



Facultad de ciencias económicas y empresariales

Screening de oportunidades para la inversión de *private equity* mid- market en España

Clave: 202105203

Autor: Carlos Espinosa de los Monteros Pérez-Brotóns

Tutora: Susana Josefa Gago Rodríguez

ÍNDICE

Relación de abreviaturas

Resumen

Abstract

1. Introducción

2. Marco teórico y contextual

2.1. El *private equity* y el proceso de *screening* de oportunidades

2.2. El mercado español y las particularidades del segmento *mid-market*

2.3. Limitaciones informativas en fases iniciales y uso de datos contables y *machine learning*

3. Datos y metodología

3.1. Planteamiento metodológico y objetivo del trabajo

3.2. Base de datos, universo de análisis y criterios de selección

3.3. Construcción de variables y diseño de los indicadores de entrada

3.4. Desarrollo de los modelos predictivos

3.4.1. Modelo de predicción de ingresos

3.4.2. Modelo de predicción de EBITDA

3.5. Criterios de evaluación y comparación de modelos

3.6. Diseño del score de priorización y construcción del ranking

4. Resultados

4.1. Resultados del modelo de predicción de ingresos

4.2. Resultados del modelo de margen EBITDA

4.3. Construcción del ranking final e integración de las predicciones

4.4. Caracterización geográfica y sectorial del Top 100

4.4.1. Distribución geográfica del Top 100 frente a la base limpia

4.4.2. Distribución sectorial del Top 100 frente a la base limpia

5. Discusión

5.1. Interpretación de la capacidad predictiva de los modelos

5.2. Encaje del ranking final con la lógica del *private equity mid-market*

5.3. Utilidad práctica, limitaciones y posibles mejoras del enfoque

6. Conclusiones

Declaración de IA

Bibliografía

Anexo 1. Top 10 compañías ranking final

Anexo 2. Proceso de extracción y depuración de la muestra

Anexo 3. Especificación técnica de los modelos

Anexo 4. Construcción del score de priorización

Anexo 5. Fragmentos de código relevantes

Relación de abreviaturas

IA: Inteligencia artificial.

CAGR: Compound Annual Growth Rate – Tasa de crecimiento anual compuesta.

CNAE: Clasificación Nacional de Actividades Económicas.

EBIT: Earnings Before Interest and Taxes – Resultado antes de intereses e impuestos.

EBITDA: Earnings Before Interest, Taxes, Depreciation and Amortization – Resultado antes de intereses, impuestos, depreciaciones y amortizaciones.

EV: Enterprise Value – Valor empresa.

GP: General Partner – Socio gestor del fondo.

ICT: Information and Communication Technologies – Tecnologías de la información y la comunicación.

IRR / TIR: Internal Rate of Return – Tasa interna de retorno.

LBO: Leveraged Buyout – Adquisición apalancada.

LP: Limited Partner – Inversor del fondo / socio comanditario.

MAE: Mean Absolute Error – Error absoluto medio.

ML: Machine Learning – Aprendizaje automático.

MOIC: Multiple on Invested Capital – Múltiplo sobre el capital invertido.

OWC: Operating Working Capital – Capital circulante operativo.

PE: Private Equity – Capital riesgo.

RMSE: Root Mean Squared Error – Raíz del error cuadrático medio.

R²: Coefficient of Determination – Coeficiente de determinación.

SABI: Sistema de Análisis de Balances Ibéricos.

TMT: Technology, Media and Telecommunications – Tecnología, medios y telecomunicaciones.

VC: Venture Capital – Capital riesgo orientado a empresas en fases iniciales.

Resumen

Este trabajo desarrolla una herramienta de *screening* orientada a la identificación y priorización de oportunidades de inversión en el *private equity mid-market* español. A partir de una base de datos de empresas no cotizadas obtenida de SABI, se construye un modelo basado en variables contables históricas para estimar el margen EBITDA y el crecimiento de ingresos, así como un score agregado que permite ordenar las compañías según su atractivo potencial.

Los resultados muestran que la información contable es más útil para estimar la rentabilidad operativa que para predecir el crecimiento, lo que justifica un mayor peso del margen EBITDA en el ranking final. La composición del Top 100 obtenido resulta coherente con las tendencias recientes del mercado, tanto a nivel sectorial como geográfico. En conjunto, el trabajo pone de manifiesto el valor de combinar herramientas cuantitativas con el criterio inversor en fases tempranas del proceso de análisis.

PALABRAS CLAVE

Private equity; screening; EBITDA; mid-market; análisis financiero; ranking de empresas

Abstract

This paper develops a screening tool aimed at identifying and prioritizing investment opportunities in the Spanish private equity mid-market. Using a dataset of private companies from SABI, a model based on historical accounting variables is built to estimate EBITDA margins and revenue growth, as well as an aggregated score to rank firms according to their potential attractiveness.

The results show that accounting information is more useful for estimating operating profitability than for predicting growth, supporting a higher weight of EBITDA margins in the final ranking. The composition of the Top 100 is broadly consistent with recent market trends, both sectorally and geographically. Overall, the paper highlights the value of combining quantitative tools with investor judgment in the early stages of the investment process.

KEYWORDS

Private equity; screening; EBITDA; mid-market; financial analysis; company ranking

1. Introducción

El proceso de inversión en *private equity* se caracteriza por la necesidad de analizar un número elevado de oportunidades en un contexto de información limitada y bajo presión de tiempo. Esta situación es especialmente relevante en el segmento *mid-market*, donde muchas compañías potencialmente invertibles no cuentan con cobertura de analistas ni con información externa suficientemente desarrollada que facilite una evaluación inicial estructurada. Como consecuencia, una parte significativa del proceso inversor se concentra en las fases tempranas de filtrado, en las que el inversor debe decidir rápidamente qué oportunidades merecen un análisis más profundo y cuáles deben descartarse.

En este contexto, el *screening* de oportunidades adquiere una importancia central. Más allá de ser una fase preliminar, constituye un punto crítico del proceso, ya que condiciona de forma directa la eficiencia del análisis posterior, la asignación de recursos del equipo y, en última instancia, la calidad de las inversiones que llegan a ejecutarse. Sin embargo, este filtrado inicial suele apoyarse en información incompleta y en criterios que, en muchos casos, no están plenamente sistematizados.

El objetivo de este trabajo es abordar este problema mediante el desarrollo de una herramienta cuantitativa de *screening* orientada a la identificación y priorización de oportunidades de inversión en el *private equity mid-market* español. A partir de una base de datos de empresas no cotizadas obtenida de SABI, se construye un modelo basado en variables contables históricas que permite, por un lado, aproximar el comportamiento futuro de determinadas magnitudes relevantes y, por otro, sintetizar esa información en un score agregado capaz de ordenar las compañías según su atractivo relativo.

La contribución del trabajo se sitúa en la intersección entre el análisis financiero y el uso de técnicas de *machine learning* en un contexto aplicado. En lugar de centrarse en la valoración de una compañía concreta, el enfoque adoptado se orienta a resolver un problema previo: cómo reducir de forma eficiente un universo amplio de empresas hasta obtener una *shortlist* manejable sobre la que concentrar el análisis detallado. En este sentido, la herramienta desarrollada no pretende sustituir el juicio del inversor, sino complementarlo mediante un proceso más estructurado y homogéneo.

El trabajo se organiza de la siguiente manera. En primer lugar, se presenta el marco teórico y contextual, en el que se describe el funcionamiento del *private equity*, las particularidades del mercado español y las limitaciones informativas propias de las fases iniciales del proceso inversor. A continuación, se detallan los datos utilizados y la metodología aplicada, incluyendo la construcción de variables, el desarrollo de los modelos predictivos y el diseño del score de priorización. Posteriormente, se exponen los principales resultados obtenidos y se analiza la composición del ranking final. Por último, se discuten las implicaciones del trabajo, sus limitaciones y posibles líneas de mejora, y se recogen las conclusiones principales.

2. Marco teórico y contextual

2.1. El *private equity* y el proceso de *screening* de oportunidades

El *private equity* puede definirse, de forma general, como la inversión con un peso accionarial relevante en compañías con el objetivo de impulsar su valor y obtener una rentabilidad en el momento de la desinversión (Kaplan & Strömberg, 2009). Dentro de este ámbito, una de las operaciones más representativas es el *leveraged buyout*, en la que la adquisición se financia mediante una combinación de capital y deuda (Kaplan & Strömberg, 2009). Frente al venture capital, que suele centrarse en empresas jóvenes o en fases iniciales de desarrollo, el *private equity* tiende a dirigirse a compañías más maduras, con modelos de negocio ya consolidados y con una mayor visibilidad sobre su capacidad de generación de caja (Cumming et al., 2023). Esta diferencia resulta relevante para este trabajo, ya que el problema que aquí se analiza no consiste en identificar startups o proyectos incipientes, sino en priorizar compañías operativas dentro de un universo amplio de potenciales objetivos de inversión.

Desde el punto de vista económico, el funcionamiento del *private equity* se articula normalmente a través de fondos cerrados gestionados por firmas especializadas, con los *General Partners* al mando, que captan capital de inversores institucionales, conocidos como *Limited Partners* (Kaplan & Strömberg, 2009). La lógica del modelo no se basa únicamente en comprar y vender empresas, sino en combinar palancas de creación de valor: el apalancamiento financiero, la mejora operativa, reflejada en un EBITDA superior al final del *holding period*, y, en algunos casos, una expansión del múltiplo (Kaplan & Strömberg, 2009). Por ello, la selección de oportunidades ocupa una posición central en el proceso inversor. Una firma puede analizar un número elevado de compañías, pero solo una parte muy reducida llegará a estudiarse en profundidad. En consecuencia, la capacidad de filtrar bien desde el inicio tiene un impacto directo sobre la eficiencia del proceso, el uso del tiempo del equipo y la calidad final de las oportunidades que avanzan a fases posteriores (Maccagni & Boiani, 2026).

Esta necesidad de seleccionar con rigor se ha intensificado en el contexto reciente del sector. Aunque la actividad de *private equity* ha mostrado cierta recuperación, el mercado sigue operando en un entorno más exigente, marcado por mayores dificultades de financiación, procesos de salida más lentos y una presión creciente sobre la creación de valor operativa (Bain & Company, 2026). Décadas atrás, tener un flujo de caja suficientemente bueno podía hacer por sí solo a una empresa un *target* atractivo, gracias principalmente a los niveles de 80/20 de *debt/equity* que se permitían en el pasado. Actualmente, los niveles de apalancamiento permitido han bajado considerablemente, lo que hace que los retornos dependan en mayor medida de la mejora operativa de la compañía, y no tanto del repago de la deuda (Kaplan & Strömberg, 2009). Bain &

Company, en su informe anual de 2026, destaca que, pese al repunte del *dealmaking*, la industria continúa enfrentándose a un escenario en el que los inversores son más selectivos y en el que la generación de crecimiento y EBITDA en cartera ha ganado todavía más importancia (Bain & Company, 2026). En la misma línea, el informe sectorial de Gain muestra una recuperación de la actividad de entrada en Europa durante el primer semestre de 2026 y confirma que sectores como *Business Services*, *Industrials* y *TMT* siguen concentrando una parte muy significativa de las operaciones (Gain, 2026). En este contexto, no basta con tener acceso a muchas oportunidades; lo verdaderamente relevante es disponer de criterios consistentes que permitan priorizar mejor, adelantarse al resto, concentrar el esfuerzo analítico en menos compañías y evitar dedicar recursos a objetivos que previsiblemente no superarán un análisis posterior más profundo.

Dentro de ese proceso, el *screening* constituye la fase inicial de filtrado y priorización de oportunidades. Se sitúa entre la recepción preliminar de información sobre una compañía y el inicio de una *due diligence* más detallada, y su función principal consiste en decidir qué oportunidades podría justificar seguir avanzando (Maccagni & Boiani, 2026). Aunque a veces pueda parecer una etapa meramente preliminar, en realidad se trata de un momento especialmente delicado, porque las decisiones se toman con poco tiempo, con información incompleta y en un contexto de elevada asimetría informativa. En estas primeras fases, el análisis suele apoyarse en documentación reducida, como un *teaser* o datos financieros resumidos, lo que favorece el uso de reglas rápidas, criterios simplificados o valoraciones apoyadas en la experiencia previa del inversor. Por ello, el *screening* no debe verse como un simple paso administrativo dentro del pipeline de inversión, sino como una fase crítica que condiciona qué compañías recibirán atención adicional y cuáles quedarán descartadas desde el inicio.

En estas etapas tempranas, los inversores suelen apoyarse sobre todo en variables financieras y contables, no porque estas recojan por sí solas toda la realidad de la empresa, sino porque en muchos casos son la información más accesible, homogénea y comparable entre compañías (Maccagni & Boiani, 2026). Magnitudes como el crecimiento, la rentabilidad, el apalancamiento, el tamaño o determinados indicadores de estructura financiera aparecen de forma recurrente como señales útiles para una evaluación inicial, ya que nos permiten formar una primera impresión sobre la calidad del activo y su posible encaje con la estrategia de inversión de una firma concreta.

El problema, sin embargo, no radica solo en qué variables observar, sino en cómo procesarlas de forma consistente cuando el universo inicial es muy amplio. Desde esta perspectiva, tiene sentido explorar herramientas basadas en datos contables que permitan ordenar, filtrar y priorizar compañías en una fase temprana. No se trata de sustituir el juicio inversor, sino de complementarlo con un apoyo más sistemático en una etapa en la que,

de otro modo, la decisión puede depender en exceso de reglas poco estructuradas o de valoraciones demasiado rápidas, además de mejorar la eficiencia y rapidez en el proceso.

2.2. El mercado español y las particularidades del segmento *mid-market*

Nuestro mercado presenta unas características que lo convierten en un contexto especialmente adecuado para un trabajo centrado en el *screening* de oportunidades de inversión. El tejido empresarial español está claramente dominado por pequeñas y medianas empresas, pymes, lo que implica la existencia de un universo muy amplio de compañías no cotizadas, con niveles muy distintos de escala, profesionalización y transparencia informativa. Las fuentes oficiales más recientes reflejan, por un lado, que a 1 de enero de 2025 había en España 3.310.824 empresas económicamente activas (INE, 2025) y, por otro, que en noviembre de 2025 se contabilizaban 2.955.770 pymes de 0 a 249 asalariados (DGPYME, 2025). Esto nos muestra la necesidad que tenemos de buscar herramientas que nos ayuden a buscar, acotar, ordenar y priorizar este amplio universo.

Esta estructura empresarial ayuda también a explicar por qué el capital privado en España ha mantenido una relación especialmente estrecha con la pyme y, dentro de ella, con el segmento *mid-market* (SpainCap, 2025). El informe de SpainCap refleja que en 2024 la inversión de *private equity* y *venture capital* en empresas españolas alcanzó los 6.292 millones de euros y se materializó en más de 785 inversiones, dirigidas en un 90% a pymes. Además, se destaca al *middle market* como el segmento más atractivo para el capital privado en España, lo que confirma que siga ocupando una posición central dentro del mercado nacional. En este sentido, el *mid-market* resulta especialmente relevante porque agrupa compañías que ya han superado la fase más inicial de desarrollo, pero que todavía conservan muchas de las características propias de una empresa no cotizada de tamaño intermedio, tanto en términos de estructura como de información disponible. Es decir, es a la vez estable y madura, pero con la capacidad de crecimiento y expansión que una cotizada no tiene.

Una de las particularidades más importantes de este segmento es que no responde a un único perfil de inversor ni a una sola forma de crear valor. Dentro del *mid-market* español conviven firmas de enfoque generalista, como Portobello Capital o ProA Capital, junto con otras firmas más especializadas o con equipos sectoriales dentro de las propias gestoras, que centran su búsqueda e inversión en sectores y verticales más concretas en las que consideran que sus capacidades diferenciales pueden traducirse en un mayor conocimiento, mejores contactos y tesis de inversión más definidas y realistas. Las propias webs corporativas de ambas firmas muestran bien esta lógica: Portobello se presenta como una firma independiente de *mid-market* que identifica compañías dinámicas y colabora activamente en su gestión, mientras que ProA define su estrategia como generalista dentro del *mid-market* no cotizado, con foco en compañías de entre 30 y 500 millones de valor

empresa. Más allá de estos ejemplos, lo relevante es que la amplitud del segmento hace que la identificación de oportunidades no sea homogénea ni sencilla.

Precisamente por eso, en el *mid-market* también conviven distintas estrategias de *value creation*. En algunos casos, el foco está en situaciones más complejas, en las que la firma inversora busca estabilizar la compañía, corregir ineficiencias, reducir costes, ordenar la estructura operativa o financiera y devolver el negocio a un camino de crecimiento y rentabilidad, las conocidas como *special situations*. En otros, la tesis de inversión se apoya menos en la reestructuración y más en planes de mejora operativa, profesionalización de la gestión, expansión comercial o crecimiento orgánico. En estos supuestos, el objetivo no es tanto arreglar una situación deteriorada como acelerar una compañía que ya funciona razonablemente bien, pero que todavía tiene margen para crecer, ganar escala o mejorar sus márgenes. Esta diferenciación encaja con el contexto actual del sector, en el que la creación de valor depende cada vez más de mejoras operativas reales y de crecimiento de EBITDA, y no solo del repago de la deuda con la caja que genera la compañía.

A ello se añade una estrategia que ha ganado todavía más peso en los últimos años, el *buy-and-build*. Esta estrategia parte de la adquisición de una compañía plataforma, es decir, una empresa con tamaño, estructura y capacidades suficientes para actuar como base sobre la que realizar adquisiciones adicionales. A partir de esa plataforma, el inversor impulsa un proceso de crecimiento inorgánico mediante la compra sucesiva de compañías más pequeñas, los llamados *add-ons* o *bolt-ons*, con el objetivo de ganar escala, ampliar cobertura geográfica, incorporar nuevas capacidades, reforzar la oferta comercial o consolidar un mercado fragmentado. La lógica económica del modelo es que la combinación de activos permita generar un grupo más grande, más eficiente y potencialmente más valioso que la suma de sus partes por separado (Borell & Heger, 2013). En sectores fragmentados, esta estrategia permite acelerar el crecimiento, capturar sinergias y construir operadores con mayor tamaño y mejor posicionamiento competitivo, lo que puede favorecer una mejor valoración a la salida y, en consecuencia, un mayor retorno para el inversor. Podemos destacar a firmas como Fremman Capital, Suma Capital o Everwood Capital que utilizan esta estrategia de forma muy recurrente para sus participadas e inversiones.

Esta heterogeneidad del *private equity mid-market* español refuerza la utilidad de herramientas de screening como la desarrollada en este trabajo, no solo para la identificación de nuevas plataformas de inversión, sino también para la localización de posibles *add-ons* o compañías complementarias dentro de estrategias de crecimiento inorgánico.

2.3. Limitaciones informativas en fases iniciales y uso de datos contables y *machine learning*

Uno de los mayores problemas en las primeras fases de una operación de *private equity* es que casi siempre hay que tomar decisiones con muy poca información (Maccagni & Boiani, 2026). Antes de entrar en una *due diligence* completa, habitualmente solo se dispone de cuentas históricas, cierta información comercial y, en el mejor de los casos, un resumen inicial de la compañía. Esto hace que el primer filtro se haga con una visión todavía muy parcial del negocio. En el *mid-market*, además, esta limitación suele ser todavía más clara, porque muchas empresas no tienen el mismo nivel de *reporting*, detalle o estandarización que una compañía grande. Además, al no ser cotizadas, no están sujetas a las exigencias de información y transparencia propias del mercado bursátil. Por eso, en esta fase inicial el inversor se enfrenta a una doble dificultad: por un lado, tiene poca información y, por otro, tiene que decidir lo antes posible cuáles merecen la pena y cuáles no.

En ese contexto, los datos contables tienen bastante valor, porque son de las pocas fuentes que permiten comparar muchas empresas de una forma relativamente homogénea. Evidentemente, unas cuentas no recogen por sí solas toda la realidad de una compañía, pero sí permiten obtener una primera fotografía de aspectos que para una firma de *private equity* son relevantes, como el crecimiento, la rentabilidad, el endeudamiento, el tamaño o la capacidad de generar caja. Además, estas variables no solo ayudan a entender cómo está la empresa hoy, sino que también sirven como punto de partida para hacerse una idea de hacia dónde puede ir en el futuro. Al final, cuando una firma analiza una oportunidad, no mira solo la foto actual, sino también si ese negocio tiene capacidad de seguir creciendo, mantener márgenes razonables o mejorar su perfil financiero y comercial con el tiempo. Desde ese punto de vista, trabajar con información contable en una primera criba no es una simplificación excesiva, sino una forma lógica de empezar a ordenar un universo muy amplio de compañías.

El verdadero problema aparece cuando ese universo inicial es demasiado grande. En la práctica, una firma puede tener miles de empresas potencialmente invertibles, pero solo unas pocas podrán analizarse a fondo, e incluso menos tener sentido como adquisición. Ahí es donde entra realmente el sentido del *screening*: no se trata de decidir desde el principio cuál será la mejor inversión, sino de evitar perder tiempo buscando una aguja en un pajar. Lo que se busca es reducir de forma rápida y razonable ese universo inicial hasta quedarse con un grupo mucho más manejable de compañías que sí merezcan una revisión más profunda. En este punto, el uso de herramientas de *machine learning* puede aportar bastante valor, porque permite procesar muchas variables al mismo tiempo, encontrar patrones que no siempre se ven a simple vista y construir una clasificación más consistente de las empresas (Maccagni & Boiani, 2026). Su utilidad, por tanto, no está en sustituir el criterio del inversor, sino en ayudarlo a filtrar mejor y a priorizar con más orden.

Desde esta perspectiva, el interés de una herramienta basada en datos contables está en que permite dos cosas a la vez. Por un lado, ayuda a construir un ranking o una priorización inicial que reduzca mucho el tiempo dedicado a revisar compañías que nunca llegarían a pasar un análisis más profundo. Por otro, permite utilizar esa misma información histórica como base para aproximarse al posible comportamiento futuro de la empresa, al menos de manera preliminar. Es sobre esta lógica que se apoya este trabajo. La idea no es reemplazar el juicio profesional ni convertir la inversión en un proceso automático, sino contar con una herramienta que permita filtrar mejor, ordenar mejor y centrar antes el esfuerzo en aquellas compañías que, por sus características financieras, parecen más interesantes dentro de un universo muy amplio.

3. Datos y metodología

3.1. Planteamiento metodológico y objetivo del trabajo

Este trabajo parte de una dificultad muy concreta dentro del proceso de inversión en *private equity*: analizar de forma eficiente un universo amplio de compañías potencialmente invertibles cuando, en una fase inicial, la información disponible es todavía limitada. Este problema es especialmente relevante en el segmento *mid-market*, donde muchas empresas pueden ser atractivas desde el punto de vista inversor, pero no cuentan con el nivel de cobertura informativa que sí existe en operaciones de mayor tamaño. En este tipo de compañías, más pequeñas y menos seguidas por el mercado, no suele haber analistas que elaboren estimaciones periódicas, ni *broker reports*, ni otra documentación externa suficientemente desarrollada que facilite una primera valoración rápida.

En la práctica, esta falta de información obliga a que el primer filtrado de oportunidades se base, en gran medida, en la interpretación de la información contable histórica. Sin embargo, cuando el universo inicial incluye varios miles de empresas, ese proceso puede convertirse en una tarea lenta y muy difícil de sistematizar. Por ello, el principal objetivo de este trabajo es diseñar una herramienta de screening que permita ordenar y priorizar compañías de forma más estructurada, reduciendo de manera significativa el número de empresas a revisar en detalle y concentrando el análisis en aquellas que, a priori, presentan un perfil más atractivo.

Junto a esa función de priorización, el trabajo incorpora una segunda dimensión. Al no existir previsiones externas fiables de este tipo de empresas, una herramienta que ayude a anticipar variables relevantes puede aportar valor real en una fase preliminar del análisis. En este sentido, se desarrollan modelos orientados a aproximar el crecimiento de ingresos y el margen EBITDA, de forma que las predicciones obtenidas complementen la información observada y además sirvan como apoyo para decidir si merece la pena profundizar o no en una determinada oportunidad de inversión. De este modo, la herramienta propuesta no sustituye el juicio del inversor, pero sí ayuda a hacerlo más ágil y homogéneo.

3.2. Base de datos, universo de análisis y criterios de selección

La base de datos utilizada en este trabajo procede de SABI, base elaborada por Informa D&B que reúne información financiera y societaria de empresas españolas. La extracción se diseñó con un criterio aplicado, buscando aproximar un universo de compañías coherente con el tipo de *target* que suele resultar relevante para una estrategia de *private equity* centrada en el *middle market* español y generalista. Por ello, la selección inicial no pretendía representar al conjunto del tejido empresarial, sino acotar desde el principio un

grupo de sociedades con un tamaño y un perfil financiero compatibles con el objetivo del estudio y con la lógica de una primera criba inversora. Así, además, la base de datos se hacía más manejable e interpretable.

En concreto, la búsqueda se limitó a compañías con ingresos de explotación comprendidos entre 5 y 250 millones de euros en todos los ejercicios del periodo 2020-2024, un EBITDA mínimo de 1 millón de euros en 2024 y una plantilla de entre 20 y 2.500 empleados. Además, se exigió que las empresas presentaran EBITDA no negativo y EBIT no negativo en todos los ejercicios entre 2020 y 2024, fondos propios positivos en 2024, deuda financiera no negativa en 2024, tesorería no negativa en 2024 y cuentas no consolidadas identificadas en SABI con código U1. Tras aplicar este conjunto de filtros, la búsqueda inicial devolvió 5.834 compañías.

La lógica de estos criterios responde al propio enfoque del trabajo. El rango de ingresos permite excluir tanto empresas demasiado pequeñas como compañías de un tamaño excesivo para el foco *mid-market*. A su vez, la exigencia de rentabilidad operativa positiva y de fondos propios no negativos ayuda a dejar fuera perfiles claramente *distress* o situaciones contables extremas, ya que el objetivo no es estudiar empresas en dificultades, sino construir una herramienta de priorización dentro de un universo de compañías operativamente viables y comparables para una primera criba inversora. Aunque el *private equity* puede incluir estrategias de *special situations* o *distress investing*, este trabajo no se centra en ese tipo de operaciones, sino en una lógica generalista de *screening* sobre compañías operativamente viables.

A partir de esta extracción se descargaron variables de identificación y clasificación, como nombre, NIF, comunidad autónoma y código primario CNAE 2009, junto con una variable de tamaño, el número de empleados en 2024. También se incorporaron las principales variables financieras necesarias para el análisis: ingresos de explotación, EBITDA y EBIT entre 2020 y 2024; deuda financiera y tesorería en 2024; y partidas de circulante, en concreto existencias, deudores y acreedores comerciales, para ese mismo periodo. Posteriormente, la base fue depurada mediante la homogeneización de nombres de variables, la conversión de las columnas relevantes a formato numérico y la revisión de posibles duplicados. Finalmente, a partir del CNAE se excluyeron las compañías pertenecientes a los códigos 64, 65 y 66, correspondientes a actividades financieras y de seguros, sectores cuyas cuentas anuales y cuya lógica contable, especialmente en la estructura de activo y pasivo, difieren sustancialmente de las del resto de empresas. Tras esta última depuración, la muestra de trabajo quedó formada por 5.798 compañías, sobre las que se desarrollan los modelos.

3.3. Construcción de variables y diseño de los indicadores de entrada

Una vez definida la muestra de trabajo, el siguiente paso consistió en transformar la información contable descargada de SABI en un conjunto de variables más útil para el análisis. La base original incluía partidas financieras en niveles absolutos, pero trabajar únicamente con esas magnitudes presentaba varias limitaciones. Por un lado, dificultaba la comparación entre empresas de distinto tamaño. Por otro, no permitía captar de forma suficientemente directa algunas dimensiones especialmente relevantes en una fase inicial de screening, como la trayectoria de crecimiento, la calidad de la rentabilidad operativa, la estabilidad histórica o la presión financiera del negocio. Por ello, antes de estimar los modelos predictivos, se construyó un bloque de indicadores derivados orientado a resumir de forma más comparable y económicamente interpretable la situación de cada compañía.

El primer bloque de variables se diseñó para recoger la evolución del negocio a través de los ingresos. En concreto, se calculó el crecimiento anual de las ventas para los ejercicios 2021, 2022 y 2023, definido como la variación porcentual de los ingresos respecto al año anterior. A su vez, se construyó una tasa compuesta de crecimiento para el periodo 2020-2023, el CAGR. La combinación de ambas aproximaciones permitía recoger, por un lado, los crecimientos interanuales, que ofrecían una lectura más inmediata de la dinámica reciente de la compañía, y, por otro, la tasa compuesta, menos dependiente de cambios puntuales y más útil para resumir la trayectoria acumulada del negocio. En las pymes, los resultados de un ejercicio aislado pueden verse afectados por factores transitorios, de modo que una medida como el CAGR ayuda a complementar la información.

El segundo bloque se centró en la rentabilidad operativa. Para ello, se calcularon márgenes EBITDA y márgenes EBIT a partir de los datos históricos disponibles, dividiendo cada magnitud operativa entre los ingresos del ejercicio correspondiente. Esta transformación permitía comparar empresas con distintos tamaños sin que las diferencias de escala distorsionaran por completo el análisis. En concreto, se utilizaron los márgenes EBITDA de 2020, 2021, 2022 y 2023, así como los márgenes EBIT de los mismos ejercicios. Además, se construyó un margen EBITDA medio para el periodo 2021-2023, con el objetivo de disponer de una medida más estable de rentabilidad histórica, no incluyendo el 2020 por lo que el COVID pudiese haber afectado. La lógica económica detrás de estas variables es clara: en una fase preliminar de análisis, no solo importa cuánto vende una empresa, sino también qué capacidad tiene para convertir esas ventas en resultado operativo. En *private equity*, esta cuestión resulta especialmente relevante porque la calidad del margen condiciona tanto la valoración del activo como su capacidad potencial de soportar apalancamiento mediante los flujos de caja que vaya generando.

Junto a crecimiento y rentabilidad, se incorporó una tercera dimensión relacionada con la estabilidad del negocio. En este caso, el objetivo era captar hasta qué punto la trayectoria

histórica de la empresa había sido consistente o, por el contrario, errática. Para ello, se construyeron dos medidas de volatilidad: una asociada al crecimiento de ingresos y otra al margen EBITDA, ambas calculadas a partir de la dispersión observada entre 2021 y 2023. La inclusión de estas variables responde a una idea sencilla: dos compañías pueden presentar un nivel similar de crecimiento o de margen en el último ejercicio, pero ofrecer perfiles de riesgo muy distintos si una muestra estabilidad y la otra una evolución mucho más irregular. En un proceso de screening inicial, esta diferencia es importante porque la estabilidad suele actuar como una señal indirecta de visibilidad operativa y de menor incertidumbre, además de ser un aspecto clave deseado para el posible *holding period* tras la adquisición.

A estas variables se añadió una medida de tamaño mediante el logaritmo de los ingresos. La utilización del logaritmo, en lugar del dato bruto de ventas, permitía incorporar el efecto del tamaño de una forma más equilibrada y evitar que las empresas más grandes dominaran excesivamente el análisis. Además, esta variable resultaba útil porque el tamaño puede influir en la dinámica operativa de la empresa: en términos relativos, no es esperable que una compañía de mayor escala crezca al mismo ritmo que otra más pequeña, del mismo modo que ciertas diferencias de margen pueden venir condicionadas por la posición alcanzada dentro de su mercado.

Una vez construidas estas variables, se realizó un tratamiento específico de los valores problemáticos. En primer lugar, los cocientes que generaban valores infinitos, normalmente por divisores muy reducidos o por situaciones contables extremas, se sustituyeron por valores perdidos. En segundo lugar, se aplicó *winsorización* en los extremos de la distribución, concretamente en el 1% inferior y superior, con el fin de limitar la influencia de observaciones atípicas sin eliminar empresas de la muestra. Esta decisión era especialmente importante en variables como crecimientos, márgenes o medidas de volatilidad, donde pueden aparecer valores extremos que no representan el comportamiento típico de la base y que, sin embargo, pueden condicionar de forma desproporcionada el ajuste de los modelos.

A partir de este proceso se configuraron los indicadores de entrada utilizados en los modelos predictivos. En el caso del modelo de crecimiento de ingresos, las variables explicativas finales fueron los crecimientos históricos, la tasa compuesta de crecimiento, el margen EBIT más reciente, la volatilidad del crecimiento y el tamaño. En el modelo de margen EBITDA, el peso principal recayó en los márgenes históricos, la volatilidad del margen, una medida de volatilidad del crecimiento y el tamaño de la empresa. Aunque ambos modelos comparten parte de la arquitectura de entrada, la selección final de variables refleja que las dos variables objetivo responden a lógicas distintas: el crecimiento es una magnitud más inestable y difícil de anticipar, mientras que la rentabilidad operativa muestra una persistencia histórica mayor.

Por último, además de las variables empleadas en la fase estrictamente predictiva, la construcción de indicadores se amplió posteriormente con algunas medidas adicionales destinadas al ranking final. Entre ellas destacan la deuda financiera neta sobre EBITDA y una medida de intensidad de capital circulante sobre ventas, calculada a partir de existencias, deudores y acreedores comerciales. Estas variables no se incorporan para anticipar directamente el comportamiento futuro de la empresa, sino para ordenar luego en el ranking teniendo en cuenta variables particularmente relevantes para los *private equity*. En conjunto, el resultado de esta fase fue una base analítica que no se limita a reproducir información contable bruta, sino que la transforma en señales más comparables, de forma que sean más interpretables y útiles.

3.4. Desarrollo de los modelos predictivos

3.4.1. *Modelo de predicción de ingresos*

Una vez definida la muestra a analizar y seleccionadas en el apartado anterior las variables con mayor capacidad explicativa, se pasó a la fase predictiva del modelo de crecimiento de ingresos. En este punto, el objetivo ya no era analizar qué variables parecían estar asociadas al crecimiento observado en 2024, sino comprobar hasta qué punto esa información histórica permitía anticiparlo fuera de muestra. Para ello, se construyó un modelo supervisado tomando como variable objetivo el crecimiento de ingresos de 2024 y utilizando como variables de entrada la versión ajustada obtenida tras la regresión explicativa previa.

En concreto, el modelo final se estimó con siete variables: el crecimiento de ingresos de 2021, 2022 y 2023, la tasa compuesta de crecimiento anual de los ingresos en el periodo 2020-2023, el margen EBIT de 2023, la volatilidad histórica del crecimiento de ingresos y el logaritmo de los ingresos de 2023. La selección es coherente con la lógica económica del problema. Por un lado, los crecimientos históricos recogen la inercia reciente del negocio y permiten observar si la trayectoria que la compañía venía mostrando era relativamente sostenida. Por otro, el CAGR resume el ritmo de crecimiento acumulado en varios ejercicios y evita depender en exceso de la variación de un único año. A su vez, el margen EBIT de 2023 introduce una medida de rentabilidad operativa, mientras que la volatilidad del crecimiento y la variable de tamaño permiten recoger diferencias de estabilidad y escala entre compañías.

Tras comprobar que no existían observaciones con información incompleta para las variables seleccionadas, la muestra final del modelo quedó formada por 5.798 empresas. A partir de esa base, se separaron las variables explicativas de la variable objetivo y se aplicó una partición del 70% para entrenamiento y del 30% para validación. Como resultado, 4.058 observaciones se destinaron al entrenamiento de los modelos y los 1.740 restantes se reservaron para evaluar su comportamiento fuera

de muestra. Esta separación permite valorar la capacidad predictiva real del modelo sobre empresas no utilizadas en el ajuste.

Sobre esta estructura se estimaron dos modelos distintos. En primer lugar, una regresión lineal, utilizada como *benchmark* inicial por su sencillez y por ofrecer una referencia clara de hasta dónde puede llegar una relación lineal entre las variables históricas y el crecimiento futuro. En segundo lugar, se entrenó un modelo de *Random Forest*, parametrizado con 200 árboles, una profundidad máxima de 8 niveles y un mínimo de 5 observaciones por nodo terminal. La inclusión de este segundo enfoque responde a que el crecimiento de ingresos es una variable más irregular y menos estable que otras magnitudes contables, por lo que tenía sentido contrastar si un algoritmo más flexible, capaz de captar relaciones no lineales e interacciones entre variables, podía mejorar el rendimiento del modelo lineal. De este modo, el desarrollo del modelo de crecimiento de ingresos quedó planteado como una comparación entre una alternativa más simple e interpretable y otra potencialmente más potente desde el punto de vista predictivo.

3.4.2. *Modelo de predicción de EBITDA*

De forma paralela, se desarrolló un segundo modelo supervisado orientado a predecir el margen EBITDA de 2024. Aunque la metodología seguida es la misma que en el caso anterior, aquí la variable objetivo es distinta y también lo es su lógica económica. Mientras que el crecimiento de ingresos suele estar más expuesto a factores externos y a una mayor variabilidad interanual, el margen EBITDA refleja de manera más directa la rentabilidad operativa del negocio y, en principio, presenta una mayor persistencia en el tiempo. Precisamente por eso, era razonable esperar que este segundo modelo fuese más dependiente de los márgenes históricos que de las variables de crecimiento.

Partiendo del modelo explicativo ajustado en el apartado anterior, se construyó con cinco variables distintas de las anteriores: el margen EBITDA de 2023, el margen EBITDA de 2021 (sin incluir el del 2022 por el alto p-valor que tenía), la volatilidad del crecimiento de ingresos, la volatilidad histórica del margen EBITDA y el logaritmo de los ingresos de 2023. La presencia de los márgenes históricos responde a que, si una empresa que ha mostrado en ejercicios anteriores una determinada capacidad de generación operativa tiende, salvo cambios relevantes, a mantener un perfil similar en el corto plazo. Junto a ello, las variables de volatilidad permiten introducir una medida de estabilidad del negocio, mientras que el tamaño vuelve a actuar como control adicional dentro de la muestra.

Después de aplicar los mismos criterios de depuración que en el modelo anterior, la muestra volvió a quedar formada por 5.798 compañías, lo que permite mantener la

comparabilidad entre ambos ejercicios predictivos. A continuación, se realizó también la partición 70/30 entre entrenamiento y test, con 4.058 observaciones destinadas al ajuste y 1.740 reservadas para validación fuera de muestra. Sobre esta base se estimaron, de nuevo, dos modelos: una regresión lineal y un *Random Forest* con la misma configuración utilizada en el caso del crecimiento de ingresos.

La lógica de esta doble estimación era comprobar si la mayor estabilidad del margen EBITDA hacía suficiente un modelo lineal sencillo o si, por el contrario, seguía existiendo margen de mejora mediante una metodología más flexible. En otras palabras, no se trataba solo de elegir el modelo con mejor ajuste, sino de ver si esa posible mejora justificaba introducir una estructura más compleja.

Adicionalmente, se exploró una aproximación sectorial, estimando modelos separados para distintos grupos de actividad con el fin de comprobar si una mayor homogeneidad entre compañías mejoraba la capacidad predictiva. No obstante, esta segmentación no produjo mejoras claras ni en la predicción del crecimiento de ingresos ni en la del margen EBITDA. En términos prácticos, la reducción del tamaño muestral dentro de cada sector restaba robustez a las estimaciones y no se traducían en una mejora suficiente de las métricas, por lo que se optó por mantener un enfoque agregado en la versión final del modelo.

3.5. Criterios de evaluación y comparación de modelos

Una vez desarrollados los modelos predictivos, el siguiente paso consiste en definir cómo se va a evaluar su rendimiento. En este trabajo, la evaluación no se plantea como un ejercicio únicamente técnico, sino como una forma de comprobar hasta qué punto las predicciones obtenidas pueden resultar útiles dentro de una herramienta de screening inicial. Por ello, el análisis se realiza sobre el conjunto de test, es decir, sobre observaciones no utilizadas durante el entrenamiento. Esta distinción es importante porque permite valorar la capacidad de generalización del modelo y evita extraer conclusiones a partir de un ajuste realizado únicamente sobre la propia muestra de estimación.

Para comparar los modelos se utilizan tres métricas complementarias: el error absoluto medio (MAE), la raíz del error cuadrático medio (RMSE) y el coeficiente de determinación (R^2). El MAE mide, en promedio, cuánto se desvía la predicción respecto al valor real en términos absolutos, por lo que ofrece una referencia sencilla e intuitiva del error medio del modelo. El RMSE también mide esa desviación, pero penaliza en mayor medida los errores más elevados, de modo que resulta útil para detectar si determinadas observaciones concentran fallos especialmente importantes. Por último, el R^2 permite valorar en qué medida el modelo consigue explicar la variabilidad observada en la variable objetivo y se utiliza como medida complementaria de ajuste.

La utilización conjunta de estas tres métricas permite obtener una visión más completa del rendimiento predictivo. Mientras que el MAE ofrece una medida más directa y fácil de interpretar, el RMSE permite captar mejor la existencia de errores extremos, y el R² aporta una referencia adicional sobre la capacidad explicativa global del modelo. En consecuencia, la comparación no se apoya en una sola métrica, sino en la lectura conjunta de las tres.

Este mismo criterio de evaluación se aplicará tanto al modelo de crecimiento de ingresos como al de margen EBITDA, así como a las dos metodologías estimadas en cada caso, Regresión Lineal y *Random Forest*. A partir de estas métricas, en el apartado de resultados se comparará qué modelo de los dos en cada caso presenta un mejor comportamiento fuera de muestra y en qué medida sus predicciones pueden resultar útiles para la fase posterior de priorización de compañías.

3.6. Diseño del score de priorización y construcción del ranking

Una vez estimados y evaluados los modelos predictivos, el siguiente paso consistió en trasladar ese análisis a una herramienta de priorización aplicable al conjunto de empresas de la muestra. El objetivo de esta fase ya no fue predecir una variable aislada, sino sintetizar distintas señales económico-financieras en un único indicador que permitiera ordenar las compañías según su atractivo relativo dentro de una primera criba de inversión. De este modo, el ranking final se planteó como una herramienta de apoyo al screening, orientada a reducir el universo inicial de análisis y a identificar, de forma más eficiente y rápida, aquellas empresas que presentaban unas características más favorables dentro del enfoque definido por una firma de *private equity*.

En una primera aproximación se valoró incorporar al score final tanto la predicción de ingresos como la predicción del margen EBITDA. Sin embargo, tras comparar el comportamiento de ambos modelos, se optó por no integrar la predicción de ingresos como componente directo del ranking. La razón fue que su capacidad predictiva resultó claramente inferior y menos robusta que la observada en el modelo de margen EBITDA, aunque ahondaremos más en ello en el siguiente apartado. Por ello, la dimensión de crecimiento se recogió finalmente a través del CAGR de ingresos 2021-2024, calculado con información observada a cierre de 2024, mientras que la dimensión predictiva del score descansó en la estimación del margen EBITDA esperado para 2025. Esta decisión permitió mantener en el trabajo la modelización de ingresos como una línea explorada y evaluada, aunque sin trasladarla al diseño final del ranking.

A partir de esta lógica, el score se construyó con seis indicadores: el CAGR de ingresos 2021-2024, la predicción del margen EBITDA de 2025, el ratio de deuda financiera neta sobre EBITDA de 2024, la variación del *operating working capital* entre 2023 y 2024 sobre ventas de 2024, y dos medidas de volatilidad histórica, una del crecimiento de

ingresos y otra del margen EBITDA. Antes de agregarlas, estas variables fueron *winsorizadas* al 1% y 99% para reducir la influencia de valores extremos. Además, en el caso de la variable de capital circulante, los valores ausentes, principalmente en la variable de *operating working capital*, se imputaron con la mediana de la muestra. Posteriormente, todas las variables se transformaron a percentiles con el fin de hacer comparables magnitudes expresadas en escalas distintas. En aquellas variables en las que un valor menor se interpretaba como más favorable, como el apalancamiento, la intensidad de capital circulante o las volatilidades, el percentil se invirtió para mantener una lectura homogénea del score.

El score final se definió como una combinación ponderada de esos percentiles. Se asignó un peso del 35% al margen EBITDA previsto, un 25% al CAGR de ingresos, un 15% al apalancamiento, un 10% a la variación del OWC y un 7,5% a cada una de las dos medidas de volatilidad. Esta ponderación no respondió a un criterio de optimización automática, sino a una elección razonada a partir de la lógica económica del *private equity*. En este tipo de operaciones, la creación de valor para el inversor suele apoyarse, de forma simplificada, en tres grandes palancas: la mejora operativa del negocio, la expansión o contracción de los múltiplos de valoración y el efecto del apalancamiento (Kaplan & Strömberg, 2009). Por ello, el mayor peso se otorgó al margen EBITDA previsto y al CAGR histórico de ingresos, ya que, como señalan Gompers, Kaplan & Mukharlyamov (2015), la creación de valor en *private equity* se apoya más en el crecimiento del negocio que en una simple reducción de costes; en ese sentido, un margen EBITDA ya elevado deja más margen para centrar la tesis en crecer. Además, un crecimiento sostenido de las ventas puede contribuir a ampliar la base de EBITDA a la salida y, además, favorecer una valoración más elevada del activo si el mercado reconoce esa trayectoria. A continuación, se asignó un peso intermedio al apalancamiento y a la variación del OWC, es decir, del capital circulante operativo, ya que una menor deuda ofrece mayor flexibilidad para estructurar la financiación de la operación y una menor absorción de caja por OWC facilita que una mayor parte del *cash flow* pueda destinarse al repago de deuda durante el periodo de inversión (Gompers, Kaplan & Mukharlyamov, 2015). Por último, las medidas de volatilidad recibieron una ponderación menor, aunque relevante, porque en operaciones apalancadas resulta especialmente importante poder construir un *business plan* creíble y ejecutable, de modo que la evolución futura de ingresos y márgenes no se aleje de forma excesiva de lo previsto.

Los detalles técnicos complementarios de implementación se recogen en los anexos, donde se incluyen únicamente los fragmentos de código más relevantes para la metodología empleada.

4. Resultados

4.1. Resultados del modelo de predicción de ingresos

Los resultados obtenidos en la predicción del crecimiento de ingresos muestran que esta es la parte más débil del ejercicio. Una vez entrenados los modelos sobre el 70% de la muestra y evaluados sobre el 30% reservado para test, la Regresión Lineal presenta un MAE de 0,1156, un RMSE de 0,1653 y un R^2 de 0,0278. Por su parte, el *Random Forest* mejora ligeramente estas métricas, con un MAE de 0,1141, un RMSE de 0,1634 y un R^2 de 0,0497.

Aunque el *Random Forest* ofrece un comportamiento algo mejor que la Regresión Lineal, la diferencia entre ambos es reducida y, sobre todo, los errores siguen siendo elevados. En términos más intuitivos, el error absoluto medio se sitúa en torno a 11-12 puntos porcentuales de crecimiento, mientras que el error cuadrático medio ronda los 16 puntos porcentuales. Para una variable tan sensible como el crecimiento anual de ingresos, estos niveles de error son demasiado altos como para considerar que la predicción es suficientemente fiable en una herramienta de priorización.

Por tanto, la conclusión principal de este apartado es que, aun existiendo una pequeña señal en algunas variables históricas, la capacidad predictiva del modelo de ingresos es limitada. El crecimiento de ventas de un año a otro no puede anticiparse con suficiente precisión a partir de la información contable utilizada, al menos no con el grado de fiabilidad que requeriría su incorporación al ranking final. Por ello, la predicción de ingresos no se utilizará posteriormente como componente directo del score de priorización, aunque sí resulta útil como evidencia de una de las principales limitaciones del enfoque planteado.

4.2. Resultados del modelo de margen EBITDA

Los resultados del modelo de margen EBITDA son claramente superiores a los obtenidos en la predicción de ingresos. En este caso, la Regresión Lineal alcanza en la muestra de test un MAE de 0,0261, un RMSE de 0,0404 y un R^2 de 0,8351. El *Random Forest* presenta unas métricas muy similares, aunque ligeramente peores, con un MAE de 0,0262, un RMSE de 0,0419 y un R^2 de 0,8225.

A diferencia de lo que ocurría con los ingresos, aquí los errores de predicción son mucho más reducidos. El error absoluto medio se sitúa en torno a 2,6 puntos porcentuales de margen EBITDA, y el error cuadrático medio ronda los 4 puntos porcentuales. Esto muestra que el comportamiento del margen EBITDA puede anticiparse con bastante más precisión a partir de la información histórica disponible. Además, el hecho de que la Regresión Lineal funcione incluso algo mejor que el *Random Forest* sugiere que, en este

caso, no es necesario recurrir a una estructura más compleja para obtener buenos resultados predictivos.

En consecuencia, sí puede afirmarse que la predicción del margen EBITDA ofrece una precisión razonable y suficientemente útil para una herramienta de screening inicial. Frente al modelo de ingresos, aquí el nivel de error es bastante más bajo y la predicción resulta mucho más consistente.

Una vez comprobado que el modelo ofrecía una capacidad predictiva razonable sobre 2024, este se utilizó para generar una estimación del margen EBITDA de 2025 para el conjunto de compañías, predicción que posteriormente se incorporó al ranking final.

4.3. Construcción del ranking final e integración de las predicciones

Una vez comparado el rendimiento de ambos modelos, el ranking final se construyó incorporando únicamente aquella información predictiva que mostraba una precisión suficiente dentro de la muestra. En la práctica, esto supuso no utilizar la predicción de ingresos como componente directo del score, ya que sus errores eran demasiado elevados para que aportase una señal fiable en una herramienta de priorización. En su lugar, la dimensión de crecimiento se recogió a través del CAGR de ingresos 2021-2024 observado, mientras que la dimensión predictiva del ranking descansó sobre la predicción del margen EBITDA para 2025, cuyo comportamiento había resultado sensiblemente más preciso y estable.

A partir de esta decisión, el score final se diseñó combinando seis indicadores: el CAGR de ingresos 2021-2024, el margen EBITDA previsto para 2025, el ratio de deuda financiera neta sobre EBITDA de 2024, la variación del *operating working capital* entre 2023 y 2024 sobre ventas de 2024, la volatilidad histórica del crecimiento de ingresos y la volatilidad histórica del margen EBITDA. Antes de agregarlas, estas variables se *winsorizaron* para reducir la influencia de valores extremos y, posteriormente, se transformaron a percentiles, con el fin de hacer comparables magnitudes expresadas en escalas distintas. En aquellas variables en las que un valor menor se interpretaba como más favorable, como el apalancamiento, la intensidad de capital circulante o las volatilidades, el percentil se invirtió para que una puntuación más alta reflejase siempre un perfil relativamente más atractivo.

El score agregado se calculó como una combinación ponderada de esos percentiles. El mayor peso se asignó al margen EBITDA previsto, con un 35%, seguido del CAGR histórico de ingresos, con un 25%. A continuación, se incorporaron el apalancamiento, con un 15%, la variación del *operating working capital*, con un 10%, y, por último, las dos medidas de volatilidad, con un 7,5% cada una. Esta ponderación no respondió a un criterio automático, sino a una elección razonada desde la lógica del *private equity*. En

este tipo de operaciones, la rentabilidad del inversor depende en gran medida de la evolución operativa del negocio, de su capacidad de crecimiento y de la posibilidad de estructurar la inversión sobre una base financiera razonable (Axelson, Jenkinson, Strömberg & Weisbach, 2013). Por ello, el mayor peso recayó sobre las variables más directamente vinculadas con la generación de EBITDA y con la expansión del negocio, mientras que el apalancamiento, el consumo de caja derivado del capital circulante y la estabilidad histórica actuaron como elementos complementarios dentro de la priorización.

Una vez calculado el score para el conjunto de la base limpia, las compañías se ordenaron de mayor a menor puntuación y se seleccionaron las 100 primeras posiciones como resultado final del ejercicio. El principal interés de esta fase es que el ranking no depende de una sola variable ni de una única predicción, sino de una combinación de señales que recogen crecimiento, rentabilidad, estabilidad, endeudamiento y consumo de caja operativa. De esta forma, el Top 100 no identifica necesariamente a las empresas con mayor crecimiento o con mayor margen de forma aislada, sino a aquellas que presentan un perfil más equilibrado y, en conjunto, más atractivo para una primera fase de screening.

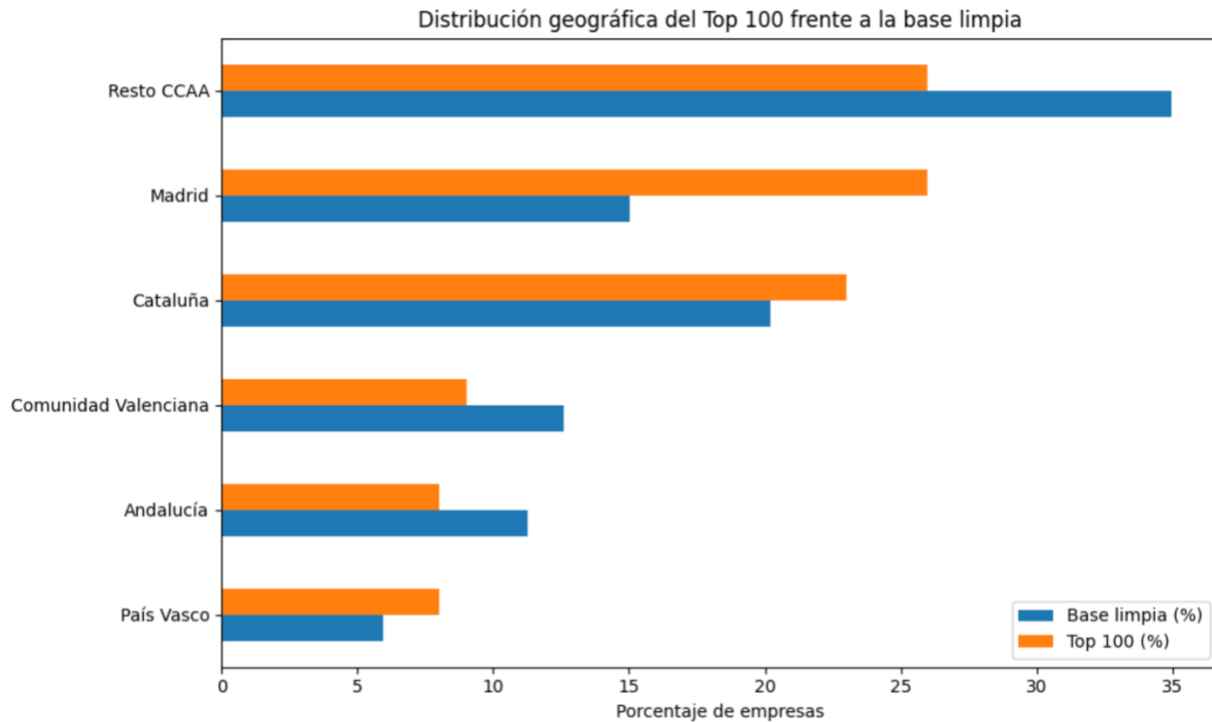
En términos prácticos, esta fase permite traducir el ejercicio analítico previo en una herramienta útil de priorización. Frente a una base inicial de 5.798 compañías, el ranking reduce de forma sustancial el universo de análisis y genera una *shortlist* mucho más manejable. Por tanto, el resultado no es solo técnico, sino también operativo: la herramienta permite pasar de un conjunto amplio y difícil de revisar manualmente a un grupo acotado de compañías sobre las que sí tendría sentido concentrar un análisis posterior más detallado.

4.4. Caracterización geográfica y sectorial del Top 100

Una vez construido el ranking final, resulta útil analizar si la *shortlist* obtenida reproduce de forma más o menos neutra la composición de la base limpia o si, por el contrario, concentra una mayor proporción de compañías en determinadas regiones y sectores. Esta comparación permite complementar la lectura del score final, ya que no solo muestra qué empresas ocupan las primeras posiciones, sino también qué tipo de perfiles aparecen con más frecuencia dentro del Top 100. Para ello, se contrasta el peso relativo de las principales comunidades autónomas y de los principales sectores de actividad de la base limpia frente a su peso dentro de la selección final.

4.4.1. *Distribución geográfica del Top 100 frente a la base limpia*

La comparación geográfica muestra que el Top 100 no reproduce de forma exacta la distribución territorial de la base limpia, sino que concentra una mayor proporción de compañías en algunas comunidades autónomas concretas.



Fuente: elaboración propia a partir de SABI.

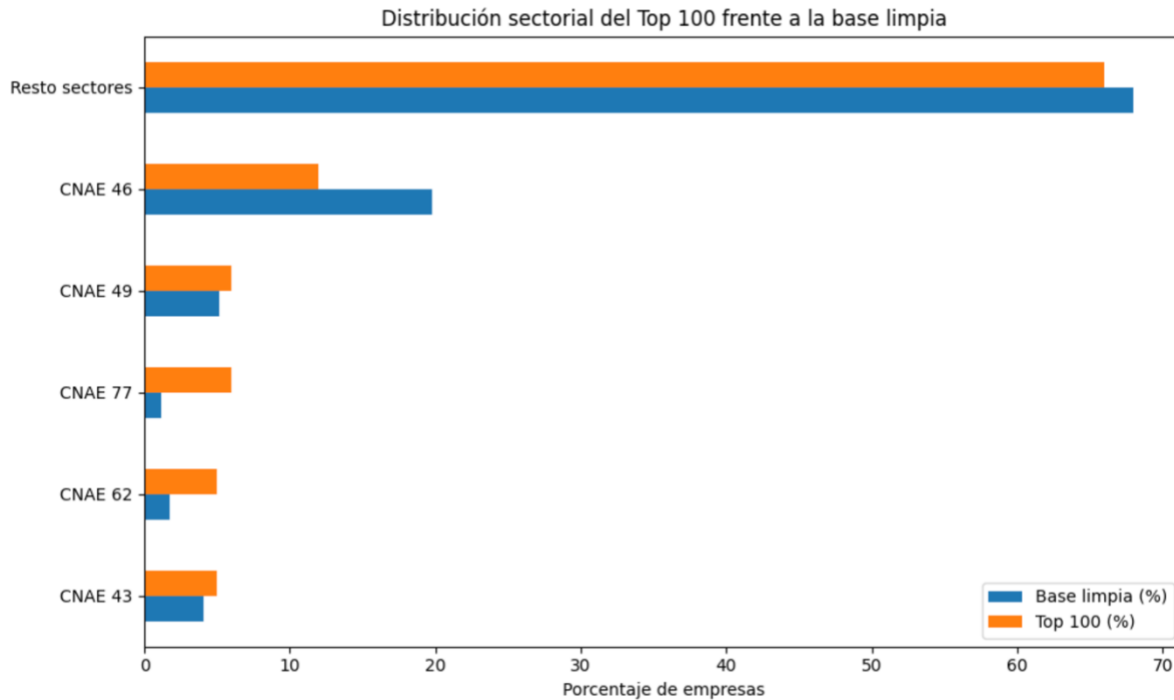
El resultado más claro es el aumento de peso de Madrid, que pasa a tener una presencia bastante mayor dentro del Top 100 que en la base limpia. También Cataluña gana algo de peso relativo, mientras que el País Vasco aparece igualmente algo más representado en la selección final. En cambio, otras regiones como Comunidad Valenciana y Andalucía pierden algo de presencia relativa, al igual que el grupo agregado del resto de comunidades autónomas.

En conjunto, esto sugiere que la *shortlist* final tiende a concentrarse más en territorios donde se localiza una parte relevante del tejido empresarial de mayor tamaño, más profesionalizado o con mayor presencia de compañías de servicios y actividades intensivas en conocimiento. En todo caso, esta lectura debe hacerse con prudencia. El gráfico no implica que unas regiones sean mejores que otras en términos absolutos, sino que, dentro del universo analizado y de acuerdo con los criterios utilizados en el ranking, ciertas comunidades concentran una proporción mayor de compañías con un perfil financiero y operativo más alineado con la lógica de priorización construida en el trabajo.

4.4.2. Distribución sectorial del Top 100 frente a la base limpia

La comparación sectorial permite observar con algo más de detalle qué tipos de actividad aparecen relativamente más o menos representados dentro de la selección

final. Igual que en la dimensión geográfica, el Top 100 no replica de forma mecánica la composición de la base limpia, sino que muestra una cierta concentración en algunos sectores concretos.



Fuente: elaboración propia a partir de SABI.

Entre los sectores que ganan peso en el ranking final destaca, en primer lugar, el de actividades de alquiler (CNAE 77), que aparece claramente más representado en el Top 100 que en la base limpia. También aumentan de forma visible las actividades de programación, consultoría y otros servicios informáticos (CNAE 62), así como, en menor medida, el transporte terrestre y por tubería (CNAE 49) y la construcción especializada (CNAE 43). Por el contrario, el comercio al por mayor e intermediarios del comercio, excepto de vehículos de motor y motocicletas (CNAE 46) pierde peso relativo en la selección final, aunque sigue manteniendo una presencia importante en términos absolutos.

Este patrón resulta relevante porque sugiere que el score final tiende a favorecer relativamente más a compañías vinculadas a actividades de servicios, negocios potencialmente más escalables o sectores con mayor capacidad de generación de margen, frente a otros segmentos más intensivos en volumen o con dinámicas operativas más ajustadas. De nuevo, esto no debe interpretarse como una conclusión absoluta sobre la calidad de un sector frente a otro, sino como una señal de que, dentro

del universo analizado, algunos sectores concentran una proporción mayor de compañías que encajan mejor con los criterios de priorización utilizados para construir el ranking final.

5. Discusión

5.1. Interpretación de la capacidad predictiva de los modelos

La comparación entre ambos modelos permite extraer una conclusión bastante clara: la información contable histórica utilizada en este trabajo resulta mucho más útil para estimar el margen EBITDA que para anticipar el crecimiento anual de los ingresos. Esta diferencia no parece casual, sino que responde a la propia naturaleza de ambas variables. Mientras que el crecimiento suele ser más cambiante y difícil de prever, las variables ligadas a la rentabilidad operativa tienden a mantenerse más estables a lo largo del tiempo (Chan, Karceski & Lakonishok, 2001).

En este sentido, que el modelo de ingresos obtenga errores elevados no significa necesariamente que el planteamiento sea incorrecto, sino que la variable que se intenta predecir es difícil de estimar con este tipo de información. La evidencia disponible sugiere que el crecimiento no suele mantenerse de forma estable en el tiempo y que, además, resulta difícil de anticipar, incluso para los analistas, especialmente cuando se analiza a varios años vista (Chan, Karceski & Lakonishok, 2001).

Además, en este trabajo la predicción de ingresos se construye únicamente a partir de datos contables históricos. Esto es importante porque el crecimiento de ventas de una empresa no depende solo de su evolución financiera previa, sino también de muchos factores externos que las cuentas anuales apenas reflejan, o directamente no recogen. Entre ellos están, por ejemplo, la posición competitiva, la capacidad comercial, la concentración de clientes, los cambios de precios, el lanzamiento de productos, las adquisiciones o la evolución del sector. Por eso, no sorprende que la información contable por sí sola no sea suficiente para predecir con precisión una variable tan cambiante como el crecimiento anual de ingresos.

A ello se añade una segunda limitación, relacionada con el tipo de información disponible en empresas privadas. La evidencia reciente sugiere que las compañías no cotizadas no siempre operan en un contexto de transparencia plena, sino que están sujetas a distintas fricciones informativas (Gassen & Muhn, 2024). Esto significa que sus cuentas pueden verse condicionadas por costes de información, limitaciones en el procesamiento de esa información y distintos incentivos para revelar más o menos datos (Gassen & Muhn, 2024). Por ello, aunque una base como SABI resulta muy útil para construir una herramienta homogénea de screening, no puede asumirse que recoja toda la realidad económica de la empresa con el detalle necesario para predecir con precisión una variable tan cambiante como el crecimiento de ingresos.

El contraste con el modelo de margen EBITDA es muy interesante. Aquí los errores son mucho menores y la predicción resulta claramente más consistente. La razón principal es

que el margen EBITDA está más ligado a la estructura operativa del negocio y, por tanto, presenta una mayor persistencia histórica que el crecimiento de ingresos. En otras palabras, la capacidad de una empresa para convertir ventas en resultado operativo suele depender de elementos más estables, como pueden ser la eficiencia, estructura de costes, mix de producto, posicionamiento o disciplina operativa, que no cambian tan bruscamente de un año para otro.

Esta idea encaja, además, con la evolución reciente del propio sector. Los informes más recientes del mercado subrayan que, en el entorno actual, las estrategias tradicionales de creación de valor apoyadas en apalancamiento y expansión de múltiplos han perdido fuerza y que la generación de valor depende cada vez más de la ejecución operativa dentro de la cartera y de la expansión del EBITDA (Biesinger, Bircan & Ljungqvist, 2023). De forma similar, también se insiste en que el mercado actual exige procesos de inversión más disciplinados, una mayor atención a la calidad del activo y una creciente presión por identificar fuentes de valor menos dependientes de la ingeniería financiera (McKinsey & Company, 2026).

Desde la lógica del *private equity*, por tanto, tiene sentido que el modelo más útil no sea el que intenta predecir mejor el crecimiento de ingresos, sino el que estima con más precisión la calidad operativa futura de la empresa. En el contexto actual, el sector da cada vez más importancia a la mejora de márgenes y a la capacidad de generar EBITDA. Por eso, en este trabajo se opta por excluir la predicción de ingresos del *score* final e incorporar únicamente la predicción del margen EBITDA. Esta decisión no responde solo a que el modelo funcione mejor técnicamente, sino también a que encaja mejor con la lógica económica del *private equity mid-market*.

5.2. Encaje del ranking final con la lógica del *private equity mid-market*

La composición del Top 100 obtenido en este trabajo encaja, en términos generales, con varias de las tendencias que vienen observándose en el *private equity* europeo y español. En particular, el mayor peso relativo de actividades ligadas a programación, consultoría e informática, así como de otros negocios de servicios, no parece un resultado extraño. En Europa, una parte muy relevante de la actividad reciente de *private equity* se ha concentrado en *Business Services*, *Industrials* y *TMT*, y dentro de ese bloque destacan especialmente subsectores vinculados a software, servicios profesionales y negocios con capacidad de escalabilidad (Gain, 2026).

Esta lectura también es compatible con la evidencia internacional reciente, según la cual tanto la estrategia de inversión como la geografía condicionan el comportamiento del *private equity*, destacando los *buyout funds* europeos como uno de los segmentos con mejores resultados relativos (Tommar, Darolles & Jurczenko, 2024).

Y la misma idea aparece también en la evidencia de Invest Europe. En 2024, *ICT (Information and Communication Technologies*, o tecnologías de la información y la comunicación) fue el sector que más inversión recibió en Europa y *business products and services* ocupó la segunda posición. Además, casi el 50% de las compañías respaldadas por *private equity* y *venture capital* se concentraron en *ICT* y *Biotech & Healthcare*, lo que confirma el peso creciente de negocios intensivos en tecnología y servicios dentro del mercado europeo (Invest Europe, 2025). Desde esta perspectiva, que en el Top 100 ganen peso las actividades de informática, consultoría y otros servicios resulta bastante coherente con la lógica reciente del mercado.

Si se baja al caso español, el encaje también es razonable. SpainCap destaca que, dentro del *middle market* en 2024, una parte importante de la actividad se concentró en Medicina/Salud, tecnologías digitales, Productos y Servicios Industriales y Transporte, tanto por volumen como por número de operaciones (SpainCap, 2025). Además, en la actividad de fondos internacionales en España también sobresalen Tecnologías digitales y Otros servicios como algunos de los sectores con mayor peso (SpainCap, 2025). Por tanto, que el ranking final favorezca relativamente más a compañías de servicios, tecnología o transporte no parece desalineado con el tipo de operaciones que realmente se están viendo en el mercado español.

Algo parecido ocurre con la distribución geográfica. En los resultados del apartado anterior se observaba que Madrid y Cataluña ganaban peso dentro del Top 100. Esta concentración tampoco resulta llamativa, ya que en 2024 estas dos comunidades continuaron situándose entre los principales focos de inversión del capital privado en España (SpainCap, 2025). En ese sentido, la herramienta parece captar bastante bien algunos de los principales polos empresariales donde el *private equity* está más activo y donde existe una mayor concentración de compañías con cierto tamaño, grado de profesionalización y presencia en sectores atractivos para el inversor.

En conjunto, por tanto, el ranking final puede considerarse razonablemente alineado con la lógica actual del *private equity mid-market*. Los informes recientes del sector insisten en que, en el contexto actual, la creación de valor depende cada vez más de la calidad del activo, de su potencial operativo y de su capacidad de generar EBITDA (Axelson et al., 2013), y menos de la simple expansión de múltiplos o del apalancamiento barato (Bain & Company, 2026). De forma similar, también se subraya que las estrategias tradicionales ya no bastan por sí solas y que el mercado exige un enfoque más disciplinado, más operativo y más apoyado en datos (KPMG, 2025). Desde ese punto de vista, que la herramienta tienda a priorizar compañías de servicios, tecnología y otros negocios con mejor perfil operativo resulta bastante coherente con la forma en que hoy se entiende el *private equity mid-market*.

5.3. Utilidad práctica, limitaciones y posibles mejoras del enfoque

Más allá de sus resultados predictivos, la principal utilidad de la herramienta desarrollada en este trabajo es práctica. Su valor no está en sustituir el juicio del inversor ni en identificar automáticamente la mejor oportunidad de inversión, sino en ayudar a reducir un universo inicial muy amplio a una *shortlist* mucho más manejable. En este caso, pasar de 5.798 compañías a un Top 100 supone una reducción muy significativa del esfuerzo de revisión y permite concentrar antes el análisis en empresas que, al menos desde una primera aproximación cuantitativa, presentan un perfil más atractivo.

Esta utilidad resulta especialmente relevante en el contexto del *private equity mid-market*, donde muchas decisiones iniciales se toman con información limitada y bajo presión de tiempo. En un entorno así, disponer de una herramienta que ordene compañías de forma más homogénea puede mejorar la eficiencia del proceso y reducir la dependencia de criterios demasiado rápidos o poco estructurados. Además, el contexto actual del sector refuerza esta necesidad, ya que la selección de activos y la creación de valor operativa han ganado peso frente a estrategias más apoyadas en apalancamiento o expansión de múltiplos.

Al mismo tiempo, el enfoque tiene limitaciones claras. La primera es que trabaja únicamente con información contable histórica, lo que deja fuera variables cualitativas y comerciales que pueden ser muy relevantes para el inversor, como la calidad del equipo directivo, la posición competitiva, la concentración de clientes o la existencia de palancas reales de crecimiento inorgánico. La segunda es que se apoya en información de empresas privadas, que no siempre refleja toda la realidad económica de la compañía con el mismo grado de detalle y transparencia que en una empresa cotizada. Por último, el propio score final responde a una ponderación razonada, pero no deja de incorporar una parte de juicio en la selección y peso de las variables.

A futuro, el modelo podría mejorarse incorporando variables no contables, información sectorial más rica o indicadores externos que permitan captar mejor el contexto competitivo y comercial de cada empresa. También tendría sentido explorar versiones más específicas por estrategia o por sector. En este último caso, en este trabajo sí se intentó una aproximación sectorial, pero no produjo mejoras claras en las métricas, tal y como se ha explicado previamente. No obstante, ello no implica necesariamente que la idea sea errónea desde el punto de vista conceptual. De hecho, agrupar compañías por sectores debería, en principio, permitir predicciones más fiables, ya que tanto los crecimientos como los márgenes EBITDA estarían siendo estimados entre empresas más comparables entre sí y expuestas de forma similar a determinados factores externos. Por ello, es razonable pensar que la falta de mejora observada responde más a una limitación de muestra —tamaño insuficiente y menor robustez de las estimaciones dentro de cada

subconjunto sectorial— que a un problema de fondo de la propia idea. En definitiva, la herramienta no elimina la necesidad de análisis posterior, pero sí puede servir como un punto de partida útil, ordenado y realista para una primera fase de screening.

6. Conclusiones

El presente trabajo ha abordado el problema del *screening* de oportunidades de inversión en el *private equity mid-market* desde una perspectiva aplicada, proponiendo una herramienta basada en datos contables históricos que permite ordenar y priorizar compañías dentro de un universo amplio de potenciales objetivos de inversión.

Los resultados obtenidos permiten extraer varias conclusiones relevantes. En primer lugar, la evidencia empírica muestra que la información contable histórica resulta significativamente más útil para estimar la rentabilidad operativa que para anticipar el crecimiento de los ingresos. Mientras que el modelo de margen EBITDA ofrece una capacidad predictiva elevada y consistente, la predicción del crecimiento presenta errores demasiado altos como para resultar útil en una herramienta de priorización. Esta diferencia no responde únicamente a una limitación técnica, sino a la propia naturaleza de ambas variables, siendo el crecimiento una magnitud más volátil y dependiente de factores externos.

En segundo lugar, el diseño del score de priorización permite trasladar el análisis predictivo a una herramienta operativa de *screening*, combinando variables de crecimiento, rentabilidad, apalancamiento, consumo de caja y estabilidad histórica. La construcción del ranking final demuestra que es posible reducir de forma significativa un universo inicial de miles de compañías a una *shortlist* mucho más manejable, manteniendo al mismo tiempo una lógica económica coherente con el proceso de inversión en *private equity*.

En tercer lugar, la composición del Top 100 obtenido resulta consistente con las tendencias recientes del mercado, tanto a nivel sectorial como geográfico. El mayor peso de actividades vinculadas a servicios, tecnología o negocios escalables, así como la concentración en determinadas regiones, refleja patrones que coinciden con la actividad observada en el *private equity* europeo y español. Esto refuerza la idea de que la herramienta no solo funciona desde un punto de vista técnico, sino que también captura de forma razonable la lógica del mercado.

No obstante, el trabajo presenta limitaciones importantes. La principal es la dependencia exclusiva de información contable histórica, que no recoge variables cualitativas clave para el inversor, como la calidad del equipo directivo, la posición competitiva o las palancas reales de creación de valor. Además, el uso de datos de empresas privadas introduce incertidumbre adicional en términos de calidad y transparencia de la información. Por último, el propio score incorpora decisiones de ponderación que, aunque razonadas, no dejan de reflejar un cierto grado de juicio.

En conjunto, el trabajo pone de manifiesto que el uso de herramientas cuantitativas puede aportar valor real en fases tempranas del proceso inversor, especialmente en contextos donde el volumen de oportunidades es elevado y la información disponible es limitada. Lejos de sustituir el criterio del inversor, este tipo de enfoques permite estructurarlo mejor, mejorar la eficiencia del análisis y centrar antes el esfuerzo en aquellas compañías con mayor potencial.

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa

Por la presente, yo, Carlos Espinosa de los Monteros Pérez-Brotóns, estudiante de Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "*Screening de oportunidades para la inversión de private equity mid-market en España*", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
3. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
4. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
5. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
6. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
7. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: Abril, 2026

Firma: Carlos Espinosa de los Monteros Pérez-Brotóns

Bibliografía

- Axelson, U., Jenkinson, T., Strömberg, P., & Weisbach, M. S. (2013). Borrow cheap, buy high? The determinants of leverage and pricing in buyouts. *The Journal of Finance*, 68(6), 2223–2267.
- Bain & Company. (2026). *Global private equity report 2026*.
- Biesinger, M., Bircan, Ç., & Ljungqvist, A. (2023). *Value creation in private equity* (Working paper).
- Borell, M., & Heger, D. (2013). *Sources of value creation through private equity-backed mergers and acquisitions: The case of buy-and-build strategies* (ZEW Discussion Paper No. 13-094).
- Chan, L. K. C., Karceski, J., & Lakonishok, J. (2001). *The level and persistence of growth rates* (NBER Working Paper No. 8282).
- Cumming, D., Kumar, S., Lim, W. M., & Pandey, N. (2023). Mapping the venture capital and private equity research: A bibliometric review and future research agenda. *Small Business Economics*, 61, 173–221.
- Dirección General de Estrategia Industrial y de la Pequeña y Mediana Empresa. (2025). *Cifras PyME. Edición noviembre 2025*.
- Gain (2026). *The State of European Private Equity Report. H1 2026*.
- Gassen, J., & Muhn, M. (2024). Financial transparency of private firms: Evidence from a randomized field experiment. *Journal of Accounting Research*. Advance online publication.
- Gompers, P., Kaplan, S. N., & Mukharlyamov, V. (2015). *What do private equity firms say they do?* (Harvard Business School Working Paper No. 15-081).
- Informa D&B. (2026). *Sistema de Análisis de Balances Ibéricos (SABI)* [Base de datos].
- Instituto Nacional de Estadística. (2025). *Directorio Central de Empresas (DIRCE). 1 de enero de 2025*.
- Invest Europe. (2025). *Investing in Europe: Private Equity Activity 2024*.
- Kaplan, S. N., & Strömberg, P. (2009). Leveraged buyouts and private equity. *Journal of Economic Perspectives*, 23(1), 121–146.
- KPMG. (2025). *Value creation in private equity: From stock-pickers to the quant PE house*.
- Maccagni, E., & Boiani, R. R. (2026). Enhancing private equity decision-making: A machine learning framework for early-stage screening. *Finance Research Open*, 2, 100103.

McKinsey & Company. (2026). *Global Private Markets Report 2026: Private equity—Clearer view, tougher terrain*.

SpainCap. (2025). *Informe 2025: Venture Capital & Private Equity en España*.

Tommar, S. A., Darolles, S., & Jurczenko, E. (2024). Private equity performance around the world. *Financial Analysts Journal*, 80(2), 99–121.

Anexos

Anexo 1: Top 10 compañías ranking final

Este anexo recoge las diez compañías mejor posicionadas en el ranking final elaborado en el trabajo, a partir del score agregado definido en la metodología. La tabla resume, para cada empresa, algunas de las variables más relevantes utilizadas en la priorización, con el objetivo de ofrecer una visión más concreta del tipo de perfil que la herramienta tiende a favorecer en la fase inicial de *screening*.

Anexo 1. Top 10 compañías del ranking final

Primeras 10 compañías del ranking final según el score agregado del modelo. Se muestran la posición, la compañía y las principales variables utilizadas en la priorización.

Posición	Empresa	Sector (CNAE 2D)	Comunidad autónoma	CAGR 2021-2024	Margen EBITDA previsto 2025	Deuda neta / EBITDA 2024	Variación OWC / ventas 2024	Score final
1	WORLD MEDICA SLU	46 - Comercio al por mayor	Madrid	14.4%	33.9%	-2.2x	-6.3%	0.875
2	OSTEOPLAC INNOVATIONS SL.	46 - Comercio al por mayor	País Vasco	25.4%	24.5%	-0.8x	-2.6%	0.840
3	EXCAVACIONES GRAVAL SOCIEDAD LIMITADA.	43 - Construcción especializada	Canarias	29.4%	50.9%	-2.1x	1.0%	0.825
4	ARFX FXCFI FNOCIA ROBOTICA SL.	47 - Comercio al por menor	Madrid	32.7%	23.2%	0.6x	-0.1%	0.821
5	ASPHALION SL	71 - Servicios técnicos de arquitectura e ingeniería	Cataluña	26.5%	24.9%	1.1x	1.0%	0.817
6	SERVAL NETWORKS SL.	61 - Telecomunicaciones	Madrid	28.3%	20.1%	0.9x	12.2%	0.814
7	CLERVIN ENERGY SERVICES SOCIEDAD LIMITADA.	33 - Reparación e instalación de maquinaria	Madrid	19.4%	41.0%	-2.6x	-4.2%	0.813
8	GAINZA FORGE SL.	25 - Fabricación de productos metálicos	País Vasco	25.6%	25.6%	-0.5x	-13.6%	0.813
9	CORPORACION H 10 HOTELS SL	82 - Actividades administrativas de oficina	Cataluña	32.3%	44.7%	-1.2x	2.3%	0.810
10	AXON PARTNERS GROUP CONSULTING SL.	70 - Consultoría de gestión empresarial	Madrid	19.0%	37.9%	-1.0x	1.0%	0.809

Nota de lectura

Una ratio Deuda neta / EBITDA negativa indica posición de caja neta. CAGR, margen EBITDA previsto y variación OWC / ventas se muestran en porcentaje.

Fuente: elaboración propia a partir de la salida final del ranking y de la base descargada de SABI.

Anexo 2: Proceso de extracción y depuración de la muestra

Este anexo resume el proceso de extracción, filtrado y depuración de la muestra utilizada en el trabajo. En él se recogen los principales criterios aplicados sobre la base descargada de SABI hasta llegar a la muestra final de 5.798 compañías empleada en el análisis.

Anexo 2. Proceso de extracción, filtros y depuración de la muestra

Resumen de los criterios aplicados a la base descargada de SABI y de la construcción de la muestra final.

Criterio / paso	Condición aplicada	Resultado u observación	Justificación económica
Rango de ingresos	Ingresos de explotación entre 5 y 250 millones en todos los ejercicios 2020-2024	Filtro inicial de tamaño	Aproxima el foco mid-market y excluye extremos.
EBITDA mínimo	EBITDA 2024 > 1 millón	Filtro inicial	Evita compañías demasiado pequeñas en capacidad operativa.
Tamaño por empleo	Entre 20 y 2.500 empleados	Filtro inicial	Mantiene una escala coherente con el universo objetivo.
Rentabilidad operativa	EBITDA no negativo y EBIT no negativo en 2020-2024	Filtro inicial	Excluye perfiles claramente distress en sentido operativo.
Solvencia contable	Fondos propios positivos en 2024	Filtro inicial	Evita situaciones contables extremas.
Consistencia financiera	Deuda financiera no negativa y tesorería no negativa en 2024	Filtro inicial	Limpia valores problemáticos y asegura coherencia de magnitudes.
Tipo de cuentas	No consolidadas, código U1 en SABI	Filtro inicial	Favorece comparabilidad entre compañías.
Empresas tras extracción inicial	Aplicación conjunta de filtros anteriores	5.834 compañías	Base inicial operativa.
Variables descargadas	Nombre, NIF, CCAA, CNAE, empleados, ingresos, EBITDA, EBIT, deuda, tesorería, existencias, deudores y acreedores	Descarga final	Permite construir indicadores, modelos y ranking.
Depuración técnica	Homogeneización de nombres, conversión a formato numérico y revisión de duplicados	Paso previo al modelado	Garantiza consistencia de la base.
Exclusión sectorial	CNAE 64, 65 y 66	Se excluyen actividades financieras y de seguros	Su lógica contable difiere sustancialmente del resto.
Muestra final	Tras exclusión sectorial	5.798 compañías	Base definitiva para modelos y ranking.

Fuente: elaboración propia.

Nombre del producto	SABI Informa	Resultado etapa resultado búsqueda	
Actualización:	319		
Versión software	171.00		
Actualización datos	04/03/2026 (nº 3190)		
Usuario	CUNEF IP-3		
Export date	08/03/2026		
Cut off date	31/03		
1. Ingresos de explotación (mil EUR): 2024, 2023, 2022, 2021, 2020, para todos los periodos seleccionados, min=5,000, max=250,000		23,297	23,297
2. EBITDA (mil EUR): 2024, min=1,000		28,807	13,490
3. Tipos de empresas: Empresas		1,998,459	13,490
4. EBITDA (mil EUR): 2024, 2023, 2022, 2021, 2020, para todos los periodos seleccionados, min=0		285,133	11,897
5. Código de consolidación: U1 (empresas sólo con cuentas no consolidadas)		1,989,486	10,788
6. Fondos propios (mil EUR): 2024, min=0		626,084	10,748
7. Deudas financieras (mil EUR): 2024, min=0		261,529	7,568
8. Tesorería (mil EUR): 2024, min=0		659,599	7,516
9. Número empleados: 2024, min=20, max=2,500		63,618	6,602
10. EBIT (mil EUR): 2024, 2023, 2022, 2021, 2020, para todos los periodos seleccionados, min=0		240,213	5,834
Búsqueda booleana : 1 Y 2 Y 3 Y 4 Y 5 Y 6 Y 7 Y 8 Y 9 Y 10			
		TOTAL	5,834

Fuente: SABI.

Anexo 3: Especificación técnica de los modelos

En este anexo se presenta, de forma sintética, la estructura técnica de los dos ejercicios predictivos desarrollados en el trabajo. Se incluyen la variable objetivo, las principales variables de entrada, la partición entre entrenamiento y test, y los modelos comparados en cada caso.

Anexo 3. Especificación técnica de los modelos y métricas

Comparación de los dos ejercicios predictivos y de sus resultados fuera de muestra.

A. Especificación técnica		
Elemento	Modelo crecimiento ingresos	Modelo margen EBITDA
Variable objetivo	Crecimiento de ingresos 2024	Margen EBITDA 2024
Muestra total	5.798 empresas	5.798 empresas
Split train/test	70% / 30%	70% / 30%
Observaciones entrenamiento	4.058	4.058
Observaciones test	1.740	1.740
Modelos comparados	Regresión lineal y Random Forest	Regresión lineal y Random Forest
Variables de entrada principales	Crecimientos 2021-2023, CAGR 2020-2023, margen EBIT 2023, volatilidad crecimiento y log Ingresos 2023	Margen EBITDA 2023, margen EBITDA 2021, volatilidad crecimiento, volatilidad margen EBITDA y log ingresos 2023
Hiperparámetros RF	200 árboles, profundidad máxima 8, mínimo 5 observaciones por nodo terminal	200 árboles, profundidad máxima 8, mínimo 5 observaciones por nodo terminal
Métricas de evaluación	MAE, RMSE y R ²	MAE, RMSE y R ²

Fuente: elaboración propia.

Anexo 4: Construcción del score de priorización

Este anexo resume la construcción del score de priorización utilizado para ordenar las compañías de la muestra y obtener el Top 100 final. En él se muestran las variables incorporadas al score, su tratamiento previo y la lógica general de ponderación aplicada.

Anexo 4. Construcción del score de priorización

Variables, pesos y lógica económica utilizados para construir el ranking final del Top 100.

Variable del score	Tratamiento previo	Percentil invertido	Peso	Sentido económico
CAGR ingresos 2021-2024	Winsorización 1%-99%	No	25.0%	Recoge la trayectoria de crecimiento observada.
Margen EBITDA previsto 2025	Predicción del modelo final	No	35.0%	Proxy de calidad operativa futura.
Deuda financiera neta / EBITDA 2024	Winsorización 1%-99%	Sí	15.0%	Menor apalancamiento = mayor atractivo relativo.
Variación OWC 2023-2024 / ventas 2024	Missings imputados con mediana	Sí	10.0%	Menor absorción de caja = perfil más favorable.
Volatilidad crecimiento ingresos	Winsorización 1%-99%	Sí	7.5%	Mayor estabilidad de crecimiento = mejor señal.
Volatilidad margen EBITDA	Winsorización 1%-99%	Sí	7.5%	Mayor estabilidad operativa = mejor señal.

Fórmula y nota metodológica

$Score = 0,35 \cdot P(\text{EBITDA previsto}) + 0,25 \cdot P(\text{CAGR}) + 0,15 \cdot P(\text{Deuda neta/EBITDA invertido}) + 0,10 \cdot P(\text{OWC invertido}) + 0,075 \cdot P(\text{Vol crecimiento invertida}) + 0,075 \cdot P(\text{Vol margen invertida})$

La ponderación no busca optimizar estadísticamente el ajuste, sino trasladar al ranking una lógica de priorización coherente con una primera fase de screening en private equity.

Fuente: elaboración propia.

Anexo 5. Fragmentos de código relevantes

Nota sobre el código empleado:

En este anexo se incluyen únicamente los fragmentos de código más relevantes para la comprensión de la metodología aplicada en el trabajo. El código completo empleado en el trabajo se pone a disposición de la tutora y del tribunal en archivo aparte, si así se solicita.

5.1 Preparación y depuración de la muestra

El siguiente fragmento recoge las operaciones iniciales de preparación de la base de datos. En concreto, incluye la limpieza de columnas auxiliares, la conversión de variables a formato numérico, la eliminación de duplicados y la exclusión de sectores no comparables dentro del universo de análisis. Su finalidad es mostrar cómo se obtiene una base de partida consistente para el resto del trabajo.

```
# Eliminar columna índice si aparece

if "Unnamed: 0" in df.columns:
    df = df.drop(columns=["Unnamed: 0"])

df.columns.tolist()

# Sustituir valores tipo texto por NaN

df = df.replace("n.d.", np.nan)
df = df.replace("n.s.", np.nan)
df = df.replace("-", np.nan)

# Renombrar columnas para trabajar más cómodo

df = df.rename(columns={
    "Nombre": "nombre",
    "Código NIF": "nif",
    "Comunidad autónoma": "region",
    "Código primario CNAE 2009": "cnae_4d",
    "Número empleados\n2024": "empleados_2024",

    "Ingresos de explotación\nmil EUR\n2024": "revenue_2024",
    "Ingresos de explotación\nmil EUR\n2023": "revenue_2023",
    "Ingresos de explotación\nmil EUR\n2022": "revenue_2022",
    "Ingresos de explotación\nmil EUR\n2021": "revenue_2021",
    "Ingresos de explotación\nmil EUR\n2020": "revenue_2020",

    "EBITDA\nmil EUR\n2024": "ebitda_2024",
    "EBITDA\nmil EUR\n2023": "ebitda_2023",
    "EBITDA\nmil EUR\n2022": "ebitda_2022",
    "EBITDA\nmil EUR\n2021": "ebitda_2021",
    "EBITDA\nmil EUR\n2020": "ebitda_2020",

    "EBIT\nmil EUR\n2024": "ebit_2024",
    "EBIT\nmil EUR\n2023": "ebit_2023",
    "EBIT\nmil EUR\n2022": "ebit_2022",
    "EBIT\nmil EUR\n2021": "ebit_2021",
    "EBIT\nmil EUR\n2020": "ebit_2020",

    "Deudas financieras\nmil EUR\n2024": "debt_2024",
    "Tesorería\nmil EUR\n2024": "cash_2024",
```

```

"Existencias\nmil EUR\n2024": "stock_2024",
"Existencias\nmil EUR\n2023": "stock_2023",
"Existencias\nmil EUR\n2022": "stock_2022",
"Existencias\nmil EUR\n2021": "stock_2021",
"Existencias\nmil EUR\n2020": "stock_2020",

"Deudores\nmil EUR\n2024": "debtors_2024",
"Deudores\nmil EUR\n2023": "debtors_2023",
"Deudores\nmil EUR\n2022": "debtors_2022",
"Deudores\nmil EUR\n2021": "debtors_2021",
"Deudores\nmil EUR\n2020": "debtors_2020",

"Acreeedores comerciales\nmil EUR\n2024": "creditors_2024",
"Acreeedores comerciales\nmil EUR\n2023": "creditors_2023",
"Acreeedores comerciales\nmil EUR\n2022": "creditors_2022",
"Acreeedores comerciales\nmil EUR\n2021": "creditors_2021",
"Acreeedores comerciales\nmil EUR\n2020": "creditors_2020"
})

# Convertir columnas numéricas a formato numérico

columnas_numericas = [
    "cnae_4d",
    "empleados_2024",

    "revenue_2024", "revenue_2023", "revenue_2022", "revenue_2021",
"revenue_2020",
    "ebitda_2024", "ebitda_2023", "ebitda_2022", "ebitda_2021", "ebitda_2020",
    "ebit_2024", "ebit_2023", "ebit_2022", "ebit_2021", "ebit_2020",

    "debt_2024", "cash_2024",

    "stock_2024", "stock_2023", "stock_2022", "stock_2021", "stock_2020",
    "debtors_2024", "debtors_2023", "debtors_2022", "debtors_2021",
"debtors_2020",
    "creditors_2024", "creditors_2023", "creditors_2022", "creditors_2021",
"creditors_2020"
]

for col in columnas_numericas:
    df[col] = pd.to_numeric(df[col], errors="coerce")

# Comprobar tipos y missing values

# Eliminar duplicados por NIF si los hubiera
df = df.drop_duplicates(subset="nif")

# Crear CNAE a 2 dígitos por si luego lo usamos
df["cnae_2d"] = df["cnae_4d"].astype("Int64").astype(str).str[:2]

# Eliminar sectores no comparables: CNAE 64, 65 y 66
df = df[~df["cnae_2d"].isin(["64", "65", "66"])].copy()

# Resetear índice
df = df.reset_index(drop=True)

```

5.2 Construcción de variables

El siguiente bloque muestra la construcción de las principales variables utilizadas en el análisis empírico. Entre ellas se incluyen medidas de crecimiento, márgenes, volatilidad y tamaño, así como el tratamiento de valores extremos mediante winsorización. Este apartado traduce a términos operativos la lógica económica y financiera sobre la que se apoya la modelización posterior.

```
# Crear variables para el modelo de EBITDA Margin 2024

# Target: margen EBITDA en 2024
df["ebitda_margin_2024"] = df["ebitda_2024"] / df["revenue_2024"]

# Crecimientos históricos de ingresos
df["rev_growth_2023"] = (df["revenue_2023"] / df["revenue_2022"]) - 1
df["rev_growth_2022"] = (df["revenue_2022"] / df["revenue_2021"]) - 1
df["rev_growth_2021"] = (df["revenue_2021"] / df["revenue_2020"]) - 1

# CAGR 2020-2023
df["revenue_cagr_2020_2023"] = ((df["revenue_2023"] / df["revenue_2020"]) **
(1/3)) - 1

# Márgenes EBITDA históricos
df["ebitda_margin_2023"] = df["ebitda_2023"] / df["revenue_2023"]
df["ebitda_margin_2022"] = df["ebitda_2022"] / df["revenue_2022"]
df["ebitda_margin_2021"] = df["ebitda_2021"] / df["revenue_2021"]
df["ebitda_margin_2020"] = df["ebitda_2020"] / df["revenue_2020"]

# Márgenes EBIT históricos
df["ebit_margin_2023"] = df["ebit_2023"] / df["revenue_2023"]
df["ebit_margin_2022"] = df["ebit_2022"] / df["revenue_2022"]
df["ebit_margin_2021"] = df["ebit_2021"] / df["revenue_2021"]
df["ebit_margin_2020"] = df["ebit_2020"] / df["revenue_2020"]

# Volatilidad histórica del crecimiento
df["rev_growth_vol_2021_2023"] = df[["rev_growth_2021", "rev_growth_2022",
"rev_growth_2023"]].std(axis=1)

# Volatilidad histórica del margen EBITDA
df["ebitda_margin_vol_2021_2023"] = df[["ebitda_margin_2021",
"ebitda_margin_2022", "ebitda_margin_2023"]].std(axis=1)

# Margen EBITDA medio histórico
df["ebitda_margin_avg_2021_2023"] = df[["ebitda_margin_2021",
"ebitda_margin_2022", "ebitda_margin_2023"]].mean(axis=1)

# Tamaño: logaritmo de revenue 2023
df["log_revenue_2023"] = np.log(df["revenue_2023"])

# Sustituir infinitos por NaN

df = df.replace([np.inf, -np.inf], np.nan)

# Revisar las variables creadas

columnas_margin = [
    "ebitda_margin_2024",
    "rev_growth_2023",
    "rev_growth_2022",
    "rev_growth_2021",
```

```

    "revenue_cagr_2020_2023",
    "ebitda_margin_2023",
    "ebitda_margin_2022",
    "ebitda_margin_2021",
    "ebitda_margin_2020",
    "ebit_margin_2023",
    "ebit_margin_2022",
    "ebit_margin_2021",
    "ebit_margin_2020",
    "rev_growth_vol_2021_2023",
    "ebitda_margin_vol_2021_2023",
    "ebitda_margin_avg_2021_2023",
    "log_revenue_2023"
]

df[columnas_margin].describe().round(3)

# Función para winsorizar

def winsorize_series(s, lower=0.01, upper=0.99):
    lower_bound = s.quantile(lower)
    upper_bound = s.quantile(upper)
    return s.clip(lower=lower_bound, upper=upper_bound)

# Variables a winsorizar

variables_winsor_margin = [
    "ebitda_margin_2024",
    "rev_growth_2021",
    "rev_growth_2022",
    "rev_growth_2023",
    "revenue_cagr_2020_2023",
    "ebitda_margin_2020",
    "ebitda_margin_2021",
    "ebitda_margin_2022",
    "ebitda_margin_2023",
    "ebit_margin_2020",
    "ebit_margin_2021",
    "ebit_margin_2022",
    "ebit_margin_2023",
    "rev_growth_vol_2021_2023",
    "ebitda_margin_vol_2021_2023"
]

# Aplicar winsorización

for col in variables_winsor_margin:
    df[col] = winsorize_series(df[col], lower=0.01, upper=0.99)

# Revisar estadísticas tras winsorizar

df[columnas_margin].describe().round(3)

```

5.3 Estimación y evaluación de los modelos

El siguiente fragmento resume el proceso de estimación de los modelos predictivos y su posterior evaluación. Se incluyen la selección de variables, la división entre muestra de entrenamiento y test, la estimación de los modelos comparados y el cálculo de las métricas de error utilizadas en el trabajo. Se trata de un bloque representativo del proceso de modelización seguido.

```
# Seleccionar variables explicativas del modelo de EBITDA margin

variables_modelo_margin = [
    "rev_growth_2023",
    "rev_growth_2022",
    "rev_growth_2021",
    "revenue_cagr_2020_2023",
    "ebitda_margin_2023",
    "ebitda_margin_2022",
    "ebitda_margin_2021",
    "ebit_margin_2023",
    "ebit_margin_2022",
    "ebit_margin_2021",
    "rev_growth_vol_2021_2023",
    "ebitda_margin_vol_2021_2023",
    "ebitda_margin_avg_2021_2023",
    "log_revenue_2023"
]

# Crear la base para el modelo de EBITDA margin

df_margin_model = df.copy()

df_margin_model = df_margin_model.dropna(subset = variables_modelo_margin +
["ebitda_margin_2024"])

# Importar statsmodels

import statsmodels.formula.api as smf
import statsmodels.api as sm

# Estimar modelo explicativo OLS para EBITDA margin 2024

model_margin_ols = smf.ols(
    "ebitda_margin_2024 ~ rev_growth_2023 + rev_growth_2022 + rev_growth_2021 +
revenue_cagr_2020_2023 + ebitda_margin_2023 + ebitda_margin_2022 + ebitda_margin_2021
+ ebit_margin_2023 + ebit_margin_2022 + ebit_margin_2021 + rev_growth_vol_2021_2023 +
ebitda_margin_vol_2021_2023 + ebitda_margin_avg_2021_2023 + log_revenue_2023",
    data = df_margin_model
).fit()

# Estimar modelo explicativo parsimonioso para EBITDA margin 2024

model_margin_ols_2 = smf.ols(
    "ebitda_margin_2024 ~ ebitda_margin_2023 + ebitda_margin_2021 +
rev_growth_vol_2021_2023 + ebitda_margin_vol_2021_2023 + log_revenue_2023",
    data = df_margin_model
).fit()

# Definir variables explicativas (X) y variable objetivo (y)

X_margin = df_margin_model[
    [
```

```

        "ebitda_margin_2023",
        "ebitda_margin_2021",
        "rev_growth_vol_2021_2023",
        "ebitda_margin_vol_2021_2023",
        "log_revenue_2023"
    ]
]

y_margin = df_margin_model["ebitda_margin_2024"]

print("Dimensión de X:", X_margin.shape)
print("Dimensión de y:", y_margin.shape)

# Dividir la muestra en training set y test set

from sklearn.model_selection import train_test_split

X_train_margin, X_test_margin, y_train_margin, y_test_margin =
train_test_split(
    X_margin,
    y_margin,
    test_size=0.30,
    random_state=42
)

print("X_train:", X_train_margin.shape)
print("X_test:", X_test_margin.shape)
print("y_train:", y_train_margin.shape)
print("y_test:", y_test_margin.shape)

# Comprobar que no haya valores nulos

print("NaN en X_train:", X_train_margin.isna().sum().sum())
print("NaN en X_test:", X_test_margin.isna().sum().sum())
print("NaN en y_train:", y_train_margin.isna().sum())
print("NaN en y_test:", y_test_margin.isna().sum())

# Entrenar modelo de Regresión Lineal

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

modelo_lr_margin = LinearRegression()
modelo_lr_margin.fit(X_train_margin, y_train_margin)

# Generar predicciones sobre el test set

pred_lr_margin = modelo_lr_margin.predict(X_test_margin)

# Evaluar el modelo

mae_lr_margin = mean_absolute_error(y_test_margin, pred_lr_margin)
rmse_lr_margin = np.sqrt(mean_squared_error(y_test_margin, pred_lr_margin))
r2_lr_margin = r2_score(y_test_margin, pred_lr_margin)

print("MAE:", round(mae_lr_margin, 4))
print("RMSE:", round(rmse_lr_margin, 4))
print("R2:", round(r2_lr_margin, 4))

# Entrenar modelo Random Forest

from sklearn.ensemble import RandomForestRegressor

```

```
modelo_rf_margin = RandomForestRegressor(  
    n_estimators=200,  
    max_depth=8,  
    min_samples_leaf=5,  
    random_state=42  
)  
  
modelo_rf_margin.fit(X_train_margin, y_train_margin)  
  
# Generar predicciones con Random Forest  
  
pred_rf_margin = modelo_rf_margin.predict(X_test_margin)  
  
# Evaluar el modelo Random Forest  
  
mae_rf_margin = mean_absolute_error(y_test_margin, pred_rf_margin)  
rmse_rf_margin = np.sqrt(mean_squared_error(y_test_margin, pred_rf_margin))  
r2_rf_margin = r2_score(y_test_margin, pred_rf_margin)  
  
print("MAE:", round(mae_rf_margin, 4))  
print("RMSE:", round(rmse_rf_margin, 4))  
print("R2:", round(r2_rf_margin, 4))
```

5.4 Construcción del score y del ranking final

El siguiente bloque recoge la construcción de la herramienta final de priorización de compañías. En él se muestran la generación de variables adicionales, su transformación a percentiles comparables, la aplicación de ponderaciones y el cálculo del score agregado que permite ordenar las empresas y obtener el ranking final. Este apartado constituye la traducción práctica del modelo de screening propuesto.

```
# Crear variables necesarias para 2025

# Revenue growth histórico
df["rev_growth_2024"] = (df["revenue_2024"] / df["revenue_2023"]) - 1
df["rev_growth_2023"] = (df["revenue_2023"] / df["revenue_2022"]) - 1
df["rev_growth_2022"] = (df["revenue_2022"] / df["revenue_2021"]) - 1

# CAGR 2021-2024
df["revenue_cagr_2021_2024"] = ((df["revenue_2024"] / df["revenue_2021"]) **
(1/3)) - 1

# Márgenes EBITDA
df["ebitda_margin_2024"] = df["ebitda_2024"] / df["revenue_2024"]
df["ebitda_margin_2023"] = df["ebitda_2023"] / df["revenue_2023"]
df["ebitda_margin_2022"] = df["ebitda_2022"] / df["revenue_2022"]

# Márgenes EBIT
df["ebit_margin_2024"] = df["ebit_2024"] / df["revenue_2024"]

# Volatilidad histórica del crecimiento
df["rev_growth_vol_2022_2024"] = df[["rev_growth_2022", "rev_growth_2023",
"rev_growth_2024"]].std(axis=1)

# Volatilidad histórica del margen EBITDA
df["ebitda_margin_vol_2022_2024"] = df[["ebitda_margin_2022",
"ebitda_margin_2023", "ebitda_margin_2024"]].std(axis=1)

# Tamaño: logaritmo de revenue 2024
df["log_revenue_2024"] = np.log(df["revenue_2024"])

# Sustituir infinitos por NaN
df = df.replace([np.inf, -np.inf], np.nan)

# Función para winsorizar

def winsorize_series(s, lower=0.01, upper=0.99):
    lower_bound = s.quantile(lower)
    upper_bound = s.quantile(upper)
    return s.clip(lower=lower_bound, upper=upper_bound)

# Variables a winsorizar para 2025

variables_winsor_2025 = [
    "rev_growth_2024",
    "rev_growth_2023",
    "rev_growth_2022",
    "revenue_cagr_2021_2024",
    "ebitda_margin_2024",
    "ebitda_margin_2023",
    "ebitda_margin_2022",
    "ebit_margin_2024",
    "rev_growth_vol_2022_2024",
    "ebitda_margin_vol_2022_2024"
```

```

]

for col in variables_winsor_2025:
    df[col] = winsorize_series(df[col], lower=0.01, upper=0.99)

# Revisar descriptivos tras winsorizar
df[variables_winsor_2025 + ["log_revenue_2024"]].describe().round(3)

# Crear base para predecir EBITDA Margin 2025

df_margin_2025 = df.copy()

df_margin_2025["ebitda_margin_2023_model"] =
df_margin_2025["ebitda_margin_2024"]
df_margin_2025["ebitda_margin_2021_model"] =
df_margin_2025["ebitda_margin_2022"]
df_margin_2025["rev_growth_vol_2021_2023_model"] =
df_margin_2025["rev_growth_vol_2022_2024"]
df_margin_2025["ebitda_margin_vol_2021_2023_model"] =
df_margin_2025["ebitda_margin_vol_2022_2024"]
df_margin_2025["log_revenue_2023_model"] = df_margin_2025["log_revenue_2024"]

variables_pred_margin_2025 = [
    "ebitda_margin_2023_model",
    "ebitda_margin_2021_model",
    "rev_growth_vol_2021_2023_model",
    "ebitda_margin_vol_2021_2023_model",
    "log_revenue_2023_model"
]

df_margin_2025 = df_margin_2025.dropna(subset=variables_pred_margin_2025)

print("Número de empresas base EBITDA margin 2025:", len(df_margin_2025))

# Preparar X y generar predicción de EBITDA Margin 2025

X_margin_2025 = df_margin_2025[variables_pred_margin_2025].copy()

X_margin_2025.columns = [
    "ebitda_margin_2023",
    "ebitda_margin_2021",
    "rev_growth_vol_2021_2023",
    "ebitda_margin_vol_2021_2023",
    "log_revenue_2023"
]

df_margin_2025["pred_ebitda_margin_2025"] =
modelo_lr_margin.predict(X_margin_2025)

# Construir la base final de ranking

df_ranking = df_margin_2025[[
    "nif",
    "nombre",
    "region",
    "cnae_4d",
    "cnae_2d",
    "revenue_2024",
    "ebitda_2024",
    "ebit_2024",
    "debt_2024",
    "cash_2024",
    "stock_2024",

```

```

    "stock_2023",
    "debtors_2024",
    "debtors_2023",
    "creditors_2024",
    "creditors_2023",
    "revenue_cagr_2021_2024",
    "rev_growth_vol_2022_2024",
    "ebitda_margin_vol_2022_2024",
    "pred_ebitda_margin_2025"
  ]].copy()

print("Número de empresas en la base final de ranking:", len(df_ranking))
df_ranking.head()

# Construir variables financieras adicionales

# Deuda financiera neta
df_ranking["net_debt_2024"] = df_ranking["debt_2024"] - df_ranking["cash_2024"]

# Ratio DFN / EBITDA
df_ranking["net_debt_ebitda_2024"] = df_ranking["net_debt_2024"] /
df_ranking["ebitda_2024"]

# Operating Working Capital 2024 y 2023
df_ranking["owc_2024"] = df_ranking["stock_2024"] + df_ranking["debtors_2024"]
- df_ranking["creditors_2024"]
df_ranking["owc_2023"] = df_ranking["stock_2023"] + df_ranking["debtors_2023"]
- df_ranking["creditors_2023"]

# Variación de OWC sobre ventas
df_ranking["delta_owc_revenue_2024"] = (
    (df_ranking["owc_2024"] - df_ranking["owc_2023"]) /
df_ranking["revenue_2024"]
)

# Sustituir infinitos por NaN
df_ranking = df_ranking.replace([np.inf, -np.inf], np.nan)

# Imputar missing values de OWC con la mediana
df_ranking["delta_owc_revenue_2024"] =
df_ranking["delta_owc_revenue_2024"].fillna(
    df_ranking["delta_owc_revenue_2024"].median()
)

# Winsorizar variables del score final

def winsorize_series(s, lower=0.01, upper=0.99):
    lower_bound = s.quantile(lower)
    upper_bound = s.quantile(upper)
    return s.clip(lower=lower_bound, upper=upper_bound)

variables_score = [
    "revenue_cagr_2021_2024",
    "pred_ebitda_margin_2025",
    "net_debt_ebitda_2024",
    "delta_owc_revenue_2024",
    "rev_growth_vol_2022_2024",
    "ebitda_margin_vol_2022_2024"
]

for col in variables_score:
    df_ranking[col] = winsorize_series(df_ranking[col], lower=0.01, upper=0.99)

```

```

# Convertir variables a percentiles

df_ranking["p_cagr"] = df_ranking["revenue_cagr_2021_2024"].rank(pct=True)
df_ranking["p_ebitda_margin"] =
df_ranking["pred_ebitda_margin_2025"].rank(pct=True)

df_ranking["p_leverage"] = 1 -
df_ranking["net_debt_ebitda_2024"].rank(pct=True)
df_ranking["p_owc"] = 1 - df_ranking["delta_owc_revenue_2024"].rank(pct=True)
df_ranking["p_rev_vol"] = 1 -
df_ranking["rev_growth_vol_2022_2024"].rank(pct=True)
df_ranking["p_margin_vol"] = 1 -
df_ranking["ebitda_margin_vol_2022_2024"].rank(pct=True)

# Construir score final ponderado

df_ranking["score_final"] = (
    0.25 * df_ranking["p_cagr"] +
    0.35 * df_ranking["p_ebitda_margin"] +
    0.15 * df_ranking["p_leverage"] +
    0.10 * df_ranking["p_owc"] +
    0.075 * df_ranking["p_rev_vol"] +
    0.075 * df_ranking["p_margin_vol"]
)

# Revisar el score final

df_ranking[[
    "nombre",
    "revenue_cagr_2021_2024",
    "pred_ebitda_margin_2025",
    "net_debt_ebitda_2024",
    "delta_owc_revenue_2024",
    "score_final"
]].head()

# Ordenar la base por score final

df_ranking = df_ranking.sort_values("score_final",
ascending=False).reset_index(drop=True)

df_ranking["ranking"] = df_ranking.index + 1

# Extraer el Top 100

top_100 = df_ranking.head(100).copy()

```