associated with a decrease of 8.3 million tonnes of greenhouse gas emissions in the United States (US) alone. Since then, the governance mechanisms of the US and the EU have continued to impose mandates on the production of transport vehicles, aimed at reducing consumption and emissions [17]. As a result of these mandates, the so-called Euro emission standards have been progressively developed, with successively more demanding requirements on consumption and emissions of transport elements, as shown in Table 1, on carbon monoxide (CO), nitrogen oxides (NOX) and particulate matter (PM).

If we focus specifically on heavy goods vehicles, Haugen and Bishop [19] establish two complete vehicle emission samplings at loading and unloading points with significant freight movements, finally determining that a reduction from an average age of 7.8 to 6 years for these vehicles results in a significant reduction in emissions of up to 87% of suspended particulate matter. A subsequent study [20] also determined that the reduction in harmful NOX emissions decreased from 38 g to 9 g per kg of diesel fuel consumed by heavy goods vehicles from 2005 to 2020, representing a reduction of 76.3% in vehicles

produced between these dates, i.e., an average reduction of 5% in emissions for each year of vehicle production during this period.

Table 1. Emission standards adopted by the EU for diesel cars and heavy goods vehicles. Source: Own elaboration on the data of [18].

Emission Standards for Diesel Cars				
Standard	Date	CO g/Km	NOX g/Km	PM g/Km
Euro 4	2005	0.50	0.30	0.025
Euro 5	2010	0.50	0.23	0.005
Euro 6	2015	0.50	0.17	0.005
Emission Standards for Heavy Goods Vehicles				
Standard	Date	CO g/KWh	NOX g/KWh	PM g/KWh
Euro IV	2005	1.50	3.50	0.020
Euro V	2008	1.50	2.00	0.020
Euro VI	2013	1.50	0.40	0.010

With possible differences in the quantification of emission improvements, there is one clear conclusion; heavy goods vehicles are a significant contributor to the emission of harmful gases, and the age of said vehicles is also a determining factor in their efficiency and sustainability.

1.2. Previous Work on Digitisation

New perspectives emerge when incorporating the digitalisation approach to road transportation. A study by a panel of 52 experts [21], highlights the value of process automation, the complete collection of data on digital information and the basis for the application of artificial intelligence for adequate performance and planning. Taking a step further in this direction, and with a clear focus on improving the management of supply chains, a study was carried out on the data collected by the EU in the Permanent Surveys of Goods Transport by Road [22]. With this purpose, a modelling was carried out on this activity, which was compiled by EUROSTAT for its member countries between 2011 and 2014. It was carried out under a Horizontal Collaboration model [23] and analysed the improvements that would have resulted from the application of Pooling techniques, similar to optimised freight groupings, and an implementation model of the so-called Physical Internet [24]. This was done by developing computer models that simulate each option, which are used to estimate a calculation of emissions. The end result is a clear advantage for the currently purely theoretical physical internet model.

More specifically, previous contributions have been made concerning the use of Artificial Intelligence (AI) in general and Machine Learning (ML) in particular, in order to address sustainability issues in road freight transport. In [25], an integrated fuzzy failure mode and effects analysis approach was proposed for the selection of risk mitigation strategies in track and trace tasks in the industry, which aimed to help managers choose a strategy considering the criticality of the risks under a limited budget. In addition, [26] presented a novel container transport optimisation model that incorporates the road network along with connectivity metrics, aiming to minimise total trip distance, truck fuel cost, container rental cost, and container movements between multiple consignees and haulers. More recently, [27] proposed a novel approach to predict the profit margin, on a customer basis, in the sustainable road freight transport sector by combining different ML methods. This helps managers to obtain useful information on strategic and sustainable development perspectives.

Differing from this previous study, the authors of this article have applied different ML techniques [28] to the same data family (see Section 2). Significant results have been obtained with respect to their link with the economic activity cycles. The economic turning points that occurred during the Great Recession in Spain have also been identified, based exclusively on various clustering techniques from the road freight transport dataset. A

strong recession was observed until the second quarter of 2012, followed by a severe depression during the following period until the second quarter of 2013, which was then followed by a gradual recovery until in the second quarter of 2015, where a clear phase of economic growth emerged.

The dataset showing the average fleet age of heavy goods road vehicles never ceased to increase at any given moment during the period in question (although it increased at different rates). It almost exactly coincided at the points of economic inflection with the quartile (Q) distribution of the data, as can be seen in Figure 1. This accelerated or sustained growth corresponds to nothing other than a pattern of over-amortisation of the means of production in times of economic recession, as a measure to protect the profitability of its economic activity.



Figure 1. Transport fleet age data series, divided into quartiles and related to the phases determined for the Great Depression. Source: Own elaboration.

We have decided to focus our studies on this data series grouped into quartiles owing to the following reasons: Firstly, it is specifically linked to the sustainability of transport activity; and secondly, its distribution of data by quartiles coincides with the economic phases. Furthermore, the study of this series, which is key to the interpretation of the sustainability of the fleet of freight transport vehicles, was neglected from the evolution and study of the rest of the series that reflected specific increases or decreases in the depression phase, which makes it advisable to apply complementary analysis techniques to it.

To address this problem, pioneer visualisation tools based on ML are put forward in this study. More specifically, clustering and Exploratory Projection Pursuit (EPP) techniques have been combined for the first time, under the frame of Hybrid Unsupervised Exploratory Plots (HUEPs), to support the visual analysis of sustainability data regarding road transportation. Likewise, the original formulation of HUEPs is extended, as well as new projection techniques being proposed and validated.

Thus far, researchers have widely studied clustering and EPP methods, concluding some of them that projection methods are not a useful in order to reduce the dimensionality of data for a following clustering [29,30]. Although this statement may be true in some cases, some other combinations of such methods have been previously proposed, different from this sequential application of methods. That is the case of [31,32], where the output of clustering methods (i.e., the assigned cluster for each data instance) is added to the projection obtained by EPP methods, that could be 2D or 3D. On the other hand, other authors have proposed [33,34] the simultaneous application of clustering and dimensionality-reduction methods. As opposed to these previous ideas, HUEPs have been recently proposed for the combination of clustering and projection methods, being independently applied.

The remaining sections of this article are organised as follows: the methods employed, together with the data on which they are applied, are described in Section 2. The results obtained in the experimental study are presented in Section 3, and the main conclusions in relation to these, as well as some proposals for future work, are presented in Section 4.

2. Materials and Methods

As previously stated, road transportation activity was researched in this study, with a specific focus on its sustainability. This was done by analysing a dataset described in Section 2.1 with the novel techniques that are presented in Section 2.2.

2.1. Dataset

Data were retrieved from two different sources:

- The Ministry of Transport, Mobility and Urban Agenda (*Ministerio de Transportes, Movilidad y Agenda Urbana*) of Spain, through its General Sub-Directorate of Economic Studies and Statistics.
- The European Road Freight Transport survey (ERFT). This survey relates to the activity of heavy goods vehicles licenced in Spain for the transport of goods. It has a sufficiently high sampling level to be of statistical representativeness for each Autonomous Region, in order to measure their transport operations. With this aim, the survey registers the movement of a single class of goods, from a departure point to a destination. The research was conducted in accordance with the corresponding regulation [35] and its subsequent revision [36]. The total number of records included on that database was 1,932,671 that has a sampling representativeness of 1,259,938,252 transport operations.

Data from between 2011 and 2017 were used. All the data represented quarterly levels of aggregation, which therefore included each variable in the study, a total of 28 values. The variables under consideration were:

- Transportation costs (B): based on a 100 percent increase above the average yearly prices in 2000, as determined by the Ministry of Development's quarterly research studies.
- Fuel costs in Spain: quarterly midpoints weighted in centimes of a euro, as indicated by the data gathered by the Ministry of Development.
- Fuel costs in the EU: quarterly midpoints weighted in centimes of a euro, as indicated by the data gathered by the Ministry of Development.
- Number of tons transported (A, B, C): weight of transported goods.
- Completed trips (A, B, C): number of transport operations and empty distance.
- Empty distance (A, B): kilometres travelled without goods.
- Maximum load for transport operations (A, B): upper weight limit for completed trips in tons.
- Maximum load for empty distance (A, B): upper weight limit for empty distance covered in tons.
- Haulage distance (A, B): kilometres travelled.
- Empty haulage distance (A, B): kilometres travelled without goods.
- Quantity of vehicles represented (A): number of vehicles represented.
- Represented load capacity (A): upper load limit of the represented vehicles.
- Tons-kms (A, B, C): total tons transported, and distance covered in each haulage operation.
- Average fleet age (A, B): average amount of years elapsed since the registration of the vehicles. As previously indicated in Section 1.1, this is an important data finding in regard to sustainability. Owing to this, it is also used in the glyph metaphor.
- Average fleet age for empty distance (A, B): average amount of years elapsed since the registration of the vehicles travelling without goods.

The data series with assigned letters were sub-divided as according:

- (A) Type of transport: (A1) All transport; (A2) Own transport; (A3) Hire or reward.
- (B) Distance range: (B1) All distances; (B2) < 50 km; (B3) 51–100 km; (B4) 101–200 km; (B5) 201–300 km; (B6) > 300 km.
- (C) Geographic catchment: (C1) All catchments; (C2) Municipal; (C3) Regional; (C4) National; (C5) Importation; (C6) Exportation; (C7) Cabotage.

113 transport data series were calculated, with the values of the 28 previously indicated quarters in each one.

As a result, a dataset with high dimensionality is required to be analysed in order to investigate the sustainability of road transportation.

2.2. Hybrid Unsupervised Exploratory Plots

Hybrid Unsupervised Exploratory Plots (HUEPs) [37] have been recently proposed as a new visualisation tool to combine the outputs of Exploratory Projection Pursuit (EPP) and clustering methods in a novel and informative way. To address the well-known “curse of dimensionality” challenge and advancing in descriptive data analysis, both EPP and clustering methods are independently applied, and their outputs combined in a new way. In particular, 3 EPP methods were put forward, commonly known as Principal Component Analysis (PCA), Maximum Likelihood Hebbian Learning (MLHL), and Cooperative MLHL (CMLHL). There are different ways of implementing such methods; in the original formulation of HUEPs, they were implemented as Artificial Neural Networks.

Additionally, an extension to this study is included to improve the visualisation capability of HUEPs. In order to generate the displays, each original x vector (from the input space) is processed as follows:

1. 2D projection of the vector is obtained by the applied EPP method (y_1^{EPP}, y_2^{EPP}).
2. The output of the clustering method (i.e., the assigned cluster number) is calculated (y_c).
3. The two previous outputs are combined in a 3D vector that is located in the output space (y_1, y_2, y_3).
4. Optionally, further information (sustainability data in the present study) is added to the visualisation by using the glyph metaphor.

Originally, HUEPs were conceived as a new way of intuitively displaying data by applying one partitional (k-means) or one hierarchical (agglomerative) clustering method together with one EPP method. As an evolution of this initial proposal, this study validates the incorporation of complementary and well-known display methods, namely Kernel-PCA (KPCA) [38] and Sammon Mapping (SM) [39].

KPCA is a non-linear extension of conventional PCA that takes the majority of kernel functions in order to obtain more interesting projections of data by extracting non-linear principal components while keeping the computation cost at a reasonable level. On the other hand, SM was proposed as a special case of the distance-based metric Multidimensional Scaling family, being itself one of the first manifold learning proposals. Furthermore, SM can be considered as the first proposed nonlinear manifold learning method. These non-linear EPP methods are proposed for the first time under the frame of HUEPs as being ones of the main non-linear EPP methods. They are analysed in this study and validated with the data previously described.

2.3. Methodology

In order to validate the proposed application of HUEPs in the present work, visualizations have been obtained by combining the projections of EPP methods (PCA, MLHL, CMLHL, KPCA, and SM) with the output of clustering methods (k -means and agglomerative). Experiments have been performed tuning each one of the methods with the following parameter values.

PCA

- Number of principal components to be obtained: 2.

MLHL

- Number of projected dimensions to be obtained: 2.
- Learning rate: [0.01, 0.05].
- p parameter: [1, 2].

CMLHL

- Number of projected dimensions to be obtained: 2.
- Learning rate: [0.01, 0.05].
- p parameter: [1, 2].
- tau parameter: [0.00000001, 0.01].

KPCA

- Number of projected dimensions to be obtained: 2, 3.
- Kernel: linear, polynomial, gaussian.

SM

- Number of projected dimensions to be obtained: 2, 3.
- k -means.
- Number of clusters: 2, 3, 4, 6, 8.
- Distance: sqEuclidean, Cityblock, Cosine, Correlation.

Agglomerative

- Number of clusters (cutoff): 2, 3, 4, 6, 8.
- Distance: Euclidean, sEuclidean, sqEuclidean, Cityblock, Hamming, Jaccard, Minkowski, Chebychev, Spearman, Cosine, Correlation.
- Linkage: Average, Centroid, Complete, Median, Single, Ward, Weighted.

3. Results

The HUEP displays obtained are shown in this section. Firstly, Figure 2 shows the HUEP display obtained by combining agglomerative clustering with different EPP techniques. As a result, the influence of the different EPP techniques on the obtained results can be compared. Due to this, additional information is not shown through the glyph metaphor in this figure, to enable the sole comparison of the projections. For the sake of brevity, the most interesting graphical displays are shown and those obtained by some of the EPP methods are not included in Figure 2.

The display obtained by KPCA can be considered the most revealing. It allows the structure of the data to be observed more clearly, as it represents the data in a more compact form and thus allows trends to be analysed with more clarity. Since it is not possible to include all the results obtained in this study, only results obtained using KPCA are shown in the rest of this section.

These results validate the main proposal of the present research: extending the original HUEP formulation by adding new EPP methods that can improve the visualization of a given dataset. For the dataset under analysis, none of the EPP methods in the original HUEP formulation provides with the best projection, but one of the new ones (KPCA) instead.

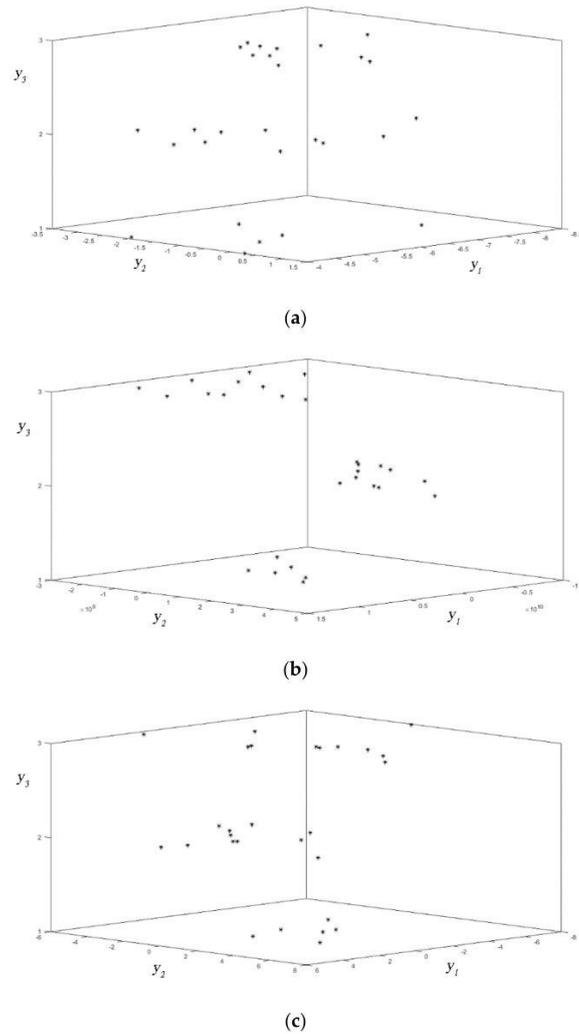


Figure 2. HUEPs obtained by applying agglomerative clustering ($k = 3$, distance = sEuclidean, linkage = average) to the analysed dataset, varying the EPP technique: (a) CMLHL, (b) KPCA, (c) SM.

Results including the Glyph Metaphor

After having selected KPCA as the EPP model that offers the best projections for the data analysed, the results using the glyph metaphor are presented in this section. It is worth mentioning that for any other dataset, this may not be the most appropriate EPP model.

In this section, additional information on the Fleet Age variable (sustainability data) is incorporated in the following graphs. The symbols for each piece of data are differentiated according to the quartile to which they belong, consistent with the value taken for that variable, in accordance with the symbols shown in Table 2.

Table 2. Legend for the graphs using the glyph metaphor according to the values of the sustainability-related feature (Average age of the vehicle fleet).

Q	Glyph
1	*
2	o
3	▷
4	+

In accordance with the above, the previously selected HUEP graph is shown (Figure 2b), although with the sustainability feature now incorporated. Additionally, the figure is enhanced by a form to guide the reader in the interpretation of the results.

In Figure 3 it is possible to see that a clear differentiation of the data quartiles of the series is obtained, and a clear line of progression can be marked on the result (dashed yellow line). The yellow line is associated to the temporal progression of data. From a practical point of view, the graph shows that the visualisation obtained is useful when determining the phases of progression of the age series; 1. the initial age, 2. the phase of over-amortisation of means of transport and 3. the phase of the new average age of the fleet.

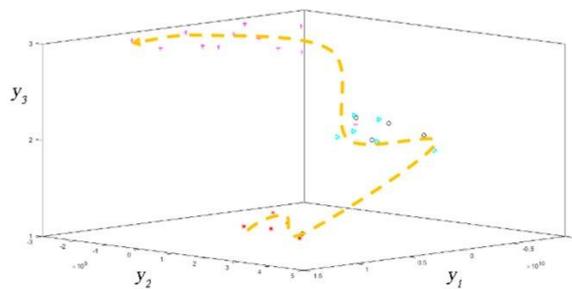


Figure 3. HUEP obtained by applying KPCA and agglomerative clustering ($k = 3$, distance = sEuclidean, linkage = average) to the analysed dataset, using the sustainability feature in the glyph metaphor. The yellow line is associated to the temporal progression of data.

To appreciate the impact of the k parameter of the clustering technique on the display, different visualisations are presented below (Figure 4), with the same EPP and clustering models, but with a varying number of clusters.

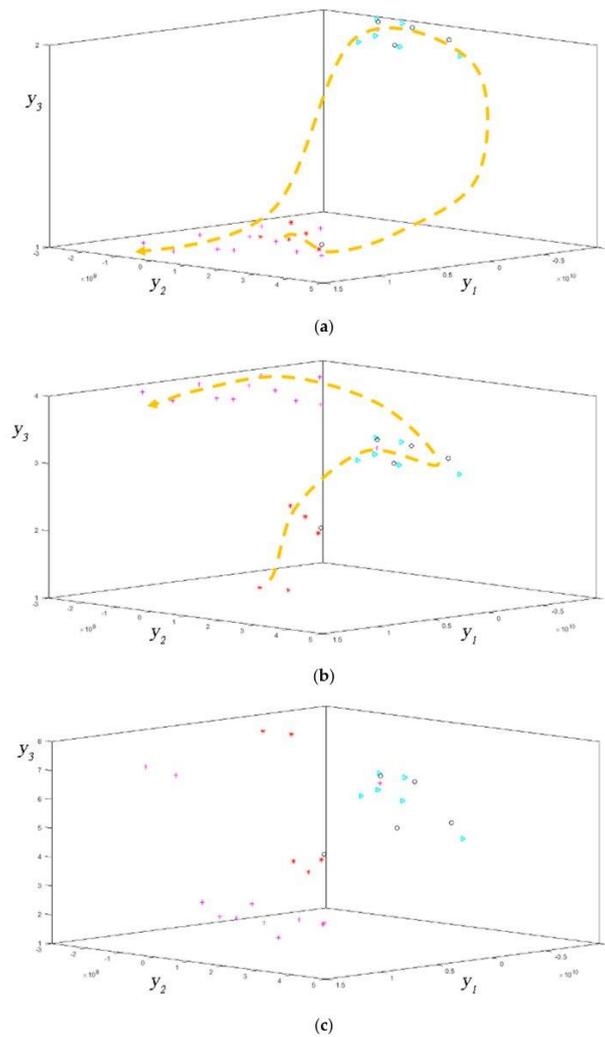


Figure 4. HUEPs obtained with KPCA and agglomerative clustering (distance = sEuclidean, linkage = average), showing the sustainability feature in the glyph metaphor and a varying number of clusters: (a) $k = 2$, (b) $k = 4$, and (c) $k = 8$.

After analysing the results shown in Figure 4, we can deduce that for $k = 2$, the visualisation is limited to differentiating between the phases of sustained amortisation and those of increasing amortisation. For $k = 4$, we can see a progression of increasing age, but compared to the significant result for $k = 3$ shown in Figure 3, it is limited to differentiating a new cluster in the initial periods of the study. The result for $k = 8$ does not seem significant. It can therefore be concluded that 3 is an appropriate number of clusters for the dataset under study.

Next, the visualisations obtained by partitional clustering are presented, with the aim of observing the influence of the clustering technique used, comparing it with the previous results (agglomerative clustering). Since 3 has been revealed as an appropriate number of clusters, different HUEPs obtained with k -means for a number of clusters close to 3 are shown in Figure 5.

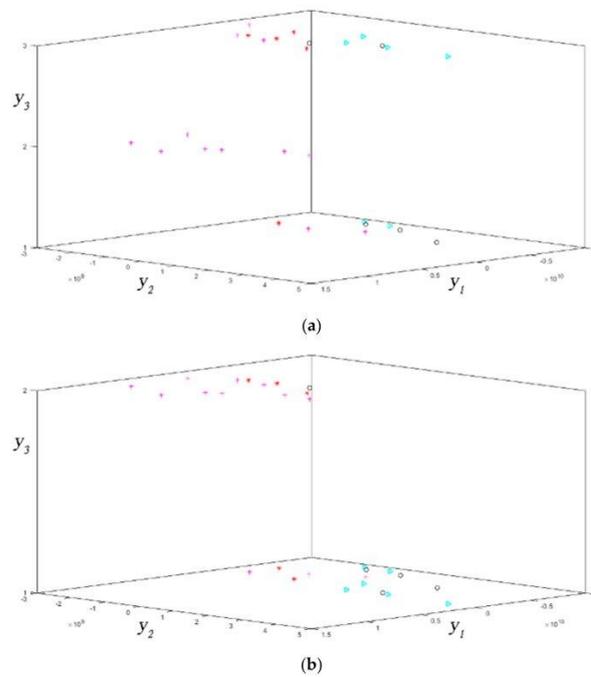


Figure 5. Cont.

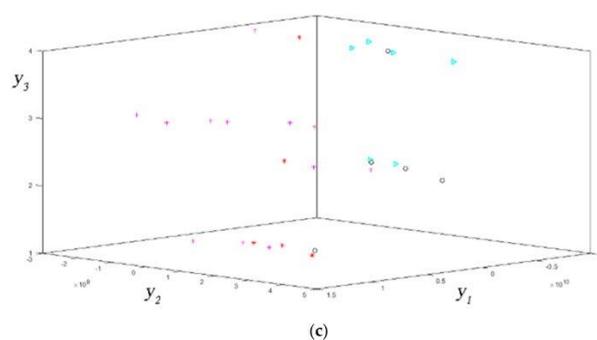


Figure 5. HUEPs obtained by applying KPCA and k-means clustering (distance = sEuclidean, linkage = average), showing the sustainability feature in the glyph metaphor and a varying number of clusters: (a) $k = 2$, (b) $k = 3$, and (c) $k = 4$.

The results shown in Figure 5 do not provide a clear or convincing display. No significant clustering can be inferred with respect to the progression of the data in terms of the age of the transport fleet in any of the tested configurations. We can therefore conclude that the application of k-means does not provide better displays of the data under study when compared to the agglomerative method.

Finally, the enhanced display obtained by the extended HUEPs is compared to that obtained by the newly applied KPCA method in order to identify what HUEPs can provide users with. The results obtained by KPCA (2D and 3D) are shown in Figure 6, comprising the glyph metaphor to support direct comparison.

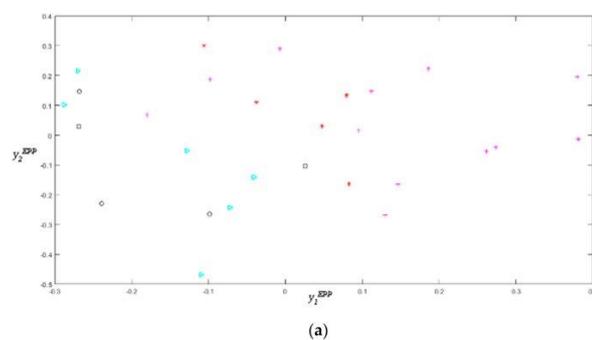


Figure 6. *Cont.*

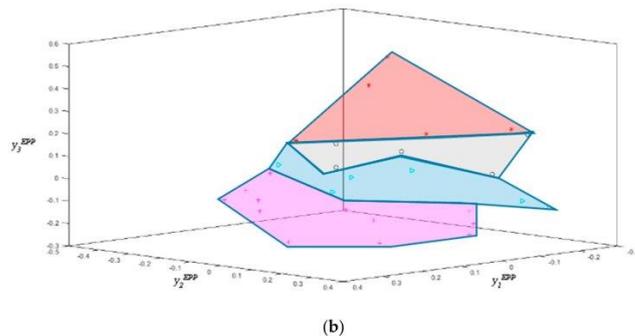


Figure 6. Projection obtained by applying KPCA in isolation to the analysed dataset, showing the sustainability feature in the glyph metaphor: (a) 2D and (b) 3D.

In Figure 6, we can see that after applying KPCA in isolation, the display is less informative than the one obtained through the HUEPs, as it does not allow for a clear identification of the structure of the data. On the other hand, a certain improvement is obtained in its 3D display as opposed to the 2D display. In the 3D view, significant groups of quartiles of fleet age are clustered in related areas of the three-dimensional space, as indicated in the figure by the coloured areas corresponding to the quartiles. Thus, it can be said that the new HUEP formulation reveals a much more interesting visualization of the dataset. Thanks to it, sustainability of the fleet (regarding the age) can be visually and intuitively analysed.

4. Conclusions and Future Work

On the one hand, it is clear that the development of road freight transport activity is intrinsically necessary for economic activity and the growing exchange of material goods between different areas of the planet, while on the other hand, its deployment entails a significant rate of GHG emissions. Additionally, it is also clear that the positive progression in the development of means of transport has led to significant reductions in the emission of gases and particles into the atmosphere with the most up-to-date equipment. As a consequence of the above, tools for the study, control and prediction of the evolution of the age of means of transport can be deemed priceless for corporate governance and management of this activity. More specifically, in their lines of commitment to sustainability, which may enable them to implement stimulus plans and social responsibility policies aimed at renewing the fleet of transport vehicles.

In this sense, the HUEPs (see Figures 3 and 4) constitute a valid, fast and intuitive visual tool to study the progression of the fleet age based on a set of data related to road freight transport activity.

Regarding the techniques employed for the dataset under study, it is worth noting that one of the new EPP techniques tested (KPCA) is the one that provides the best displays (see Figure 2), further confirming the proposed extension of the original formulation of the HUEPs. It has also been proven that the combination of KPCA with a clustering method provides better results than the 3D displays generated only with the EPP technique (see Figures 3 and 6). As for the clustering techniques, the suitability of the agglomerative method versus k-means has been confirmed, although it is necessary to adjust the number of clusters (k parameter) to the data (see Figures 4 and 5). As previously mentioned, this is the best performing EPP method for the dataset under analysis. It may vary for other datasets.

For future work, it is important to highlight that there are also other factors related to the study of the sustainability of road freight transport activity, which could be the subject of a study of the application of these and other techniques linked to ML, which would probably yield valuable conclusions. A clear example would be the study of the occupation of means of transport in the development of their activity, in addition to scaling the dimension of the data, considering the whole of Europe. As for ML techniques, the application of HUEPs in this field is proposed, adding other clustering techniques such as those based on density.

Author Contributions: Conceptualization, C.A.d.A., S.G. and Á.H.; methodology, C.A.d.A., D.U. and R.A.; software, D.U. and Á.H.; validation, C.A.d.A. and S.G.; formal analysis, R.A., Á.H.; investigation, C.A.d.A., D.U. and Á.H.; data curation, C.A.d.A. and D.U.; writing—original draft preparation, C.A.d.A. and Á.H.; writing—review and editing, C.A.d.A., D.U. and Á.H.; visualization, D.U. and Á.H.; supervision, C.A.d.A. and Á.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from the Spanish Ministry of Transport and the Environment.

Acknowledgments: We are grateful for the complete datasets of the Permanent Survey of Goods Transport by Road (*Encuesta Permanente de Transporte de Mercancías por Carretera*), facilitated by the General Sub-Directorate of Economic Studies and Statistics of the Ministry of Transport and the Environment (*Subdirección General de Estudios Económicos y Estadísticas del Ministerio de Transportes y Medio Ambiente*) of Spain.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. OECD/IFT. *International Transport Forum: Global Dialogue for Better Transport The Carbon Footprint of Global Trade Tackling Emissions from International Freight Transport*; Organisation for Economic Co-operation and Development-IFT: Leipzig, Germany, 2015.
2. Ministerio para la Transición Ecológica y el Reto demográfico. Secretaría de Estado De Medio Ambiente Dirección General De Calidad Y Evaluación Ambiental. Centro virtual de publicaciones del Ministerio de Transporte. In *Inventario Nacional de Emisiones a la Atmósfera Emisiones de Gases de Efecto Invernadero; Movilidad y Agenda Urbana*: Madrid, Spain, 2021.
3. European-Commission. Reducing CO₂ Emissions from Heavy-Duty Vehicles | Climate Action. Available online: https://ec.europa.eu/clima/policies/transport/vehicles/heavy_en (accessed on 16 June 2021).
4. Consejo-Europeo. *Reglamento del Parlamento Europeo y del Consejo Por el Que se Establecen Normas de Comportamiento en Materia de Emisiones de CO₂ Para Vehículos Pesados Nuevos*; Official Journal of the European Union: Brussels, Belgium, 2019; Volume 198.
5. Tob-Ogu, A.; Kumar, N.; Cullen, J.; Ballantyne, E.E.F. Sustainability Intervention Mechanisms for Managing Road Freight Transport Externalities: A Systematic Literature Review. *Sustainability* **2018**, *10*, 1923. [CrossRef]
6. Shankar, R.; Pathak, D.K.; Choudhary, D. Decarbonizing freight transportation: An integrated EFA-TISM approach to model enablers of dedicated freight corridors. *Technol. Forecast. Soc. Change* **2019**, *143*, 85–100. [CrossRef]
7. Agbo, A.A.; Li, W.; Zheng, L.; Zhang, Y.; Atombo, C. Optimisation of intermodal freight transport network. *Eur. Transp. Trasp. Eur.* **2017**, *63*, 1.
8. Kelle, P.; Song, J.; Jin, M.; Schneider, H.; Claypool, C. Evaluation of operational and environmental sustainability tradeoffs in multimodal freight transportation planning. *Int. J. Prod. Econ.* **2019**, *209*, 411–420. [CrossRef]
9. Meyer, T. Decarbonizing road freight transportation—A bibliometric and network analysis. *Transp. Res. Part D Transp. Environ.* **2020**, *89*, 102619. [CrossRef]
10. Rossi, T.; Pozzi, R.; Pirovano, G.; Cigolini, R.; Pero, M. A new logistics model for increasing economic sustainability of perishable food supply chains through intermodal transportation. *Int. J. Logist. Res. Appl.* **2021**, *24*, 346–363. [CrossRef]
11. De Campos, R.S.; Simon, A.T.; Martins, F.d. Assessing the impacts of road freight transport on sustainability: A case study in the sugar-energy sector. *J. Clean. Prod.* **2019**, *220*, 995–1004. [CrossRef]
12. Macioszek, E. Essential techniques for fastening loads in road transport. *Zesz. Nauk. Transp. Politech. Śląska* **2021**, *110*, 97–104. [CrossRef]
13. Macioszek, E. Oversize cargo transport in road transport ↓ Problems and issues. *Sci. J. Silesian Univ. Technol. Ser. Transp.* **2020**, *108*, 133–140. [CrossRef]
14. Corlu, C.G.; de la Torre, R.; Serrano-Hernandez, A.; Juan, A.A.; Faulin, J. Optimizing Energy Consumption in Transportation: Literature Review, Insights, and Research Opportunities. *Energies* **2020**, *13*, 1115. [CrossRef]

15. Hassani, A.; Safavi, S.R.; Hosseini, V. A comparison of light-duty vehicles' high emitters fractions obtained from an emission remote sensing campaign and emission inspection program for policy recommendation. *Environ. Pollut.* **2021**, *286*, 117396. [CrossRef]
16. Ang-Olson, J.; Schroerer, W. Energy Efficiency Strategies for Freight Trucking: Potential Impact on Fuel Use and Greenhouse Gas Emissions. *Transp. Res. Rec. J. Transp. Res. Board* **2002**, *1815*, 11–18. [CrossRef]
17. Comisión Europea. *Directiva 2002/80/CE de la Comisión de 3 de Octubre de 2002 por la que se Adapta al Progreso Técnico la Directiva 70/220/CEE del Consejo Relativa a Las Medidas Que Deben Adoptarse Contra la Contaminación Atmosférica Causada Por Las Emisiones de Los Vehículos*; OPOCE: Brussels, Belgium, 2002.
18. Comisión Europea. Emisiones de los vehículos pesados (Euro VI): Normas de certificación Diario Oficial de la Comisión Europea; 2005. Available online: <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=LEGISSUM:mi0029&from=ES> (accessed on 30 July 2021).
19. Haugen, M.J.; Bishop, G.A. Long-Term Fuel-Specific NOx and Particle Emission Trends for In-Use Heavy-Duty Vehicles in California. *Environ. Sci. Technol.* **2018**, *52*, 6070–6076. [CrossRef] [PubMed]
20. Yu, K.A.; McDonald, B.C.; Harley, R.A. Evaluation of Nitrogen Oxide Emission Inventories and Trends for On-Road Gasoline and Diesel Vehicles. *Environ. Sci. Technol.* **2021**, *55*, 6655–6664. [CrossRef] [PubMed]
21. Pernestål, A.; Engholm, A.; Bemler, M.; Gidofalvi, G. How Will Digitalization Change Road Freight Transport? Scenarios Tested in Sweden. *Sustainability* **2021**, *13*, 304. [CrossRef]
22. Mangina, E.; Narasimhan, P.K.; Saffari, M.; Vlachos, I. Data analytics for sustainable global supply chains. *J. Clean. Prod.* **2020**, *255*, 120300. [CrossRef]
23. Soysal, M.; Bloemhof-Ruwaard, J.M.; Haijema, R.; van der Vorst, J.G. Modeling a green inventory routing problem for perishable products with horizontal collaboration. *Comput. Oper. Res.* **2018**, *89*, 168–182. [CrossRef]
24. Montreuil, B. Toward a Physical Internet: Meeting the global logistics sustainability grand challenge. *Logist. Res.* **2011**, *3*, 71–87. [CrossRef]
25. Dadsena, K.K.; Sarmah, S.P.; Naikan, V.N.A. Risk evaluation and mitigation of sustainable road freight transport operation: A case of trucking industry. *Int. J. Prod. Res.* **2019**, *57*, 6223–6245. [CrossRef]
26. Wong, E.Y.C.; Tai, A.H.; So, S. Container drayage modelling with graph theory-based road connectivity assessment for sustainable freight transportation in new development area. *Comput. Ind. Eng.* **2020**, *149*, 106810. [CrossRef]
27. Budak, A.; Sarvari, P.A. Profit margin prediction in sustainable road freight transportation using machine learning. *J. Clean. Prod.* **2021**, *314*, 127990. [CrossRef]
28. Alonso de Armiño, C.; Manzanedo, M.Á.; Herrero, Á. Analysing the intermeshed patterns of road transportation and macroeconomic indicators through neural and clustering techniques. *Pattern Anal. Appl.* **2020**, *23*, 1059–1070. [CrossRef]
29. Desarbo, W.; Jedidi, K.; Cool, K.; Schendel, D. Simultaneous multidimensional unfolding and cluster analysis: An investigation of strategic groups. *Mark. Lett.* **1991**, *22*, 129–146. [CrossRef]
30. Vichi, M.; Kiers, H.A.L. Factorial k-means analysis for two-way data. *Comput. Stat. Data Anal.* **2001**, *37*, 49–64. [CrossRef]
31. Sánchez, R.; Herrero, Á.; Corchado, E. Visualization and clustering for SNMP intrusion detection. *Cybern. Syst.* **2013**, *44*, 505–532. [CrossRef]
32. Arroyo, Á.; Herrero, Á.; Corchado, E.; Tricio, V. A hybrid intelligent system for the analysis of atmospheric pollution: A case study in two European regions. *Log. J. IGPL* **2017**, *25*, 915–937. [CrossRef]
33. Bolton, R.J.; Krzanowski, W.J. Projection Pursuit Clustering for Exploratory Data Analysis. *J. Comput. Graph. Stat.* **2003**, *12*, 121–142. [CrossRef]
34. Yatracos, Y.G. Detecting Clusters in the Data from Variance Decompositions of Its Projections. *J. Classif.* **2013**, *30*, 30–55. [CrossRef]
35. Consejo-Europeo. *Reglamento (CE) N° 1172/98 del Consejo de 25 de Mayo de 1998 Sobre la Relación Estadística de los Transportes de Mercancías por Carretera*; OPOCE: Brussels, Belgium, 1998.
36. Consejo-Europeo. *Reglamento (UE) No 70/2012 del Parlamento Europeo y Del Consejo de 18 de Enero de 2012 Sobre la Relación Estadística de Los Transportes de Mercancías Por Carretera*; Official Journal of the European Union: Brussels, Belgium, 2012.
37. Herrero, Á.; Jiménez, A.; Bayraktar, S. Hybrid Unsupervised Exploratory Plots: A Case Study of Analysing Foreign Direct Investment. *Complexity* **2019**, *2019*, 6271017. [CrossRef]
38. Schölkopf, B.; Smola, A.; Müller, K.-R. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Comput.* **1998**, *10*, 1299–1319. [CrossRef]
39. Sammon, J.W. A Nonlinear Mapping for Data Structure Analysis. *IEEE Trans. Comput.* **1969**, *100*, 401–409. [CrossRef]

B.- Artículo 2



Non-linear Neural Models to Predict HRC Steel Price in Spain

Roberto Alcalde¹ (✉), Daniel Urda², Carlos Alonso de Armiño², Santiago García², Manuel Manzanedo², and Álvaro Herrero²

¹ Universidad de Burgos, Pza. de la Infanta Dña. Elena S/N, 09001 Burgos, Spain
radelgado@ubu.es

² Universidad de Burgos, Av. Cantabria s/n, 09006 Burgos, Spain
{durda, caap, lgpineda, ahcosio}@ubu.es, mms0133@alu.ubu.es

Abstract. Steel is a raw material widely used in industry due to its advantages over other alternatives, such as cost, fast and environmentally friendly recycling, ease of use, high strength, different finishes and qualities. Forecasting steel prices has been an important and challenging task that has traditionally been tackled with econometric, stochastic-Gaussian and time series techniques. Advancing from previous work on this open challenge, in the present paper some Artificial Neural Networks are applied for the first time to forecast the price of hot rolled steel in Spain. More precisely, some non-linear neural networks are applied to several different input time series. The target of this research is twofold; on the one hand, identify which of the neural models outperforms the other ones when predicting steel prices and, on the other hand to validate different data series for such prediction. The main outcomes of this research, after validating the neural models on real data from last 7 years, greatly contribute to this field as novel and relevant conclusions are obtained.

Keywords: Neural networks · NIO · NAR · NARX · Forecasting · Steel price

1 Introduction and Previous Work

Steel is an alloy of iron ore with other elements, mainly carbon, but also other minerals to modify its properties. Besides using these minerals, steel can also be produced using steel scrap, making it a highly recyclable material that represents a true circular economy today. The discovery of steel and its subsequent improvement in steel production technology was part of the first and second industrial revolutions, so steel has been essential to the modernisation of the world and continues to be so today.

Global steel production exceeded 1875 million tonnes in 2019, according to the International Steel Association¹, and it is the most energy-intensive industrial sector in the world [1].

Therefore, steel is of great importance for world industry, as it is used in different formats for a large number of industrial sectors (construction, household appliances,

¹ <https://worldsteel.org/>

vehicles, etc.), and it is a material with an excellent relationship between its physical properties (elasticity, strength, resistance) for its cost. It is also a versatile material that is used in new and varied ways everyday thanks to the development of new alloys and coatings with improved properties [2].

Initially, steel was obtained in a furnace, to which iron ore and coking coal were added to melt these minerals with fire. Air blowing was used to drive the furnace fire to force the combustion gases through and mix the raw materials to obtain pig iron, which is an alloy that melts at a lower temperature than steel. The pig iron is then minced and ground, and mixed with coal in a crucible, which is sealed to prevent oxidation and slowly cooled to obtain the piece of steel, which is then used for forging. Later, various technological advances were made in the process (Bessemer process, Gilchrist-Thomas process, Siemens-Martin), to produce large quantities at low cost. Today, electric mini-furnaces are also used to produce steel from scrap (ferrous scrap and iron ore fines Fe 62%). There are two different routes in the steelmaking process, the primary or BOS (Basic Oxygen Furnace), which mainly uses iron ore and coal, and the secondary route or EAF (Electric Arc Furnace), which uses steel scrap [3].

A distinction can be made between different types of steel depending on their chemical composition (carbon, stainless, electrical and special steel), their physical form (coil, flat and long steel), their processing stage (hot-rolled, cold-rolled, galvanised steel), as well as their finish, which means that steel is not a homogeneous product [3].

The price of different types of steel depends on many factors, such as raw materials (iron ore, coal), production process, production capacity, investments, energy costs, demand, supply, political factors [4]. In addition, it appears that there may be some correlation between the price of steel and the share price of the steelmaking companies [5]. On the other hand, there are studies that suggest a one-way correlation between some variables, such as the price of oil influencing the price of steel, but not vice versa [6]. And, metal prices have been considered as relevant indicators for markets due to their ability to adjust to macroeconomic and speculative conditions [7].

Due to the widespread use of steel in many sectors, the price of steel is a major factor influencing the competitiveness of many companies. Therefore, predicting price trends can help companies to make decisions on buying and selling steel in order to increase their profitability.

In general, the prices of steel products all follow the same pattern, but Hot-rolled Sheet Coil (HRC) is usually considered as the benchmark because it accounts for the largest volume of steel exports [8].

Although the steel market is global, the price of steel can differ from country to country, due to different reasons, such as tariffs, market demand, taxes, among others. Nevertheless, the evolution of steel prices is parallel in each country [9–11]. In addition, the price of commodities, including steel, seems to experience relatively different correlated fluctuations for each country, which may be due to similar international terms of trade [12].

For the prediction of steel price, multiple models have been used, such as the Auto-Regressive Integrated Moving Average (ARIMA) model [13, 14], as well as Long Short-Term Memory LSTM [15]. On the other hand, the applied non-linear neural networks have been previously applied to some other problems such as logistics [16] or precision

agriculture [17] among others. Differentiating from this previous work in the present paper some non-linear neural networks are applied to this problem for the first time. Additionally, novel data series are used as the input data for such predicting models.

This study focuses on the prediction of the HRC price for Spain, and it includes a comparison of several neural network models (Described in Sect. 2), namely: Non-linear Input-Output (NIO), Non-linear Auto-Regressive (NAR), and Non-linear Auto-Regressive with Exogenous Input (NARX)). Such supervised-learning models are applied for steel price time series forecasting. Furthermore, the impact of the parameter tuning for these models is also assessed. Additionally, some different data series (described in Sect. 3) are studied and compared in order to identify which one contributes to the most precise prediction.

2 Materials and Methods

In the present section both the used data (Subsect. 2.1) and the applied models (Subsect. 2.2) are introduced.

2.1 Dataset

As previously mentioned, the present research focuses on predicting the price of hot-rolled sheet coil (HRC) for Spanish market. The source for this data is CRU-Commodities Research Unit² which is a leading independent supplier of steel market information. In order to predict the price, three groups of data are studied:

- Series 1: Stock market prices of three major global steel producing companies, which are Posco, ArcelorMittal and NipponSteel, whose data source is NYSE – Nasdaq Real Time Price in USD³.
- Series 2: Economic situation in Spain, including data related to three issues: a) Price (Index 2010 = 100); Consumer Price Index Harmonized, Consumer Price Index All items, Producer Price Index All Commodities Index; b) Economic Activity- Industrial Production (Index 2010 = 100): Total Index, Manufacturing Index, Mining Index; c) External Trade Goods (US Dollars): Value of Imports (Cost, Insurance, Freight CIF), Value of Exports (Free on Board FOB). (Data source from International Monetary fund⁴).
- Series 3: Price of other commodities: price of Aluminum; Brent Crude; Cobalt; Copper; Dubai Crude; Gold; Lead; Molybdenum; Silver; Zinc (Data source from International Monetary fund).

The data series for all features corresponds to a monthly periodicity during the period 2013–2019. A monthly periodicity is used for the data source as this is the time unit generally established in steel price indexation and is sufficient to reflect the volatility of the steel price [18]. As a sample of the target data to be forecasted, the maximum price of HRC in the series under analysis was identified in.

² <https://www.crugroup.com/>

³ <https://finance.yahoo.com/>

⁴ <https://www.imf.org>

2.2 Non-linear Neural Models

In order to predict the price of HRC steel, the data described in previous subsection has been used as the input data for several neural models for non-linear time-series forecast [19], namely: Non-linear Input-Output (NIO), Non-linear Autoregressive (NAR) and Non-linear Autoregressive with Exogenous Input (NARX). All these three can be defined as feedforward Neural Networks (NN) that includes a tap delay associated to the input weight. The models have a finite dynamic response to time series input data thanks to such delay.

Although these three models have similar characteristics, there are also some differences, being mainly the data the model is provided with. Firstly, there is the NIO models that tries to predict a data series only taking into account previous values of other data series but not the target one. That is, the data series to be predicted is not used itself as an input. Oppositely, the NAR model, as stated in its name just models a relation between previous values of the data series whose future values is trying to predict. Hence, only the series to be predicted is used by this model. These two approaches are combined in the NARX model, as it predicts the values of a data series by using the previous values of it (endogenous input) while considering some other data series (exogenous inputs).

In a more formal, way, these models can be mathematically defined as follows, considering $y(t)$ the feature to be predicted in time instant t and $f(\cdot)$ the function to be approximated by the time-series forecasting model. Mainly consisting of a model that predicts a series (y) only using n previous values of it, the NAR model can be expressed as:

$$y(t) = f(y(t-1), \dots, y(t-n_y)) \quad (1)$$

From an opposed perspective, the NIO model tries to predict a series (y) using n previous values of a different one (x):

$$y(t) = f(x(t-1), \dots, x(t-n_x)) \quad (2)$$

Finally, the NARX is a model that combines previous values of both the series to be predicted (y) and those from a different one (x) that is considered the exogenous one:

$$y(t) = f(y(t-1), \dots, y(t-n_y), x(t-1), \dots, x(t-n_x)) \quad (3)$$

3 Experiments and Results

This section presents information about the performed experiments, as well as the obtained results. When defining the experiments and based on previous results, the parameters of these models has been tuned with combinations of the following values:

- Training algorithm: {1 – Levenberg-Marquardt, 2 – Batch Gradient Descent, 3 – Gradient Descent with Momentum, 4 – Adaptive Learning Rate Backpropagation, 5 – Gradient Descent with Momentum and Adaptive Learning Rate, 6 – Scaled Conjugate Gradient, 7 – Broyden-Fletcher-Goldfarb-Shanno Backpropagation}

190 R. Alcalde et al.

- Number of hidden neurons: {1, 5, 10, 15, 20}
- Number of input delays: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
- Number of output delays: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}

In order to obtain more statistically significant results, 10 executions have been performed for the same configuration of the parameters. Due to the combination of all these values, 350 executions have been carried out for the NIO and NAR models per each one of the data series. In the case of NARX, as it combines a varying number of both input and output delays, 3,500 executions have been carried out per each one of the data series.

The averaged Mean Squared Error (MSE) for the 10 executions is provided in this section. In each one of the tables, the lowest error value per column is in bold.

Initially, the averaged results obtained by the time-series models are presented per the number of input delays in Table 1.

Table 1. MSE of the results obtained by the NAR, NIO, and NARX neural models. Averaged results for the 3 data series are shown per the number of input delays.

N input delays	NAR	NIO	NARX
1	44777.62	34768.56	30731.95
2	36811.43	33635.41	1259633.09
3	35092.46	36092.17	31739.07
4	37317.69	30805.89	31725.67
5	35941.02	34324.44	34231.69
6	32765.64	34953.35	34211.34
7	29586.15	35951.07	34099.02
8	35067.41	42498.11	36024.10
9	35714.85	41546.65	37190.99
10	27439.20	39195.92	39736.90

From the results in Table 1, it can be said that opposed results have been obtained for some of the models: in the case of NAR, the lowest error is obtained with the highest number of input delays (10) while in the case of NARX it has been obtained with the lowest one (1). In the case of NIO, an intermediate value (4) for input delays is the one with the lowest error.

Results obtained by the neural models are also presented per the number of hidden neurons in Table 2.

When considering the number of neurons in the hidden layer of the neural models, opposing results are found: when only the price data series is used (NAR model), the lowest error is obtained with the biggest hidden layer (20 neurons). However, when the endogenous data series are used (NIO and NARX models), the lowest error is obtained with the smallest hidden layer (1 neuron).

Table 2. MSE of the results obtained by the NAR, NIO, and NARX neural models. Averaged results for the 3 data series are shown per the number of hidden neurons.

Neurons	NAR	NIO	NARX
1	35092.46	10662.66	10453.78
5	36205.07	21073.80	20342.18
10	34353.33	33933.62	646951.91
15	32326.78	52494.22	47046.96
20	31577.02	63721.49	59867.09

In Table 3, the obtained results are presented per the training algorithm.

Table 3. MSE of the results obtained by the NAR, NIO, and NARX neural models. Averaged results for the 3 data series are shown per the training algorithm.

Training algorithm	NAR	NIO	NARX
1	2029.26	12615.28	11009.04
2	115329.59	94843.10	947446.69
3	117558.44	95041.44	89026.89
4	3321.75	15365.87	12867.00
5	3845.27	13214.13	16612.29
6	1448.84	11720.95	10919.50
7	1826.28	11839.34	10645.26

As it can be seen in Table 3, there are strong differences regarding the errors obtained when training the models with different algorithms, being the best and worst results clearly identified. The two algorithms associated to best results (smallest error) are "Scaled Conjugate Gradient" (6) and "Broyden-Fletcher-Goldfarb-Shanno Back-propagation" (7) for all the three models. On the other hand, the worst results (highest error rates) are associated to the "Batch Gradient Descent" (2) and Gradient Descent with Momentum (3) for all the three models.

As one of the targets of the present research is identifying the best time series for predicting the price, averaged results are also presented in Table 4 per the used data series (only NIO and NARX models as the NAR model is only provided with the price itself).

When only using the endogenous data series to make the prediction (NIO), it is the data series 3 (Price of other commodities) the one associated to best results and the data series 1 (Stock market prices) the second best. When combining the endogenous data series with the price itself (NARX), it is the other way round: the data series 1 is the one associated to best results and the data series 3 is the second best. For both models,

Table 4. MSE of the results obtained by the NIO, and NARX neural models. Averaged results are shown per the data series.

Data series	NIO	NARX
1	29775.95	26434.71
2	52250.05	417830.09
3	27105.46	26532.35

the data series 2 (Economic situation in Spain) is the one associated to worst results. Regarding the neural model that is applies, it can be said that each one of the data series, the best result is always obtained by the NARX model, being NIO in all cases the second best.

Finally, in order to clearly identify the best single results, Table 5 presents the lowest error rates for each parameter combination. In this case, the error has been averaged only for the 10 executions run with exactly the same values for the different parameters.

Table 5. MSE of the results obtained by the NAR, NIO, and NARX neural models. Best single results (averaged only for the 10 executions) are shown.

Input data series	NAR	NIO	NARX
Price itself	566,63	-	-
1	-	789,75	460,41
2	-	1150,34	796,86
3	-	664,81	582,63

For this individual results (error is not averaged for all the parameter combinations), similar results to those in Table 4 can be seen: the best result is obtained with data series 3 in the case of the NIO model while it is the data series 3 in the case of the NARX models. In both cases, the worst results are obtained with data series 2. Regarding the neural models, for all the endogenous data series, NARX clearly outperforms NIO in all cases.

All in all, the best single result (10 executions) is obtained when combining the NARX model and the data series 1. The second best is obtained by the NAR model, hence using only the price data series.

4 Conclusions and Future Work

This study has presented empirical results obtained with three datasets (stock market, economic and other commodity) on the performance of several non-linear models (NIO, NAR, NARX) for steel price forecasting. The performance of the different neural models

and their parameters has been compared. The used models have not been previously applied for steel price forecasting.

After analyzing the individual results, it can be concluded without any doubt that best predictions of the HRC steel price are obtained with the NARX model and data series 1, showing superior performance to the other models and data series. So, it can be concluded there is a strong relationship between the share price of steel companies and the price of steel.

From the perspective of parameter tuning, it can be concluded that for some of the parameters there is not a consensus affecting all the models. The best values for the number of input delays and hidden neurons varies from one model to the other one so it must be adjusted case by case. However, in the case of the algorithm used to train the model, it can be concluded that the "Scaled Conjugate Gradient" and "Broyden-Fletcher-Goldfarb-Shanno Back-propagation" are the most advisable ones for all the three models.

When considering the input data series in general terms, the "Price of other commodities" and "Stock market prices" are those associated to best results. Worst results have always been obtained when using the "Economic situation in Spain" data series.

From a business perspective, the result obtained shows that the stock market performance of the shares of steel producing companies is correlated with the evolution of the price of steel. This connection is consistent with the effect that increases in the selling price of a company's material, with equal demand, have a positive influence on the profitability of the company, and therefore the value of the share is increased by investors.

In future studies, new soft-computing techniques and other data series can be used to improve the existing predictive models.

References

1. Conejo, A.N., Birat, J.P., Dutta, A.: A review of the current environmental challenges of the steel industry and its value chain. *J. Environ. Manag.* **259** (2020). <https://doi.org/10.1016/j.jenvman.2019.109782>
2. Gutierrez, J.P., Vianna, A.C.: Price effects of steel commodities on worldwide stock market returns. *North Amer. J. Econ. Fin.* **51**(April) (2020). <https://doi.org/10.1016/j.najef.2018.09.007>
3. European Commission: Towards competitive and clean European steel (2021)
4. Malanichev, A.G., Vorobyev, P.V.: Forecast of global steel prices. *Stud. Russ. Econ. Dev.* **22**(3), 304–311 (2011). <https://doi.org/10.1134/S1075700711030105>
5. Manu, M.V., Brătășanu, V., Vasile, I.: Challenging the status quo: steel producer case study on the enterprise value for M&A. *Manag. Dyn. Knowl. Econ.* **7**(2), 207–228 (2019). <https://doi.org/10.25019/mdke7.2.05>
6. Ming-Tao Chou, S.-C.C., Yang, Y.L.: Review of Economics & Finance A Study of the Dynamic Relationship between Crude Oil Price. Better Advances Press, Canada in its Journal Review of Economics & Finance., vol. 2, no. May (2012)
7. Omura, A., Todorova, N., Li, B., Chung, R.: Steel scrap and equity market in Japan. *Resour. Policy* **47**, 115–124 (2016). <https://doi.org/10.1016/j.resourpol.2016.01.001>
8. Rodríguez Liboreiro, P.: Competencia, rendimientos crecientes y exceso de capacidad: la industria siderúrgica mundial (2000–2014). *Cuadernos de Economía* **38**(76) (2019). <https://doi.org/10.15446/cuad.econ.v38n76.61257>

9. Gutierrez, J.P., Vianna, A.C.: Price effects of steel commodities on worldwide stock market returns. *North Am. J. Econ. Fin.* **51**(Oct) (2020). <https://doi.org/10.1016/j.najef.2018.09.007>
10. Pauliuk, S., Wang, T., Müller, D.B.: Steel all over the world: Estimating in-use stocks of iron for 200 countries. *Resour. Conserv. Recycl.* **71** (2013). <https://doi.org/10.1016/j.resconrec.2012.11.008>
11. Panasiyk, D., Laratte, B., Remy, S.: Steel stock analysis in Europe from 1945 to 2013. *Proc. CIRP* **48**, 348–351 (2016). <https://doi.org/10.1016/j.procir.2016.04.084>
12. Adewuyi, A.O., Wahab, B.A., Adeboye, O.S.: Stationarity of prices of precious and industrial metals using recent unit root methods: implications for markets' efficiency. *Resour. Policy* **65**(Jul), 2020 (2019). <https://doi.org/10.1016/j.resourpol.2019.101560>
13. Zola, P., Carpita, M.: Forecasting the steel product prices with the arima model. *Stat. e Appl.* **14**(1) (2016)
14. Adli, K.A.: Forecasting steel prices using ARIMAX model: a case study of Turkey. *Int. J. Bus. Manag. Technol.* **4**(5), 62–68 (2020)
15. Cetin, K., Aksoy, S., Iseri, I.: Steel price forecasting using long short-term memory network model (2019). <https://doi.org/10.1109/UBMK.2019.8907015>
16. Alonso de Armiño, C., Manzanedo, M.Á., Herrero, Á.: Analysing the intermeshed patterns of road transportation and macroeconomic indicators through neural and clustering techniques. *Pattern Anal. Appl.* **23**(3), 1059–1070 (2020). <https://doi.org/10.1007/s10044-020-00872-x>
17. Yartu, M., Cambra, C., Navarro, M., Rad, C., Arroyo, Á., Herrero, Á.: Humidity forecasting in a potato plantation using time-series neural models. *J. Comput. Sci.* **59** (2022). <https://doi.org/10.1016/j.jocs.2021.101547>
18. Guo, S., Li, H., An, H., Sun, Q., Hao, X., Liu, Y.: Steel product prices transmission activities in the midstream industrial chain and global markets. *Resour. Pol.* **60**(Nov), 56–71 (2019). <https://doi.org/10.1016/j.resourpol.2018.11.014>
19. Leontaritis, I.J., Billings, S.A.: Input-output parametric models for non-linear systems Part I: Deterministic non-linear systems. *Int. J. Control* **41**(2) (1985). <https://doi.org/10.1080/0020718508961129>

Article

Analysis of the Economic Sustainability of the Supply Chain Sector by Applying the Altman Z-Score Predictor

Roberto Alcalde ^{1,*}, Carlos Alonso de Armiño ² and Santiago García ²

- ¹ Departamento de Economía y Administración de Empresas, Facultad de Ciencias Económicas y Empresariales, Universidad de Burgos, Pza. de la Infanta Dña. Elena S/N, 09001 Burgos, Spain
- ² Departamento de Ingeniería de Organización, Escuela Politécnica Superior, Universidad de Burgos, Av. Cantabria S/N, 09006 Burgos, Spain; caap@ubu.es (C.A.d.A.); lgpineda@ubu.es (S.G.)
- * Correspondence: radelgado@ubu.es

Abstract: This paper fills the gap in the financial perspective of supply chain performance measurement, related to the lack of a bankruptcy probability indicator, and proposes a predictor which is the eighth-model of the Altman Z-Score Logistic Regression. Furthermore, a bankruptcy probability ranking is established for the companies' supply chains, according to the industry to which they belong. Moreover, the values are set to establish three categories of companies according to predictor. The probability of bankruptcy is analysed and studied for the supply chain of different industries. The building industry is revealed to have the highest probability of bankruptcy.

Keywords: bankruptcy; risk prediction; supply chain; Altman's Z-score



Citation: Alcalde, R.; Alonso de Armiño, C.; García, S. Analysis of the Economic Sustainability of the Supply Chain Sector by Applying the Altman Z-Score Predictor. *Sustainability* **2022**, *14*, 851. <https://doi.org/10.3390/su14020851>

Academic Editors: Paulo Afonso, Antonio Zanin and Orlando Durán

Received: 5 December 2021

Accepted: 5 January 2022

Published: 12 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The definition of “supply chain” has been widely established in the state of the art as a network of materials, information and services linked in three states: procurement, production and distribution. The most complex supply chain is the one called the ultimate supply chain, which includes the entire sector of companies from different industries that pertain to all flows of products, services, finance and information from the ultimate suppliers to the ultimate customers [1,2].

The sector of the supply chain called the ultimate supply chain represents the alignment of multiple companies that bring products or services to the ultimate consumers in the market. The industries to which the companies in the ultimate supply chain belong are: manufacturers, suppliers, transporters, warehouses, wholesalers, retailers and other intermediaries [1].

In recent years, a variety of supply chain performance measurement (SCPM) systems have been published to quantify the efficiency and effectiveness of supply chain processes, their relationships and the multiple functions of organisations and companies to enable supply chain orchestration [3–6]. Companies can benchmark themselves against others through the use of supply chain maturity models. In both maturity models and SCPM, aspects related to environmental regulation are included in order to understand that natural resources are limited, so that companies adopt sustainable natural resource practices [7]. In addition to this environmental perspective, recent studies include company metrics from other perspectives that can influence their supply chain, such as: internal process, learning and growth, customer, social and financial [8]. On a financial perspective, neither the maturity models nor the supply chain improvement measurement models include an indicator of the economic sustainability of the company, such as the probability of company bankruptcy, to measure the continuity of the company and its influence on the supply chain.

Economic sustainability is important to pursue, as financial difficulties and failures of critical suppliers trigger announcements of supply chain disruptions that severely penalize

business operations [9–11]. Therefore, companies in a supply chain must maintain a healthy financial state to achieve the successful integration of their activities and proper coordination of all their key processes. Otherwise, the uncertainty and risks of bankruptcy of one company in the supply chain may threaten the operations and activities of the other companies in the supply chain [12].

Predicting the risk of bankruptcy is important for the early detection of signs of deteriorating financial conditions and can enable corrective action to be taken in the operation of the supply chain [13,14]. Consequently, companies with a stronger financial position can assume the liquidity necessary to maintaining their supply chain activity by accepting higher temporal payment terms or paying in advance to increase trade credit to companies with a weaker financial position and thus complement the financial health of these companies [15,16].

In this context, the research questions (RQ) considered in this investigation are two. RQ1: Which are the industries in the supply chains of Spanish companies with the highest probability of bankruptcy? RQ2: Which values allow the probability of bankruptcy to be delimited into three zones for the financial classification of a company?

In addition to the introduction, this article is divided into four sections. Initially, there is a review of the theoretical concept of the supply chain and the probability of bankruptcy, as well as Altman's bankruptcy Z-predictor. Then, the methodology of the study is presented. This is followed by the empirical study to answer the research questions. Finally, the discussion and conclusions of the study are presented, highlighting the supply chain industries with the highest probability of risk.

2. Theoretical Framework

2.1. Supply Chain

Supply chain management facilitates the flow of goods, information and money. Of these three elements, the least studied in the state of the art is the flow of money in the supply chain, which is called supply chain finance (SCF) [15].

Due to the increased dependency between companies in the ultimate supply chain, the bankruptcy of one company may cause other members of the chain to fall into financial difficulties [17].

The provision of liquidity to the supply chain helps to ensure its financial stability. The reduction of the cash conversion cycle increases the liquidity of companies in the supply chain, and therefore reduces the probability of bankruptcy of these companies [15]. The cash conversion cycle can be reduced by focusing on the optimization of working capital in terms of accounts payable, accounts receivable, inventories and sometimes even fixed asset financing [18]. The trade credit financing is another option to increase liquidity, which may increase the level of risk and bankruptcy, as it may stimulate more orders to be placed [19,20].

The ultimate supply chain, which includes service only supply chains (SOSCs) and the product service supply chains (PSSCs) [21], is different for each industry, so the causes of bankruptcy may vary for each industry [22].

At the industry level, the market demand, sales growth and sustainable growth of each company in the supply chain defines the company's strategic behaviour in various categories (expansive, dominant, shrinking, restructuring, in decline and unfocused) that can be used to predict companies' profitability [23].

The visibility of a supply chain and integration of suppliers to share information to detect the bankruptcy of a supplier in a supply chain enables risk management for intentional supply chain disruptions [24].

Therefore, the measurement of the probability of bankruptcy is important for the economic sustainability of each industry's supply chain. Since there is no such indicator in SCPM systems, there is a gap that is addressed in this research with a proposal to introduce a bankruptcy predictor and the analysis of bankruptcy in each industry.

2.2. Predictor of Bankruptcy: The Z-Score Model by Altman

A pioneering work of scientific research on the prediction of business failure dates back to the 1930s, when Fitzpatrick (1932) and Smith et al. (1935) applied a very basic univariate analysis based on the study of the evolution of financial ratios [25,26].

The contributions of Beaver (1966) incorporated a univariate discriminant analysis into the research with a separate evaluation of the predictive capacity of each one of the ratios considered, which meant a qualitative leap, although it was soon abandoned by multivariate techniques [27].

Altman (1968) was the pioneer in applying techniques of multivariate discriminant analysis (MDA) to the study of the prediction of corporate bankruptcy, replacing the previous contribution of Beaver [28]. With this research and others that followed, good results were achieved with minor classification errors [29].

Since then, this research line has been enriched by numerous contributions by different authors with new approaches and the incorporation of other techniques, such as artificial intelligence, neural networks, self-organized maps, multidimensional scales, and the Logit technique (which makes it possible to estimate the likelihood of failure for a set of constraints or attributes) [30–32].

Despite these other contributions, this Altman tool is commonly accepted to anticipate financial failures and bankruptcy risk in various industry sectors or companies within the same supply chain [33–35].

2.2.1. Altman Z-Score Model

The initial model was obtained in 1968 with a sample of 66 companies, of which 33 filed for bankruptcy and another 33 did not. All of the companies belonged to the US manufacturing sector, were medium-sized and were listed on the stock market for 20 years (1946–1965). The 66 financial statements were obtained from Moody's industrial manual, and from them Altman extracted 22 financial indices related to: profitability, indebtedness, activity, liquidity and solvency. He grouped these indicators into five variables, obtained from seven data from accounting and financial results [28], as can be seen in Table 1.

As a result of applying the multiple discrimination analysis technique to the sample of companies, Equation (1) was obtained:

$$Z = 1.2 X_1 + 1.4 X_2 + 3.3 X_3 + 0.6 X_4 + 1.0 X_5 \quad (1)$$

Companies in Edward Altman's sample that scored below 1.81 ended up in bankruptcy. While companies that scored above 2.99 were in a healthy financial zone. Consequently, those companies with scores between 1.81 and 2.99 were considered to be in the caution zone.

This model was criticized for being applicable only to capital-market-listed manufacturing companies and not to commercial and service companies, listed or unlisted. For this reason, Altman agreed to revise the model, although it is true that initially he resisted, arguing that it was not necessary. Finally, he published two new versions of the model, the Z' and the Z'' [36–38].

Despite this, recent academic articles can be found that still use this model instead of the latest update [39–42].

2.2.2. Altman Z' -Score Model

In 1983, an adjustment was made to the original model so that it could be applied to unlisted manufacturing companies. In particular, in the calculation of the X_4 ratio, it replaced the numerator with the book value of the equity instead of the market value of shares and divided by the book value of total liabilities [43]. Consequently, the new Equation (2) was obtained:

$$Z' = 0.717 X_1 + 0.847 X_2 + 3.107 X_3 + 0.42 X_4 + 0.998 X_5 \quad (2)$$

Table 1. Key figures of the initial Z-Altman model [28].

Variable	Name	Formula	Explanation
X_1	Liquidity ratio	WCTA = working capital/ total assets	The part of the current assets that is financed using permanent resources is called the working capital. The relationship between the working capital and total assets is related to the capacity of a company to comply with its short-term financial obligations, that is to say, liquidity. Altman distinguished X_1 as one of the most important ratios, since it is proportional to the company's liquidity and therefore to its financial health.
X_2	Accumulated return ratio	RETA = accumulated reserves/total assets	The age of the company is quite relevant to predicting bankruptcy and insolvency, since a younger company is more likely not to survive compared to a company with an established operation or economic activity. Normally, older operating companies have a higher level of accumulated reserves. When the business accumulates losses, reserves may become negative. However, when it accumulates profits, the reserves will be positive, evidencing a capacity for growth and reinvestment to self-finance its projects.
X_3	Profitability ratio	EBITTA = profitability before tax and interest (EBIT)/total assets	If a company fails to consolidate its operation to generate sufficient resources, it will subsequently disappear from one year to the next because it is unable to meet its payment obligations. The higher the value of this variable, the lower the probability of bankruptcy.
X_4	Financial structure ratio	BVETD = market value/ total liabilities	For quoted companies, the market value can be calculated using the stock market price. Its division into total liabilities allows it to be related to the book value. The result is a variable related to the financial structure. The higher the value of this variable, the lower the probability of bankruptcy of the company, since the market is valuing the company above its book value.
X_5	Asset turnover rate	SALTA = net sales/ total assets	This indicator is associated with the speed of movement of assets, i.e., the administrative and commercial capacity of the company in relation to competition from other companies in the sector in which it operates. The result is interpreted as the number of times sales contain the asset.

In this model, results below 1.23 are considered likely to fail, while those above 2.99 are considered healthy. So, between 2.99 and 1.23 is a precautionary zone.

Altman did not validate this model with a second sample due to the lack of private companies in the database. He considered that the X_5 ratio (net sales/total assets) was influenced by the company's sector or industry, so he decided to make a new model valid for all industries or sectors, which he called the Z'' -score [37]. Nevertheless, in recent academic articles, this Z' model is likely to continue to be used instead of the latest update [9,44].

2.2.3. Altman Z'' -Score Model

In order to obtain a model suitable for use with all categories of companies, the variable X_5 from the above Equation (2) is excluded, since it is a value that varies significantly from one company to another in different sectors or industries, which causes a distortion and can influence the model. So, Equation (3) is obtained:

$$Z'' = 3.25 + 6.56 X_1 + 3.26 X_2 + 6.72 X_3 + 1.05 X_4 \quad (3)$$

In this other model Z'' , a firm with a score below 1.1 is considered a high bankruptcy likelihood, while if it is higher than 2.6, it is considered healthy. Thus, companies with a value between 2.6 and 1.1 are considered to be in a financial caution zone [36,37].

As before, instead of the latest update, recent academic articles can be found using this model [45–47].

2.2.4. Model Z-Score of Altman 2014

From the year 2000 to 2014 there have been published numerous scientific articles mentioning the Z-score Altman model. From a selection of 33 articles, 16 of them used the model to measure companies' financial strength or bankruptcy, 14 other studies verified and modified the model, and in 3 cases it was used to validate its robustness. Although this model has been published for many years and is widely accepted, it seems logical to think that it requires a revision, in which its parameters are re-estimated with more recent data and new estimation techniques are used.

For this reason, Altman and others have made a new version of Z-score model [43], with a new data sample of 2,640,778 companies (2,602,563 non-bankruptcy and 38,215 bankruptcy) from USA, China, Colombia and 31 European countries. New variables (country, industry, size and age) are also included and grouped into seven hypotheses to improve the performance of the model [48], which are shown in Table 2.

Table 2. The seven hypotheses considered by Altman in the updated model.

Hypothesis	Description
H1	Obsolescence of the coefficients. In the previous model, the same sample of companies was used as those in the initial Altman model carried out many years earlier.
H2	Estimation method. In the previous model, the method used for estimation is the MDA, which is based on the least squares method and requires the data to comply with the principles of multinormality, homoeasticity and linearity, which is not always possible.
H3	Bankruptcy year. Year in which the bankruptcy occurs. In contrast to the previous model, whose sample corresponds to a long period of time, specifically 1946–1965, the year is considered.
H4	Size of the company. The previous model did not include data from very small or very large companies.
H5	Age of the company. This is an additional contribution to previous model, since it does not explicitly consider the age of the company, even though it does influence the bankruptcy. Bankruptcy is more likely for young companies.
H6	Company sector or industry. The previous model was made for the productive industrial sector. The sector or industry influences the financial analysis and bankruptcy of companies [49].
H7	Country of origin. It allows the model to be adjusted to the country, since samples of companies from various countries have been used (USA, China, Colombia and 31 European countries)

Logistic regression expression was used to create this new Z-score model, since it does not require the data to comply with homoscedasticity, normality, and collinearity, which is required in previous Z-score models because they are obtained by MDA. The score of the new Z-score model is obtained with Equation (4), whose value represents the likelihood of failure, where the value 1 is the maximum and 0 the minimum.

$$Z = \frac{1}{1 + e^{-L}} \tag{4}$$

The value of the variable L in Equation (4) is obtained from a linear expression that depends on the hypotheses considered in the model. There are seven hypotheses, one for each model. Equation (5) represents the expression for the set of all the hypotheses at once, which has been called the eighth model of the Altman Z-score logistic regression.

$$L = -13.302 - 0.459 X_1 - 1.160 X_2 - 1.682 X_3 - 0.013 X_4 - 0.034 D_1 - 0.150 D_2 - 0.631 D_3 + 1.837 S_1 - 0.061 S_2 + 0.186 A_1 - 0.099 A_2 - 0.628 I_1 + 0.365 I_2 - 0.157 I_3 - 0.176 I_4 + 0.095 I_5 - 0.472 I_6 - 0.915 I_7 - 0.014 C_1 \quad (5)$$

The variables X_i in Equation (5) correspond to the Table 1. The dummy variable D_i refers to the latest year of available data in the accounting statement of the company, so that the values of D_i are, if the year is equal to or earlier than 2008 ($D_1 = 1, D_2 = 0, D_3 = 0$), if it is 2009 ($D_1 = 0, D_2 = 1, D_3 = 0$), and if it is equal to or greater than 2010 ($D_1 = 0, D_2 = 0, D_3 = 1$). The dummy variables I_i are zero unless the company corresponds to one of these industries: restaurants and hotels ($I_1 = 1$), building ($I_2 = 1$); wholesale and retailing ($I_3 = 1$); agriculture ($I_4 = 1$); manufacturing ($I_5 = 1$); energy and water production ($I_6 = 1$); or information technology ($I_7 = 1$). The dummy variable C_1 refers to the country risk of the company's home country and is based on Standard & Poor's country risk rating. This variable rates the payment risk of a country and implicitly assesses the economic situation of the country. Based on the S&P country risk value, the value of C_1 is: AAA = 1, AA+ = 2, AA = 3, AA- = 4, A+ = 5, A = 6, A- = 7, BBB+ = 8, BBB = 9, BBB- = 10, BB+ = 11, BB = 12, BB- = 13, B+ = 14, B = 15, B- = 16, CCC+ = 17, CCC = 18, CCC- = 19, CC = 20, C = 21, D = 22.

From the publication of the first Z-score model, 50 years have passed, and the Altman Z-score model has become a standard for estimating the probability of bankruptcy to enable analysis for credit granting, investments and mergers and acquisitions purposes [29].

A review of the academic literature indicates that the original model was developed for quoted companies, while the model Z' was focused on private manufacturing companies, and both models (Z' and the Z-score 2014) were tailored to different companies (private and quoted) and different business purposes (manufacturing or non-manufacturing) [50].

The Altman model has been used in many academic papers. However, a review of the latest scientific papers reveals that some of the older Z-score models are used, rather than the latest version which is more complete and updated. It is uncertain why the scientific community continues to use an older model, when the renewed model has been proven to be more accurate [51].

A possible explanation could be the industry typology of the companies in the study, but the latest Z-Score 2014 model, which is valid for all company typologies, should be used in any case. Another possible cause could be the lack of values that allow the categorisation of the outcome of the probability of bankruptcy risk into several zones, as the previous models have. This is one of the contributions of this article.

3. Methodology

The methodology consists of a statistical study of the latest version of the Altman bankruptcy predictor on a dataset of Spanish firms from different industries over a period of time.

The sample of dataset is based on 1379 Spanish companies (73% are public limited companies and 27% are limited liability companies) that have accounting data for the annual period between 2010 and 2013. The data for this sample have been obtained from the SABI (Iberian Balance Sheet Analysis System) database.

The SABI database has classified the main activity of the companies according to the Spanish National Classification of Economic Activities (CNAE 2009). In order to apply the Z-Score model [43], the classification of this model must be used, so an equivalence has been made between both codifications. Table 3 shows this equivalence between the CNAE

2009 classification and the Z-score. This equivalence is simple, since there is an almost univocal equivalence between both codifications. Thus, the companies in the study are grouped into seven main industries, and the remaining companies are grouped into the “others” category.

Table 3. Table of equivalences of economic activities between CNAE and Z-score Altman 2014. Source: own elaboration.

CNAE 2009		Z-Score Altman 2014 [43]	
I	Restaurants and hotels	1	Restaurants and hotels
F	Building	2	Building
G	Wholesale and retailing	3	Wholesale and retailing
A	Agriculture, farming, forestry and fishing	4	Agriculture
C	Manufacturing industry	5	Manufacturer
D, E	Supply of electricity, gas, steam and air conditioning; water supply, sanitation, waste management and decontamination activities	6	Energy and water
J	Information and communications	7	Information Technology
B, H, K, L, M, N, O, P, Q, R, S, T, U	Extractive industries; financial and insurance activities; real estate activities; professional, scientific and technical activities; administrative and support service activities; public administration and defence; compulsory social security; education; human health and social work activities; arts, entertainment and recreation; service activities; activities of households as employers of domestic personnel; activities of households as producers of goods and services for own use; activities of extraterritorial organisations and bodies	8	Others

4. Empirical Study of the Eighth Model of the Altman Z-Score Logistic Regression

In total there is a large sample of data, consisting of 5516 financial statements of Spanish companies from different industries on which the calculations have been performed to obtain the values for the Z"-score model [36,37] and the eighth model of the Altman Z-score [43].

The bankruptcy financial classification of the companies according to the Z" version has been used to obtain the values of the companies for each of the classifications with the eighth model of the Altman Z-Score Logistic Regression. The results of these calculations are contained in Table 4.

Table 4. Table of the descriptive values of the eighth-model of the Altman Z-Score Logistic Regression of the research sample based on the bankruptcy financial classification of Z"-score model. Source: own elaboration.

Financial Classification of the Company Z"-Score Model [36,37]	N	Average	Standard		The 95% Confidence Interval for the Mean		Min.	Max.
			Deviation	Error	Lower	Upper		
Likely bankruptcy	100	0.315033	0.1340710	0.0134071	0.288431	0.341636	0.0833	0.5865
Caution	419	0.274204	0.1140356	0.0055710	0.263254	0.285155	0.0583	0.5665
Healthy	4997	0.222384	0.1051514	0.0014875	0.219468	0.225300	0.0020	0.5958
Total	5516	0.228000	0.1079477	0.0014535	0.225151	0.230849	0.0020	0.5958

The average values of the probability of bankruptcy, according to the latest Z score of each grouping of financial statements of the companies, have allowed us to obtain the boundaries of each of the three categories (likely bankruptcy, caution and healthy) that have allowed us to answer the research question RQ2.

Table 5 shows the summary obtained, which establishes the equivalence of the groupings for both bankruptcy models. Although the probability of bankruptcy's value ranges from 0 to 1, a company enters the probability of bankruptcy category from a low value of 0.315 onwards.

Table 5. Equivalence of the financial classification between the Z"-Score model and the renewed measurement of the eighth model of the Altman Z-Score. Source: own elaboration.

Financial Classification of the Company	Z"-Score [36,37]	The Eighth Model of the Altman Z-Score [43]
Healthy	$2.6 < Z''\text{-score}$	$0 \leq Z\text{-score} < 0.2223$
Caution	$1.1 \leq Z''\text{-score} \leq 2.6$	$0.2223 \leq Z\text{-score} \leq 0.3150$
Likely bankruptcy	$Z''\text{-score} < 1.1$	$0.3150 < Z\text{-score} < 1$

The statistical results of the eighth model of the Altman Z-Score Logistic Regression are shown in Table 6. This table is ordered from the lowest to highest probability of failure for each industry in order to answer research question RQ1.

Table 6. Measurement of the eighth model of the Altman Z-score for different industries. Source: own elaboration.

Industries	N	Average	Standard		95% of the Confidence Interval for the Mean		Min.	Max.
			Deviation	Error	Lower	Upper		
Information technology	236	0.14	0.06	0.00	0.13	0.14	0.02	0.33
Restaurants and hotels	124	0.18	0.07	0.01	0.17	0.20	0.08	0.42
Agriculture	56	0.20	0.06	0.01	0.18	0.21	0.11	0.39
Wholesale and retailing	1344	0.21	0.09	0.00	0.21	0.22	0.05	0.47
Energy and water	176	0.25	0.07	0.00	0.24	0.26	0.10	0.39
Manufacturer	2376	0.27	0.09	0.00	0.27	0.27	0.06	0.59
Building	316	0.37	0.10	0.01	0.36	0.38	0.10	0.60
Others	888	0.12	0.05	0.00	0.11	0.12	0.00	0.31
Total	5516	0.23	0.11	0.00	0.23	0.23	0.00	0.60

Companies in the construction industry supply chain have the highest probability of bankruptcy, which is in the probable bankruptcy category. Next, the manufacturing and energy and water industries are in the financial caution category. Then, in the healthy category, according to the probability of risk, are the following industries: wholesale and retailing, agriculture, restaurant and hotels and information technology industries, respectively.

The grouping of industries categorized as others has the lowest probability of bankruptcy of all.

A comparison of the maximum values for each industry shows that the ranking remains the same, although it can be concluded that within each industry there is a deviation close to the value 0.1, which means that there are very different financial situations for each one.

5. Discussion

This study introduces the convenience of including a new indicator for measuring the performance of the supply chain (SCPM) to assess the probability of company bankruptcy. These SCPM indicators can be shared among supply chain companies to improve supply chain operations and allow for a more efficient supply chain. In addition, significant differences in the probability of bankruptcy have been found for each of the industries.

The information technology industry's supply chain is comprised of technology services that are experiencing rapid progress in cloud computing to reduce costs, as well as technology products that enable end customers to gain cost efficiencies in their information flows [52,53]. This increased efficiency for the ultimate customer could be leading to greater liquidity and financial credit for companies in this industry, which reduces the likelihood of company bankruptcy and the disruption of the company's industry supply chain.

The restaurant and hotel industry supply chain is related to the economic power of the final consumers and the economic activity of other industries [54,55]. Due to the fact that the data corresponds to a growth economic moment, after the crisis of 2008, that enhanced the efficiency of the companies in the industry, which justifies the healthy situation in Table 6.

In the agricultural industry, the supply chain has been found to have more stable relationships over time, which have positive effects on chain upgrading that justify a healthy probability of bankruptcy [56–58].

The wholesale and retail supply chain has high cash requirements and therefore financial risk, and can be protected by insurance and mechanisms to enhance financial decisions with commercial and bank credits. So, an increase in bank lending rates is an indicator of the reducing the profits of these companies. The period of years of the sample of companies corresponds to low interest rates, which improves the profits of this industry and provides it with a healthy financial situation [59,60].

The energy and water supply chain involves high investment in infrastructure regulated by the country's government, requiring constant adaptations and changes in regulations that require additional capital investment. The increased investment, and the lack of updated government-regulated prices for these companies caused the margins of companies in this industry to shrink sharply. Therefore, the probability of bankruptcy increased, which caused this industry to be in the caution zone [61–63].

The manufacturing industry is composed of supply chains that have to be particularly attentive to risks in operations, as this influences the efficiency of the supply chain and potentially poses risks. Therefore, besides the product, the quality of the information and service must be improved to reduce the supply risk and avoid the likelihood of bankruptcy. This need to manage supply risk caused the industry to be in a cautionary zone [64,65].

The building industry's supply chain has numerous risks, which if not properly managed can contribute to an increased probability of bankruptcy. Moreover, managers in this industry try to use their expertise rather than analytical tools to reduce risk. The industry's specific characteristics, such as the culture and fragmentation of its supply chains, among others, have led to a reluctance to improve and reduce the complexity of supply chain management. As a result, the industry was in the zone of probable bankruptcy [66–68].

Consequently, the new indicator for SCPM recommended as a predictor is the Altman Z-score, which, despite being simple, has been shown to be capable of reliably predicting businesses' failure or bankruptcy and is currently being used by investors [69–72].

The previous Altman models (Z , Z' and Z'') were based on the multi-parameter multiple discriminant analysis (MDA) method. This method requires that the independent

variables of the parameters satisfy some requirements (normality, linearity, homoscedasticity), which are not fulfilled for the financial variables used.

The logistic or logit regression technique has been applied to the renewed Altman model [43], eliminating the limitation of the MDA models, which require that the data satisfy homoscedasticity, normality, and collinearity. Additionally, new variables have been considered that provide a more accurate model [43,73]. Despite this, academic publications are more likely to use the earlier versions of Altman, as they are more familiar and easier to use.

The new revision of Altman's Z-score model is considered to have some parameters that are already outdated. For example, a dummy variable refers to the last year with data available from the accounting status of the company, and is only referenced for the years 2007 to 2010, so if a study is made at present this variable will be zero for all companies, since many of them will have accounting statements for 2010 unless they were newly created. In the same way, the variable for considering the company's country of origin is supported by Standard & Poor's country risk rating, but it has not been established what to do if this entity decides to introduce new ratings, as happened in the last crisis of 2008.

The introduction into the Altman's 2014 Z-score model of a variable reflecting the age of the company and another variable reflecting the country risk is considered positive, as it overcomes the myopia of discriminatory methods on the time and macroeconomic axes.

This research is one of the earliest to apply Altman's 2014 Z-score model [43], to perform a ranking of the probability of bankruptcy for supply chain firms of different industries. Furthermore, this study provides for the first time a ranking of the financial probability of bankruptcy of companies for this new z-score model.

The main limitations of this study are that it was conducted for Spanish companies during a period of time after the 2008 crisis.

6. Conclusions

This article fills the gap in the financial perspective of supply chain improvement performance measurement (SCPM), related to the lack of a bankruptcy predictor, which allows the company to summarize the probability of bankruptcy and compare itself with other companies in its industry and in the supply chain sector of this company. For SCPM, the inclusion of an indicator to predict the probability of bankruptcy of the company, such as the eighth model of the Altman Z-score, is recommended.

In addition, a ranking of the bankruptcy probability by industry has been established, based on a study of a large sample of financial statements of Spanish companies, which allows us to answer research question RQ1. This ranking can be used to benchmark companies in an industry's supply chain, since it can be compared with the average value obtained for the probability of bankruptcy for each case. A very large empirical sample of Spanish firms has been used, which has allowed for greater precision in the results obtained.

Furthermore, this paper is a pioneer in the classification of firms into three categories (healthy, cautious and likely bankruptcy) for the prediction of financial distress using Altman's new Z-score model, which has answered research question RQ2.

Future studies can be suggested, such as the analysis of the evolution of the companies studied and which ones became bankrupt after 2014. Similarly, the strategic behaviour of each company can be reviewed to determine the influence of the company's bankruptcy on the supply chain of the industry to which it belongs.

Author Contributions: Conceptualization, R.A., C.A.d.A. and S.G.; methodology, R.A.; software, R.A.; validation, R.A., C.A.d.A. and S.G.; formal analysis, R.A.; investigation, R.A.; resources, R.A.; data curation, R.A.; writing—original draft preparation, R.A.; writing—review and editing, R.A., C.A.d.A. and S.G.; visualization, R.A.; supervision, R.A., C.A.d.A. and S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from the SABI (Iberian Balance Sheet Analysis System) database.

Acknowledgments: The authors would like to thank the editors and anonymous reviewers of this journal.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Mentzer, J.T.; DeWitt, W.; Keebler, J.S.; Min, S.; Nix, N.W.; Smith, C.D.; Zacharia, Z.G. Defining Supply Chain Management. *J. Bus. Logist.* **2001**, *22*, 1–25. [\[CrossRef\]](#)
- Min, S.; Zacharia, Z.G.; Smith, C.D. Defining Supply Chain Management: In the Past, Present, and Future. *J. Bus. Logist.* **2019**, *40*, 44–55. [\[CrossRef\]](#)
- Akyuz, G.A.; Erkan, T.E. Supply Chain Performance Measurement: A Literature Review. *Int. J. Prod. Res.* **2009**, *48*, 5137–5155. [\[CrossRef\]](#)
- Gopal, P.; Thakkar, J. A Review on Supply Chain Performance Measures and Metrics: 2000–2011. *Int. J. Prod. Perform. Manag.* **2012**, *61*, 518–547. [\[CrossRef\]](#)
- Maestrini, V.; Luzzini, D.; Maccarrone, P.; Caniato, F. Supply Chain Performance Measurement Systems: A Systematic Review and Research Agenda. *Int. J. Prod. Econ.* **2017**, *183*, 299–315. [\[CrossRef\]](#)
- Khan, S.A.R.; Yu, Z.; Golpira, H.; Sharif, A.; Mardani, A. A State-of-the-Art Review and Meta-Analysis on Sustainable Supply Chain Management: Future Research Directions. *J. Clean. Prod.* **2021**, *278*, 123357. [\[CrossRef\]](#)
- Santos, D.D.A.; Quelhas, O.L.G.; Gomes, C.F.S.; Zotes, L.P.; França, S.L.B.; De Souza, G.V.P.; De Araújo, R.A.; Santos, S.D.S.C. Proposal for a Maturity Model in Sustainability in the Supply Chain. *Sustainability* **2020**, *12*, 9655. [\[CrossRef\]](#)
- Santos, D.D.A.; Quelhas, O.L.G.; Gomes, C.F.S.; Filho, J.R.D.F. Theoretical Proposal for an Integrated Sustainability Performance Measurement System in the Supply Chain. *Front. Sustain. Food Syst.* **2021**, *2*, 78. [\[CrossRef\]](#)
- Jahani, H.; Alavifard, F.; Ivanov, D.; Ghasemishabankareh, B. Managing the risk of supply chain bankruptcy in supply chain network redesign. *IFAC-PapersOnLine* **2019**, *52*, 2431–2436. [\[CrossRef\]](#)
- Shi, Y.; Li, X. An Overview of Bankruptcy Prediction Models for Corporate Firms: A Systematic Literature Review. *Intang. Cap.* **2019**, *15*, 114–127. [\[CrossRef\]](#)
- Yang, S.A.; Birge, J.; Parker, R. The Supply Chain Effects of Bankruptcy. *Manag. Sci.* **2015**, *61*, 2320–2338. [\[CrossRef\]](#)
- Craighead, C.W.; Hult, G.T.M.; Ketchen, D.J. The Effects of Innovation-Cost Strategy, Knowledge, and Action in the Supply Chain on Firm Performance. *J. Oper. Manag.* **2009**, *27*, 405–421. [\[CrossRef\]](#)
- Prusak, B. Review of Research into Enterprise Bankruptcy Prediction in Selected Central and Eastern European Countries. *Int. J. Financ. Stud.* **2018**, *6*, 60. [\[CrossRef\]](#)
- Karan, M.B.; Ulucan, A.; Kaya, M. Credit Risk Estimation Using Payment History Data: A Comparative Study of Turkish Retail Stores. *Cent. Eur. J. Oper. Res.* **2012**, *21*, 479–494. [\[CrossRef\]](#)
- Zhang, T.; Zhang, C.Y.; Pei, Q. Misconception of Providing Supply Chain Finance: Its Stabilizing Role. *Int. J. Prod. Econ.* **2019**, *213*, 175–184. [\[CrossRef\]](#)
- Alikhani, R.; Torabi, S.A.; Altay, N. Strategic Supplier Selection under Sustainability and Risk Criteria. *Int. J. Prod. Econ.* **2019**, *208*, 69–82. [\[CrossRef\]](#)
- Xu, X.; Sun, Y.; Hua, Z. Reducing the Probability of Bankruptcy Through Supply Chain Coordination. *IEEE Trans. Syst. Man Cybern. Part. C Appl. Rev.* **2009**, *40*, 201–215. [\[CrossRef\]](#)
- Gelsomino, L.M.; Mangiaracina, R.; Perego, A.; Tumino, A. Supply Chain Finance: A Literature Review. *Int. J. Phys. Distrib. Logist. Manag.* **2016**, *46*, 348–366. [\[CrossRef\]](#)
- Yan, N.; He, X. Optimal Trade Credit with Deferred Payment and Multiple Decision Attributes in Supply Chain Finance. *Comput. Ind. Eng.* **2020**, *147*, 106627. [\[CrossRef\]](#)
- Huang, R.; Ren, S. Do Credit Ratings Incorporate Business Linkages Along the Supply Chain? *SSRN Electron. J.* **2017**. [\[CrossRef\]](#)
- Wang, Y.; Wallace, S.W.; Shen, B.; Choi, T.-M. Service Supply Chain Management: A Review of Operational Models. *Eur. J. Oper. Res.* **2015**, *247*, 685–698. [\[CrossRef\]](#)
- Liu, W.; Bai, E.; Liu, L.; Wei, W. A Framework of Sustainable Service Supply Chain Management: A Literature Review and Research Agenda. *Sustainability* **2017**, *9*, 421. [\[CrossRef\]](#)
- Alcalde-Delgado, R.; Sáiz-Bárcena, L.; Olmo, R.; Pérez, C.A.D.A. Empirical Study of the Business Growth Strategy Related to the Added Value by Intellectual Capital. *Int. J. Prod. Manag. Eng.* **2020**, *8*, 1–12. [\[CrossRef\]](#)
- DuHadway, S.; Carnovale, S.; Hazen, B. Understanding Risk Management for Intentional Supply Chain Disruptions: Risk Detection, Risk Mitigation, and Risk Recovery. *Ann. Oper. Res.* **2019**, *283*, 179–198. [\[CrossRef\]](#)
- FitzPatrick, P.J. A Comparison of the Ratios of Successful Industrial Enterprises With Those of Failed Companies. *Certif. Public Account.* **1932**, *12*, 598–729.
- Elwell, F.H. Changes in the Financial Structure of Unsuccessful Industrial Corporations (Book). *Account. Rev.* **1936**, *11*, 87–88.

27. Beaver, W.H. Financial Ratios As Predictors of Failure. *J. Account. Res.* **1966**, *4*, 71–111. [[CrossRef](#)]
28. Altman, E.I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *J. Financ.* **1968**, *23*, 589–609. [[CrossRef](#)]
29. Altman, E.I. Applications of Distress Prediction Models: What Have We Learned After 50 Years from the Z-Score Models? *Int. J. Financ. Stud.* **2018**, *6*, 70. [[CrossRef](#)]
30. Campillo, J.P.; Serer, G.L.; Ferrer, E.V. Validez de la Información Financiera en los Procesos de Insolvencia. Un Estudio de la Pequeña Empresa Española. *Cuad. Econ. Dir. Empresa* **2013**, *16*, 29–40. [[CrossRef](#)]
31. Du Jardin, P. A Two-Stage Classification Technique for Bankruptcy Prediction. *Eur. J. Oper. Res.* **2016**, *254*, 236–252. [[CrossRef](#)]
32. Mai, F.; Tian, S.; Lee, C.; Ma, L. Deep Learning Models for Bankruptcy Prediction Using Textual Disclosures. *Eur. J. Oper. Res.* **2019**, *274*, 743–758. [[CrossRef](#)]
33. Ellinger, A.E.; Natarajarithnam, M.; Adams, F.G.; Gray, J.B.; Hofman, D.; O'Marah, K. Supply Chain Management Competency and Firm Financial Success. *J. Bus. Logist.* **2011**, *32*, 214–226. [[CrossRef](#)]
34. Calandro, J. Considering the Utility of Altman's Z-Score as a Strategic Assessment and Performance Management Tool. *Strat. Leadersh.* **2007**, *35*, 37–43. [[CrossRef](#)]
35. Ashraf, S.; Félix, E.G.S.; Serrasqueiro, Z. Do Traditional Financial Distress Prediction Models Predict the Early Warning Signs of Financial Distress? *J. Risk Financ. Manag.* **2019**, *12*, 55. [[CrossRef](#)]
36. Altman, E.I. *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1993; ISBN 978-04-7155-253-6.
37. Altman, E.I.; Hotchkiss, E. *Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt*, 3rd ed.; John Wiley & Sons Inc.: New York, NY, USA, 2011; ISBN 978-04-7169-189-1.
38. Siddiqui, S.A. Business Bankruptcy Prediction Models: A Significant Study of the Altman's Z-Score Model. *Asian J. Manag. Res.* **2012**, *3*, 212–219. [[CrossRef](#)]
39. Panigrahi, A. Validity of Altman's 'Z' Score Model in Predicting Financial Distress of Pharmaceutical Companies. *J. Econ. Public Policy* **2019**, *4*, 65–73.
40. Alali, M.S. Examining the Effect of Altman's Zeta Model Score on the Share Price of Healthcare Companies Listed at Kuwait Stock Exchange. *Int. J. Econ. Financ.* **2019**, *11*, 25. [[CrossRef](#)]
41. Csikósová, A.; Janošková, M.; Čulková, K. Prediction of Developments in the Textile and Clothing Industry in Slovakia by Selected Indicators of Financial Analysis. *Fibres Text. East. Eur.* **2019**, *27*, 9–16. [[CrossRef](#)]
42. Nandi, A.; Sengupta, P.P.; Dutta, A. Diagnosing the Financial Distress in Oil Drilling and Exploration Sector of India through Discriminant Analysis. *Vis. J. Bus. Perspect.* **2019**, *23*, 364–373. [[CrossRef](#)]
43. Altman, E.I.; Iwanicz-Drozdowska, M.; Laitinen, E.K.; Suvas, A. Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Soc. Sci. Electron. Publ.* **2014**, 1–48. [[CrossRef](#)]
44. González-Martín, J.M.; Sánchez-Medina, A.J.; Alonso, J.B.; González-Martín, J.M. Optimización de la Predicción de Problemas Financieros en Empresas Sanitarias Privadas Españolas Aplicando Algoritmos Genéticos. *Gac. Sanit.* **2018**, *33*, 462–467. [[CrossRef](#)]
45. Hosaka, T. Bankruptcy Prediction Using Imaged Financial Ratios and Convolutional Neural Networks. *Expert Syst. Appl.* **2019**, *117*, 287–299. [[CrossRef](#)]
46. Muñoz-Izquierdo, N.; Laitinen, E.K.; Camacho-Miñano, M.; Pascual-Ezama, D. Does Audit Report Information Improve Financial Distress Prediction over Altman's Traditional Z-Score Model? *J. Int. Financ. Manag. Account.* **2019**, *31*, 65–97. [[CrossRef](#)]
47. Alfaro, L.; Asis, G.; Chari, A.; Panizza, U. Corporate Debt, Firm Size and Financial Fragility in Emerging Markets. *J. Int. Econ.* **2019**, *118*, 1–19. [[CrossRef](#)]
48. Barboza, F.; Kimura, H.; Altman, E. Machine Learning Models and Bankruptcy Prediction. *Expert Syst. Appl.* **2017**, *83*, 405–417. [[CrossRef](#)]
49. Smith, M.; Liou, D. Industrial Sector and Financial Distress. *Manag. Audit. J.* **2007**, *22*, 376–391. [[CrossRef](#)]
50. Altman, E.I.; Hotchkiss, E.; Wang, W. *Corporate Financial Distress, Restructuring, and Bankruptcy*, 4th ed.; John Wiley & Sons, Inc.: New York, NY, USA, 2019; ISBN 978-11-1954-192-9.
51. Lin, H. Default Prediction Model for SME's: Evidence from UK Market Using Financial Ratios. *Int. J. Bus. Manag.* **2015**, *10*, 81. [[CrossRef](#)]
52. Ferguson, D.F.; Hadar, E. Optimizing the IT Business Supply Chain Utilizing Cloud Computing. In Proceedings of the 2011 8th International Conference & Expo on Emerging Technologies for a Smarter World, Long Island, NY, USA, 2–3 November 2011; pp. 1–6.
53. Reyes, P.; Raisinghani, M.S.; Singh, M. Global Supply Chain Management in the Telecommunications Industry: The Role of Information Technology in Integration of Supply Chain Entities. *J. Glob. Inf. Technol. Manag.* **2002**, *5*, 48–67. [[CrossRef](#)]
54. Chen, Y.J.; Tsai, H.; Liu, Y.-F. Supply Chain Finance Risk Management. *Tour. Econ.* **2018**, *24*, 593–614. [[CrossRef](#)]
55. Alberca, P.; Parte, L. Operational Efficiency Evaluation of Restaurant Firms. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 1959–1977. [[CrossRef](#)]
56. van Bergen, M.; Steeman, M.; Reindorp, M.; Gelsomino, L. Supply Chain Finance Schemes in the Procurement of Agricultural Products. *J. Purch. Supply Manag.* **2019**, *25*, 172–184. [[CrossRef](#)]

57. Ye, H.; Hu, X.; Xie, M. Research on Supply Chain Finance Model based on Agricultural Logistics Park Information Platform. In Proceedings of the 2018 International Symposium on Social Science and Management Innovation (SSMI 2018), Xi'an, China, 15–16 December 2018; Atlantis Press: Xi'an, China, 2019; pp. 226–230.
58. Sun, Y.; Liu, Z.; Yang, H. How Does Suppliers' Fairness Affect the Relationship Quality of Agricultural Product Supply Chains? *J. Food Qual.* **2018**, *2018*, 1–15. [[CrossRef](#)]
59. Xing, Y.; Zhongkai, W. Supply Chain Financing Model under a New Mechanism of Bankruptcy Guarantee. *Econ. Comput. Econ. Cybern. Stud. Res.* **2020**, *54*, 243–262. [[CrossRef](#)]
60. Lin, Q.; Xiao, Y. Retailer Credit Guarantee in a Supply Chain with Capital Constraint under Push & Pull Contract. *Comput. Ind. Eng.* **2018**, *125*, 245–257. [[CrossRef](#)]
61. Ameyaw, E.E.; Chan, A.P.C.; Owusu-Manu, D.-G.; Edwards, D.J.; Dartey, F. A Fuzzy-Based Evaluation of Financial Risks in Build-Own-Operate-Transfer Water Supply Projects. *J. Infrastruct. Syst.* **2017**, *23*, 04017033. [[CrossRef](#)]
62. Jiang, Z.-Z.; He, N.; Xiao, L.; Sheng, Y. Government Subsidy Provision in Biomass Energy Supply Chains. *Enterp. Inf. Syst.* **2019**, *13*, 1367–1391. [[CrossRef](#)]
63. Ye, F.; Xie, Z.; Cai, Z.; Lin, Q. Optimization of the Biofuel Supply Chain With Capital-Constrained Farmers Under Government Subsidies. *IEEE Access* **2019**, *8*, 8178–8192. [[CrossRef](#)]
64. Truong, H.Q.; Hara, Y. Supply Chain Risk Management: Manufacturing-and Service-Oriented Firms. *J. Manuf. Technol. Manag.* **2018**, *29*, 218–239. [[CrossRef](#)]
65. Benedettini, O.; Swink, M.; Neely, A. Examining the Influence of Service Additions on Manufacturing Firms' Bankruptcy Likelihood. *Ind. Mark. Manag.* **2017**, *60*, 112–125. [[CrossRef](#)]
66. Shojaei, P.; Haeri, S.A.S. Development of Supply Chain Risk Management Approaches for Construction Projects: A Grounded Theory Approach. *Comput. Ind. Eng.* **2019**, *128*, 837–850. [[CrossRef](#)]
67. Panova, Y.; Hilletoth, P. Managing Supply Chain Risks and Delays in Construction Project. *Ind. Manag. Data Syst.* **2018**, *118*, 1413–1431. [[CrossRef](#)]
68. Aloini, D.; Dulmin, R.; Mininno, V.; Ponticelli, S. Supply Chain Management: A Review of Implementation Risks in the Construction Industry. *Bus. Process. Manag. J.* **2012**, *18*, 735–761. [[CrossRef](#)]
69. Giroux, G.; Cassell, C. Changing Audit Risk Characteristics in the Public Client Market. *Res. Account. Regul.* **2011**, *23*, 177–183. [[CrossRef](#)]
70. Chanos, J. Short-Lived Lessons from an Enron Short. Available online: <https://www.wsj.com/articles/SB114894232503965715> (accessed on 10 May 2021).
71. Swalih, M.M.; Vinod, M.S. Application Of Altman Z Score on BSE-Greenex Companies. *J. Appl. Manag. Invest.* **2017**, *6*, 205–215.
72. Srebro, B.; Mavrenski, B.; Arsić, V.B.; Knežević, S.; Milašinović, M.; Travica, J. Bankruptcy Risk Prediction in Ensuring the Sustainable Operation of Agriculture Companies. *Sustainability* **2021**, *13*, 7712. [[CrossRef](#)]
73. Laitinen, E.K.; Suvas, A. International Applicability of Corporate Failure Risk Models Based on Financial Statement Information: Comparisons across European Countries. *J. Financ. Econ.* **2013**, *1*, 1–26. [[CrossRef](#)]