



Máster en Ingeniería Industrial

Trabajo de Fin de Máster

Predicción de quiebras financieras utilizando Machine Learning

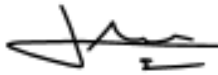
Autor: Luis Sierra Bonilla

Director: Javier Fuentes Ibáñez

Madrid

Julio de 2025

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
Predicción de quiebras financieras utilizando Machine Learning
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el
curso académico 2024-2025 es de mi autoría, original e inédito y
no ha sido presentado con anterioridad a otros efectos. El Proyecto no es
plagio de otro, ni total ni parcialmente y la información que ha sido tomada
de otros documentos está debidamente referenciada.



Fdo.: Luis Sierra Bonilla

Fecha: 11/ 07/ 2025

Autorizada la entrega del proyecto
EL DIRECTOR DEL PROYECTO



Fdo.: Javier Fuentes Ibáñez

Fecha: 11/ 07/ 2025



Máster en Ingeniería Industrial

Trabajo de Fin de Máster

Predicción de quiebras financieras utilizando Machine Learning

Autor: Luis Sierra Bonilla

Director: Javier Fuentes Ibáñez

Madrid

Julio de 2025

Predicción de quiebras financieras utilizando Machine Learning

Autor: Sierra Bonilla, Luis

Director: Fuentes Ibáñez, Javier

Entidad colaboradora: ICAI – Universidad Pontificia Comillas

Resumen del proyecto

El proyecto que se presenta a continuación constituye una iniciativa de emprendimiento con un marcado enfoque ingenieril. Su objetivo es estudiar la viabilidad, tanto técnica como económica y empresarial, de una herramienta que integra dos sectores de gran relevancia e impacto global: el sector financiero y el sector tecnológico, representado por una solución basada en inteligencia artificial.

Este trabajo permite evidenciar no solo la importancia de la innovación y de la incorporación de tecnologías avanzadas en sectores tradicionales como el financiero, sino también el proceso completo de desarrollo de un producto de última generación y su planificación estratégica para su introducción y comercialización en el mercado.

En conjunto, el proyecto ofrece una visión integral del ciclo de vida de una iniciativa emprendedora de carácter ingenieril, abarcando desde la concepción técnica del producto hasta su análisis comercial, con una orientación clara hacia la generación de valor económico y social.

Para el desarrollo de este proyecto se pueden distinguir tres grandes bloques:

1. Marco teórico

En este apartado se presentan los dos pilares fundamentales sobre los que se sustenta el proyecto: las quiebras financieras y machine learning. A través del análisis de ambos conceptos, se identifica la oportunidad de integrar estas dos áreas críticas para dar lugar a una herramienta innovadora: un sistema de predicción de quiebras financieras basado en machine learning.

El marco teórico permite comprender en profundidad qué son las quiebras financieras y el gran impacto que generan, tanto a nivel económico como social. Esta comprensión pone de manifiesto la necesidad de desarrollar herramientas que ayuden a las empresas a anticiparse y protegerse frente a estas situaciones. La innovación tecnológica, en particular el uso de inteligencia artificial se perfila como la vía más eficaz para lograrlo. Machine learning, por sus características, se presenta como la herramienta óptima para desarrollar este tipo de soluciones, convirtiéndose en el puente ideal entre el ámbito tecnológico y el financiero.

2. Desarrollo técnico

Este apartado se centra en la validación técnica de la propuesta mediante el desarrollo de un prototipo funcional. A través de este prototipo, se analiza el comportamiento del modelo, sus funcionalidades y rendimiento, permitiendo obtener conclusiones valiosas que servirán como base para su futura implementación definitiva y comercialización.

Este estudio técnico no solo demuestra la viabilidad práctica del producto, sino que también permite visualizar cómo sería su funcionamiento en un entorno real, identificando las mejores metodologías y algoritmos a emplear. Además, refuerza el valor de aplicar inteligencia artificial al sector financiero como vía para mejorar su eficiencia y capacidad de anticipación.

3. Estudio empresarial

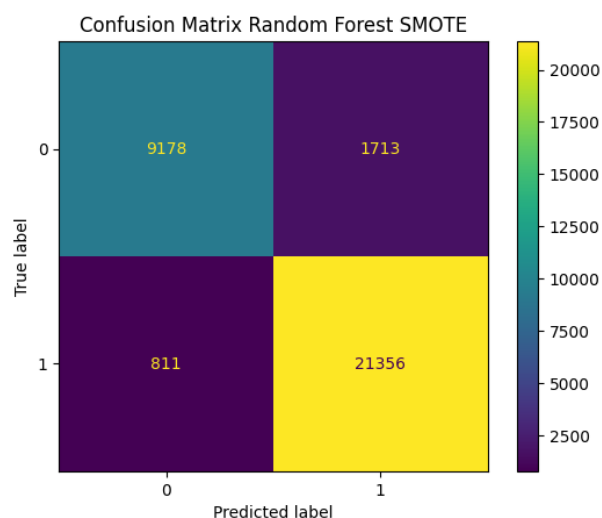
Una vez validado el modelo desde el punto de vista técnico, se desarrolla un plan de negocio que contempla la estrategia de comercialización y posicionamiento del producto. En este apartado se profundiza en el diseño del modelo de negocio mediante el desarrollo del Business Model Canvas y se incluye un plan financiero básico que permite evaluar la viabilidad económica del proyecto.

Este análisis empresarial proporciona una visión clara de cómo el producto podría entrar al mercado, cuáles serían sus principales segmentos de clientes, canales de distribución, fuentes de ingresos y estructura de costes. En conjunto, este bloque consolida una perspectiva integral del proyecto como iniciativa de emprendimiento de base tecnológica e ingenieril, orientada a resolver un problema real con alto impacto potencial.

La herramienta desarrollada, tal como se ha explicado previamente, consiste en un sistema de predicción de quiebras financieras basado en machine learning. Para construir el prototipo de esta solución, se ha utilizado la plataforma Google Colab y el lenguaje de programación Python, sobre el cual se ha implementado un modelo capaz de predecir la insolvencia de empresas a partir de una base de datos compuesta por registros de compañías solventes e insolventes.

Con el objetivo de evaluar diferentes enfoques, se han aplicado múltiples algoritmos de machine learning, como Random Forest, SVM, Gradient Boosting, redes neuronales y Bagging. Esto ha permitido comparar el comportamiento y rendimiento de cada método, identificando cuál se adapta mejor a la tarea específica de predecir quiebras empresariales.

En la siguiente imagen, extraída directamente de la herramienta desarrollada, se muestra uno de los resultados obtenidos al aplicar el modelo de Random Forest. El sistema analiza la información financiera de las empresas y clasifica cada caso como 0 (empresa en quiebra) o 1 (empresa solvente). Esta visualización representa de forma clara y comprensible cómo opera la herramienta, permitiendo evaluar de manera visual su eficacia en un caso concreto.



En la siguiente tabla se presentan los resultados obtenidos a partir de las distintas técnicas empleadas, según el tipo de datos utilizados en cada caso. Esta comparación permite evaluar el desempeño de cada modelo a través de diferentes métricas, facilitando así la identificación del enfoque más eficaz para la predicción de quiebras financieras.

Datos/Algoritmo	Resultados	Random Forest	SVM	Gradient Boosting	Redes Neuronales 3 capas	Redes Neuronales 5 capas	Redes Neuronales 10 capas	Bagging RF	Bagging SVM	Bagging GB	Bagging RN (mejor caso)
Datos SMOTE	Accuracy	0.9236	0.6710	0.9136	0.7019	0.7285	0.7094	0.9903	0.6713	0.9785	0.6869
	Precision	0.9257	0.6709	0.9218	0.7160	0.7410	0.7219	0.9909	0.6711	0.9782	0.6902
	Recall	0.9634	0.9997	0.9520	0.9130	0.9102	0.9151	0.9946	0.9997	0.9899	0.9672
	F1 Score	0.9442	0.8029	0.9366	0.7978	0.7991	0.7983	0.9927	0.8031	0.9841	0.8055
	Media	0.9392	0.7861	0.9310	0.7821	0.7947	0.7861	0.9921	0.7863	0.9826	0.7874
Datos mejorados	Accuracy	0.9325	0.9316	0.9156	0.9345	0.9344	0.9338	0.9316	0.9316	0.8982	0.9316
	Precision	0.9326	0.9316	0.9348	0.9347	0.9345	0.9338	0.9316	0.9316	0.9321	0.9316
	Recall	0.9998	1	0.9775	0.9997	0.9999	1	1	1	0.9606	1
	F1 score	0.9650	0.9646	0.9557	0.9660	0.9661	0.9657	0.9646	0.9646	0.9461	0.9646
	Media	0.9574	0.9569	0.9459	0.9587	0.9587	0.9583	0.9569	0.9569	0.9342	0.9569

Asimismo, el análisis del plan de negocio revela que la entrada de este producto en el mercado presenta ciertos desafíos, debido a su alto nivel de especialización técnica y a un entorno competitivo en el que los clientes requieren soluciones de gran precisión. En este contexto, la incorporación de inteligencia artificial, concretamente a través de técnicas de machine learning, se posiciona como el principal elemento diferenciador del producto frente a otras herramientas, especialmente las de enfoque tradicional.

La capacidad de innovación continua y los resultados avanzados que puede ofrecer esta tecnología constituyen uno de los pilares fundamentales del proyecto. Por ello, una estrategia de marketing sólida y una imagen de marca confiable serán aspectos clave para posicionar el producto en el mercado. A su vez, se ha optado por una estrategia de precios competitivos con el objetivo de facilitar la adopción entre pequeñas y medianas empresas

españolas, especialmente aquellas con recursos limitados, que encontrarán en esta herramienta una solución accesible, innovadora y eficaz a una necesidad crítica del sector financiero.

Las principales conclusiones que se extraen son las siguientes:

- Los datos utilizados, así como el volumen de información disponible, representan un factor clave en el desarrollo del estudio y en la calidad de los resultados obtenidos.
- Los resultados obtenidos en el prototipo evidencian la complejidad del problema de la predicción de quiebras financieras.
- El algoritmo de Random Forest se posiciona como el modelo con mejor rendimiento, especialmente cuando se potencia mediante la técnica de bagging.
- Para alcanzar una versión final del producto viable, será imprescindible contar con un equipo altamente cualificado, capaz de desarrollar una solución que cumpla con los estándares técnicos y de rendimiento exigidos.
- La comercialización del producto también presenta desafíos importantes, debido a la necesidad de penetrar en un mercado muy especializado, con una competencia consolidada.
- La principal ventaja competitiva que debe aprovechar este producto es el uso de machine learning como elemento diferencial frente a las soluciones tradicionales, así como una política de precios competitiva orientada a pequeñas y medianas empresas.
- El plan financiero proyectado refleja que será necesaria una inversión significativa, y que es poco probable alcanzar beneficios netos durante los primeros cinco años.

En resumen, se trata de un producto con un alto grado de complejidad técnica, en el que el rendimiento y la fiabilidad son aspectos críticos para los potenciales clientes. A pesar de los retos técnicos y financieros identificados, el valor añadido que ofrece la solución, especialmente para pequeñas y medianas empresas en el entorno español, queda demostrado, posicionándola como una iniciativa con alto potencial si se logra una correcta ejecución y estrategia de posicionamiento.

Palabras clave: predicción, quiebras, financieras, machine, learning, emprendimiento.

Prediction of Financial Bankruptcies Using Machine Learning
Author: Sierra Bonilla, Luis
Supervisor: Fuentes Ibáñez, Javier
Collaborating Entity: ICAI – Universidad Pontificia Comillas

Project Summary

The project presented below is an entrepreneurial initiative with a strong engineering focus. Its objective is to assess the technical, economic, and business feasibility of a tool that integrates two highly relevant and globally impactful sectors: the financial sector and the technology sector, represented here by a solution based on artificial intelligence.

This work highlights not only the importance of innovation and the integration of advanced technologies into traditional industries such as finance, but also the complete development process of a cutting-edge product and the strategic planning required for its introduction and commercialization in the market.

Overall, the project offers a comprehensive view of the lifecycle of an engineering-driven entrepreneurial initiative, from the technical conception of the product to its commercial analysis, with a clear orientation toward generating economic and social value.

The development of this project is structured into three main sections:

1. Theoretical Framework

This section presents the two fundamental pillars on which the project is based: financial bankruptcies and machine learning. Through the analysis of both concepts, the opportunity to integrate these two critical areas is identified, giving rise to an innovative tool: a financial bankruptcy prediction system based on machine learning.

The theoretical framework allows for a deep understanding of what financial bankruptcies are and the significant impact they generate, both economically and socially. This understanding reveals the need to develop tools that help companies anticipate and protect themselves from such situations. Technological innovation, particularly the use of artificial intelligence, emerges as the most effective way to achieve this. Machine learning, due to its characteristics, is presented as the optimal tool for developing this kind of solution, becoming the ideal bridge between the technological and financial domains.

2. Technical Development

This section focuses on the technical validation of the proposal through the development of a functional prototype. This prototype allows for the analysis of the model's behaviour, functionalities, and performance, offering valuable conclusions that will serve as a foundation for future implementation and commercialization.

This technical study not only demonstrates the practical feasibility of the product but also provides a clear view of how it would operate in a real-world setting, identifying the best methodologies and algorithms to employ. Additionally, it reinforces the value of applying artificial intelligence to the financial sector as a way to enhance efficiency and predictive capabilities.

3. Business Study

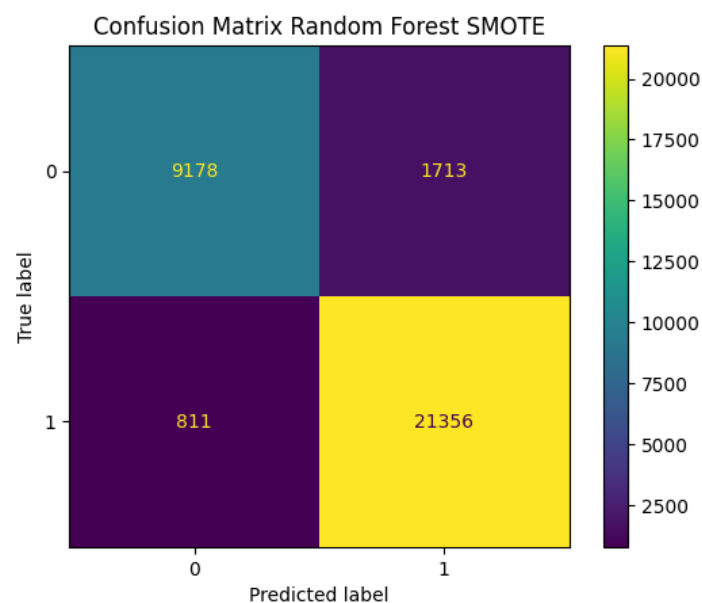
Once the model has been technically validated, a business plan is developed that includes the commercialization and market positioning strategy. This section delves into the business model design using the Business Model Canvas and includes a basic financial plan to evaluate the project's economic viability.

This business analysis offers a clear vision of how the product could enter the market, its main customer segments, distribution channels, revenue streams, and cost structure. Altogether, this section consolidates a comprehensive perspective of the project as a technology- and engineering-based entrepreneurial initiative aimed at solving a real problem with high potential impact.

As previously explained, the developed tool consists of a financial bankruptcy prediction system based on machine learning. To build the prototype, Google Colab and the Python programming language were used, implementing a model capable of predicting business insolvency using a dataset composed of records from both solvent and insolvent companies.

To evaluate various approaches, multiple machine learning algorithms were applied, including Random Forest, SVM, Gradient Boosting, Neural Networks and Bagging. This allowed for a comparison of each method's behaviour and performance, identifying the most suitable one for the specific task of bankruptcy prediction.

In the following image, taken directly from the developed tool, one of the results obtained using the Random Forest model is shown. The system analyses the financial data of companies and classifies each case as 0 (bankrupt) or 1 (solvent). This visualization clearly and effectively demonstrates how the tool operates, enabling a visual evaluation of its accuracy in a specific case.



The table below presents the results obtained from the different techniques used, depending on the type of data employed in each case. This comparison enables the evaluation of each model's performance across various metrics, facilitating the identification of the most effective approach for financial bankruptcy prediction.

Datos/Algoritmo	Resultados	Random Forest	SVM	Gradient Boosting	Redes Neuronales 3 capas	Redes Neuronales 5 capas	Redes Neuronales 10 capas	Bagging RF	Bagging SVM	Bagging GB	Bagging RN (mejor caso)
Datos SMOTE	Accuracy	0.9236	0.6710	0.9136	0.7019	0.7285	0.7094	0.9903	0.6713	0.9785	0.6869
	Precision	0.9257	0.6709	0.9218	0.7160	0.7410	0.7219	0.9909	0.6711	0.9782	0.6902
	Recall	0.9634	0.9997	0.9520	0.9130	0.9102	0.9151	0.9946	0.9997	0.9899	0.9672
	F1 Score	0.9442	0.8029	0.9366	0.7978	0.7991	0.7983	0.9927	0.8031	0.9841	0.8055
	Media	0.9392	0.7861	0.9310	0.7821	0.7947	0.7861	0.9921	0.7863	0.9826	0.7874
Datos mejorados	Accuracy	0.9325	0.9316	0.9156	0.9345	0.9344	0.9338	0.9316	0.9316	0.8982	0.9316
	Precision	0.9326	0.9316	0.9348	0.9347	0.9345	0.9338	0.9316	0.9316	0.9321	0.9316
	Recall	0.9998	1	0.9775	0.9997	0.9999	1	1	1	0.9606	1
	F1 score	0.9650	0.9646	0.9557	0.9660	0.9661	0.9657	0.9646	0.9646	0.9461	0.9646
	Media	0.9574	0.9569	0.9459	0.9587	0.9587	0.9583	0.9569	0.9569	0.9342	0.9569

Likewise, the business plan analysis reveals that entering the market with this product presents certain challenges due to its high level of technical specialization and a competitive environment in which clients demand highly precise solutions. In this context, the incorporation of artificial intelligence, specifically through machine learning techniques, stands out as the product's key differentiating factor, particularly compared to traditional tools.

The ability for continuous innovation and the advanced results that this technology can offer constitute one of the fundamental pillars of the project. Therefore, a solid marketing strategy and a strong brand image will be essential for positioning the product in the market. A competitive pricing strategy has also been adopted to facilitate adoption among small and medium-sized Spanish enterprises, especially those with limited resources, who will find in this tool an accessible, innovative, and effective solution to a critical need in the financial sector.

The main conclusions that can be derived are as follows:

- The data used, as well as the volume of available information, are key factors in the development of the study and the quality of the results obtained.
- The results obtained in the prototype highlight the complexity of the bankruptcy prediction problem.
- The Random Forest algorithm stands out as the best-performing model, especially when enhanced with the bagging technique.
- To develop a viable final version of the product, it will be essential to have a highly qualified team capable of delivering a solution that meets technical and performance standards.
- Commercializing the product also presents important challenges, as it must penetrate a highly specialized market with well-established competitors.

- The product's main competitive advantage lies in its use of machine Learning as a differentiator compared to traditional solutions, alongside a competitive pricing policy aimed at small and medium-sized enterprises.
- The projected financial plan shows that a significant investment will be required, and it is unlikely to achieve net profits within the first five years.

In summary, this is a highly technical product in which performance and reliability are critical for potential clients. Despite the technical and financial challenges identified, the added value that the solution offers, particularly for small and medium-sized enterprises in the Spanish market, is clearly demonstrated, positioning it as a high-potential initiative if executed with a sound implementation and market strategy.

Key words: prediction, bankruptcies, financial, machine, learning, entrepreneurship.

Índice de contenido

1.	Introducción	16
2.	Objetivo del proyecto	19
3.	Estructura del proyecto	19
4.	Marco teórico.....	21
4.1	Quiebras financieras	21
4.2	Machine Learning	25
5.	Modelo de predicción de quiebras financieras mediante Machine Learning	31
5.1	Metodología empleada en el modelo.....	31
5.2	Datos empleados	34
5.3	Funcionamiento del modelo	42
6.	Resultados	48
6.1	Resultados con SMOTE.....	48
6.2	Resultados con datos iniciales mejorados.....	57
7.	Plan de negocio.....	71
7.1	Resumen ejecutivo	71
7.2	Definición del plan de negocio	73
7.3	Análisis del mercado	73
7.4	Plan de marketing.....	77
7.5	Business model canvas	81
7.6	Hipótesis de desarrollo del producto.....	89
7.7	Plan financiero	92
8.	Conclusiones.....	99
9.	Alineamiento con los Objetivos de Desarrollo Sostenible (ODS).....	103
10.	Referencias.....	105
11.	Anexos.....	111

Índice de figuras

Figura 1: Empresas concursadas en España, primer trimestre 2025	22
Figura 2: Variación empresas concursadas en España.....	23
Figura 3: Proceso machine learning	26
Figura 4: Datos iniciales	34
Figura 5: Matriz de correlación.....	36
Figura 6: Boxplots	37
Figura 7: Datos sin outliers.	38
Figura 8: Boxplots sin outliers	39
Figura 9: Relación año-status.	40
Figura 10: SMOTE vs ADASYN.	42
Figura 11: Parámetros random forest.	45
Figura 12: Matriz confusión random forest SMOTE.	48
Figura 13: Curva precision-recall random forest SMOTE.	49
Figura 14: Factores más relevantes random forest SMOTE	50
Figura 15: Matriz confusión SVM SMOTE.	51
Figura 16: Matriz confusión gradient boosting SMOTE.	52
Figura 17: Curva precision-recall GB SMOTE.	53
Figura 18: Factores más relevantes GB SMOTE.....	54
Figura 19: Matriz confusión random forest R.....	58
Figura 20: Curva precision-recall random forest R.	59
Figura 21: Factores más relevantes random forest R.....	60
Figura 22: Matriz confusión random forest R con pesos.	61
Figura 23: Curva precision-recall random forest R con pesos	62
Figura 24: Matriz confusión SVM R	64
Figura 25: Matriz confusion gradient boosting R.....	65
Figura 26: Curva precision-recall GB R.	66
Figura 27: Factores más relevantes GB R.	67
Figura 28: Casos de quiebra España & UE	74
Figura 29: Business Model Canvas.....	88

Índice de tablas

Tabla 1: SMOTE vs ADASYN	43
Tabla 2: Resultados random forest SMOTE.	48
Tabla 3: Resultados SVM SMOTE.....	51
Tabla 4: Resultados Gradient Boosting SMOTE.	52
Tabla 5: Resultados redes neuronales SMOTE.	54
Tabla 6: Resultados Bagging SMOTE.	55
Tabla 7: Resultados random forest.....	57
Tabla 8: Random forest con pesos.	60
Tabla 9: Resultados SVM.	63
Tabla 10: Resultados Gradient Boosting.....	64
Tabla 11: Resultados redes neuronales.....	67
Tabla 12: Resultados Bagging.	68
Tabla 13: Resumen resultados.....	70
Tabla 14: Cronograma.	79
Tabla 15: Costes fijos.	93
Tabla 16: Costes variables.....	94
Tabla 17: Plan financiero.	97

1. Introducción

“En un entorno económico caracterizado por su alta volatilidad e incertidumbre, la detección temprana de riesgos financieros se ha convertido en una prioridad para empresas, inversores y entidades reguladoras. La predicción de quiebras empresariales, tradicionalmente abordada mediante modelos estadísticos convencionales, encuentra en las herramientas de inteligencia artificial —y en particular en el aprendizaje automático (Machine Learning)— un enfoque más robusto, dinámico y preciso.

Este Trabajo de Fin de Máster se centra en el desarrollo de un modelo predictivo de quiebras financieras mediante técnicas de machine learning, aplicando una perspectiva netamente ingenieril que integra análisis de datos, optimización de algoritmos y visión sistémica del problema. La investigación se concibe desde una óptica emprendedora, proponiendo una solución innovadora y escalable que puede ser implementada tanto en startups financieras como en departamentos de análisis de riesgo de grandes corporaciones.

La introducción de metodologías basadas en inteligencia artificial en el sector financiero no solo mejora la capacidad predictiva y la toma de decisiones, sino que también abre nuevas oportunidades para la automatización de procesos críticos, reduciendo errores humanos y fortaleciendo la resiliencia de las organizaciones. Este proyecto aspira a contribuir, desde el campo de la ingeniería industrial, al desarrollo de herramientas tecnológicas con impacto real en la sostenibilidad y eficiencia del sistema económico actual.”[1]

Con tan solo el siguiente guion es posible desarrollar una introducción como esta en cuestión de unos segundos, empleando la inteligencia artificial ChatGPT de OpenAI, como se presenta a continuación:

Eres un estudiante de ingeniería industrial que está realizando un trabajo de fin de máster de ingeniería industrial. El TFM trata de la predicción de quiebras financieras utilizando machine learning. Tienes que realizar lo siguiente:

- Realiza la introducción para este trabajo de fin de máster
- La introducción no debe ser más de 20 líneas
- Emplea un lenguaje formal Resalta la importancia de la introducción de herramientas de inteligencia artificial en el sector financiero
- Subraya el carácter de emprendimiento del proyecto, así como la relevancia de aportar una visión y carácter ingenieril.

Tal como se ha podido comprobar, mediante unas simples instrucciones, la inteligencia artificial es capaz de generar una introducción formal y estructurada que refleja los elementos clave previamente indicados: la relevancia de su aplicación en el sector financiero, el carácter emprendedor del proyecto, y la necesidad de destacar su enfoque ingenieril.

Esta breve demostración pone de manifiesto la disrupción tecnológica que representa la inteligencia artificial, cuyo avance ha sido especialmente acelerado en los últimos años. Su capacidad para automatizar tareas complejas y generar contenido con estructura lógica

y sentido contextual evidencia su enorme potencial y justifica su creciente integración en procesos clave de diversos sectores.

La aplicación de herramientas de última generación como machine learning no solo representa una mejora tecnológica, sino también un cambio de paradigma. En sectores estratégicos como el financiero, donde las decisiones tienen impacto directo en la estabilidad económica global, el uso de estas tecnologías se vuelve no solo oportuno, sino esencial. Este proyecto, por tanto, busca explorar y demostrar el valor real de la inteligencia artificial aplicada a la predicción de quiebras, un campo donde la ingeniería y la innovación pueden marcar una diferencia significativa.

Las quiebras financieras representan uno de los mayores riesgos a los que puede enfrentarse cualquier empresa. Sus consecuencias no solo afectan directamente a la organización afectada, que puede perder en cuestión de días todo lo que ha construido, sino que también generan un impacto sistémico que puede extenderse a todo el tejido económico, tanto a nivel local como global. Un ejemplo reciente, es la quiebra del banco de inversión estadounidense Lehman Brothers, la quiebra de este banco se considera como el detonante de la crisis financiera mundial que se vivió entre 2008 y 2011 (en España). La quiebra de Lehman Brothers fue un pieza más que contribuyo a la desencadenamiento de la crisis pero sus implicaciones fueron tremendas, como indica el informe del banco de España :“INFORME SOBRE LA CRISIS FINANCIERA Y BANCARIA EN ESPAÑA, 2008-2014”: “Así, por ejemplo, en los días posteriores a la quiebra del banco estadounidense, las primas en el mercado interbancario en ese país para operaciones a tres meses se multiplicaron por cuatro, pasando de menos de 1 pp a más de 3,5 pp, el índice bursátil S&P 500 descendió un 28% y su volatilidad se triplicó.” [2]. Evidenciando las consecuencias que este tipo de eventos tienen.

El caso de Lehman Brothers constituye un ejemplo emblemático de las consecuencias devastadoras que puede acarrear la quiebra financiera de una entidad de gran magnitud. La caída de este banco de inversión estadounidense no solo tuvo efectos a nivel local, sino que generó un impacto global que afectó a millones de personas en todo el mundo, alterando profundamente los mercados financieros y la economía real.

Sin embargo, las quiebras financieras no se limitan a episodios excepcionales protagonizados por grandes corporaciones cada cierto número de años. Se trata de un fenómeno que ocurre con frecuencia en empresas de todo tamaño, generando efectos significativos en sus cadenas de valor, en el empleo y en la estabilidad del tejido empresarial y económico, tanto a escala nacional como internacional.

Las quiebras financieras son circunstancias las que las empresas se enfrentan a diario, en un artículo de El Economista se exponía : “España es el país de la Unión Europea donde más han crecido porcentualmente las declaraciones de bancarrota de empresas desde 2020, año en el que se declaró la pandemia por coronavirus.” [3]. Lo que en números se traduce según el artículo en: “En concreto, el total de empresas que han manifestado entrar en quiebra ha crecido un 163%. En número, casi se han triplicado, pasando de 3.658 hace tres años a 9.606 al cierre del año pasado. Suman 19.134 negocios en los últimos tres años.” [3].

A la luz de estos datos, resulta evidente la necesidad de contar con herramientas capaces de anticipar situaciones de insolvencia, dada la magnitud del impacto que estas pueden

tener sobre las empresas, la economía y, en última instancia, la sociedad en su conjunto. Desarrollar métodos eficaces para detectar y predecir quiebras financieras no solo es deseable, sino imprescindible en un entorno cada vez más volátil e interconectado.

En este sentido, la inteligencia artificial, y en particular machine learning, se presenta como una herramienta de enorme potencial. Tal como se ha ilustrado en esta introducción, su capacidad para procesar grandes volúmenes de información, identificar patrones ocultos y aprender de forma autónoma, supera con creces las capacidades humanas en tareas complejas, permitiendo abordar problemas como la predicción de insolvencias con mayor precisión, rapidez y eficiencia.

Por lo tanto, este proyecto tiene como objetivo el desarrollo de un producto capaz de predecir quiebras financieras mediante la utilización de una herramienta de inteligencia artificial como es el aprendizaje automático. No solo se aborda la creación de una solución que contribuye a mitigar uno de los principales riesgos a los que se enfrentan las empresas, sino que también se demuestra su viabilidad técnica a través del diseño y validación del prototipo.

Además, el proyecto adopta un enfoque integral, incorporando el análisis de su potencial de comercialización, lo que permite abordar no solo la dimensión tecnológica, sino también su proyección empresarial, alineando así el trabajo con una visión de emprendimiento con marcado carácter ingenieril.

2. Objetivo del proyecto

El objetivo de este proyecto es desarrollar un programa capaz de predecir quiebras financieras utilizando técnicas de machine learning, y demostrar su viabilidad tanto técnica como comercial. Se trata de una iniciativa de emprendimiento dividida en dos fases fundamentales:

- Fase técnico-científica:

En esta primera etapa se valida la viabilidad del producto desde una perspectiva técnica. Para ello, se ha programado un prototipo funcional que emplea algoritmos de machine learning para detectar situaciones de insolvencia empresarial. El desarrollo se ha basado en una base de datos obtenida de la plataforma Kaggle (adjunta en el anexo), sobre la cual se han aplicado distintas técnicas de aprendizaje automático para construir un modelo predictivo eficaz.

- Fase de planificación empresarial:

Una vez demostrada la viabilidad técnica, el proyecto se orienta a la elaboración de un plan de negocio que permita comercializar el producto en el mercado real. Esta fase incluye el análisis de mercado, definición del modelo de negocio, estrategia de marketing, estructura de precios y evaluación financiera.

En definitiva, este proyecto de emprendimiento presenta la posibilidad real de desarrollar un software predictivo de quiebras financieras, respaldado por tecnologías avanzadas, y establece las bases necesarias para su introducción y escalado en el mercado a través de una estrategia empresarial sólida.

3. Estructura del proyecto

La estructura de este proyecto sigue el enfoque habitual utilizado en iniciativas de emprendimiento tecnológico, dividiéndose en tres bloques fundamentales:

1. Análisis del estado del arte y marco teórico

En esta primera etapa se establece el fundamento conceptual del proyecto, abordando los dos pilares clave sobre los que se construye la propuesta: las quiebras financieras y machine learning.

- Se analizará el concepto de quiebra financiera y su impacto en el contexto económico actual, especialmente en el tejido empresarial.
- Se presentará machine learning como una rama de la inteligencia artificial capaz de aprender de forma autónoma a partir de datos, destacando su aplicabilidad y valor en el sector financiero.

- Finalmente, se establecerá la relación entre ambos conceptos, subrayando las oportunidades y ventajas de aplicar tecnologías de aprendizaje automático para anticipar situaciones de insolvencia.

Este análisis permitirá comprender la relevancia del problema abordado, así como la pertinencia de la solución tecnológica propuesta.

2. Desarrollo técnico del producto

La segunda parte del proyecto está centrada en la fase científico-técnica, en la que se desarrolla una herramienta funcional para la predicción de quiebras financieras mediante machine learning.

- Se utilizará una base de datos extraída de Kaggle (ver anexo), sobre la cual se implementarán distintas técnicas de aprendizaje automático.
- Se presentará el prototipo inicial del software, explicando su funcionamiento, las metodologías aplicadas y las funcionalidades desarrolladas hasta el momento.
- El objetivo de esta sección es demostrar la viabilidad técnica del producto, evidenciar su utilidad práctica y anticipar su potencial de evolución.

Este apartado constituye la base sobre la que se construye todo el proyecto de emprendimiento, ya que valida que la solución propuesta es técnicamente posible y funcional.

3. Desarrollo del plan de negocio

Una vez validada la viabilidad técnica, el proyecto se orienta hacia su comercialización mediante un plan de negocio estructurado. Esta sección contempla:

- Una descripción del producto en términos de propuesta de valor.
- Un análisis del mercado y de los clientes potenciales, que permita demostrar el conocimiento del entorno y de los actores relevantes.
- La estrategia de posicionamiento y comercialización, incluyendo canales, precios, modelo de negocio y acciones de marketing.
- Por último, se presentará un análisis financiero preliminar, que evalúe la viabilidad económica del proyecto y su potencial de crecimiento.

Esta estructura permite abordar el proyecto desde una doble perspectiva: la validación técnica de la solución propuesta y su posible escalabilidad comercial en el mercado real.

De este modo, el proyecto ofrece una visión integral del desarrollo y la comercialización de una herramienta innovadora, demostrando en la práctica lo que implica llevar a cabo un proyecto de emprendimiento con un marcado carácter técnico e ingenieril.

4. Marco teórico

En este apartado se presentará el estado del arte de los dos pilares fundamentales del proyecto, con el objetivo de establecer la relación que existe entre ambos y poner en valor los beneficios que su integración aporta a la sociedad. Esta revisión permitirá contextualizar el desarrollo del proyecto dentro del marco teórico y tecnológico actual, destacando su relevancia y aplicabilidad.

4.1 Quiebras financieras

En primer lugar, se presentará la definición del concepto de quiebra, con el fin de establecer un marco conceptual claro y contextualizar su uso a lo largo del proyecto. Según la Real Academia Española (RAE), la definición es la siguiente:

1. “f. Der. Juicio por el que se incapacita patrimonialmente a alguien por su situación de insolvencia y se procede a ejecutar todos sus bienes en favor de la totalidad de sus acreedores.”[4].

A continuación, se presentan los procedimientos que la Unión Europea establece como respuesta ante situaciones de quiebra o insolvencia. En este caso, se expondrán específicamente aquellos métodos aplicables al ámbito empresarial, por ser los de mayor interés y relevancia para el desarrollo del presente proyecto. La Unión Europea expone lo siguiente:

“Si una empresa o un empresario tuvieran dificultades financieras o no pudieren pagar sus deudas, existen procedimientos específicos en cada país para hacer frente a la situación de forma inclusiva, con la participación de todos los acreedores (partes a las que se debe dinero).”[5].

Los procedimientos de insolvencia difieren en función de sus objetivos:

Empresas

- Si la empresa puede salvarse o el negocio es viable, sus deudas pueden ser restructuradas (generalmente con el acuerdo de los acreedores). Con ello se pretende proteger a la empresa y mantener los puestos de trabajo.
- Si el negocio no puede salvarse, la empresa debe ser liquidada («quiebra»).[6].

Como puede comprobarse, existen dos procedimientos principales cuando una empresa se enfrenta a una situación de insolvencia. En primer lugar, se encuentra una fase preventiva o de recuperación, en la que se considera que la empresa aún tiene posibilidades de superar sus dificultades financieras. En este escenario, se activan mecanismos como la reestructuración de deuda o el concurso de acreedores, con el objetivo de restablecer la viabilidad económica de la organización.

Por otro lado, si se determina que la situación ha alcanzado un punto irreversible y la empresa ya no es capaz de cumplir con sus obligaciones, se considera que ha entrado en estado de quiebra. En este caso, se procede a la liquidación de sus activos, con el fin de satisfacer, en la medida de lo posible, las obligaciones que tenga.

En el caso de la regulación en España se presentan los siguientes tres tipos de procedimientos en los casos de insolvencia según “el Texto Refundido de la Ley Concursal (TRLR), aprobado por el Real Decreto Legislativo 1/2020, de 5 de mayo.”. Estableciendo lo siguiente: “La nueva regulación de la insolvencia prevé tres tipos de procedimientos de insolvencia:

1. El concurso de acreedores, regulado en el libro I del TRLR
2. Los procedimientos de reestructuración de deudas, con su posible comunicación previa del inicio de las negociaciones, previsto en el libro II.
3. Y el procedimiento especial de microempresas, regulado en el libro III.”[7]

Para comprender la magnitud de este fenómeno, que ocurre de forma habitual en el ámbito empresarial, se presentan a continuación los datos correspondientes al primer trimestre de 2025 en España. Estas cifras, publicadas por el Instituto Nacional de Estadística (INE), recogen el número de deudores concursados por naturaleza jurídica, desglosados por comunidades autónomas. La definición que ofrece el INE de esta termino es la siguiente:

Deudor concursado: “Persona física o jurídica declarada legalmente en concurso de acreedores”.[8]. La visualización de estos datos permite dimensionar el impacto real de la insolvencia empresarial en el tejido económico del país, facilitando una mejor comprensión del alcance de las quiebras financieras en el contexto nacional.

	EMPRESAS		
	S. Limitadas	S. Anónimas	Otras
TOTAL	1.545	78	43
Andalucía	168	5	5
Aragón	21	4	2
Asturias	30	1	0
Baleares, Illes	36	3	0
Canarias	24	3	1
Cantabria	8	2	0
Castilla y León	23	3	2
Castilla-La Mancha	39	3	2
Cataluña	415	23	12
Comunitat Valenciana	199	6	1
Extremadura	24	0	1
Galicia	55	1	0
Madrid	354	13	8
Murcia	57	1	5
Navarra	17	0	0
País Vasco	69	9	4
La Rioja	4	1	0
Ceuta y Melilla	2	0	0

Figura 1: Empresas concursadas en España, primer trimestre 2025

Fuente: INE

Como se puede observar en los datos presentados, el número de empresas deudoras concursadas en España durante el primer trimestre de 2025 asciende a 1.666 empresas. De estas, más del 92 % corresponden a sociedades limitadas, lo que podría reflejar que esta estructura jurídica es la más vulnerable a situaciones de insolvencia, un dato relevante a la hora de analizar la relación entre la insolvencia y las empresas.

En cuanto a la distribución geográfica, las comunidades autónomas con mayor número de empresas concursadas son Madrid y Cataluña, lo cual resulta coherente dado que ambas regiones concentran una proporción significativa del tejido empresarial nacional.

A continuación, con el objetivo de ampliar el marco de análisis y comprender estos datos de manera más contextualizada, se presenta la variación trimestral, anual y anual acumulada correspondiente al primer trimestre de 2025. Este enfoque comparativo permite evaluar no solo el volumen absoluto de empresas concursadas, sino también su evolución a lo largo del tiempo.

Los datos muestran una tendencia al alza en el número de empresas en situación concursal, no solo respecto al trimestre anterior, sino también en comparación con el mismo periodo del año anterior. En particular, se observa un incremento medio anual del 11,8 % en el número de empresas concursadas, considerando las tres principales formas jurídicas analizadas. Este crecimiento confirma que el fenómeno de la insolvencia empresarial continúa en expansión, lo que pone de manifiesto la necesidad de desarrollar herramientas avanzadas capaces de analizar y predecir este tipo de fenómenos, con el objetivo de anticiparlos y prevenir sus consecuencias más negativas, es decir, prevenir que las empresas lleguen a esta situación y estos números no continúen en alza.

	TOTALES	% Variación		
		Trimestral	Anual	Anual acumulada
EMPRESAS - S. Limitadas	1.545	19,4%	9,0%	9,0%
EMPRESAS - S. Anónimas	78	18,2%	6,8%	6,8%
EMPRESAS - Otras	43	16,2%	19,4%	19,4%

Figura 2: Variación empresas concursadas en España

Fuente: INE

Otros datos relevantes proporcionados por el Instituto Nacional de Estadística (INE) permiten profundizar en la comprensión del fenómeno de la insolvencia empresarial en España. Según esta información, las tres ramas de actividad con mayor número de empresas concursadas, en orden descendente, son:

1. Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas.
2. Construcción.

3. Industria manufacturera.

Además, se observa que el mayor volumen de empresas concursadas corresponde a aquellas con un volumen de negocio inferior a 0,25 millones de euros, lo que pone de manifiesto la vulnerabilidad de las empresas de menor tamaño.

Conociendo el significado del término y comprendiendo la situación actual en España, resulta igualmente importante analizar las consecuencias e implicaciones que conlleva la quiebra financiera. Como presenta la Asociación Española de Afectados por la Deuda [9], tres consecuencias que se pueden dar, especialmente a nivel empresarial, cuando se llega a una situación de insolvencia en la que la empresa está en riesgo de quiebra financiera son los siguientes:

1. Pérdida de activos y patrimonio debido a la insolvencia
2. Efectos en la reputación y confianza de los stakeholders
3. Impacto en la capacidad de acceso a financiamiento futuro

Sin embargo, las consecuencias de una quiebra financiera no se limitan únicamente al ámbito económico de la empresa. El impacto social también puede ser significativo y, en muchos casos, devastador. Las quiebras suelen conllevar la pérdida de puestos de trabajo, la interrupción de la actividad productiva, y la reducción del valor social generado por las empresas, afectando directamente a empleados, proveedores, clientes y comunidades locales.

Por lo tanto, aunque este proyecto aborda la quiebra financiera desde una perspectiva económica y empresarial, es fundamental tener presente que sus efectos se extienden también al plano social, reputacional e institucional,

Una vez comprendido qué es la quiebra financiera y cuáles son los procedimientos existentes ante situaciones de insolvencia, es posible dimensionar la magnitud de este fenómeno y sus profundas implicaciones para la sociedad, especialmente sobre el tejido empresarial. A la luz de los datos presentados y del análisis realizado, se hace evidente la necesidad de desarrollar herramientas de última generación que permitan, en la medida de lo posible, anticipar este tipo de situaciones.

Contar con sistemas capaces de predecir episodios de insolvencia proporcionaría a los actores del sector información más completa y precisa, facilitando la toma de decisiones estratégicas y contribuyendo a reducir riesgos operativos y financieros. De esta necesidad surge la idea de desarrollar un programa que, mediante la incorporación de tecnologías avanzadas como la inteligencia artificial, y en particular su rama de machine learning, permita llevar a cabo este análisis de forma más precisa, eficiente y profesional

4.2 Machine Learning

Como se ha descrito previamente, machine learning es la herramienta seleccionada para el desarrollo de este proyecto, ya que permite potenciar al máximo el rendimiento y la eficacia de la solución propuesta. Por ello, resulta fundamental establecer un marco teórico sólido que permita comprender en qué consiste esta tecnología, cuáles son sus principales métodos y algoritmos, así como algunos de los usos más relevantes que se le están dando en la actualidad en distintos sectores, incluyendo el financiero.

La definición que proporciona IBM sobre machine learning, traducida al español, es la siguiente:

“El aprendizaje automático (Machine Learning, ML) es una rama de la inteligencia artificial (IA) centrada en permitir que los ordenadores y las máquinas imiten la forma en que aprenden los seres humanos, realicen tareas de manera autónoma y mejoren su rendimiento y precisión a través de la experiencia y la exposición a más datos.”[10].

En esta definición se pueden identificar claramente los pilares fundamentales del aprendizaje automático:

- En primer lugar, se trata de una tecnología que forma parte del campo de la inteligencia artificial (IA). Tal como señala SAP al explicar la relación entre ambos conceptos:
“El machine learning y sus componentes de deep learning y redes neuronales, todos encajan como subconjuntos concéntricos de la IA. La IA procesa datos para llegar a decisiones y proyecciones. Los algoritmos de machine learning permiten que la IA no solo procese esos datos, sino que los use para aprender y ser más inteligente, sin necesidad de programación adicional. La inteligencia artificial es la madre de todos los subconjuntos de machine learning subyacentes. Dentro del primer subconjunto está el machine learning; dentro de él está el deep learning, dentro del cual están las redes neuronales.”[11].
Por lo tanto, gracias a la explicación proporcionada por SAP, resulta más sencillo comprender la relación entre machine learning y la inteligencia artificial (IA). Se observa cómo el aprendizaje automático forma parte del conjunto de herramientas que integran la IA, y cómo permite que el análisis de datos no se limite únicamente a una evaluación y estudio, sino que se transforme en un proceso dinámico de aprendizaje y evolución continua, sin necesidad de trabajo adicional.
- Se trata de una herramienta diseñada para imitar el comportamiento humano, mejorando incluso sus capacidades de aprendizaje gracias al uso de inteligencia artificial, lo que permite una comprensión y procesamiento a una escala mucho superior a las capacidades humanas.
- Es una tecnología autónoma que, una vez desarrollado y activado el modelo, continúa aprendiendo y evolucionando por sí misma, mejorando su rendimiento a través del aprendizaje continuo sin necesidad de intervención humana constante.

En la siguiente imagen se representa de forma visual y simplificada el proceso de trabajo de un modelo de machine learning. A continuación, se describen las principales etapas de este flujo:

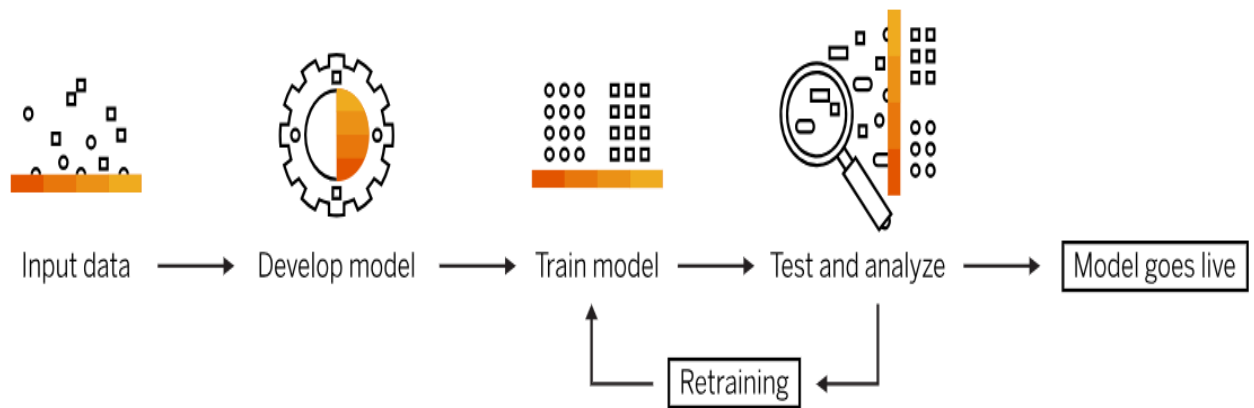


Figura 3: Proceso machine learning

Fuente: SAP

1. Obtención e introducción de datos:

El primer paso consiste en recolectar los datos que servirán como base para entrenar el modelo. La calidad, cantidad y relevancia de estos datos son factores determinantes para el rendimiento final del sistema.

2. Desarrollo del modelo:

Una vez recopilados los datos, se procede al diseño del modelo. En esta etapa es fundamental seleccionar el algoritmo de machine learning más adecuado, así como el método de aprendizaje óptimo, teniendo en cuenta el tipo de problema, la naturaleza de los datos y los objetivos del análisis. Existen numerosos algoritmos y métodos de aprendizaje, cada uno con sus ventajas y limitaciones, que serán descritos en detalle más adelante.

3. Entrenamiento del modelo:

Con el modelo definido, se inicia la fase de entrenamiento, en la que el sistema ajusta sus parámetros internos para aprender patrones a partir de los datos. Durante esta etapa, se evalúa el comportamiento del modelo y se ajustan variables clave para optimizar los resultados obtenidos.

4. Evaluación y mejora continua (retraining):

Tras el entrenamiento inicial, el modelo se pone a prueba con datos nuevos o separados del conjunto original. Se analizan los resultados y, en caso necesario, se realiza un ajuste de parámetros. Este proceso genera un ciclo de retroalimentación (retraining), donde el modelo continúa aprendiendo y mejorando su rendimiento a medida que recibe nuevos datos y se optimizan sus configuraciones.

5. Implementación y actualización:

Una vez validado, el modelo se implementa en un entorno real, donde es capaz de recibir nuevos inputs de forma continua. Gracias a su capacidad de aprendizaje automático, el modelo puede seguir ajustándose y evolucionando, manteniéndose actualizado y adaptado a posibles cambios en los datos o el contexto.

Como se puede observar, se trata de un proceso en el que los datos se utilizan para alimentar un modelo, cuya estructura y comportamiento pueden variar según la metodología empleada. Este modelo entra en un ciclo iterativo que permite ajustar sus parámetros y características hasta alcanzar un funcionamiento óptimo. Una vez entrenado, el modelo es capaz de continuar su aprendizaje de forma autónoma, evolucionando y mejorando su rendimiento de manera continua.

De acuerdo con Oracle [12], las principales características del aprendizaje automático son las siguientes:

- Detección automática de patrones
- Predicción de resultados probables
- Creación de información procesable
- Capacidad para analizar volúmenes de datos muy grandes

Al observar las características clave del aprendizaje automático señaladas por Oracle, es posible comprender con mayor claridad su gran utilidad y su aplicación práctica dentro del proceso previamente descrito. La capacidad de procesar grandes volúmenes de datos, propia de la inteligencia artificial, lo vincula directamente con el ámbito del Big Data, permitiendo transformar datos masivos en información procesable y valiosa.

Entre sus aportaciones más destacadas se encuentra la detección automática de patrones, fundamental en cualquier proceso de análisis, especialmente cuando se gestionan grandes cantidades de información. Asimismo, la generación de conocimiento a partir de datos permite descubrir nuevas relaciones y enfoques que enriquecen el proceso de toma de decisiones. Por último, la predicción de resultados probables se presenta como una de las funcionalidades más potentes de esta tecnología.

Estas características demuestran por qué el aprendizaje automático es la herramienta idónea para el presente proyecto. Dado que el objetivo es predecir situaciones de quiebra financiera, utilizar una tecnología cuya fortaleza radica precisamente en la predicción de resultados a partir de patrones históricos y una gran cantidad de datos, pudiendo ofrecer información clara, refuerza la solidez y lógica de su elección para llevar a cabo este proyecto.

Dentro de los métodos de aprendizaje automático se presentan cuatro tipos principales [11][12]:

- Aprendizaje automático supervisado
- Aprendizaje automático no supervisado
- Aprendizaje automático semisupervisado

- Aprendizaje automático por refuerzo

Aunque pudiera pensarse que la principal diferencia entre los tipos de machine learning radica en el grado de intervención humana, es decir, que la supervisión se refiere a la participación directa del ser humano en el proceso de aprendizaje, en realidad, esta no es la distinción clave. Como se ha indicado previamente, el aprendizaje automático no requiere de intervención humana constante. La verdadera diferencia entre los distintos enfoques radica en el tipo de datos empleados en el entrenamiento del modelo.

- En el caso del aprendizaje automático supervisado, los datos han sido previamente etiquetados antes de ser introducidos al modelo. Aplicado al contexto de este proyecto, esto implicaría proporcionar al sistema los datos financieros de diversas empresas, junto con la etiqueta que indica si la empresa ha quebrado o no. El modelo aprende a identificar patrones en función de esas etiquetas, lo que le permite hacer predicciones precisas sobre nuevos casos.
- En cambio, el aprendizaje no supervisado utiliza datos no etiquetados. Es decir, se alimenta al modelo con información, por ejemplo, los datos financieros de diferentes empresas, pero sin indicar si cada empresa está en quiebra o no. El modelo, por sí mismo, debe descubrir estructuras ocultas, relaciones y agrupaciones dentro de los datos, clasificando de forma autónoma a las empresas según las similitudes detectadas.
- Por su parte, el aprendizaje semisupervisado constituye un enfoque intermedio. En este caso, solo una parte de los datos está etiquetada, mientras que el resto permanece sin etiquetar. El modelo utiliza la información de los datos etiquetados para guiar su aprendizaje y mejorar la clasificación de los datos no etiquetados, logrando un equilibrio entre eficiencia y precisión.
- Finalmente, el aprendizaje por refuerzo se considera una modalidad particular del aprendizaje supervisado. En lugar de proporcionarle al modelo las respuestas correctas, se le indican las reglas del entorno o “reglas del juego”. A partir de ahí, el sistema aprende a tomar decisiones a través de prueba y error, recibiendo recompensas o penalizaciones en función de sus resultados. Aplicado al caso de estudio, esto implicaría explicarle al modelo qué características definen una situación de quiebra o de solvencia, y luego permitirle interactuar con múltiples conjuntos de datos para ir ajustando su comportamiento hasta alcanzar un rendimiento óptimo.

Para todos estos métodos de aprendizaje automático, también es fundamental considerar el algoritmo que se va a emplear en cada caso. Existen numerosos tipos de algoritmos, cada uno con sus propias características, ventajas y limitaciones, que pueden adaptarse en mayor o menor medida según la naturaleza del problema, el tipo de datos disponibles y los objetivos específicos que se persiguen. Para entender mejor lo que es un algoritmo de machine learning se ofrece la siguiente definición de IBM:

“Un algoritmo de Machine Learning es un conjunto de reglas o procesos utilizados por un sistema de IA para realizar tareas, la mayoría de las veces para descubrir nuevos

conocimientos y patrones de datos, o para predecir valores de salida a partir de un conjunto determinado de variables de entrada. Gracias a los algoritmos, el Machine Learning (ML) puede aprender.”[13].

Algunos de los algoritmos más utilizados en aprendizaje automático, incluidos algunos de los que se han aplicado en el presente proyecto, son los siguientes:

- Random forest
- SVM
- Impulso de gradiente (Gradient Boosting)
- Redes neuronales
- Bagging

Entre los usos actuales más relevantes del machine learning se encuentran múltiples aplicaciones en diversos campos. Algunas de estas tecnologías ya forman parte del día a día de la mayoría de las personas, mientras que otras se aplican en contextos más específicos y especializados, como los siguientes:

Medicina: Iberdrola expone un caso en el que investigadores del MIT ya utilizan actualmente machine learning para la detección temprana del cáncer de mama.[14]. IBM también se hace eco de esta realidad, destacando lo siguiente: “Los médicos que evalúan las mamografías para detectar el cáncer de mama pasan por alto el 40 % (enlace externo a [ibm.com](https://www.ibm.com)) de los cánceres, y el ML puede mejorar esa cifra.”[15]

Se trata de un uso altamente especializado en el ámbito médico, donde la aplicación de esta herramienta de última generación puede salvar vidas, ya que, como se menciona, las capacidades humanas no siempre logran detectar una gran parte de los casos de esta enfermedad.

Marketing, comunicación y redes: Un caso muy común, citado por numerosas compañías al explicar las aplicaciones del machine learning, es el uso de esta tecnología para la detección automática de spam en los correos electrónicos. Mediante algoritmos de ML, los sistemas son capaces de identificar y clasificar mensajes no deseados, mejorando así la seguridad y eficiencia del correo electrónico. [15][16].

Tecnología: Santander destaca el uso del machine learning en áreas clave como la ciberseguridad y la mejora en la eficacia de los resultados de búsqueda. Estas aplicaciones están presentes de forma cotidiana en nuestra vida diaria, por ejemplo, al realizar búsquedas en internet o al desbloquear teléfonos móviles de última generación mediante reconocimiento facial, tecnologías que se apoyan en algoritmos de ML para ofrecer mayor precisión, seguridad y rapidez.[17].

Transporte: Como expone Telefónica, machine learning es capaz de analizar datos históricos y otros factores relevantes para determinar la ruta óptima en tiempo real. Esta herramienta, cada vez más integrada en los sistemas de navegación de los automóviles, contribuye a reducir los tiempos de tránsito y a optimizar el consumo de combustible, ofreciendo así una solución eficiente tanto en términos de movilidad como de sostenibilidad.[18].

Finanzas: Machine learning es ampliamente utilizado en el sector financiero, ya que, gracias a las características mencionadas anteriormente, se presenta como un aliado ideal para afrontar los retos de esta industria. AWS destaca su aplicación en proyectos orientados a la mejora de la regulación y el análisis de riesgos. Un ejemplo concreto es el caso de NerdWallet, una empresa estadounidense dedicada a las finanzas personales, que utiliza machine learning para ayudar a los usuarios a comparar productos financieros.[19].

El uso del machine Learning en el sector financiero es muy amplio, como se destaca en el trabajo de Teresa Francés, “IMPACTO DEL MACHINE LEARNING EN EL SISTEMA FINANCIERO”.[20]. Donde se presentan algunos casos representativos como:

- Detección de fraude
- Predecir riesgo de crédito
- Mercados bursátiles (algo-trading)
- Chatbots
- Roboadvisors
- Análisis de documentos

5. Modelo de predicción de quiebras financieras mediante Machine Learning

A continuación, en este apartado se presentará el modelo desarrollado. En primer lugar, se expondrá la metodología utilizada para su elaboración. Posteriormente, se describirá de forma detallada y paso a paso el funcionamiento del modelo construido.

5.1 Metodología empleada en el modelo

El modelo que se presenta a continuación se fundamenta en los conceptos expuestos previamente. Se trata de un prototipo inicial cuyo objetivo principal es demostrar la viabilidad técnica de desarrollar una herramienta capaz de predecir quiebras financieras mediante el uso de machine learning. Tras haber definido qué son las quiebras financieras, su impacto en la sociedad, y las ventajas que aporta el uso de tecnologías avanzadas como machine learning, se introduce este modelo como ejemplo práctico.

Es importante destacar que no se pretende presentar una versión final del producto, con todas las funcionalidades que podría incorporar en su lanzamiento comercial. El propósito es ofrecer una visión preliminar, que permita entender cómo se podría estructurar y desarrollar este tipo de solución. En este sentido, se exploran aspectos como las posibles funcionalidades, metodologías a emplear, algoritmos más adecuados, así como los resultados que cabría esperar en un producto de estas características.

Debe entenderse que este no es el producto definitivo, sino una fase temprana de desarrollo, útil para visualizar su funcionamiento, su aplicación práctica y el valor que podría aportar al mercado. Esta aproximación permite, además, contextualizar el proyecto como un caso de emprendimiento de base tecnológica e ingenieril, facilitando así la comprensión de los apartados posteriores, en los que se analiza su viabilidad desde el punto de vista empresarial y financiero.

Conocer el producto desde esta perspectiva técnica y funcional ayuda a comprender mejor aspectos clave como: el tipo de clientes al que podría estar dirigido, los canales de distribución más adecuados, las plataformas de implementación, los costes estimados de desarrollo completo, y los beneficios económicos esperables tras su comercialización.

Por lo tanto, este apartado del proyecto cumple con un doble objetivo:

- Demostrar la viabilidad técnica de desarrollar un producto de estas características, explorando las distintas posibilidades que ofrece machine learning y presentando resultados preliminares que resultan útiles como base para su posible implementación final.
- Proporcionar un marco de referencia para el desarrollo del plan de negocio, permitiendo una comprensión más completa y práctica del producto. Esto facilita tanto su análisis comercial como la definición de una estrategia adecuada para su introducción en el mercado.

Para alcanzar estos objetivos se implementa la siguiente metodología en el proyecto:

1. Obtención de la base de datos y análisis de los datos

Para el desarrollo del proyecto, se ha seleccionado una base de datos suficientemente amplia y representativa, que permita realizar un análisis exhaustivo sin que la cantidad de datos suponga una limitación. Con este objetivo, se ha recurrido a Kaggle, una plataforma de referencia en el ámbito del machine learning y el análisis de datos, ampliamente utilizada por la comunidad científica y técnica.

En concreto, se ha optado por el dataset titulado “US Company Bankruptcy Prediction Dataset”, el cual resulta especialmente adecuado para este proyecto, ya que contiene una variedad de variables financieras, así como información relativa al estado de las empresas (en bancarrota o no) y el año correspondiente. Este conjunto de datos ofrece una base sólida para estudiar el comportamiento financiero de las compañías en relación con su riesgo de insolvencia. Todos los detalles y características del dataset seleccionado se encuentran disponibles en el anexo del presente trabajo.

Una vez elegida la base de datos, se ha realizado un análisis exploratorio inicial, una práctica común en estudios estadísticos y de data science, que permite conocer a fondo las características de los datos y evaluar su idoneidad para el análisis posterior. Esta fase incluye:

- Revisión del tipo de variables incluidas.
- Cálculo de medidas estadísticas básicas.
- Identificación de posibles valores atípicos (outliers).
- Estudio preliminar de correlaciones entre variables.

Este análisis inicial sienta las bases del estudio posterior, asegurando la calidad y la relevancia de los datos empleados en el desarrollo del modelo de predicción.

2. Desarrollo del modelo , aplicación de diferentes algoritmos:

En este apartado se presenta el desarrollo del modelo, exponiendo distintas metodologías de análisis de los datos mediante la aplicación de diversos algoritmos de machine learning. El objetivo es identificar las estrategias más adecuadas para maximizar el rendimiento del modelo y establecer las bases para un desarrollo final del producto con las mejores garantías de eficacia y fiabilidad.

Se aborda el problema desde diferentes enfoques, utilizando varios algoritmos y ajustando sus parámetros con el fin de determinar cuáles ofrecen los resultados más sólidos y útiles. Para mejorar la calidad de las predicciones, se implementan técnicas complementarias como la optimización del dataset y la eliminación de valores atípicos (outliers), lo que permite trabajar con datos más realistas y consistentes.

Los algoritmos utilizados, así como los distintos estudios realizados sobre el comportamiento del modelo en cada caso, se describen detalladamente en un apartado posterior dedicado al funcionamiento del modelo.

3. Resultados:

Este aspecto se abordará en el siguiente apartado del proyecto, en el cual se presentarán los resultados obtenidos. Se evaluará el rendimiento (performance) de cada uno de los modelos desarrollados, permitiendo compararlos entre sí y determinar cuáles representan las opciones óptimas para el desarrollo futuro del producto. Todo este análisis se realizará teniendo en cuenta las limitaciones previamente expuestas al tratarse de una etapa inicial del prototipo del producto.

5.2 Datos empleados

Los datos empleados en este proyecto constan de 78.682 filas y 21 columnas. Cabe señalar que esto no implica la existencia de 78.682 compañías diferentes, sino que el conjunto de datos recoge registros históricos de las empresas a lo largo del período de estudio, pudiendo aparecer una misma empresa en distintos años y en diferentes estados financieros (solvente o insolvente).

Tras un primer análisis exploratorio, se verifica que no existen valores nulos en el conjunto de datos. Asimismo, se identifica que:

- Existen dos variables de tipo object, el nombre de la compañía y su status (solvente/insolvente).
- Hay una variable de tipo int, el año.
- El resto de las columnas, correspondientes a variables financieras, son de tipo float.

Los datos empleados son los siguientes:

	status_label	year	Current assets	COGS	Depreciation & Amortization	EBITDA	Inventory	Net income	Total receivables	Market value	Net sales
count	78682.000000	78682.000000	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04
mean	0.933657	2007.506317	2.460414e+05	4.346629e+05	2.786804e+04	7.978594e+04	5.787418e+04	1.598756e+04	7.756495e+04	2.056855e+07	6.311807e+05
std	0.248882	5.742768	7.567481e+05	2.119847e+06	1.033214e+05	3.075114e+05	2.370688e+05	2.144773e+05	2.760120e+05	1.239569e+08	2.499401e+06
min	0.000000	1999.000000	-2.498000e+03	-3.666450e+05	0.000000e+00	-4.512135e+06	0.000000e+00	-1.605370e+07	-6.000000e-03	1.000000e-04	-1.964999e+06
25%	1.000000	2002.000000	3.705000e+03	3.935000e+03	3.740000e+00	-8.110000e-01	0.000000e+00	-4.286750e+03	3.284250e+02	1.612678e+05	5.546250e+03
50%	1.000000	2007.000000	3.679500e+04	3.396300e+04	2.042000e+03	2.525000e+03	1.415000e+03	1.533500e+01	7.229000e+03	1.262922e+06	5.740400e+04
75%	1.000000	2012.000000	1.933258e+05	2.339265e+05	1.562525e+04	4.577150e+04	2.309500e+04	1.188175e+04	5.005000e+04	7.812800e+06	4.071972e+05
max	1.000000	2018.000000	3.132258e+07	1.632456e+08	4.440438e+06	1.256672e+07	1.191851e+07	7.521731e+06	1.154206e+07	7.900501e+09	1.679396e+08

	Total assets	Long term debt	EBIT	Gross profit	Current liabilities	Retained earnings	Total revenues	Total liabilities	Total operating expenses
count	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04	7.868200e+04
mean	6.901281e+05	1.734601e+05	5.056774e+04	1.978397e+05	1.411126e+05	5.720740e+04	6.311807e+05	3.988577e+05	5.468320e+05
std	2.202777e+06	7.635129e+05	2.392709e+05	6.386401e+05	5.191615e+05	8.702102e+05	2.499401e+06	1.418661e+06	2.287409e+06
min	1.000000e-03	-2.300000e-02	-5.043795e+06	-4.367571e+06	1.000000e-03	-2.148494e+07	-1.964999e+06	1.000000e-03	-3.171970e+05
25%	9.338500e+03	0.000000e+00	-1.637000e+03	1.182763e+03	2.506000e+03	-4.413100e+04	5.546250e+03	4.540575e+03	7.963342e+03
50%	7.354200e+04	8.700000e+02	6.620000e+02	1.944400e+04	1.588250e+04	-2.450000e+00	5.740400e+04	2.681650e+04	5.934750e+04
75%	4.414575e+05	2.653100e+04	2.658700e+04	1.432210e+05	8.735050e+04	3.726475e+04	4.071972e+05	2.027002e+05	3.460062e+05
max	6.723494e+07	2.168753e+07	9.783515e+06	2.786440e+07	2.786969e+07	3.095666e+07	1.679396e+08	4.042171e+07	1.657008e+08

Figura 4: Datos iniciales

Fuente: Kaggle

Se puede comprobar de todas las variables de estudio los siguientes parámetros estadísticos:

- Cantidad de datos
- Media
- Desviación estándar
- Mínimo
- 25% - Primer cuartil
- 50% - Segundo cuartil (Mediana)
- 75% - Tercer cuartil
- Máximo

Todos los valores analizados resultan relevantes para el estudio, ya que permiten dimensionar las variables y comprender mejor su comportamiento, lo cual es fundamental para el desarrollo del modelo posterior.

Es especialmente significativo el caso de la variable `status_label`, que indica si una empresa está en quiebra o no: un valor de 0 representa una empresa en quiebra, mientras que un valor de 1 indica solvencia. Llama la atención que la media de esta variable es 0,9336, lo que implica que aproximadamente el 93,36 % de las observaciones corresponden a empresas solventes.

Esta fuerte desproporción revela un desequilibrio de clases, donde la mayoría de los datos pertenecen a empresas solventes. Esta asimetría puede constituir un desafío importante para el modelo, ya que podría sesgar los resultados, reduciendo su capacidad de identificar correctamente los casos de quiebra, que son precisamente los de mayor interés para el análisis.

A continuación, se presenta la matriz de correlación de las variables utilizadas en el estudio. Esta herramienta permite visualizar de forma clara y concisa las posibles relaciones existentes entre las distintas variables. En particular, resulta especialmente útil para identificar la correlación entre la variable `status_label`, que indica la situación de solvencia o quiebra de las empresas, y las variables financieras consideradas en el análisis. Esta información servirá como base para orientar el desarrollo del modelo y comprender mejor los resultados.

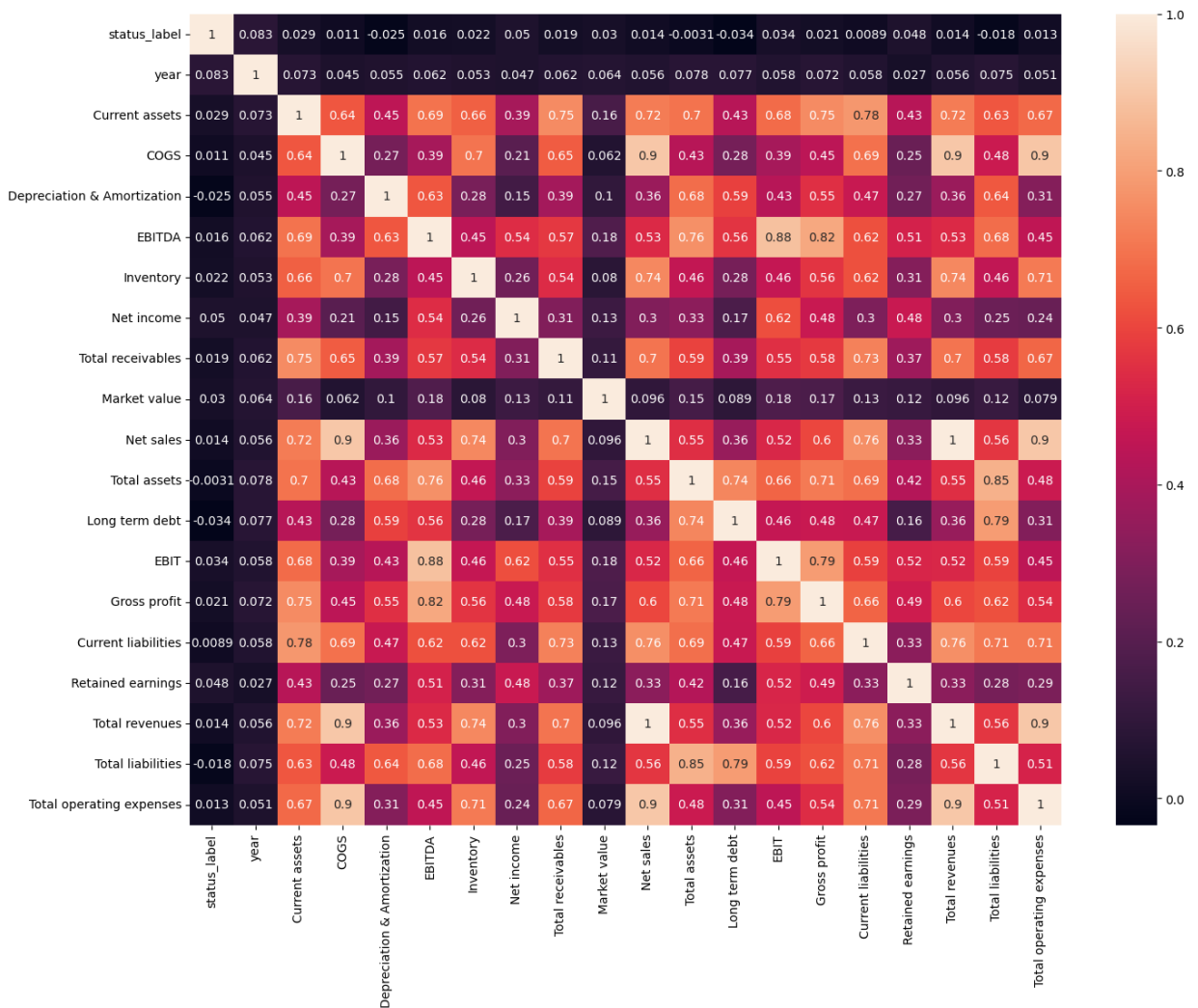


Figura 5: Matriz de correlación

Fuente: Elaboración propia

Las principales conclusiones que se pueden extraer de la matriz de correlación son las siguientes:

- Baja correlación entre el estado de solvencia y las variables financieras: Se observa una correlación muy débil entre la variable status (indicadora de quiebra o solvencia) y las distintas variables financieras analizadas, sin que ninguna supere un coeficiente de correlación de 0,1. Lo que sugiere que, en esta fase exploratoria, no se identifican relaciones sólidas entre las variables financieras estudiadas y la probabilidad de insolvencia.
- Alta correlación entre variables financieras relacionadas: Se destaca, por ejemplo, la correlación entre las variables total revenues y total sales (coeficiente de 1), lo que indica que ambas están directamente relacionadas similares dentro del comportamiento financiero de las empresas.

En la siguiente imagen se presentan los boxplots correspondientes a las variables del estudio. Esta representación gráfica permite visualizar de forma efectiva la distribución de los datos, facilitando la interpretación de su magnitud y, especialmente, la identificación de posibles valores atípicos (outliers) en cada una de las variables analizadas. Esta información es relevante para la preparación del conjunto de datos antes del desarrollo del modelo.

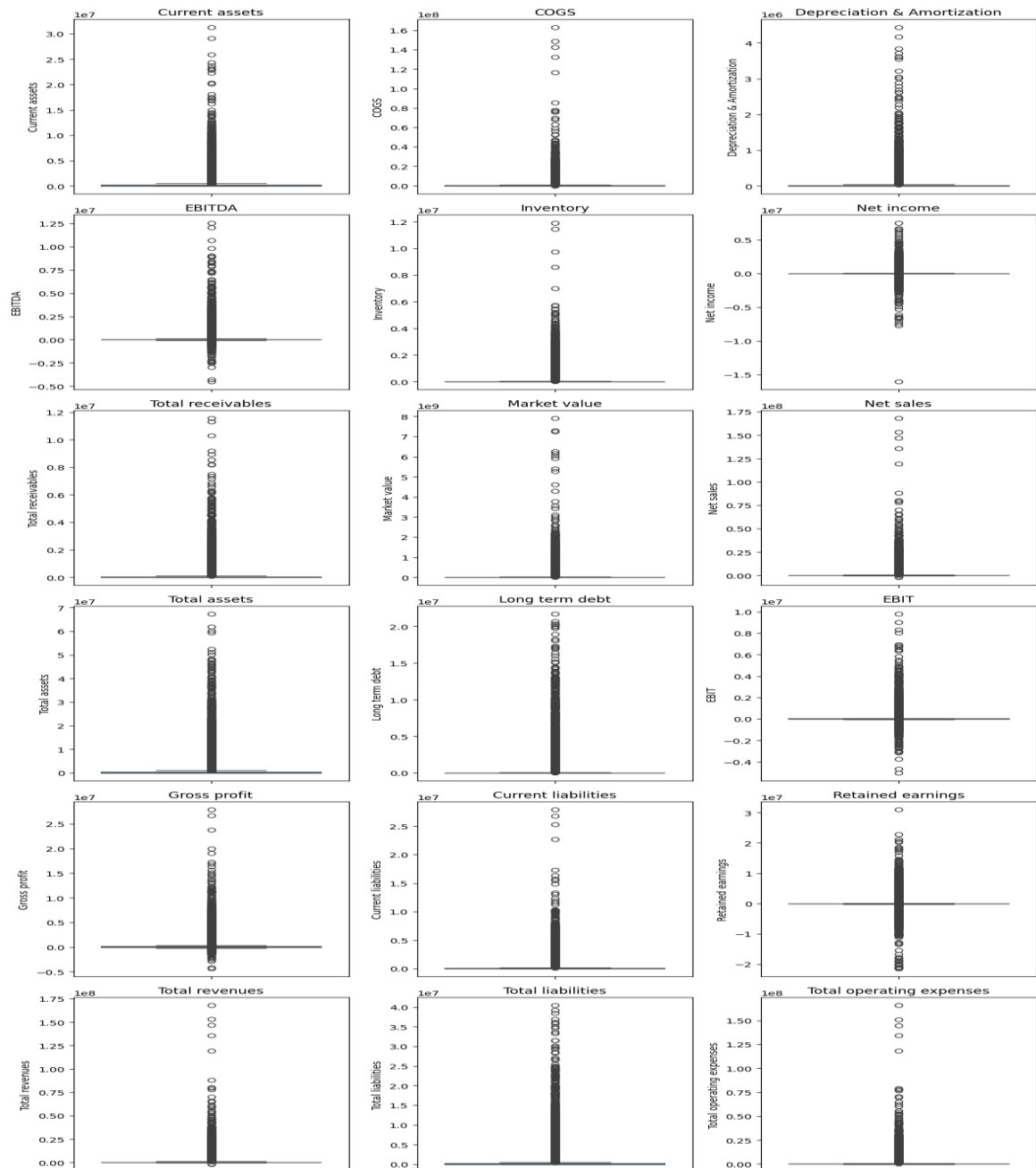


Figura 6: Boxplots

Fuente: Elaboración propia

A partir de los boxplots, se observa que todas las variables financieras presentan un número significativo de outliers, siendo mayoritariamente outliers positivos, aunque también se identifican algunos casos negativos. Ante esta situación, resulta necesario analizar el impacto de estos valores atípicos en el estudio y considerar la posibilidad de eliminar los outliers más extremos. Este paso permitirá evaluar cómo afecta esta depuración a la calidad del análisis y determinar la forma más adecuada de proceder en el desarrollo del modelo.

Se ha realizado un análisis eliminando los outliers mediante el método del IQR (Rango Inter cuartílico), una técnica ampliamente reconocida y empleada en estudios estadísticos. Tal como lo describe Oracle:

“El rango intercuartílico (IQR) es una medida de variabilidad, basada en la división de un conjunto de datos en cuartiles. Los cuartiles dividen un conjunto de datos clasificados en cuatro partes iguales. T1, T2 y T3. IQR se define como $Q3 - Q1$, y los datos que quedan fuera de $Q3 + 1,5IQR$ o de $Q1 - 1,5IQR$ se consideran valores atípicos.”[21].

Aplicando esta metodología, se obtienen los siguientes resultados.

	status_label	year	Current assets	COGS	Depreciation & Amortization	EBITDA	Inventory	Net income	Total receivables	Market value	Net sales
count	9307.000000	9307.000000	9307.000000	9307.000000	9307.000000	9307.000000	9307.000000	9307.000000	9307.000000	9.307000e+03	9307.000000
mean	0.954980	2007.287096	2781.121769	2786.954738	92.147843	-360.582164	513.524717	-766.428803	561.206451	3.945971e+05	4378.456582
std	0.207359	5.680331	4255.524889	3521.275968	339.519926	1528.684781	1181.971119	1778.428004	1045.569638	8.692351e+05	5242.035586
min	0.000000	1999.000000	-1124.000000	-0.018000	0.000000	-12579.000000	0.000000	-8743.000000	0.000000	1.000000e-04	-1093.000000
25%	1.000000	2002.000000	0.679000	0.669000	0.044000	-1058.000000	0.000000	-1483.000000	0.090000	1.306950e+04	0.859000
50%	1.000000	2007.000000	1236.000000	1506.000000	0.212000	-0.145000	0.222000	-0.491000	0.717000	6.186600e+04	2233.000000
75%	1.000000	2012.000000	3839.050000	4262.000000	0.966000	114.500000	285.750000	0.661000	727.250000	2.339570e+05	6907.000000
max	1.000000	2018.000000	41163.000000	34616.000000	7369.000000	7988.000000	9279.000000	5907.000000	9151.000000	4.667436e+06	24257.000000

Figura 7: Datos sin outliers.

Fuente: Elaboración propia

En la figura anterior, al compararla con la correspondiente al conjunto de datos original, se pueden apreciar las siguientes observaciones:

- Se ha producido una reducción significativa en el número de observaciones, pasando de 78.682 a 9.307 registros.
- La media de las variables ha disminuido, al igual que la desviación estándar, lo que indica una menor dispersión en los datos.
- Los valores máximos y mínimos han perdido su carácter extremo, lo que evidencia una menor presencia de datos atípicos.

A continuación, se presentan los boxplots correspondientes a este nuevo conjunto de datos, lo cual permite una comparación visual con la situación inicial. En esta representación gráfica se puede observar una clara reducción del número de outliers y

una mayor homogeneidad entre los valores, lo que indica la reducción en las diferencias existentes en los datos.

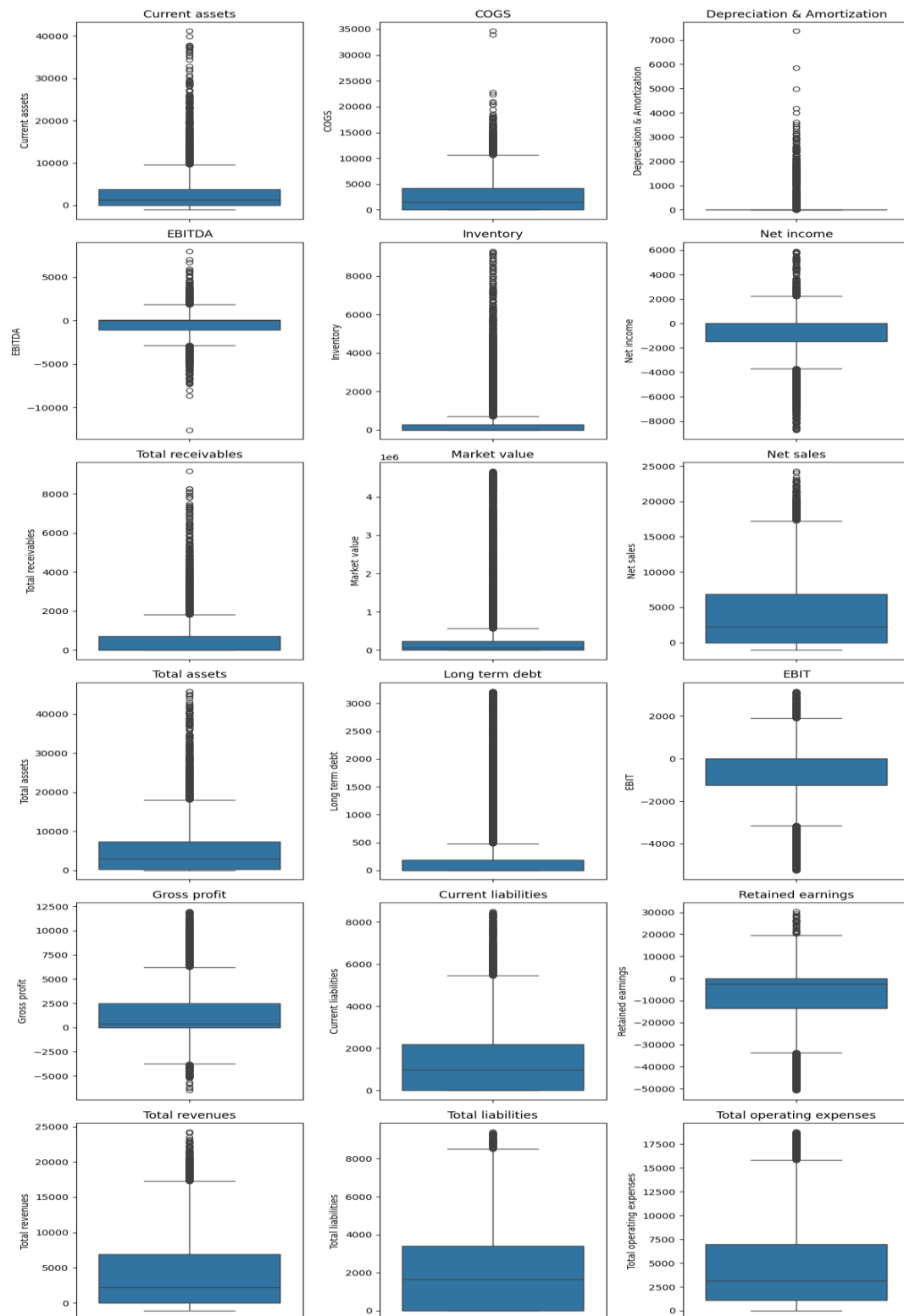


Figura 8: Boxplots sin outliers

Fuente: Elaboración propia

Sin embargo, aunque se observa que la eliminación de outliers reduce significativamente la dispersión de los datos y mejora su homogeneidad, este procedimiento podría distorsionar la naturaleza del estudio. En el contexto de la predicción de quiebras, los valores atípicos pueden representar justamente aquellas situaciones extremas que anteceden o caracterizan un proceso de insolvencia, por lo que su exclusión podría suponer la pérdida de información crítica.

Este efecto se refleja, por ejemplo, en la variable status label, que indica si una empresa es solvente (valor 1) o se encuentra en quiebra (valor 0). Tras la eliminación de los outliers, se observa un aumento en la media de esta variable, lo que evidencia que se han eliminado casos relevantes de empresas en quiebra. Este fenómeno confirma que los datos atípicos pueden estar estrechamente vinculados a las situaciones de insolvencia que se busca predecir.

Por tanto, para el desarrollo del modelo definitivo, se ha decidido no eliminar los outliers. Su exclusión alteraría de forma significativa la estructura del conjunto de datos y comprometería la calidad y precisión del modelo, reduciendo su capacidad predictiva en los casos más críticos, que son precisamente los de mayor interés en este estudio.

Por último, se analiza la posible relación entre la variable “año” y el status de las empresas, ya que durante el estudio se observan patrones que podrían indicar que esta relación no debe considerarse como un factor determinante. El análisis sugiere que el año en sí mismo no aporta información relevante sobre la probabilidad de quiebra, por lo que su inclusión como variable predictiva podría introducir ruido en el modelo sin mejorar el rendimiento del modelo.

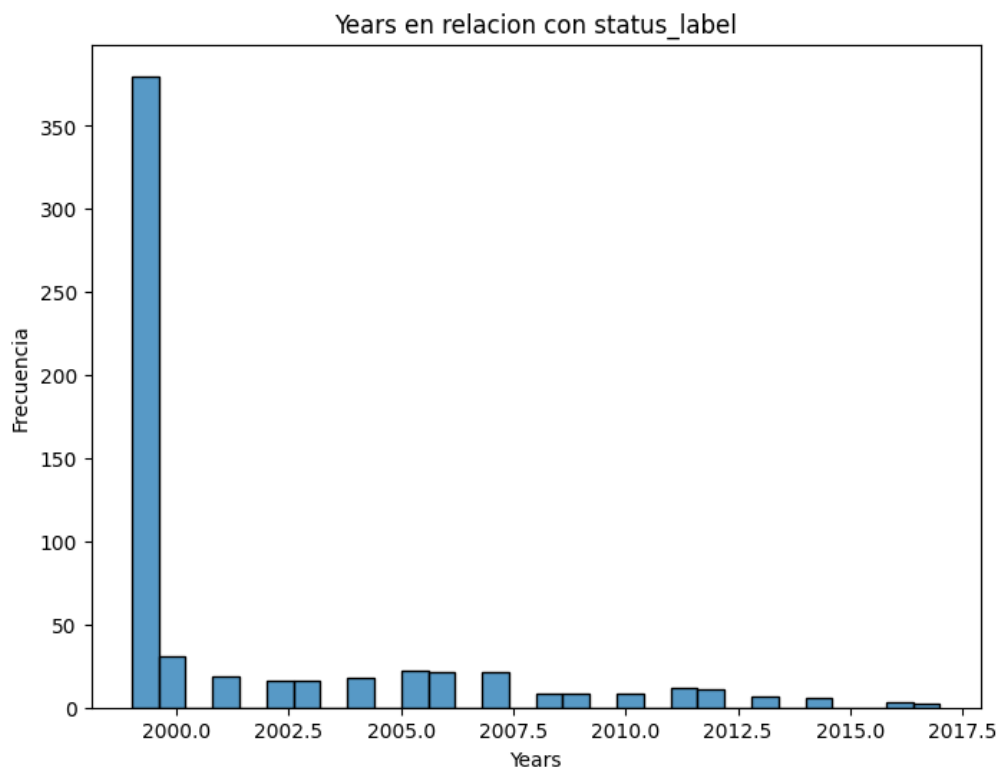


Figura 9: Relación año-status.

Fuente: Elaboración propia

Se observa que la mayoría de los casos en los que se produce un cambio en el status de las empresas están concentrados en años específicos, especialmente en 1999. Esta concentración sugiere que la variable “año” no aporta valor predictivo generalizable y, por tanto, no resulta útil para el análisis. Debido a esta distribución anómala, se decide excluir esta variable del estudio, ya que su inclusión podría sesgar los resultados al basarse en un comportamiento puntual no extrapolable al conjunto del modelo

5.3 Funcionamiento del modelo

Ante los resultados obtenidos en el análisis exploratorio de los datos, se decide, antes de iniciar el desarrollo del modelo, estudiar la aplicación de diferentes técnicas destinadas a mitigar el desbalance existente entre clases, concretamente entre el número de empresas solventes y aquellas en situación de quiebra. Con el objetivo de incrementar la representación de la clase minoritaria (empresas en quiebra), se emplean dos métodos ampliamente utilizados en el ámbito del oversampling: SMOTE y ADASYN.

Ambas técnicas tienen como finalidad la generación sintética de nuevos ejemplos correspondientes a la clase minoritaria, permitiendo así equilibrar el conjunto de datos y mejorar la capacidad del modelo para aprender patrones relevantes en ambas clases. La definición técnica, según el trabajo de Joaquín García Abad “Comparativa de técnicas de balanceo de datos. Aplicación aun caso real para la predicción de fuga de clientes”[22], es la siguiente:

- Synthetic Minority Oversampling Technique (SMOTE): “La propuesta de SMOTE (Chawla et al., 2002) se basa en la réplica de instancias de la clase minoritaria tomando en cada caso e introduciendo nuevos ejemplos sobre las líneas de unión de sus k-vecinos más cercanos”
- Adaptive Synthetic Sampling (ADASYN): “La aplicación del método es la misma (refiriéndose al caso de SMOTE), pero ADASYN añade una distorsión a los datos para hacerlos más “realistas”. Esta distorsión parte de la introducción de un ratio basado en el número de vecinos de la clase mayoritaria. Es decir, ADSYN pretende tener en cuenta la clase mayoritaria a la hora de generar instancias sintéticas”

En la siguiente imagen, incluida en el estudio, se puede observar de forma clara y visual el proceso de generación de instancias sintéticas que realizan ambos métodos, tanto SMOTE como ADASYN. Esta representación gráfica permite comprender de manera intuitiva cómo se crean nuevos datos en el espacio de características para mejorar la representación de la clase minoritaria.

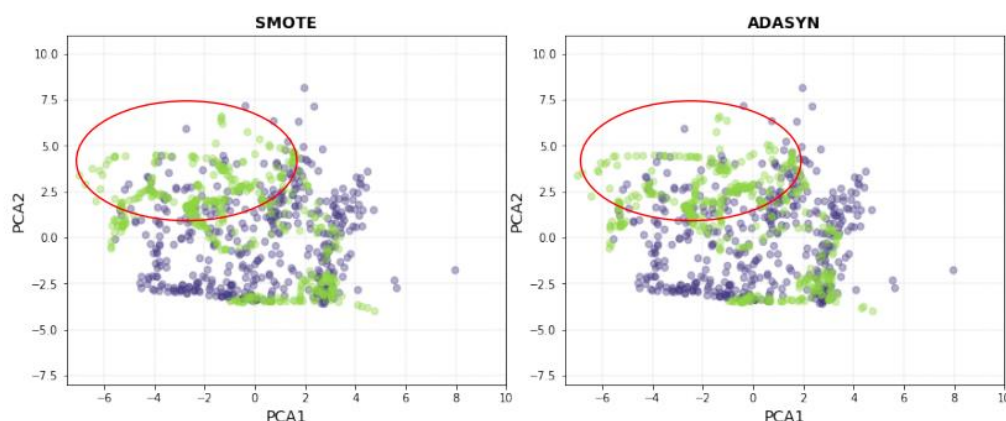


Figura 10: SMOTE vs ADASYN.

Fuente: Joaquín García Abad. Comparativa de técnicas de balanceo de datos. Aplicación acun caso real para la predicción de fuga de clientes (TFM).

Una vez comprendido el funcionamiento y propósito de ambas técnicas, se procede a su análisis comparativo con el objetivo de evaluar el rendimiento que ofrecen en distintos parámetros clave. Este estudio permite determinar cuál de las dos metodologías resulta más adecuada para ser aplicada en el desarrollo del modelo. A continuación, se presentan los resultados obtenidos tras la implementación de ambas técnicas:

Técnica/resultados	Accuracy	Precision	Recall	F1_score
SMOTE	0.6746	0.6760	0.9883	0.8028
ADASYN	0.5692	0.6250	0.3381	0.4388
Original	0.9314	0.9318	0.9994	0.9644

Tabla 1: SMOTE vs ADASYN

Fuente: Elaboración propia

A la vista de los resultados obtenidos en este estudio preliminar, se ha decidido emplear en el desarrollo del modelo los datos generados mediante la técnica SMOTE, así como el conjunto de datos original complementado con variables adicionales que permitan enriquecer el análisis. Esta elección se fundamenta en que, tal como se puede observar claramente, los datos tratados con SMOTE presentan un rendimiento superior en las cuatro métricas evaluadas en comparación con los obtenidos mediante ADASYN.

A continuación, se presentan los algoritmos de machine learning que se emplearán en el desarrollo del modelo. Previamente, se introducen los parámetros de evaluación que se utilizarán para medir el comportamiento y el rendimiento de cada algoritmo, permitiendo así una comparación objetiva y rigurosa entre ellos. Las definiciones y explicaciones proporcionadas por Google Machine Learning y Microsoft sobre estos términos son las siguientes [23] [24]:

- Accuracy: “La exactitud es la proporción de todas las clasificaciones que fueron correctas, ya sean positivas o negativas”. Esto se representa matemáticamente mediante la siguiente fórmula:

$$\frac{\text{Verdaderos Positivos} + \text{Verdaderos Negativos}}{\text{Verdaderos Positivos} + \text{Verdaderos Negativos} + \text{Falsos Positivos} + \text{Falsos Negativos}}$$

- Precision: “La precisión es la proporción de todas las clasificaciones positivas del modelo que realmente son positivas”. Esto se representa matemáticamente mediante la siguiente fórmula:

$$\frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos Positivos}}$$

- Recall: “mide la capacidad del modelo para predecir clases positivas reales. Es la relación entre los verdaderos positivos previstos y lo que realmente se ha etiquetado.” Esto se representa matemáticamente mediante la siguiente fórmula :

$$\frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos Negativos}}$$

- F1_score: “La puntuación F1 es una función de precisión y coincidencia. Es necesaria cuando se busca un equilibrio entre precisión y coincidencia.” Esto se representa matemáticamente mediante la siguiente fórmula:

$$\frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

A continuación, se presentan los algoritmos que se emplearán en el modelo, acompañados de una breve explicación de su funcionamiento y características principales. Esto permitirá una mejor comprensión de su comportamiento, facilitando el entendimiento del modelo desarrollado y ayudando a identificar qué tipo de algoritmo resulta más adecuado para el análisis de los datos empleados en este estudio. La definición que proporciona IBM para cada una de ellas es la siguiente:

- Random forest:” El bosque aleatorio es un algoritmo de aprendizaje automático de uso común, registrado por Leo Breiman y Adele Cutler, que combina el resultado de múltiples árboles de decisión para llegar a un resultado único. Su facilidad de uso y flexibilidad han impulsado su adopción, ya que maneja problemas de clasificación y regresión.”[25].
- SVM: “SVM funciona correlacionando datos a un espacio de características de grandes dimensiones de forma que los puntos de datos se puedan categorizar, incluso si los datos no se puedan separar linealmente de otro modo. Se detecta un separador entre las categorías y los datos se transforman de forma que el separador se puede extraer como un hiperplano. Tras ello, las características de los nuevos datos se pueden utilizar para predecir el grupo al que pertenece el nuevo registro.”[26].
- Gradient Boosting:” Gradient boosting es un algoritmo de aprendizaje de conjunto que genera predicciones precisas mediante la combinación de múltiples árboles de decisión en un único modelo. Este enfoque algorítmico para la modelización predictiva, introducido por Jerome Friedman, utiliza modelos base para reforzar sus fortalezas, corrigiendo errores y mejorando la capacidad predictiva. Al captar patrones complejos en los datos, gradient boosting destaca en diversas tareas de modelización predictiva.”[27].
- Redes neuronales: “Una red neuronal es un programa o modelo de machine learning que toma decisiones de forma similar al cerebro humano, al emplear

procesos que imitan la forma en que las neuronas biológicas trabajan juntas para identificar fenómenos, sopesar opciones y llegar a conclusiones.”[28].

- Bagging: “El bagging, también conocido como agregación de bootstrap, es el método de aprendizaje por conjuntos que se utiliza habitualmente para reducir la varianza en un conjunto de datos ruidoso.”[29].

En todos los casos presentados es necesario definir el tamaño de la muestra destinada al entrenamiento, así como el random state.

Según lo expuesto por Telefónica [30], el test size corresponde al conjunto de datos utilizado para verificar que el modelo desarrollado funciona correctamente. Este modelo se entrena con el resto de los datos, es decir, los datos de entrenamiento.

Habitualmente, se utiliza una proporción estándar del 70 % para entrenamiento y el 30 % restante para prueba. Dado que esta división es comúnmente aceptada, se adopta dicha proporción en el desarrollo del modelo.

Mientras que el random state lo describe Kaggle como: “El random state es un valor de semilla que se utiliza para inicializar el generador de números aleatorios.”[31]. En este proyecto, se emplea en todos los casos el valor 42, ya que, tras una revisión, se ha comprobado que es un valor comúnmente utilizado en estudios de este tipo.

- Para el caso de Random forest:

Se desarrolla el modelo utilizando el algoritmo de Random Forest, ajustando los parámetros necesarios para obtener los resultados óptimos. Este proceso incluye la optimización de la profundidad de los árboles de decisión y el número de estimadores (árboles) que conforman el modelo.

Para ello, se realiza un estudio preliminar mediante un modelo de prueba con el objetivo de identificar los parámetros más adecuados. A partir de este análisis, se obtienen los valores óptimos que serán empleados en el modelo definitivo.

Mejores parámetros: {'max_depth': 30, 'n_estimators': 300}

Figura 11: Parámetros random forest.

Fuente: Elaboración propia

- Para el caso de SVM:

Para el caso del algoritmo SVM, es necesario seleccionar el kernel que se empleará. El kernel es una función que permite transformar los datos a espacios de mayor dimensión, facilitando así la separación entre clases.[32].

En este estudio, se ha optado por utilizar un SVM con kernel de función de base radial (RBF), también conocida como función gaussiana, ya que tras diversas pruebas ha demostrado ofrecer las mejores prestaciones para el estudio desarrollado. No obstante, para el desarrollo completo del modelo en fases posteriores, sería recomendable analizar y comparar otros tipos (como el lineal, polinomial...) con el fin de identificar la opción óptima en función de la naturaleza final del conjunto de datos y los objetivos del proyecto.

- Para el caso de Gradient boosting:

En este caso, tras realizar diversas pruebas y comparar los resultados obtenidos, se ha decidido emplear los siguientes parámetros, los cuales permiten alcanzar un buen equilibrio del modelo. Al igual que en el caso anterior, para un desarrollo más completo y riguroso del modelo sería recomendable realizar un estudio exhaustivo de combinaciones posibles entre los distintos parámetros, con el objetivo de identificar la configuración óptima que maximice el rendimiento del modelo.

Los parámetros seleccionados para este estudio son:

- Número de estimadores: 150
- Profundidad máxima (max_depth): 8

- Para el caso de redes neuronales:

En el caso de las redes neuronales, es necesario ajustar varios parámetros clave para el correcto funcionamiento del modelo. El primero de ellos es el número de capas que se van a utilizar. Dado que este algoritmo se inspira en el funcionamiento de las redes neuronales humanas, se debe definir cuántas capas ocultas compondrán la red. Para este prototipo y con el objetivo de realizar una comparación eficaz, se han desarrollado tres configuraciones distintas: una con 3 capas, otra con 5 capas y una última con 10 capas.

Además, es necesario establecer la densidad de cada capa, es decir, el número de neuronas por capa, así como el tipo de función de activación que se aplicará. Esta función matemática es la encargada de activar las neuronas y puede adoptar diferentes formas (por ejemplo: linear, tanh, sigmoid...). Para definir estos parámetros, se han realizado diversas pruebas, manteniendo configuraciones similares entre los casos, empleando una mayor densidad en las capas iniciales y seleccionando funciones de activación similares.

También se ajustan otros parámetros como el número de ciclos de entrenamiento (epochs) y el batch size, tamaño del lote de datos. Tras varias pruebas, se ha determinado que en todos los casos se utilizarán 100 epochs y un batch size de 16, valores que han mostrado un buen comportamiento y equilibrio para el modelo.

En este caso se incorpora la técnica conocida como early stopping, un mecanismo que detiene automáticamente el entrenamiento del modelo cuando se detecta que no se producen mejoras significativas en los resultados. Dado que el modelo ha sido configurado para realizar hasta 100 ciclos (epochs), el early stopping evita ciclos innecesarios si se alcanza previamente el mejor rendimiento posible. Para este estudio, se establece un valor de 'paciencia' de 2 ciclos, lo que significa que, si el modelo no muestra mejoras en dos iteraciones consecutivas, el entrenamiento se interrumpe.[33].

Para el caso de bagging:

Se aplica la técnica de bagging a los demás algoritmos analizados, con el objetivo de mejorar los resultados obtenidos en los casos anteriores utilizando los mismos datos.

Como se ha indicado previamente, en el siguiente apartado se exponen los resultados obtenidos a partir del modelo desarrollado. Se presentan los dos enfoques definidos anteriormente: por un lado, el uso de la técnica SMOTE para equilibrar la base de datos, y por otro, el empleo de los datos originales complementados con un mayor número de variables, entre las que se incluyen:

- Ratios de rentabilidad:
 - Equity
 - ROA
 - ROE
 - Profit margin

- Ratios de liquidez
 - Current ratio

- Ratios de apalancamiento financiero
 - Total debt ratio
 - Debt equity ratio
 - Equity multiplier

La incorporación de estos ratios tiene como objetivo enriquecer el conjunto de datos, aportando variables adicionales que pueden ofrecer información valiosa para determinar si una compañía es solvente o se encuentra en quiebra. Esto permitirá mejorar el rendimiento del modelo y facilitar su proceso de aprendizaje.

Añadiendo estos ratios se pretende mejorar el conjunto de datos al contar con más datos que pueden ofrecer información valiosa para determinar si una compañía es solvente o está en quiebra, mejorando el rendimiento del modelo y ayudando a su aprendizaje. Se adjunta el estudio previo de estos nuevos datos en el anexo del proyecto.

6. Resultados

6.1 Resultados con SMOTE

A continuación, se presentan los resultados obtenidos del modelo en cada uno de los casos de estudio utilizando los datos tratados con la técnica de SMOTE.

- Random forest

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
Random forest	0.9236	0.9257	0.9634	0.9442

Tabla 2: Resultados random forest SMOTE.

Fuente: Elaboración propia

Se observa que todos los indicadores de rendimiento superan el 92% de precisión, destacando especialmente el valor de recall, que presenta el mejor resultado. Esto indica que el modelo identifica correctamente la mayoría de los verdaderos positivos. Por otro lado, la métrica de accuracy muestra el valor más bajo entre los indicadores, lo que sugiere que aún existe margen de mejora en términos de exactitud general del modelo.

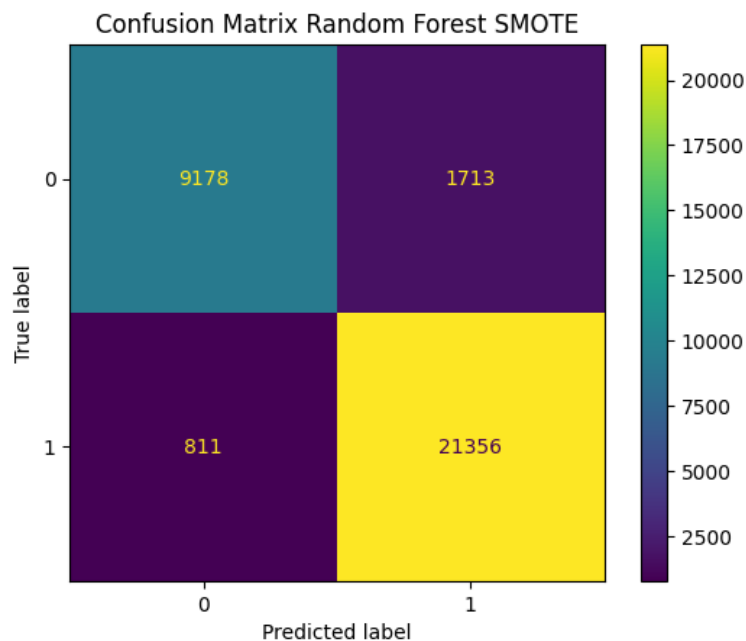


Figura 12: Matriz confusión random forest SMOTE.

Fuente: Elaboración propia

En la matriz de confusión se pueden observar los resultados de rendimiento del modelo. Se verifica que 9.178 casos de quiebra han sido correctamente clasificados, mientras que 1.713 empresas en situación de quiebra han sido erróneamente etiquetadas como solventes. Por otro lado, 811 empresas solventes fueron clasificadas incorrectamente como en quiebra.

Es especialmente relevante el caso de los 1.713 falsos negativos, es decir, empresas en quiebra que el modelo identifica erróneamente como solventes. Este tipo de error es crítico, ya que puede conducir a interpretaciones erróneas en la toma de decisiones financieras.

En la gráfica de la curva precision-recall se aprecia que el punto de equilibrio óptimo se sitúa aproximadamente cuando ambas métricas se encuentran en torno a 0.90. Asimismo, se observa que intentar maximizar al 100% cualquiera de las dos métricas conllevaría una caída significativa en la otra: un recall perfecto reduciría la precisión a menos de 0.70, y en el otro caso recall sería 0.

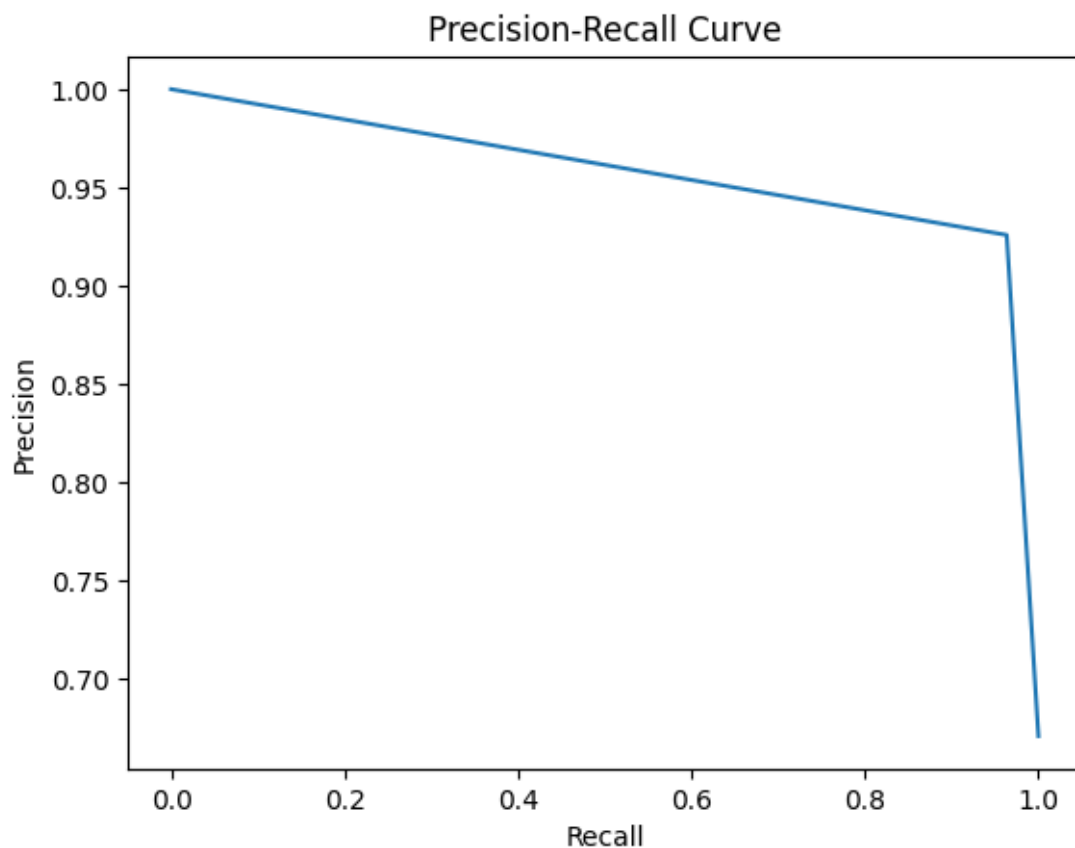


Figura 13: Curva precision-recall random forest SMOTE.

Fuente: Elaboración propia

La siguiente imagen se incluye con el objetivo de proporcionar información adicional sobre el estudio realizado, siendo especialmente relevante para identificar cuáles son los factores que el modelo considera más determinantes.

Se observa que las variables más influyentes en el modelo son el beneficio neto (net income), la deuda a largo plazo (long term debt) y el inventario (inventory). En contraste, otras variables financieras habitualmente consideradas clave, como los activos totales, los ingresos totales (total revenues) y las ventas netas (net sales), presentan una menor relevancia en el análisis.

Conocer el peso relativo de cada variable permite mejorar el diseño y la optimización del modelo, centrándose en aquellas variables que realmente influyen en la predicción de la solvencia o quiebra de una empresa. Además, este análisis aporta información valiosa para futuras investigaciones y para empresas interesadas en comprender qué indicadores financieros son más representativos del riesgo de insolvencia. Por tanto, sería recomendable ampliar este estudio en etapas posteriores del desarrollo del producto, dada su utilidad práctica y su potencial valor añadido.

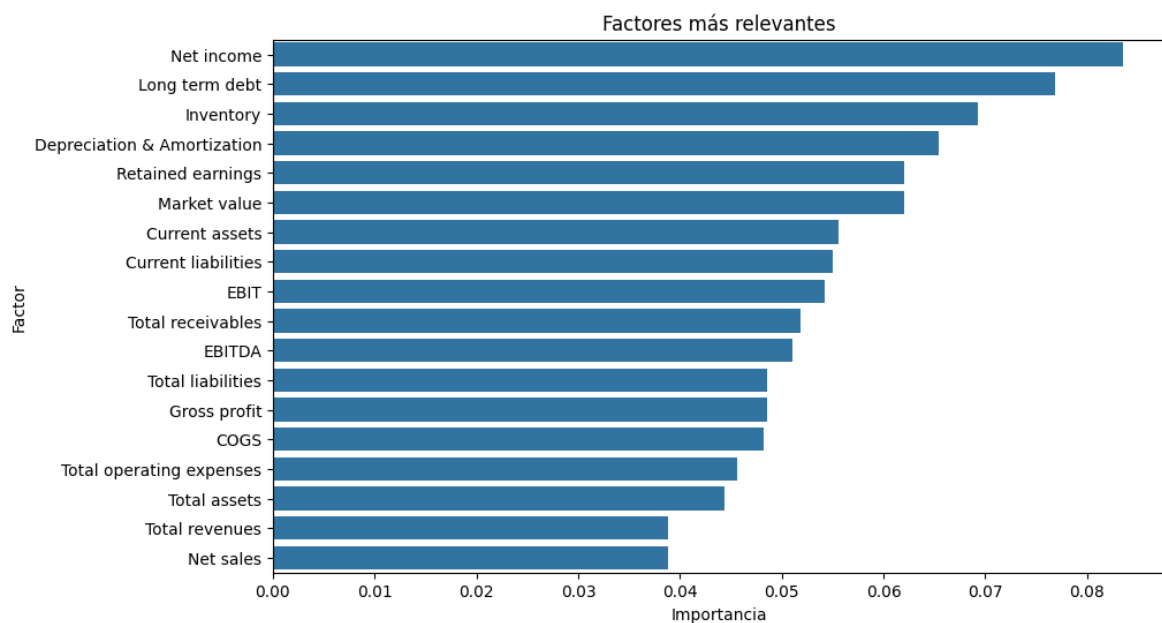


Figura 14: Factores más relevantes random forest SMOTE .

Fuente: Elaboración propia

- SVM

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
SVM	0.6710	0.6709	0.9997	0.8029

Tabla 3: Resultados SVM SMOTE.

Fuente: Elaboración propia

Se presentan los resultados obtenidos para el caso del algoritmo SVM. Las métricas evidencian un rendimiento inferior en comparación con otros modelos empleados, especialmente en los parámetros de accuracy y precision, ambos con valores inferiores a 0.7. Por otro lado, el recall muestra un rendimiento notable, cercano al 100 %.

Estos resultados se interpretan con mayor claridad al analizar la matriz de confusión. Se observa que los casos positivos, es decir, las empresas solventes, han sido correctamente clasificados como tales en todos los casos excepto en seis, lo cual explica el elevado valor del recall. Sin embargo, en el caso crítico de las empresas en quiebra, de especial interés en este estudio, únicamente 21 han sido correctamente identificadas, mientras que 10.870 casos de empresas en quiebra han sido erróneamente clasificados como solventes.

Esta elevada tasa de error en la clasificación de quiebras explica el bajo rendimiento en términos de accuracy y precision, y pone de manifiesto las limitaciones del algoritmo SVM para abordar este tipo de problemas.

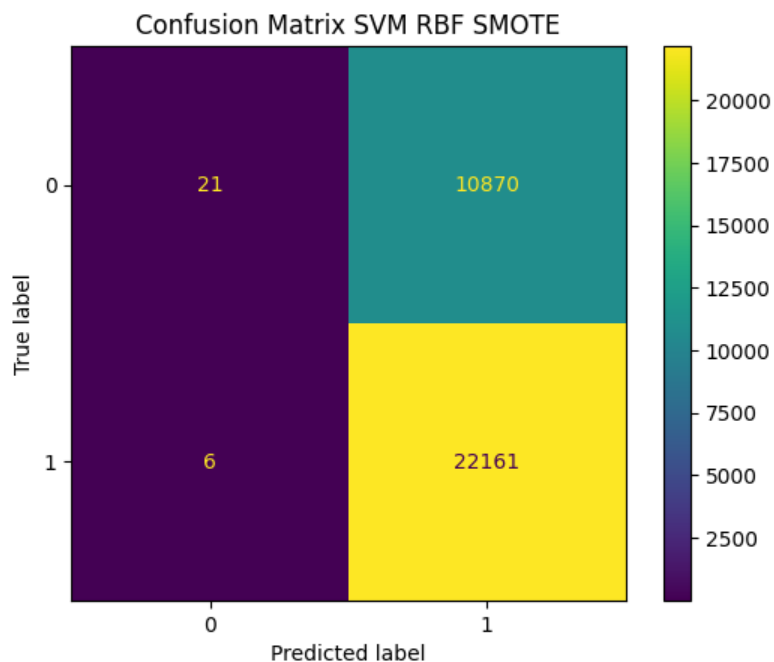


Figura 15: Matriz confusión SVM SMOTE.

Fuente: Elaboración propia

- Gradient Boosting

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
Gradient Boosting	0.9136	0.9218	0.9520	0.9366

Tabla 4: Resultados Gradient Boosting SMOTE.

Fuente: Elaboración propia

Los resultados obtenidos con el modelo de Gradient Boosting muestran un mayor equilibrio en comparación con el algoritmo SVM, presentando un comportamiento similar al de Random Forest. Todas las métricas evaluadas superan el 0.9, destacando nuevamente el recall como el parámetro con mejor desempeño.

En términos de predicción, el modelo identifica correctamente 9.102 casos de empresas en quiebra, mientras que 1.789 son erróneamente clasificadas como solventes. En cuanto a las empresas solventes, 1.064 son incorrectamente etiquetadas como en quiebra.

Estos resultados son comparables a los obtenidos con Random Forest y significativamente mejores que los alcanzados con SVM. No obstante, se observa un ligero aumento tanto en los falsos positivos como en los falsos negativos respecto a Random Forest.

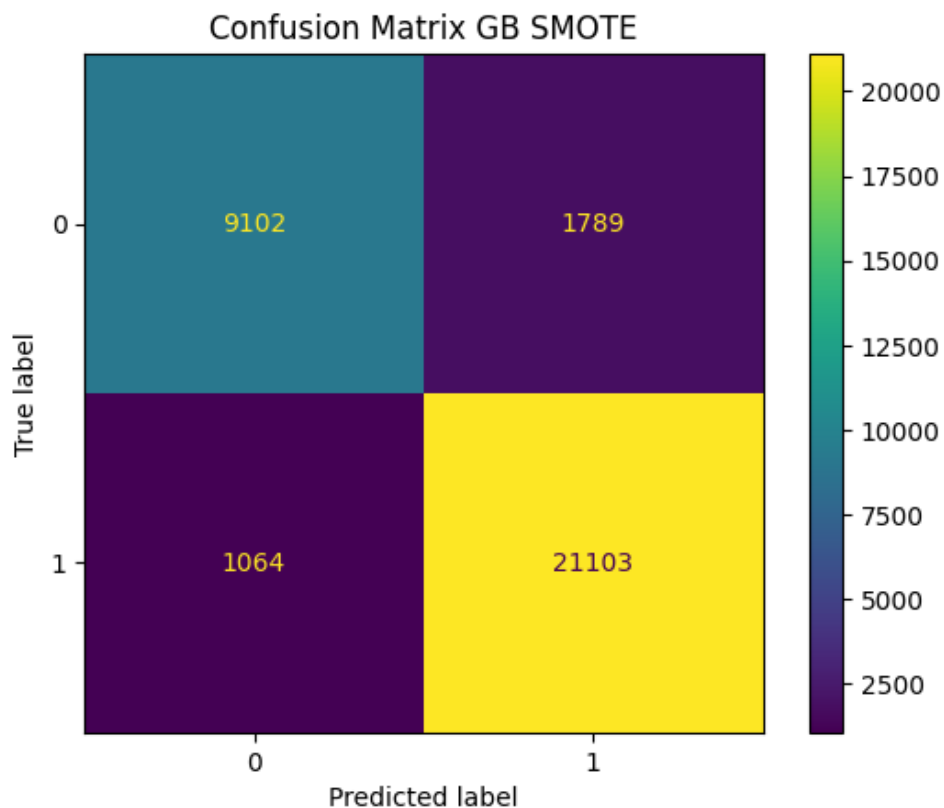


Figura 16: Matriz confusión gradient boosting SMOTE.

Fuente: Elaboración propia

Se presenta la curva Precision-Recall, la cual muestra una forma y unos valores muy similares a los obtenidos en el caso de Random Forest. Se observa un rendimiento consistente, con un pico de desempeño situado en torno a un valor de 0.9.

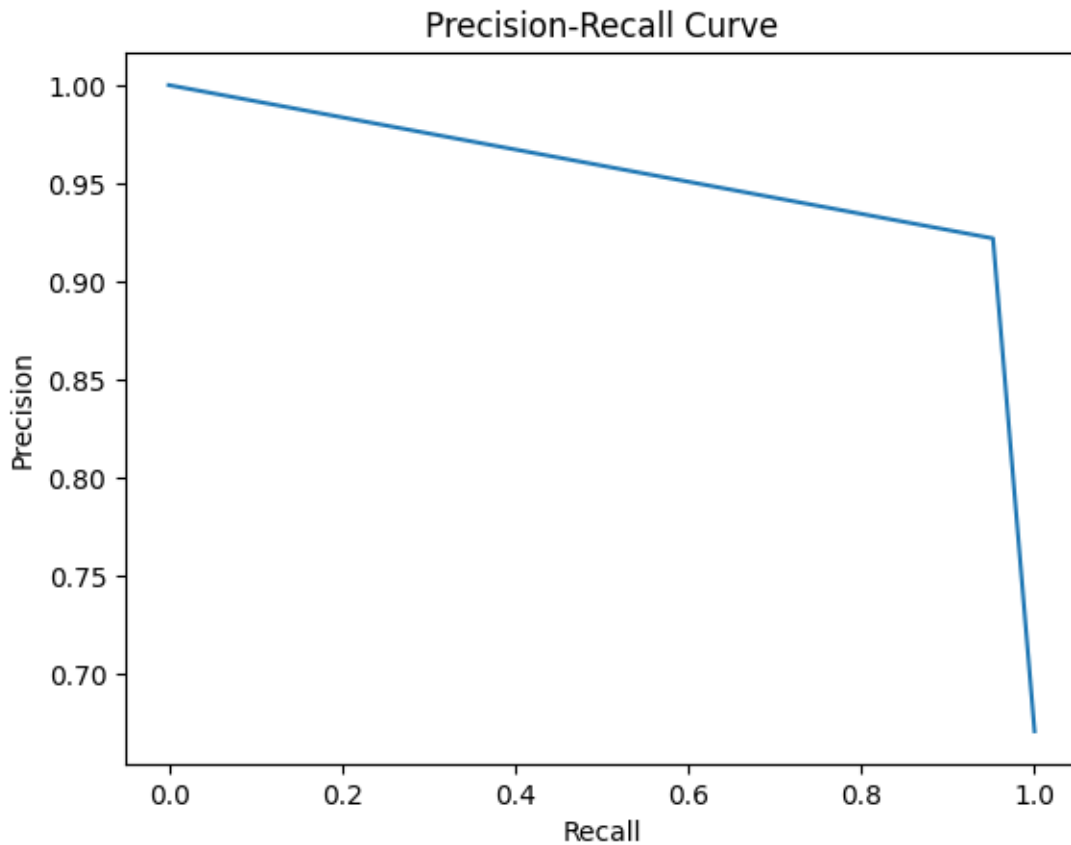


Figura 17: Curva precision-recall GB SMOTE.

Fuente: Elaboración propia

Asimismo, se presentan los factores más relevantes identificados por el modelo. Al igual que en el caso de Random Forest, se observa que las tres variables más importantes se repiten, al igual que las variables financieras menos relevantes. Las primeras adquieren un mayor peso, mientras que las últimas tienen una influencia mucho menor. Resulta especialmente significativo que, en dos modelos distintos, coincidan tanto las variables más determinantes como aquellas con menor impacto. Esto sugiere patrones consistentes en los datos que podrían ser de gran utilidad. Para una futura versión del producto final, sería recomendable profundizar en este análisis, ya que podría proporcionar información

valiosa para mejorar la precisión predictiva y optimizar el rendimiento del modelo, como se ha indicado previamente.

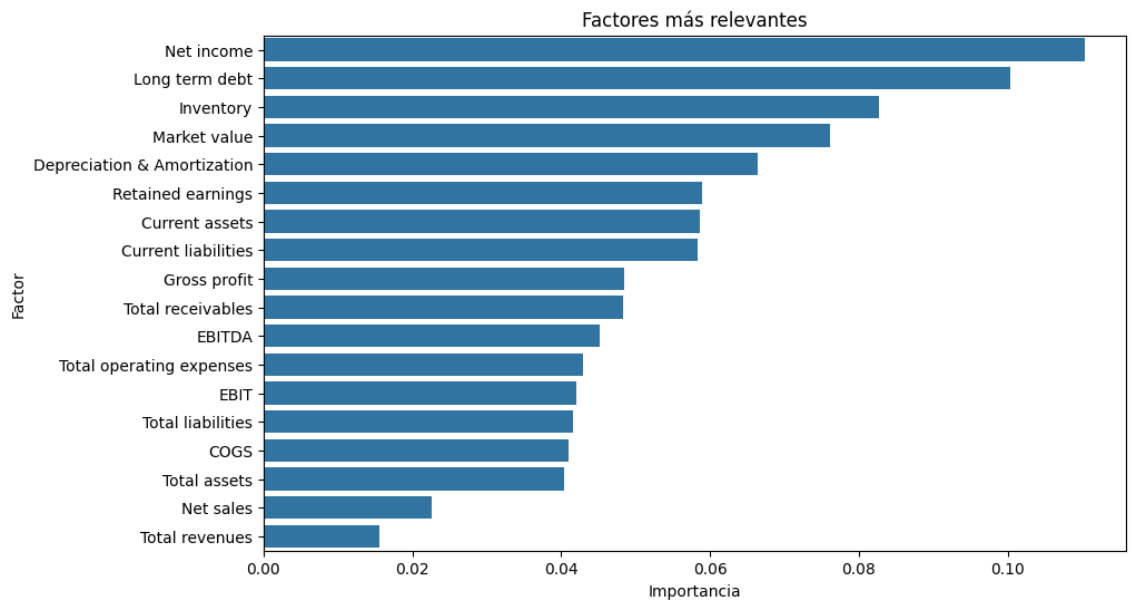


Figura 18: Factores más relevantes GB SMOTE

Fuente: Elaboración propia

- Redes neuronales

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
Redes neuronales 3 capas	0.7019	0.7160	0.9130	0.7978
Redes neuronales 5 capas	0.7285	0.7410	0.9102	0.7991
Redes neuronales 10 capas	0.7094	0.7219	0.9151	0.7983

Tabla 5: Resultados redes neuronales SMOTE.

Fuente: Elaboración propia

Se observan resultados similares en los tres casos de redes neuronales analizados. Se constata que, para este estudio y conjunto de datos, aumentar el número de capas no implica una mejora significativa en el rendimiento del modelo. De hecho, la red neuronal con cinco capas ofrece los mejores resultados en la mayoría de las métricas, salvo en el caso del recall, donde la red con diez capas presenta una ligera mejora.

Los valores de accuracy y precision, que oscilan entre 0.70 y 0.75 en todos los casos, indican , al compararlos con los modelos anteriores, que existe una proporción

considerable de empresas en quiebra que no están siendo correctamente clasificadas como tal, un comportamiento similar al observado con el modelo SVM.

Por lo tanto, los resultados sugieren que las redes neuronales, en este contexto específico, no son el enfoque más efectivo para la predicción de quiebras. Aunque presentan un buen rendimiento en la clasificación de empresas solventes (según el recall), su rendimiento en la identificación de compañías en situación de quiebra sigue siendo insuficiente, lo cual representa una limitación crítica para los objetivos de este estudio.

- Bagging

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
Bagging RF	0.9903	0.9909	0.9946	0.9927
Bagging SVM	0.6713	0.6711	0.9997	0.8031
Bagging GB	0.9785	0.9782	0.9899	0.9841
Bagging Redes neuronales 3 capas	0.6684	0.6812	0.9501	0.7935
Bagging Redes neuronales 5 capas	0.6869	0.6902	0.9672	0.8055
Bagging Redes neuronales 10 capas	0.6811	0.6853	0.9694	0.8030

Tabla 6: Resultados Bagging SMOTE.

Fuente: Elaboración propia

La técnica de Bagging se aplica a cada uno de los modelos previamente presentados con el objetivo de mejorar su rendimiento. A continuación, se detallan los efectos observados en cada caso:

- Bagging Random Forest (RF):

Se observa una mejora notable en todos los parámetros, superando el 0.99 en todas las métricas. Destacan especialmente los avances en accuracy y precision, parámetros fuertemente relacionados con la correcta predicción de casos de quiebra. Estos resultados convierten a Bagging RF en el enfoque con mejor desempeño en todo el estudio.

- Bagging SVM:

La aplicación de Bagging apenas genera mejoras significativas. Aunque se detectan ligeros aumentos en algunos parámetros, estos son marginales (en el orden de milésimas), lo que indica que esta técnica no resulta especialmente efectiva para este modelo.

- Bagging Gradient Boosting (GB):

Se produce una mejora generalizada en todos los parámetros respecto al modelo original de Gradient Boosting. Los resultados son equilibrados y sólidos, con avances notables en

accuracy y precision, lo que sugiere una mejor capacidad del modelo para detectar casos de quiebra.

- Bagging Redes Neuronales (3, 5 y 10 capas):

Se agrupan los tres casos ya que los resultados son similares en todos ellos. La técnica de Bagging no muestra eficacia en combinación con redes neuronales, independientemente del número de capas utilizadas. Se observa una leve mejora en recall, pero esta mejora se ve contrarrestada por un empeoramiento en el resto de las métricas, especialmente en accuracy y precision. En algunos casos, se registra una ligera mejora en el F1-score, en consonancia con el aumento de recall, aunque no compensa la pérdida de rendimiento general.

6.2 Resultados con datos iniciales mejorados

A continuación, se presentan los resultados obtenidos del modelo en cada uno de los casos de estudio utilizando los datos iniciales mejorados con datos sintéticos.

- Random forest

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
Random forest	0.9325	0.9326	0.9998	0.9650

Tabla 7: Resultados random forest.

Fuente: Elaboración propia

Se observa en estos resultados que el rendimiento de los parámetros evaluados mejora en comparación con el caso de SMOTE, superando en todos ellos el umbral del 0.93. Una vez más, el recall es el parámetro con mejor comportamiento. Sin embargo, al analizar la matriz de confusión, se evidencia una importante limitación: tan solo 25 casos de empresas en quiebra han sido correctamente clasificados, mientras que 1.589 han sido erróneamente etiquetados como solventes.

En contraste, para las empresas solventes, la clase mayoritaria del conjunto de datos, solo 4 casos han sido mal clasificados, frente a 21.987 correctamente predichos. Esta desproporción en la distribución de clases genera métricas globales aparentemente positivas, pero enmascara el verdadero problema: el modelo falla en identificar adecuadamente los casos de quiebra, que son precisamente el foco central de este proyecto.

En consecuencia, aunque los resultados globales mejoran respecto a SMOTE, esta mejora se debe principalmente a la elevada precisión en la clasificación de empresas solventes. La escasa representatividad de los casos de quiebra en el conjunto de datos reduce su impacto en las métricas generales, lo que pone de manifiesto la necesidad de ajustar el enfoque para mejorar la detección de esta clase crítica

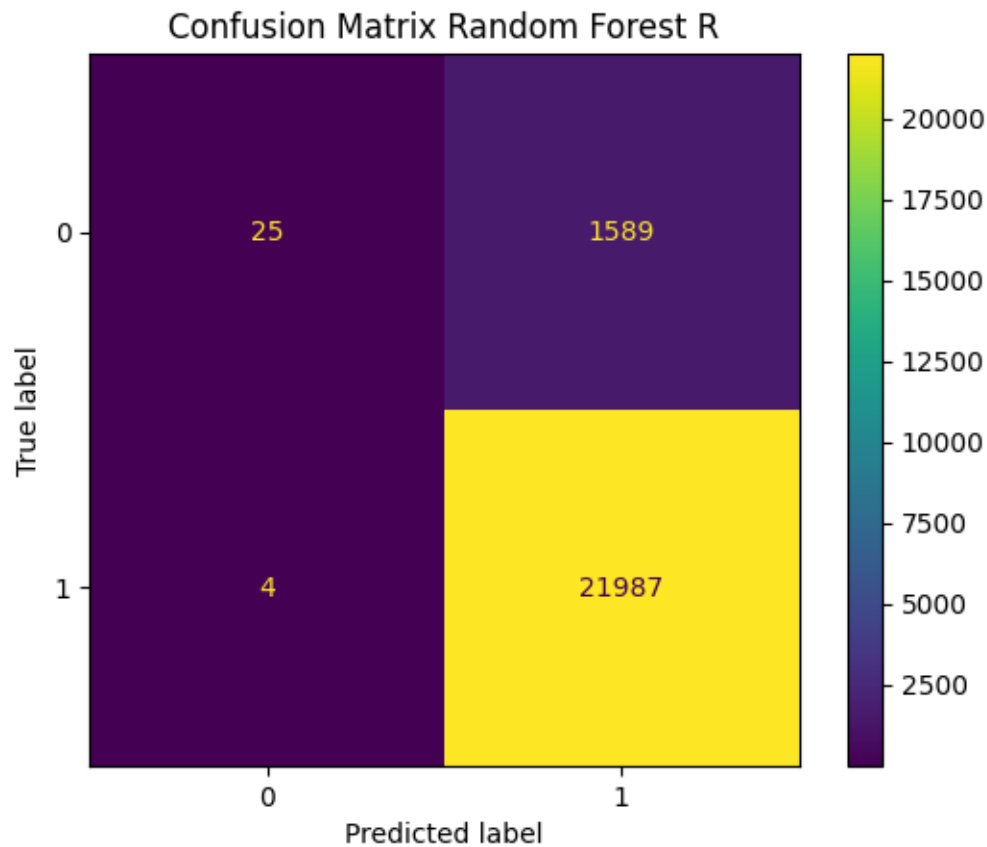


Figura 19: Matriz confusión random forest R

Fuente: Elaboración propia

En la curva Precision-Recall se observa claramente la relación inversa entre ambas métricas: un incremento en la precisión implicaría una disminución significativa en el recall, que en este caso es el parámetro con mejor desempeño. Por tanto, si se busca mejorar la precisión del modelo, ello conllevaría un coste notable en la capacidad del modelo para identificar correctamente los casos positivos.

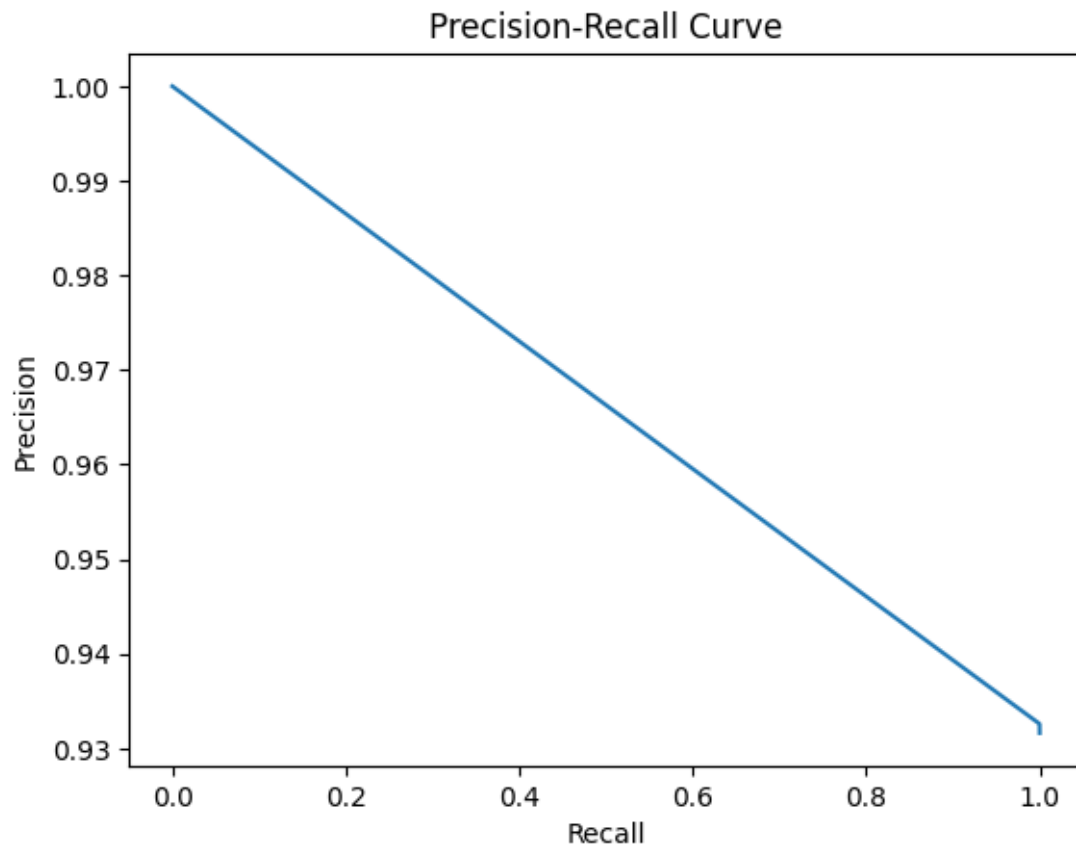


Figura 20: Curva precision-recall random forest R.

Fuente: Elaboración propia

En el análisis de los factores más relevantes, se observa un cambio significativo con respecto al caso de SMOTE. En este escenario, las variables con mayor peso, aunque con niveles de importancia inferiores a los observados anteriormente, son el valor de mercado, la depreciación y amortización, y los beneficios retenidos. Por otro lado, las variables con menor relevancia vuelven a ser las ventas totales y las ganancias totales, en consonancia con los resultados obtenidos con SMOTE. Sin embargo, destaca el hecho de que la deuda a largo plazo, que anteriormente figuraba entre las variables más influyentes, ahora se posiciona como una de las menos relevantes.

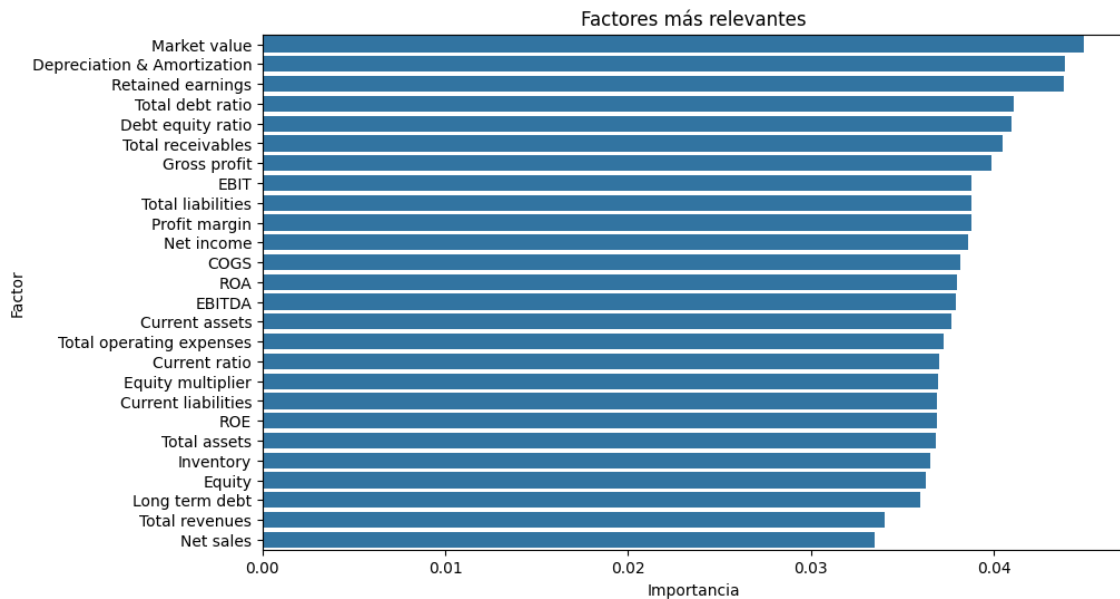


Figura 21: Factores más relevantes random forest R.

Fuente: Elaboración propia

Se busca analizar el impacto de ajustar el modelo para mejorar la capacidad de predicción de los casos de quiebra, es decir, de la clase minoritaria. Para ello, se decide aplicar unos pesos diferentes a las clases, asignando un peso de 0:100 para las empresas en quiebra (clase 0) y 1:1 para las empresas solventes (clase 1). Esto significa que los errores en la clasificación de empresas en quiebra tendrán una importancia cien veces mayor que los errores en la clasificación de empresas solventes. Con esta estrategia se pretende aumentar la sensibilidad del modelo hacia los casos de quiebra, asumiendo que ello podría afectar negativamente la precisión en la predicción de empresas solventes. A continuación, se presentan los resultados obtenidos con esta configuración:

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
Random forest pesos	0.6468	0.9599	0.6480	0.7737

Tabla 8: Random forest con pesos.

Fuente: Elaboración propia

Se puede observar que el rendimiento general del modelo ha empeorado en todos los parámetros evaluados, excepto en el caso de la precisión. En los otros tres indicadores de rendimiento (accuracy, recall y F1-score), se evidencia un deterioro significativo. En

particular, el valor de recall, que anteriormente era el parámetro con mejor resultado, desciende considerablemente, situándose cerca de 0.65, al igual que la accuracy. Para comprender con mayor profundidad estos resultados, es necesario analizar la matriz de confusión y la curva precision-recall.

En la matriz de confusión se aprecia que la predicción de los casos de quiebra ha mejorado notablemente: 1.019 casos han sido correctamente identificados como quiebra, frente a los 25 que se detectaban anteriormente. Sin embargo, 595 casos siguen siendo erróneamente clasificados como solventes. Esta mejora en la identificación de la clase minoritaria ha tenido un coste significativo en la precisión de la clase mayoritaria: 7.740 empresas solventes han sido clasificadas erróneamente como en quiebra, mientras que 14.251 han sido correctamente etiquetadas.

Tal y como se anticipaba al aplicar pesos diferenciados a las clases, se logra una mejora en la detección de los casos de quiebra, pero a costa de un deterioro considerable en la clasificación de empresas solventes. Esto tiene un fuerte impacto negativo en los resultados globales y, a pesar de la mejora relativa en la clase minoritaria, no se alcanzan aún niveles óptimos de predicción en los casos de quiebra

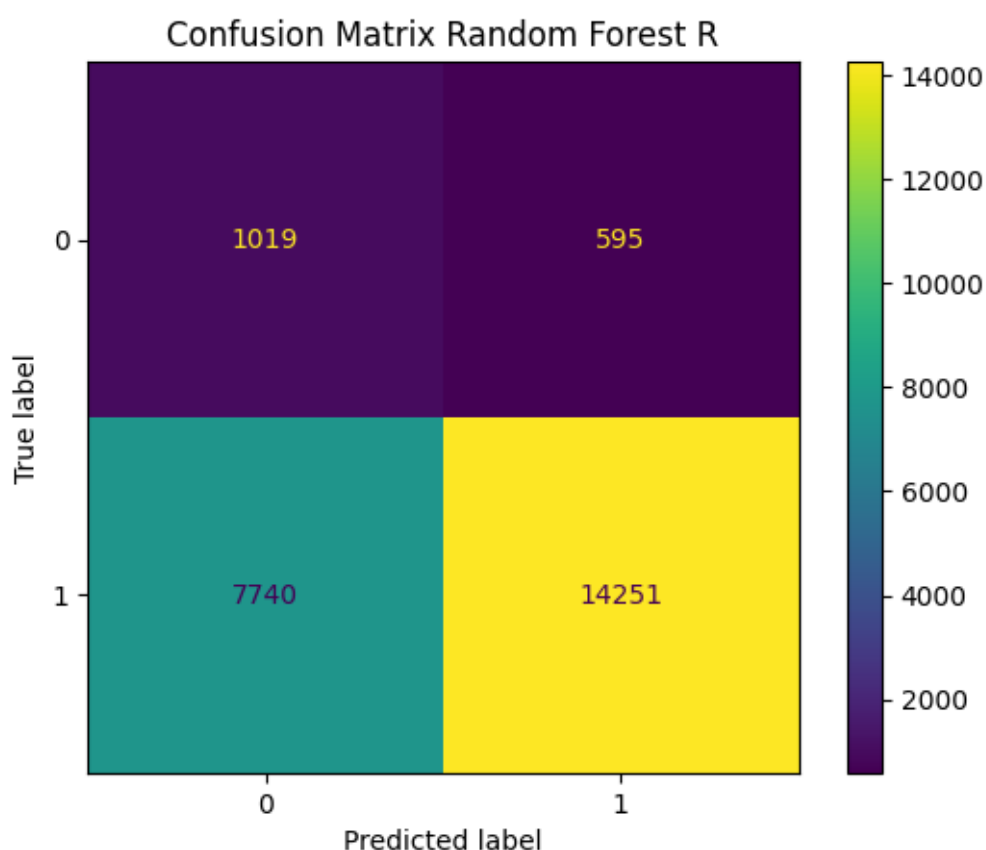


Figura 22: Matriz confusión random forest R con pesos.

Fuente: Elaboración propia

En la curva precision-recall se puede apreciar un cambio notable, identificándose un punto clave en torno a un valor de precision de 0,96 y un recall de aproximadamente 0,60. Este punto refleja de forma representativa los resultados obtenidos, tal y como se observa en la tabla de métricas del modelo, pudiendo comprobar la relación que ambos parámetros tienen y comprobándose lo explicado cuanto mayor el recall menor la precision y especialmente cuanto mayor la precision menor será el parámetro recall.

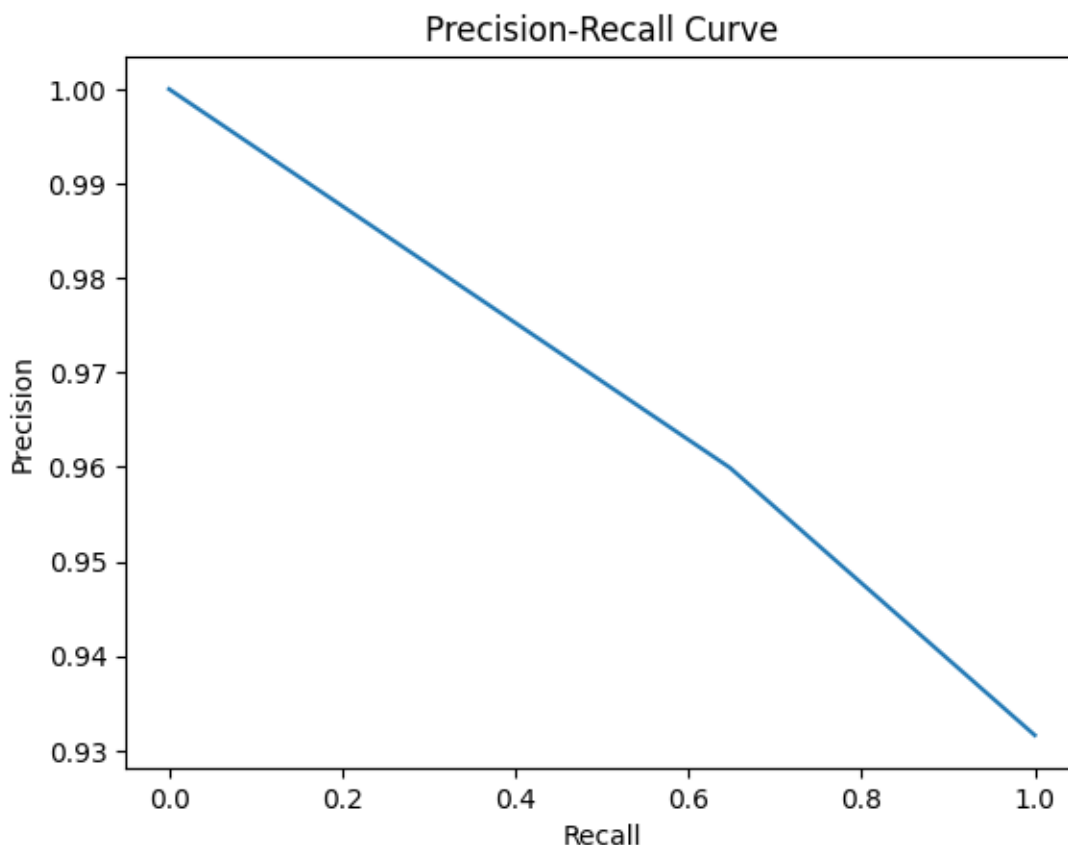


Figura 23: Curva precision-recall random forest R con pesos

Fuente: Elaboración propia

Observando los resultados obtenidos en el caso en el que se ha asignado mayor peso a los casos de quiebra, se aprecia una mejora en la capacidad del modelo para predecir este tipo de situaciones. Sin embargo, el rendimiento global ha disminuido de forma significativa y la predicción de quiebras aún no alcanza un nivel óptimo. Además, esta mejora ha supuesto una pérdida considerable en el rendimiento del modelo respecto a los casos de empresas solventes.

Por tanto, sería recomendable, de cara al desarrollo de un producto final, profundizar en este enfoque para identificar el punto de equilibrio óptimo que permita maximizar tanto el rendimiento global del modelo como la precisión en la detección de quiebras.

En los siguientes apartados, se presentarán los resultados obtenidos para el caso general, es decir, sin modificar los pesos de las clases. No obstante, como se ha señalado, sería valioso seguir explorando esta línea de ajuste para futuras versiones del modelo.

- SVM

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
SVM	0.9316	0.9316	1	0.9646

Tabla 9: Resultados SVM.

Fuente: Elaboración propia

Se observan resultados superiores al 0.93 en todos los parámetros, con una puntuación perfecta en el caso de recall. No obstante, al analizar la matriz de confusión, se detecta un problema importante que permite entender el porqué de estos valores, especialmente en el caso del recall.

El valor de recall es 1 porque el modelo clasifica correctamente todos los casos de empresas solventes, sin cometer ningún error. Sin embargo, ocurre lo contrario con las empresas en quiebra: el modelo clasifica erróneamente todos estos casos como solventes.

Debido al fuerte desequilibrio en el conjunto de datos, donde la mayoría de las empresas son solventes, el modelo optimiza sus resultados clasificando todos los casos como solventes. De esta forma, obtiene métricas aparentemente excelentes, especialmente en recall, a pesar de no identificar ni un solo caso de quiebra de manera correcta.

Este comportamiento refleja una clara limitación del modelo: aunque sus métricas generales son altas, su utilidad práctica es nula, ya que el objetivo principal de este proyecto es predecir correctamente las quiebras. Un modelo que no es capaz de identificar estos casos críticos, aunque acierte en la mayoría, no cumple su propósito.

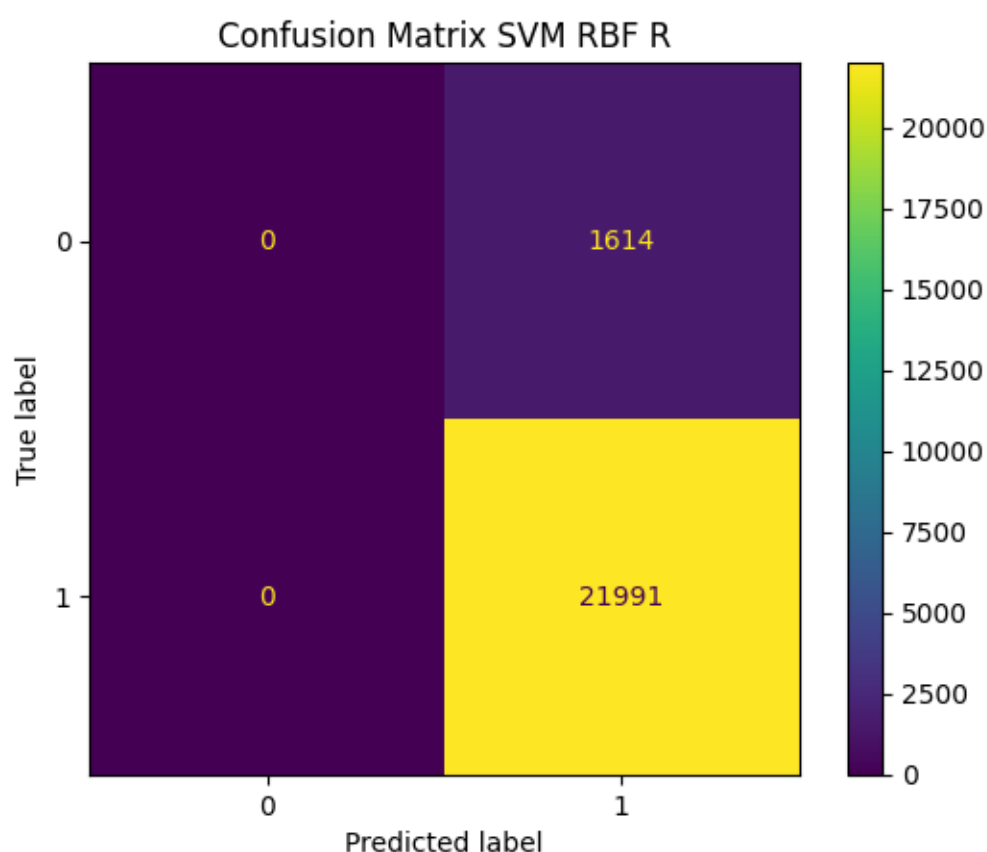


Figura 24: Matriz confusión SVM R

Fuente: Elaboración propia.

- Gradient Boosting

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
Gradient Boosting	0.9156	0.9348	0.9775	0.9557

Tabla 10: Resultados Gradient Boosting.

Fuente: Elaboración propia

Se obtienen resultados con Gradient Boosting ligeramente inferiores en comparación con los casos anteriores, manteniéndose recall como el parámetro con mejor rendimiento.

En este caso, ocurre una situación similar a la de modelos anteriores, aunque de forma menos pronunciada. Se observa que el modelo ha logrado predecir correctamente algunos casos de quiebra, concretamente 116, mientras que 1.498 han sido erróneamente clasificados como solventes.

En cuanto a las empresas solventes, la mayoría han sido clasificadas correctamente (21.498 casos), aunque 493 han sido etiquetadas como en quiebra.

Este comportamiento explica por qué el recall no alcanza el valor perfecto, aunque el modelo logra un mejor rendimiento en el parámetro precision. Esta relación entre ambos indicadores, en la que la mejora de la precision empeora el recall, se refleja claramente en la curva precision-recall que se muestra a continuación de la matriz de confusión.

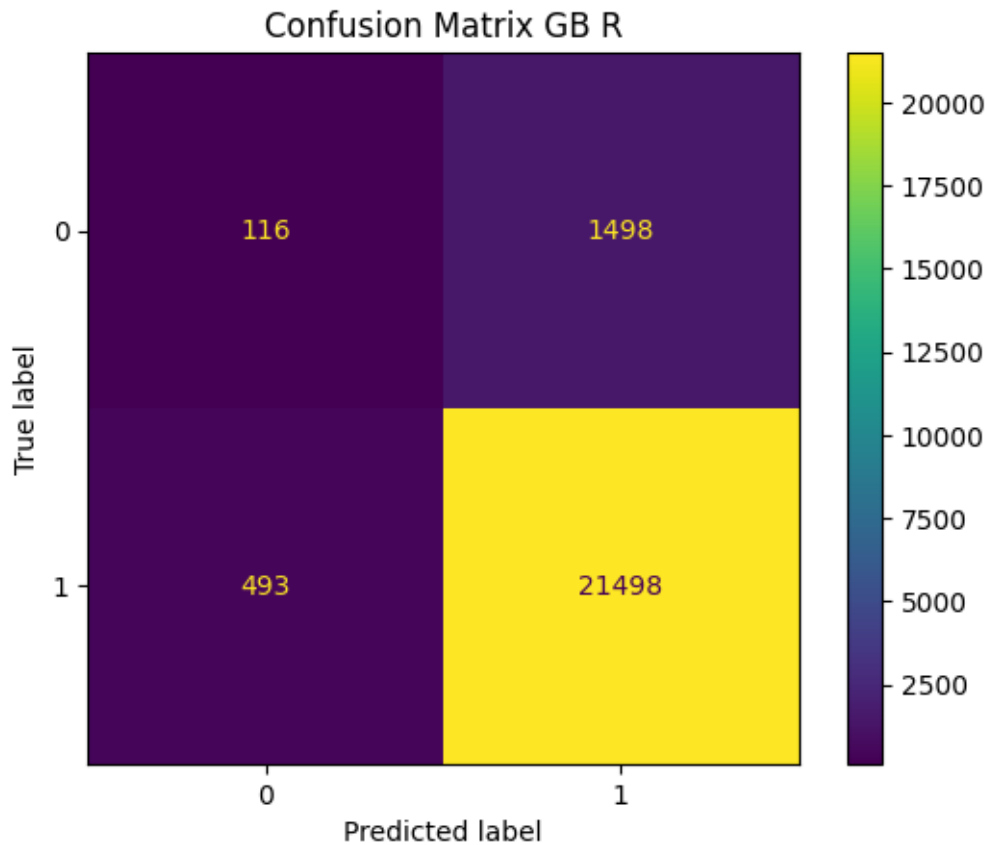


Figura 25: Matriz confusion gradient boosting R

Fuente: Elaboración propia

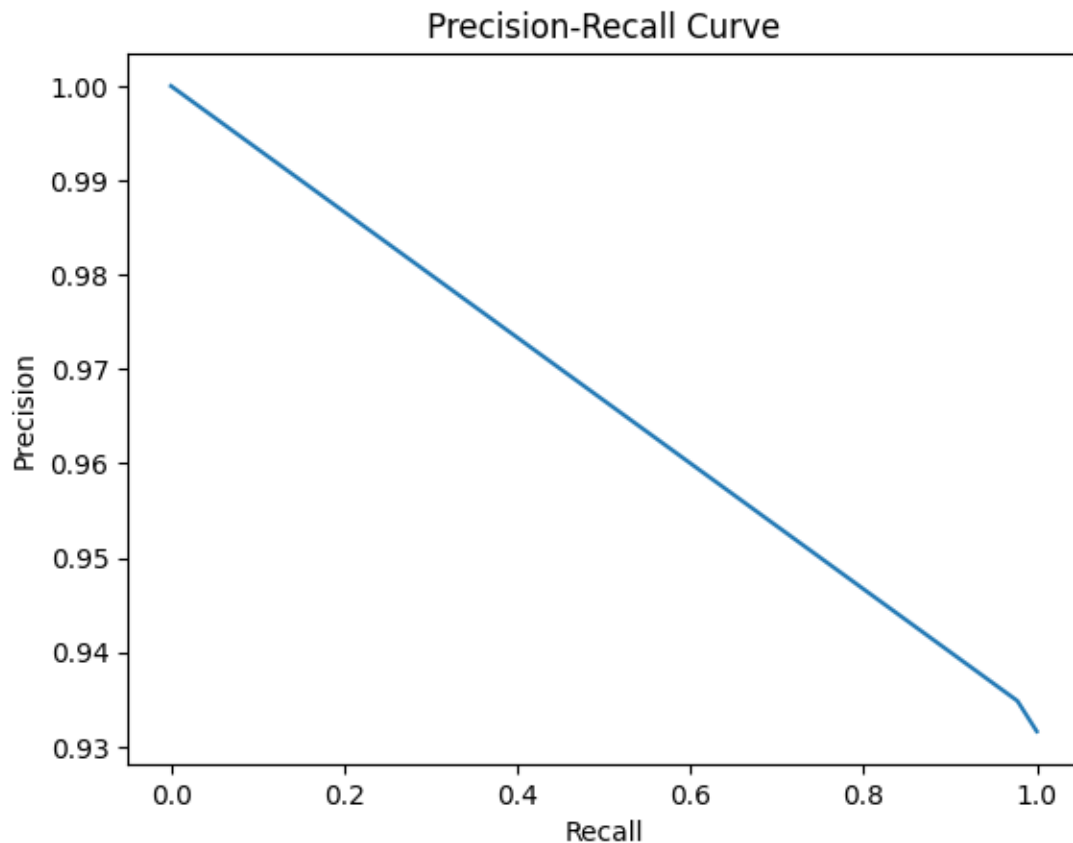


Figura 26: Curva precision-recall GB R.

Fuente: Elaboración propia

También resulta interesante analizar los factores más relevantes identificados por el modelo. Se observa que, entre las variables menos representativas, continúan apareciendo las ganancias totales y las ventas netas, a las que se suma el current ratio, uno de los ratios sintéticos introducidos en esta fase del estudio.

Por otro lado, entre las variables más influyentes para el modelo destacan dos que ya figuraban como relevantes en el caso de Random Forest: el valor de mercado y las ganancias retenidas. A estas se suma un nuevo ratio financiero, el debt-equity ratio.

Llama la atención cómo algunas variables clave se mantienen consistentes entre distintos modelos, mientras que ciertos datos sintéticos introducidos en esta etapa aparecen tanto entre las variables más relevantes como entre las menos influyentes, lo que sugiere una necesidad de análisis más profundo.

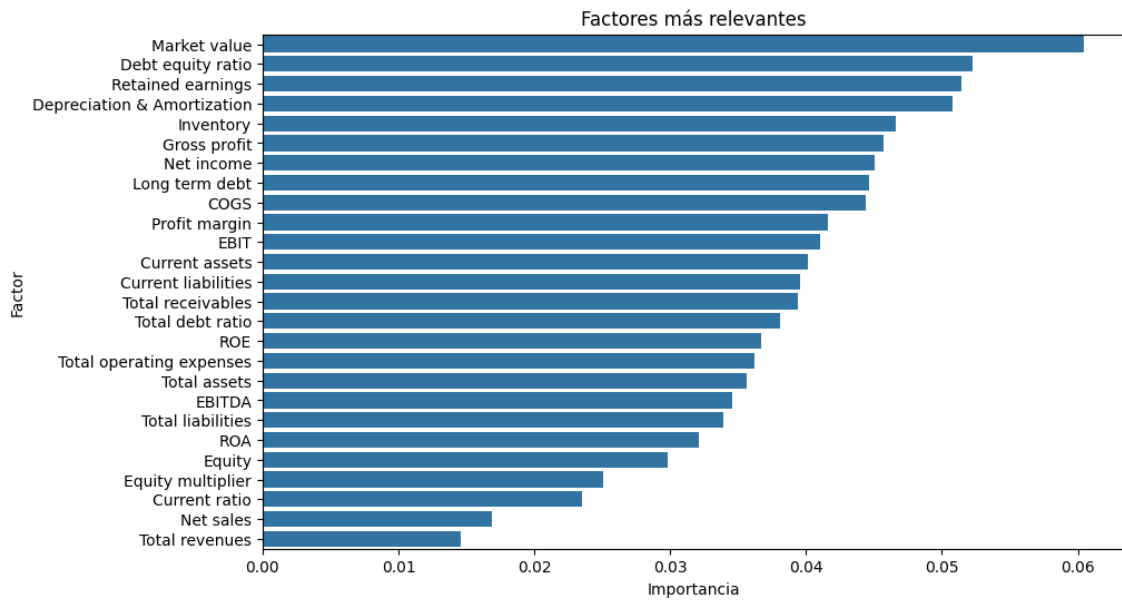


Figura 27: Factores más relevantes GB R.

Fuente: Elaboración propia

- Redes neuronales

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
Redes neuronales 3 capas	0.9345	0.9347	0.9997	0.9660
Redes neuronales 5 capas	0.9344	0.9345	0.9999	0.9661
Redes neuronales 10 capas	0.9338	0.9338	1	0.9657

Tabla 11: Resultados redes neuronales.

Fuente: Elaboración propia

Al observar los resultados, se constata que no hay un caso de los estudiados que presente un rendimiento claramente superior respecto a los otros dos. Para cada uno de los casos, en función del número de capas utilizadas, se aprecia que cada configuración destaca ligeramente en algún parámetro específico, aunque las diferencias entre ellos son mínimas, en la mayoría de los casos limitadas al tercer decimal.

Se confirma una vez más que recall es el parámetro con mejor rendimiento en los tres escenarios, comportamiento ya observado en los modelos anteriores. El hecho de que el recall alcance valores cercanos a 1 sugiere que los casos positivos, es decir, las predicciones de empresas solventes han sido clasificados correctamente en prácticamente todos los casos.

No obstante, basándonos en los análisis previos y en resultados similares observados anteriormente, es probable que estos buenos resultados generales estén enmascarando un bajo rendimiento en la predicción de las empresas en quiebra. En casos anteriores se ha demostrado que cuando recall es extremadamente alto, suelen verse comprometidos otros indicadores como precision y accuracy, especialmente en lo relativo a la clase minoritaria. Por lo tanto, es razonable suponer que en este caso podría estar ocurriendo lo mismo y que no se estén prediciendo los casos de quiebra y se estén clasificando como empresas solventes, mientras que los resultados permanecen muy positivos.

- Bagging

Algoritmo/Resultados	Accuracy	Precision	Recall	F1 score
Bagging RF	0.9316	0.9316	1	0.9646
Bagging SVM	0.9316	0.9316	1	0.9646
Bagging GB	0.8982	0.9321	0.9606	0.9461
Bagging Redes neuronales 3 capas	0.9316	0.9317	0.9999	0.9646
Bagging Redes neuronales 5 capas	0.9315	0.9316	0.9999	0.9645
Bagging Redes neuronales 10 capas	0.9316	0.9316	1	0.9646

Tabla 12: Resultados Bagging.

Fuente: Elaboración propia

Resultados del uso de Bagging con datos mejorados:

- Bagging Random Forest (RF):

En comparación con el modelo de Random Forest sin Bagging, se observa un ligero empeoramiento en casi todas las métricas, excepto en recall, que alcanza su valor máximo con esta técnica. Sin embargo, la disminución en accuracy, precision y F1-score impide una mejora efectiva en la detección de quiebras, que ya había demostrado ser limitada en este conjunto de datos (datos con ratios).

- Bagging SVM:

La aplicación de Bagging no produce ninguna variación significativa en las métricas de rendimiento. Este comportamiento coincide con lo observado en el estudio con datos SMOTE, donde Bagging tampoco aportó mejoras relevantes.

- Bagging Gradient Boosting (GB):

Se aprecia un empeoramiento generalizado en todas las métricas con respecto al modelo de Gradient Boosting sin Bagging. Estos resultados contrastan con los obtenidos con datos SMOTE, donde sí se había registrado una mejora. La tendencia de degradación del rendimiento refuerza la idea de una baja capacidad predictiva para los casos de quiebra en este conjunto de datos con ratios.

- Bagging Redes Neuronales (3, 5 y 10 capas):

Al igual que en el estudio con SMOTE, se agrupan los tres casos ya que los resultados y conclusiones son similares. Se observa un deterioro de las métricas principales (accuracy, precision y F1-score) en comparación con los modelos sin Bagging, aunque recall mejora ligeramente. Esto confirma que el uso de Bagging no resulta beneficioso en combinación con redes neuronales, independientemente del número de capas empleadas.

A continuación, se presenta una tabla resumen en la que se muestran los resultados obtenidos por cada uno de los algoritmos utilizados, comparando su rendimiento con los dos tipos de datos empleados en el modelo.

Datos/Algoritmo	Resultados	Random Forest	SVM	Gradient Boosting	Redes Neuronales 3 capas	Redes Neuronales 5 capas	Redes Neuronales 10 capas	Bagging RF	Bagging SVM	Bagging GB	Bagging RN (mejor caso)
Datos SMOTE	Accuracy	0.9236	0.6710	0.9136	0.7019	0.7285	0.7094	0.9903	0.6713	0.9785	0.6869
	Precision	0.9257	0.6709	0.9218	0.7160	0.7410	0.7219	0.9909	0.6711	0.9782	0.6902
	Recall	0.9634	0.9997	0.9520	0.9130	0.9102	0.9151	0.9946	0.9997	0.9899	0.9672
	F1 Score	0.9442	0.8029	0.9366	0.7978	0.7991	0.7983	0.9927	0.8031	0.9841	0.8055
	Media	0.9392	0.7861	0.9310	0.7821	0.7947	0.7861	0.9921	0.7863	0.9826	0.7874
Datos mejorados	Accuracy	0.9325	0.9316	0.9156	0.9345	0.9344	0.9338	0.9316	0.9316	0.8982	0.9316
	Precision	0.9326	0.9316	0.9348	0.9347	0.9345	0.9338	0.9316	0.9316	0.9321	0.9316
	Recall	0.9998	1	0.9775	0.9997	0.9999	1	1	1	0.9606	1
	F1 score	0.9650	0.9646	0.9557	0.9660	0.9661	0.9657	0.9646	0.9646	0.9461	0.9646
	Media	0.9574	0.9569	0.9459	0.9587	0.9587	0.9583	0.9569	0.9569	0.9342	0.9569

Tabla 13: Resumen resultados.

Fuente: Elaboración propia.

7. Plan de negocio

7.1 Resumen ejecutivo

Predictor de Quiebras Financieras mediante Machine Learning

El producto es una herramienta de software basada en machine learning diseñada para predecir la quiebra financiera de empresas. El sistema analiza tanto datos financieros históricos como actuales para detectar patrones asociados a situaciones de insolvencia.

Gracias a su capacidad de aprendizaje automático, esta solución permite a entidades del sector financiero anticiparse a riesgos, mejorar sus modelos de análisis y tomar decisiones estratégicas más seguras y fundamentadas.

A diferencia de los métodos tradicionales de análisis financiero, este software se basa en algoritmos de aprendizaje automático, lo que le permite aprender de forma continua a partir de nuevos datos y experiencias. Esta capacidad constante evolución le permite adaptarse a cambios en el entorno económico, mejorar su precisión predictiva con el tiempo y mantenerse constantemente actualizado frente a los cambios de los mercados.

Esta característica convierte al sistema en una herramienta dinámica y viva, avanzada respecto a los modelos tradicionales, cuya utilidad tiende a degradarse sin actualizaciones y mejoras frecuentes.

El producto está orientado a fondos de inversión, consultoras financieras y boutiques de inversión, que requieren herramientas avanzadas para analizar la salud financiera de compañías antes de tomar decisiones de inversión o adquisición. Aunque se trata de un mercado de nicho, está compuesto por organizaciones con alto poder adquisitivo y gran interés en soluciones tecnológicas avanzadas.

Se ofrecerá bajo un esquema de licencias por suscripción, con diferentes planes según el número de usuarios, volumen de análisis o integración con otras plataformas. Esto permite generar ingresos recurrentes y escalar el negocio progresivamente. Una de las fortalezas clave del predictor es su alta escalabilidad, tanto técnica como funcional. La arquitectura del sistema permite que el modelo se desarrolle continuamente, incorporando nuevos datos y capacidades analíticas. Además, puede integrarse con otras herramientas financieras, lo que abre la puerta a futuras funcionalidades como ofrecer el porcentaje de riesgo de quiebra. Esta escalabilidad asegura que el producto no se limite a un único propósito, sino que evolucione hacia una solución integral de análisis financiero basada en inteligencia artificial.

Nuestra solución combina tecnología avanzada con una propuesta altamente especializada. Existen herramientas de scoring financiero, pero pocas integran machine learning con capacidad de autoaprendizaje y actualización continua. Esta capacidad incrementa la precisión y confiabilidad del sistema con el uso, generando un valor

agregado significativo. Si bien existen en el mercado diversas herramientas que calculan el riesgo financiero y ofrecen scoring empresarial, la mayoría de ellas se basan en modelos estáticos y límites predefinidos. Nuestra solución se diferencia al incorporar machine learning, lo que la convierte en un producto inteligente, adaptable y en constante evolución.

Esta capacidad de autoaprendizaje permite al sistema ajustarse a cambios económicos, sectores emergentes y nuevas variables financieras, mejorando continuamente su precisión y utilidad. A diferencia de los competidores, que dependen de actualizaciones y mejoras de los modelos, esta herramienta se mantiene siempre actualizada y alineada con las condiciones reales del mercado.

Esto representa una ventaja competitiva clave para usuarios que requieren información precisa y actual para tomar decisiones críticas de inversión o gestión de riesgo. Considerando que el público principal al que está dirigido este producto son fondos de inversión, consultoras financieras y otras entidades dedicadas a la toma de decisiones de alto impacto económico, contar con una herramienta precisa y actualizada representa un valor estratégico clave.

En este tipo de organizaciones, una decisión basada en información incompleta o desactualizada puede traducirse en pérdidas millonarias o riesgos financieros significativos. Por eso, disponer de un sistema inteligente capaz de analizar múltiples parámetros y aprender continuamente del comportamiento del mercado no solo mejora la calidad del análisis, sino que puede marcar la diferencia entre una inversión exitosa y un fracaso financiero.

Esta necesidad crítica de contar con información confiable, actualizada y automatizada es lo que posiciona a nuestra solución como una herramienta indispensable dentro del ecosistema de análisis financiero moderno.

El proyecto se encuentra en una fase inicial de desarrollo. Ya se ha construido un prototipo funcional que ha demostrado la viabilidad técnica del modelo de predicción. Este primer desarrollo valida la capacidad del sistema para analizar datos financieros y generar predicciones con un nivel aceptable de precisión.

Actualmente, el enfoque está en validar la propuesta de valor con clientes potenciales, incorporar datasets más amplios y diversos para mejorar la robustez del modelo y avanzar hacia el desarrollo de una versión comercial más completa y escalable.

7.2 Definición del plan de negocio

El negocio se basa en el desarrollo de un software predictivo de quiebras financieras, impulsado por aprendizaje automático. Esta herramienta analiza grandes volúmenes de datos financieros para anticipar posibles situaciones de insolvencia empresarial antes de que ocurran, gracias a su capacidad de autoaprendizaje y mejora continua esta predicción es cada vez más precisa y fiable.

El producto está dirigido principalmente a consultoras financieras, fondos de inversión y boutiques de análisis, que necesitan información precisa y actualizada para minimizar riesgos y optimizar decisiones estratégicas. La principal ventaja competitiva del sistema reside en su fiabilidad, adaptabilidad y evolución constante, elementos que lo distinguen de otras soluciones tradicionales, más estáticas y menos precisas.

Este desarrollo surge en un contexto clave: la inteligencia artificial está transformando múltiples sectores y el financiero no es la excepción. Incorporar tecnologías como machine learning en el análisis de riesgo empresarial representa una oportunidad estratégica que puede traducirse en decisiones más acertadas y diferencias económicas significativas. Al igual que en sectores como la medicina, donde instituciones como el CSIC destacan: “La combinación de inteligencia artificial y tecnologías de aprendizaje automático permite augurar un futuro con diagnósticos más fiables y personalizados para tratar con mayor eficacia enfermedades como el cáncer, el alzhéimer y los ictus cerebrales”. [34]. El sector financiero puede beneficiarse enormemente de su aplicación inteligente y responsable.

7.3 Análisis del mercado

Según un informe de Market Research Future (MRFR) [35], este mercado fue valorado en 5.71 mil millones de dólares (para 2023) y se proyecta un crecimiento sostenido con una tasa anual compuesta (CAGR) del 8.61% entre 2024 y 2032. Aunque otras fuentes presentan estimaciones algo más conservadoras en cuanto al tamaño total del mercado, todas coinciden en un punto clave: se trata de un mercado en expansión, con una tasa de crecimiento estable cercana al 9%. Estos datos reflejan no solo el tamaño significativo del mercado, sino también su atractivo estratégico para nuevas soluciones tecnológicas.

Esta tendencia se ve reforzada por datos reales del entorno económico, que se exponen a continuación:



Figura 28: Casos de quiebra España & UE

Fuente: Trading Economics, INE

En España los casos de quiebra empresarial muestran una clara tendencia ascendente en la última década: los registros han pasado de 3.020 en enero de 2016 a 3.910 en enero de 2025. En la Unión Europea se observa un patrón similar, con unas tasas de crecimiento positivas en los casos de insolvencia desde 2021 hasta 2025 especialmente.

Este aumento de quiebras a nivel nacional y europeo subraya la necesidad creciente de herramientas que permitan anticipar estos eventos, mejorar la gestión del riesgo y evitar pérdidas millonarias. Cuanto mayor es el impacto del problema, mayor será la demanda de soluciones predictivas avanzadas, como el software propuesto, lo que refuerza la oportunidad comercial en este nicho específico.

Dentro del mercado en crecimiento del software predictivo financiero, este producto está dirigido a los principales actores del sector financiero, en particular: fondos de inversión, consultoras financieras, boutiques especializadas en análisis de riesgo...

Inicialmente, el enfoque se centrará en consultoras y empresas financieras de pequeño y mediano tamaño, que ya utilizan herramientas de análisis, pero enfrentan limitaciones en cuanto a precisión, actualización y personalización. Estas organizaciones suelen tener interés en incorporar tecnologías avanzadas como la inteligencia artificial, pero carecen de los recursos técnicos o financieros que poseen las grandes firmas.

Este grupo representa a los early adopters ideales del producto, ya que están altamente motivados por la necesidad de mejorar su competitividad y acceder a soluciones tecnológicas de vanguardia. Mientras que las grandes entidades pueden desarrollar internamente herramientas avanzadas, las firmas más pequeñas ven en este software una

oportunidad de nivelar el terreno de juego, accediendo a una herramienta potente, basada en aprendizaje automático, que mejora continuamente su rendimiento.

De este modo, el producto no solo cubre una necesidad técnica, sino que democratiza el acceso a soluciones predictivas de alta calidad, permitiendo que empresas medianas puedan competir en igualdad de condiciones con actores de mayor escala.

Por lo tanto, lo que se ofrece al cliente es una solución tecnológica accesible y de alto rendimiento, que le permite competir en igualdad de condiciones con grandes compañías, gracias al uso de inteligencia artificial aplicada a la predicción de quiebras.

A través de este producto, empresas con recursos limitados pueden acceder a niveles avanzados de precisión, fiabilidad y automatización, que de otro modo estarían fuera de su alcance. La incorporación de machine learning hace posible un sistema que aprende continuamente, se adapta al entorno financiero y proporciona predicciones actualizadas y fiables.

En definitiva, el software se convierte en una herramienta esencial para el análisis de riesgo, alineada con las demandas del mercado actual y capaz de aportar un valor diferencial real en la toma de decisiones estratégicas.

La creciente demanda de herramientas de análisis de riesgo financiero ha impulsado la aparición de numerosos competidores en el sector, muchos de ellos desarrollados por grandes multinacionales. Entre los más destacados se encuentran plataformas como SAP Risk Management o IBM OpenPages GRC Platform, ambas enfocadas en la evaluación integral de riesgos empresariales. [36]

IBM, por ejemplo, ha comenzado a incorporar inteligencia artificial en sus soluciones de gestión de riesgos, marcando una clara señal de hacia dónde se dirige el mercado. Además, existen múltiples plataformas que utilizan modelos de scoring financiero para medir el riesgo de quiebra y otros indicadores de solvencia empresarial.[37].

Sin embargo, la mayoría de estos sistemas continúan basándose en modelos tradicionales, con escasa capacidad de adaptación y evolución. Es aquí donde nuestro producto representa un salto cualitativo en el sector: al integrar machine learning como núcleo del sistema, ofrece una solución dinámica y constantemente actualizada, que responde a las nuevas exigencias del entorno financiero.

Esta capacidad de autoaprendizaje y mejora continua constituye el valor diferencial clave frente a los competidores existentes, anticipándose a una tendencia que otras compañías aún están comenzando a explorar. Así, nuestro software no solo responde a la necesidad actual del mercado, sino que marca el camino hacia la nueva generación de herramientas de análisis de riesgo.

El ingreso al mercado de software para análisis de riesgo financiero presenta diversas barreras de entrada relevantes, que deben ser consideradas estratégicamente para garantizar el éxito del producto:

1. Competencia consolidada

Una de las principales barreras es la presencia de grandes multinacionales que ya operan en el sector con productos consolidados, como SAP o IBM, que ofrecen soluciones avanzadas de análisis de riesgo financiero. Estas empresas cuentan con marcas

reconocidas, recursos técnicos amplios y relaciones sólidas con clientes del sector, lo que dificulta el posicionamiento inicial de nuevos actores.

2. Confianza del cliente

Dado que las decisiones que se toman a partir de estos sistemas pueden tener consecuencias económicas millonarias, la credibilidad del producto es un factor decisivo para su adopción. EL software, al estar en una fase temprana de desarrollo, aún requiere ser validado en entornos reales. Por ello, uno de los retos más importantes será demostrar su fiabilidad, precisión y eficacia desde el primer momento, incluso siendo una solución nueva en el mercado.

La construcción de una imagen de marca sólida y profesional, así como la generación de casos de éxito en clientes piloto, serán elementos fundamentales para superar esta barrera.

3. Desarrollo técnico

Otra barrera crítica es el grado de desarrollo técnico necesario para competir al nivel de los actores ya establecidos. Actualmente, el producto se encuentra en fase inicial, por lo que es imprescindible completar su desarrollo funcional y ajustarlo a las necesidades específicas del mercado, garantizando que cumpla con todos los requisitos técnicos, normativos y de usabilidad.

Lograr un producto robusto, preciso, adaptable y fácil de integrar será determinante para que pueda competir “de tú a tú” con las grandes soluciones existentes, sin que se cuestione su fiabilidad ni su capacidad tecnológica.

Si bien el mercado del análisis de riesgo financiero es un sector establecido y regulado, no se prevén barreras legales significativas para la entrada del producto en términos generales. Sin embargo, el uso de inteligencia artificial en la toma de decisiones financieras puede plantear desafíos legales, éticos y de responsabilidad. Será esencial garantizar que el sistema cumpla con las normativas vigentes y que exista una política clara sobre los límites del uso automatizado en procesos críticos de decisión, acompañada de mecanismos que refuercen estos aspectos

Para hacer frente a las principales barreras identificadas se plantean las siguientes estrategias:

- **Enfocar la entrada en compañías pequeñas y medianas:**
Como se ha mencionado previamente, el producto se orientará inicialmente a consultoras financieras y empresas de tamaño pequeño y medio. Este enfoque permitirá demostrar la fiabilidad, eficacia y competitividad de la solución en un entorno real, generando así casos de éxito y construyendo una reputación sólida que facilite el acceso a clientes de mayor tamaño en fases posteriores.
- **Desarrollar al máximo el potencial técnico del producto:**
Se priorizará un desarrollo técnico robusto y completo que elimine cualquier posible duda sobre la calidad y funcionalidad del software. La incorporación de machine learning no solo debe ser un elemento distintivo, sino una ventaja técnica clara frente a las soluciones tradicionales del mercado.

7.4 Plan de marketing

Habiendo definido previamente los clientes objetivo y los early adopters, se reconoce que este producto está orientado inicialmente a empresas españolas y posteriormente europeas del sector financiero, como consultoras, fondos de inversión y boutiques que exigen herramientas de alta precisión, fiabilidad y usabilidad. Estos clientes buscan soluciones que les permitan tomar decisiones estratégicas bien fundamentadas y operar con la máxima eficiencia, simplificando procesos sin renunciar a rigor técnico.

Aunque muchas de estas empresas ya emplean herramientas de scoring financiero y análisis de riesgos, el producto que presentamos representa un salto cualitativo, gracias al uso de tecnología de última generación basada en inteligencia artificial y aprendizaje automático.

En el mercado actual, encontramos tanto grandes multinacionales como actores emergentes que ofrecen soluciones similares. Por un lado, firmas como SAP integran herramientas de gestión de riesgos dentro de sistemas ERP robustos, lo que limita su accesibilidad para empresas pequeñas o medianas. Por otro lado, startups y proveedores más pequeños ofrecen servicios útiles, pero no especializados en la detección predictiva de quiebras financieras, como es el caso de la Startup española Embat.[38].

El enfoque de este producto es más directo, especializado y accesible. Se ofrece un software diseñado exclusivamente para la predicción de quiebras, dirigido específicamente a empresas de menor tamaño, que reconocen la importancia de estas herramientas, pero no cuentan con los recursos para adoptar soluciones más complejas o costosas.

Por lo tanto, lo que más valorarán los clientes es:

- Accesibilidad económica y técnica: una herramienta diseñada para empresas que no pueden asumir grandes inversiones.
- Especialización funcional: un software enfocado exclusivamente en la predicción de quiebras, sin funcionalidades genéricas innecesarias.
- Innovación constante: gracias al uso de IA, el sistema se actualiza y mejora de forma autónoma, eliminando la necesidad de costosos mantenimientos o desarrollos adicionales.

Es por ello que este producto transmite un mensaje claro de fiabilidad, innovación y adaptación constante, facilitando el acceso a tecnología avanzada para todo tipo de empresas del sector financiero.

Dado que se trata de un producto altamente especializado, dirigido a un público profesional y técnico, se plantea una estrategia de marketing enfocada en la segmentación y selección precisa, mediante:

1. Participación en eventos financieros y tecnológicos

La presencia en congresos, ferias y eventos del sector financiero permitirá dar a conocer el producto a una audiencia profesional, así como establecer relaciones directas con potenciales clientes y obtener visibilidad entre actores clave.

2. Contenido especializado y marketing personalizado

Mediante la creación de contenidos especializados, resulta más sencillo destacar el valor diferencial del producto frente a los clientes clave previamente identificados. Este enfoque permite desarrollar una estrategia de marketing altamente personalizada, adaptada a las necesidades y características a cada cliente potencial. A través de casos de uso concretos y demostraciones, se puede mostrar de forma directa la utilidad, eficacia y aplicabilidad real de la solución en contextos concretos.

3. Colaboraciones estratégicas

Con las dos acciones anteriores como potenciadores y facilitadores, se buscará establecer alianzas y colaboraciones estratégicas con compañías del sector que se hayan detectado como perfectas para el uso de este producto. A su vez, se destaca que como se ha expuesto se llevarán a cabo demostraciones y ofertas para atraer a los clientes.

Para fijar el precio del producto se tendrá en cuenta el precio de los softwares financieros que se ofrecen hoy día. A su vez para determinar el modelo de pago se tiene en cuenta como se realiza de forma habitual en el sector. Es por ello que se escoge un modelo de suscripción mensual/anual en el que se ofrece todas las funcionalidades del programa al cliente que cuente con un usuario. Esto es algo común en diferentes herramientas financieras como puede ser Bloomberg o Morningstar Investor [39] el cual ofrece una membresía anual o mensual. Mediante este sistema se logra que los clientes puedan utilizar la herramienta de mes a mes según sus necesidades a la vez que contar con una versión anual con la que obtendrían un descuento. El precio que pagar por la suscripción anual se determinará en los siguientes apartados financieros.

Para fijar el precio del producto, se tendrá en cuenta la estructura de precios vigente en el mercado de software financiero, así como las prácticas habituales del sector. En línea con esto, se ha optado por implementar un modelo de suscripción mensual y/o anual, que permite al cliente acceder a todas las funcionalidades del programa mediante una licencia de usuario.

Este modelo es ampliamente utilizado por herramientas consolidadas como Bloomberg o Morningstar Investor [39], que ofrecen acceso completo bajo licencia. Este modelo permite que los clientes adapten el uso de la herramienta a sus necesidades operativas, pagando de forma flexible mes a mes, o beneficiándose de un descuento por compromiso anual.

El precio definitivo de la suscripción anual se detallará en el apartado financiero del proyecto, una vez completado el análisis de costes, competencia y posicionamiento estratégico del producto.

Para la implementación de estas acciones, se ha diseñado el siguiente cronograma de trabajo:

Plazo / Acciones	Primer año (desarrollo)	Primer año (lanzamiento)	Tres años	Cinco años
Desarrollo prototipo final				
Desarrollar campana marketing inicial				
Desarrollo del equipo e infraestructura inicial				
Disponer de una base inicial de clientes, así como de un equipo y una infraestructura técnica consolidados				
Triplicar número de clientes respecto al inicio				
Mantener tasa de crecimiento				
Alcanzar la solvencia				

Tabla 14: Cronograma.

Fuente: Elaboración propia

Para evaluar el cumplimiento de los objetivos y los plazos definidos en el cronograma, se han establecido una serie de indicadores clave de rendimiento (Key Performance Indicators, KPIs). Estos permitirán hacer un seguimiento riguroso del progreso del proyecto en sus distintas fases.

Fase previa al lanzamiento del producto

- Finalizar el desarrollo completo del programa, con todas las funcionalidades activas, en un plazo máximo de 12 meses.
- Participar en los principales eventos financieros del sector durante el proceso de desarrollo.
- Realizar al menos 100 demostraciones del producto a potenciales clientes antes del lanzamiento oficial.
- Establecer un mínimo de 20 alianzas o colaboraciones estratégicas con empresas del sector antes de la comercialización.

Fase posterior al lanzamiento

- Retener al 75% de los clientes a los que se les realizó una demostración y con quienes se estableció una colaboración previa.
- Alcanzar un crecimiento mensual de clientes del 5% a lo largo del tiempo.
- Obtener un mínimo de 20 clientes en el primer año de comercialización.
- Alcanzar el beneficio mensual antes del quinto año.
- Lograr un total de clientes activos mensuales superior a 70 para el tercer año.
- Mantener una tasa de retención mensual de clientes del 80 % durante el mismo periodo.
- Mantener los costes variables con una tasa asociada al crecimiento de la empresa.
- Posicionar la solución como una de las plataformas líderes emergentes en predicción de quiebras para el quinto año.
- Conseguir presencia activa en al menos una gran entidad financiera, banco o consultora internacional, para el quinto año.
- Se espera posicionar el producto como una solución confiable y de alta calidad en el mercado. Este objetivo se medirá a través de valoraciones en plataformas especializadas, reseñas y artículos del sector, con una calificación promedio superior a 4 sobre 5.
- Reputación impecable entre los clientes y en el entorno profesional. La percepción directa del cliente será clave. Para ello, se implementarán encuestas de satisfacción post-uso, cuyo resultado esperado será una calificación media superior a 4 sobre 5.

7.5 Business model canvas

En este apartado se presenta el Business Model Canvas desarrollado para ofrecer una visión detallada y estructurada del producto propuesto en este proyecto. A continuación, se explicará en profundidad cada uno de los bloques que lo componen, con el objetivo de facilitar la comprensión del modelo de negocio y su aplicabilidad en el mercado.

1. Socios clave (Key partners):

Los socios clave necesarios para que el producto pueda lanzarse al mercado con éxito y consolidarse en él son los siguientes:

- **Equipo de desarrollo informático:**
Es fundamental contar con un equipo de especialistas en software financiero responsables del diseño, desarrollo y mantenimiento del producto. Este equipo no solo debe crear una versión final robusta y funcional, sino también encargarse del soporte técnico y la resolución de cualquier incidencia que pueda surgir durante su uso.
- **Especialistas financieros:**
Profesionales con experiencia en el sector financiero que aporten el conocimiento necesario para validar el funcionamiento del software. Su rol es garantizar que la herramienta esté alineada con las necesidades del mercado y cumpla con los estándares del análisis financiero y la predicción de riesgos.
- **Empresas usuarias:**
Las propias empresas, como usuarias finales, también se consideran socios clave. Son ellas quienes validan el producto, aportan retroalimentación esencial para su mejora y actúan como promotoras del mismo en su entorno, contribuyendo a su posicionamiento y consolidación en el mercado.
- **Gestores de bases de datos y servicios en la nube:**
Serán fundamentales para garantizar el funcionamiento óptimo del producto, así como para la adecuada organización y almacenamiento de la información. Al tratarse de un producto completamente digital y operado en línea, contar con soluciones eficientes de almacenamiento en la nube resulta esencial.
- **Colaboraciones estratégicas:**
En este apartado se incluyen aceleradoras, incubadoras y otras entidades que puedan impulsar el desarrollo del proyecto emprendedor. También se consideran aquellas empresas que puedan aportar datos valiosos o facilitar el posicionamiento y crecimiento del producto en el mercado.

2. Actividades clave (Key activities):

Las actividades fundamentales para alcanzar la propuesta de valor planteada son las siguientes:

- **Análisis del mercado y definición del producto:**
Antes del lanzamiento al mercado, es imprescindible realizar un estudio exhaustivo tanto del sector como del producto a desarrollar. Es necesario identificar las necesidades reales de los potenciales clientes y comprender en profundidad el entorno en el que operará la herramienta. Esto permitirá diseñar un producto alineado con dichas necesidades, capaz de ofrecer una solución efectiva y diferenciada.
- **Desarrollo de la versión final del producto:**
A partir del prototipo inicial, es fundamental avanzar hacia una versión final optimizada que cumpla con todos los requisitos funcionales y técnicos identificados durante el análisis previo. Este proceso debe contemplar una mejora continua, así como fases de pruebas, validaciones y ajustes que garanticen tanto su viabilidad técnica como su potencial comercial.
Además, será indispensable contar con una infraestructura digital sólida y una base de datos robusta, que permitan al software operar de manera óptima. Esto incluye disponer de todas las herramientas técnicas necesarias para asegurar un desempeño eficiente, fiable y alineado con los estándares del sector.
- **Estrategia de comercialización y mantenimiento para evolución constante:**
Para lanzar el producto al mercado será imprescindible ejecutar una estrategia de marketing sólida, orientada a generar visibilidad y captar clientes. Esta estrategia combinará:
 - Participación en eventos del ámbito tecnológico y financiero, donde se presentará la herramienta y se establecerán alianzas estratégicas.
 - Marketing digital especializado, que incluirá campañas en redes sociales, email marketing segmentado y colaboraciones con medios del sector.

El objetivo es posicionar el producto como una solución innovadora y de alto valor añadido, fomentando su adopción entre los potenciales usuarios.

Paralelamente, será crucial garantizar un mantenimiento técnico eficiente, que asegure la evolución continua del software y una experiencia óptima para el cliente. Este enfoque permitirá solucionar incidencias con rapidez, mantener el sistema actualizado y facilitar su aprendizaje constante, optimizando así su rendimiento a lo largo del tiempo.

3. Recursos clave (Key resources):

El recurso central del proyecto es el software informático: un predictor de quiebras financieras basado en inteligencia artificial. Este sistema se sustenta en técnicas

de machine learning, las cuales permiten analizar datos financieros y generar predicciones precisas sobre la viabilidad económica de las empresas.

Para su correcto funcionamiento, son necesarios varios recursos complementarios:

- Infraestructura tecnológica:
 - Una plataforma digital (portal web) desde la cual los clientes podrán informarse sobre el producto, adquirir licencias de uso, registrarse e interactuar con el equipo de soporte. Esta plataforma también facilitará el acceso al software, permitiendo su utilización en línea tras la correspondiente configuración. Además, contará con una base de datos integrada que almacenará y gestionará toda la información con la que opera el sistema, asegurando su disponibilidad mediante soluciones en la nube.
- Recursos humanos especializados:

Se requiere un equipo multidisciplinar compuesto por:

- Desarrolladores informáticos, encargados de crear y optimizar la versión final del producto.
- Especialistas financieros y de regulación, que validen la lógica del sistema y garanticen que las funcionalidades estén alineadas con las necesidades y requerimientos del sector.
- Equipo de soporte técnico y atención al cliente, que asegure el mantenimiento del software y una experiencia de usuario eficiente y satisfactoria.

Estos recursos son fundamentales para garantizar la operatividad del producto, su evolución constante y una relación óptima con los clientes.

4. Propuesta de valor (value proposition):

Se trata de un producto innovador y diferencial que permite a empresas de todos los tamaños afrontar una necesidad crítica dentro de un sector tan específico como el financiero, concretamente en el área de análisis de riesgos y predicción de quiebras empresariales.

El valor del producto radica en varios aspectos clave:

- Accesibilidad: Su diseño está orientado a que cualquier empresa, independientemente de su tamaño o capacidad económica, pueda acceder a esta herramienta gracias a un modelo de precios competitivo.

- **Tecnología avanzada:** Su principal elemento diferenciador es el uso de inteligencia artificial, concretamente técnicas de machine learning, que lo posicionan por encima de las herramientas tradicionales de análisis financiero. Esta tecnología permite identificar patrones complejos, ofrecer predicciones más precisas y mejorar su rendimiento de forma continua a medida que se alimenta de nuevos datos.
- **Adaptabilidad y evolución constante:** Al estar basado en aprendizaje automático, el software no solo ofrece un alto nivel de precisión en sus predicciones, sino que también evoluciona de forma dinámica según los cambios del entorno económico y financiero, manteniendo su eficacia y utilidad a lo largo del tiempo.

En conjunto, esta propuesta de valor convierte al producto en una solución eficaz, moderna y escalable para empresas que buscan herramientas tecnológicas que les permitan anticiparse a situaciones críticas como la quiebra financiera.

5. Relación con clientes (Customer relationships):

La relación con los clientes será completamente digital, en línea con la naturaleza del producto ofrecido. El modelo de atención estará basado en dos niveles:

- **Soporte técnico especializado:** Se contará con un equipo de profesionales informáticos capacitados para atender y resolver de forma eficiente los problemas técnicos y dudas más relevantes que puedan surgir durante el uso del producto.
- **Asistencia automatizada mediante inteligencia artificial:** Como parte del valor diferencial del producto, se implementará un sistema de atención al cliente basado en IA, que actuará como primera línea de contacto. Este sistema, a través de herramientas como chatbots, ofrecerá asistencia inmediata y disponible las 24 horas, permitiendo a los usuarios resolver consultas básicas de manera ágil y eficiente.

Esta combinación entre soporte humano e inteligencia artificial garantiza una experiencia de atención integral, accesible y escalable, que se adapta a las necesidades de los clientes y refuerza la propuesta de valor del producto.

6. Canales (channels):

Al tratarse de un producto digital, su adquisición y uso serán completamente online. Los clientes podrán adquirirlo a través de una plataforma web, de forma similar a como se comercializan la mayoría de los softwares financieros. Una vez completada la compra, el producto estará disponible para su uso inmediato, previa configuración y registro por parte del cliente.

El acceso al mercado y la captación de clientes se llevará a cabo principalmente a través de los siguientes canales:

- Participación en eventos tecnológicos y financieros, donde se podrá presentar el producto a potenciales clientes y establecer alianzas estratégicas.
- Marketing especializado digital, que incluirá acciones dirigidas como campañas en redes sociales, email marketing y otras herramientas de captación online enfocadas al público objetivo.
- Colaboraciones estratégicas: colaboraciones con empresas que permitan aumentar la visibilidad y ayuden al posicionamiento del producto.

Este enfoque permite una distribución ágil, escalable y adaptada al perfil de los clientes potenciales.

7. Segmentos de clientes (Customer segments):

El producto está inicialmente enfocado al sector financiero, con especial atención a España y posteriormente a la Unión Europea. Dada su propuesta de valor y la naturaleza del servicio ofrecido, se trata de una solución dirigida específicamente al análisis de riesgos y la predicción de quiebras empresariales cuyo objetivo principal es ser accesible para todas aquellas empresas que necesiten este tipo de herramienta, especialmente pequeñas y medianas empresas con limitaciones económicas, que suelen carecer de recursos suficientes para acceder a soluciones financieras avanzadas o costosas.

Los early adopters más probables serán fondos de inversión, consultoras financieras y boutiques especializadas en análisis de riesgo de pequeño y mediano tamaño, quienes pueden beneficiarse inmediatamente del valor que ofrece esta herramienta.

A medida que el producto demuestre su eficacia y fiabilidad, el espectro de clientes podría ampliarse hacia grandes empresas del sector financiero que también requieran herramientas de predicción avanzadas. No obstante, en su etapa inicial, el enfoque estará centrado en aquellas organizaciones con mayores barreras de acceso a soluciones similares, proporcionando así un valor diferencial y democratizando el acceso a tecnología de análisis financiero avanzada.

8. Estructura de costes (Cost structure):

Los costes asociados al desarrollo y lanzamiento del producto se pueden dividir en dos fases diferenciadas:

Primera fase – Desarrollo del producto:

En esta etapa se centra en la creación de la versión final del software. Los principales costes incluyen:

- Equipo de desarrolladores informáticos especializados, responsables de implementar todas las funcionalidades necesarias y asegurar que el producto cumple con los requisitos técnicos y operativos.
- Especialistas financieros y de regulación, encargados de validar el correcto funcionamiento del software desde el punto de vista del análisis financiero, garantizando que los resultados del modelo sean fiables y adecuados al mercado, cumpliendo con todas las normativas.

Segunda fase – Comercialización y mantenimiento:

Una vez que el producto se encuentra desarrollado y listo para su lanzamiento comercial, los costes principales son:

- Equipo de soporte y mantenimiento técnico, que se encargará de asegurar el funcionamiento continuo y óptimo del software, así como de atender incidencias y resolver las dudas de los usuarios.
- Marketing y adquisición de clientes, especialmente relevante en la etapa inicial de introducción en el mercado. Se prevé una inversión significativa en campañas de posicionamiento, participación en eventos tecnológicos y financieros, marketing especializado dirigido a empresas del sector y colaboraciones estratégicas que ayuden a validar y promocionar el producto.
- Infraestructura digital y ciberseguridad: Uno de los gastos más relevantes será la infraestructura digital necesaria para el desarrollo, implementación y correcto funcionamiento del programa. Esto incluye los costes asociados al alojamiento en la nube, la gestión de bases de datos y, especialmente, la ciberseguridad, que es fundamental para garantizar la protección de la información sensible. Estos elementos representan una inversión significativa dentro del proyecto.

Esta estructura de costes refleja las necesidades clave para asegurar tanto el éxito técnico del software como su penetración efectiva en el mercado objetivo.

9. Fuente de ingresos (Revenue streams):

La principal fuente de ingresos del producto será el modelo de suscripción. Los clientes pagarán una cuota por el uso del software financiero durante un período de tiempo determinado. Es decir, los ingresos se generarán a través de la venta de licencias que permitirán a las empresas acceder y utilizar la herramienta.

Se ofrecerán diferentes planes de suscripción en función de la duración del servicio (mensual o anual), así como posibles descuentos, ofertas o promociones adaptadas a distintos perfiles de clientes o campañas comerciales específicas.

Este enfoque permite una alta flexibilidad para adaptarse a las necesidades de distintos tipos de empresas, favoreciendo tanto la accesibilidad como la fidelización del cliente. En concreto, se trata de un modelo de suscripción mediante licencias escalables.

Además, al contar con un portal web, se ofrece la posibilidad de acceder a una versión gratuita de prueba, permitiendo a los usuarios explorar las funcionalidades de la herramienta y familiarizarse con su uso antes de adquirir la licencia completa.

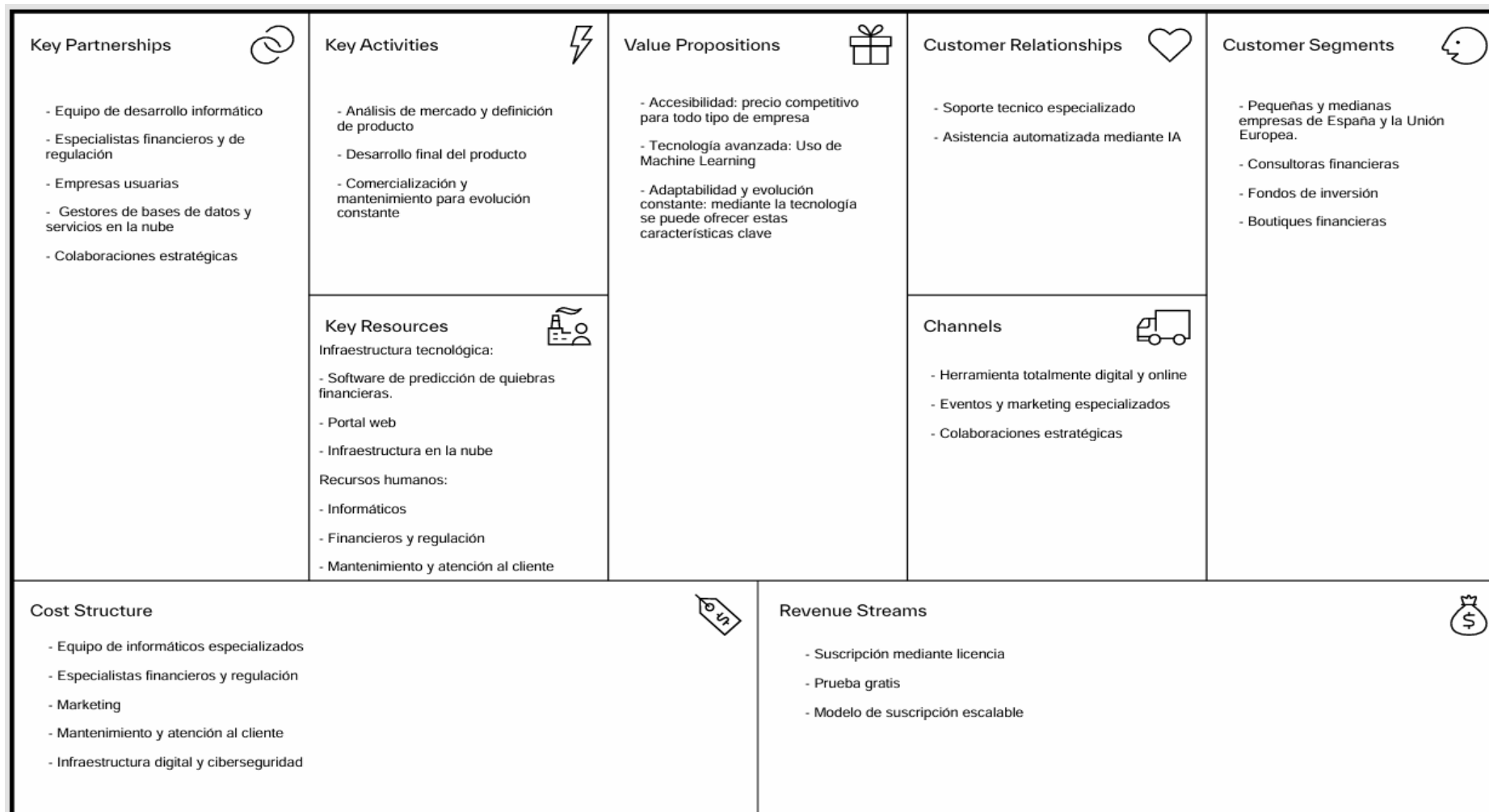


Figura 29: Business Model Canvas.

Fuente: Elaboración propia.

7.6 Hipótesis de desarrollo del producto

Para el desarrollo del producto presentado en este proyecto, se llevó a cabo un proceso de análisis y reflexión en el que se identificó un problema real existente en el sector financiero. A partir de esta problemática, definida en el apartado anterior, se formularon una serie de hipótesis que permitieron definir y acotar una posible solución, así como el enfoque adecuado para implementarla.

Una vez consolidada la idea, se determinó que la mejor forma de validarla era a través del desarrollo del modelo técnico previamente expuesto, el cual demuestra la viabilidad técnica de la propuesta. Por otro lado, su utilidad y aplicabilidad se han argumentado en el marco teórico, mientras que el plan de negocio incluye un análisis financiero que estudia la viabilidad económica del proyecto.

El proceso seguido se puede resumir en los siguientes pasos:

Se identifica un problema concreto que afecta a uno de los sectores más relevantes para la sociedad: el sector financiero. Existe un elevado número de empresas que enfrentan situaciones de insolvencia, lo cual genera un impacto negativo significativo no solo a nivel económico, con pérdidas millonarias, sino también a nivel social, afectando el empleo, la estabilidad del mercado y la confianza en el sector financiero.

Ante esta realidad, se detecta la necesidad de buscar una solución que permita mitigar estas consecuencias. El objetivo es contribuir a reducir el número de insolvencias o, al menos, anticiparlas para minimizar sus efectos. A partir de esta identificación, se desarrolla el Business Model Canvas para definir el plan a seguir, ayudando a plantear una serie de hipótesis orientadas a desarrollar una herramienta eficaz que permita abordar este problema.

1. Las empresas desean mantener su solvencia, y la insolvencia representa un desafío constante en su gestión diaria.
2. La amenaza de quiebra afecta a empresas de todos los tamaños, sin importar su escala.
3. La quiebra constituye uno de los principales riesgos a los que se enfrenta cualquier organización.
4. La mayoría de las empresas no dispone de herramientas adecuadas para evaluar el riesgo de insolvencia.
5. Evaluar con precisión la probabilidad de quiebra resulta complejo para cualquier empresa.
6. Actualmente, predecir una quiebra implica costes elevados y recursos especializados.

7. Mientras que las grandes empresas disponen de sistemas avanzados de análisis de riesgos, las pequeñas y medianas no tienen acceso a herramientas similares.
8. Contar con herramientas que permitan anticipar situaciones financieras adversas resulta claramente beneficioso para las empresas.
9. La capacidad de predecir quiebras permitiría a las empresas más pequeñas competir en mejores condiciones.
10. Anticipar escenarios financieros críticos refuerza la toma de decisiones y la estabilidad empresarial.
11. En España se registra anualmente un elevado número de casos de quiebra empresarial.
12. La disponibilidad de herramientas adecuadas podría reducir significativamente este volumen de quiebras.
13. Existe un alto interés por parte de las empresas en acceder a herramientas que permitan prever situaciones de insolvencia.
14. Si se ofreciera una herramienta eficaz y a un precio competitivo, las pequeñas y medianas empresas mostrarían gran interés en adoptarla.
15. Es viable desarrollar una solución accesible para todo tipo de empresas, manteniendo altos estándares de eficiencia y rendimiento.
16. El coste es una de las principales barreras para la adopción de este tipo de productos por parte de las pymes.
17. Esta herramienta puede desarrollarse completamente en formato digital y operar de manera online.
18. Las herramientas tradicionales de predicción de quiebras están quedando obsoletas frente a los avances tecnológicos.
19. Es urgente impulsar la modernización tecnológica en un sector tan estratégico como el financiero.
20. La incorporación de inteligencia artificial en este tipo de soluciones representa un avance significativo.
21. El uso de técnicas de machine learning es una opción óptima para mejorar la predicción de quiebras.
22. La capacidad de anticipar insolvencias genera beneficios tanto para el tejido empresarial como para la sociedad en su conjunto.

23. Hay numerosos actores de la sociedad interesados en esta idea.
24. Los elevados costes y el nivel de perfeccionamiento técnico y profesional requerido para este tipo de productos pueden representar la principal barrera para su desarrollo.
25. La aplicación de inteligencia artificial puede constituir la ventaja clave para superar estas barreras y facilitar su implementación.

Por lo tanto, a través del Business Model Canvas y las hipótesis presentadas, se considera que el plan de negocio queda estructurado. Esto permite identificar las principales características del producto a desarrollar, así como reconocer las hipótesis clave derivadas del análisis, las cuales deben ser abordadas y validadas para su correcta implementación.

Mientras que, para la validación del proyecto, como se ha mencionado previamente, se optó por abordarla desde una perspectiva técnica, en lugar de realizar entrevistas o encuestas con profesionales del sector, se decidió incorporar este enfoque en el marco teórico. En dicho apartado se justifica la utilidad y el valor diferencial que ofrece el producto.

Dado que se trata de un proyecto de emprendimiento con un fuerte componente ingenieril, se consideró fundamental el desarrollo de un primer prototipo funcional. Este prototipo permite mostrar las funcionalidades clave que podría incorporar la herramienta, así como realizar un estudio preliminar que aporta resultados concretos. A partir de estos resultados, es posible extraer conclusiones relevantes desde una perspectiva técnica que respaldan la viabilidad y el potencial del producto.

7.7 Plan financiero

Se ha elaborado un plan financiero inicial en el que se detallan los ingresos y costes estimados para la etapa inicial del proyecto. Este plan incluye una estimación de la inversión necesaria para desarrollar la versión final del producto, contemplando tanto los costes fijos como los variables que se prevé asumir.

Posteriormente, se realiza una proyección de ingresos basada en el número estimado de clientes que utilizarán el producto, así como en el precio que se establecerá para su comercialización. A partir de esta información, se construye un flujo de caja que permitirá evaluar la liquidez del proyecto y determinar, en caso de existir, el punto de equilibrio a partir del cual el emprendimiento comenzará a generar beneficios.

Para este análisis financiero se establece un horizonte temporal de cinco años, con el objetivo de ofrecer una visión realista y estratégica sobre la viabilidad económica del proyecto a medio plazo.

- Salario de un Machine Learning Engineer: alrededor de 40.000 € anuales en España.[40]. El equipo con el que se pretende contar estará formado por dos ingenieros informáticos especializados en machine learning, responsables del desarrollo de la versión final del producto, así como del portal web donde se integrarán todas las funcionalidades. Además, estos profesionales asumirán tareas de mantenimiento del sistema y atención técnica al cliente. En la práctica, es habitual contratar empresas externas para este tipo de servicios, aunque en este caso se representará como la contratación de dos trabajadores a tiempo completo con carácter anual.
- Salario de un analista financiero: en torno a 100.000 € anuales en España para perfiles con experiencia en puestos directivos.[41]. Se requerirá un único profesional con una sólida trayectoria en análisis de riesgos y profundo conocimiento del marco regulatorio. Aunque estas funciones suelen ser externalizadas a firmas especializadas, en este caso se contemplará como la contratación de una persona a tiempo completo con carácter anual.
- Costes de servicios en la nube y ciberseguridad: estos varían considerablemente en función de los requisitos técnicos específicos, el volumen de datos, las horas de uso, la capacidad de procesamiento, la frecuencia de entrenamiento del modelo, el número de predicciones, y el proveedor seleccionado. Se estima que consultando diferentes webs el coste medio de que la empresa y el producto sea plenamente digital supondrá un coste anual de 20.000 euros.[42][43]. El coste estimado de los servicios de ciberseguridad para una pequeña empresa del sector financiero oscila entre 10.000 y 30.000 euros anuales. Dado que se trata de una empresa completamente digital y enfocada a un sector altamente sensible como el financiero, se considera esencial garantizar un nivel elevado de protección. Por ello, se ha estimado un gasto en ciberseguridad de 20.000 euros, valor que

representa un equilibrio entre la dimensión de la empresa y la criticidad de los datos manejados.[44].

- Costes de marketing: Investigando sobre el coste de marketing para un software financiero, se toma como referencia lo siguiente publicado por Business Plans: “Los costos de marketing típicos para el lanzamiento del software de contabilidad pueden comprender 10-15% de los gastos operativos anuales e incluyen publicidad digital, marketing de contenido y generación de leads.”. Además: “Los gastos de marketing y adquisición de clientes también son vitales. En promedio, una estrategia de marketing sólida, que incluye campañas digitales y análisis de clientes impulsados por la IA, puede consumir sobre 20-30% de los ingresos.”. Por ello, se considera un gasto de marketing variable e inicial del 15% de los gastos durante el primer mes, así como un coste fijo posterior del 20% de los ingresos. En estos costes de marketing se incluyen otros gastos relacionados con el crecimiento. [45]
- Costes inventario inicial: En este apartado de costes se incluyen los gastos asociados al inventario inicial, como la adquisición de ordenadores, equipos y herramientas necesarias para el desarrollo de la actividad. Asimismo, se contemplan otros gastos iniciales, tales como los costes de registro y constitución de la empresa. Según estimaciones basadas en diversas fuentes consultadas, estos gastos se calculan en torno a los 5.000 euros. Adicionalmente, se incluye un coste inicial estimado de 200.000 euros destinado al desarrollo del prototipo inicial. Este importe contempla la contratación del equipo de ingenieros y del especialista financiero necesario, así como otros gastos asociados al desarrollo y lanzamiento del producto en su fase inicial.[46] [47].

Al tratarse de un producto completamente digital, no es necesario disponer de una infraestructura física, al menos durante la fase inicial contemplada en este estudio.

Costes fijos:

- Salario Machine Learning engineers
- Salario financiero especializado
- Costes infraestructura digital y ciberseguridad
- Costes marketing

Costes fijos	Coste anual	Coste mensual
Salario Machine Learning engineers	80.000,00 €	6.666,67 €
Salario financiero especializado	80.000,00 €	6.666,67 €
Coste infraestructura digital y ciberseguridad	40.000,00 €	3.333,33 €
Marketing		20 % ingresos

Tabla 15: Costes fijos.

Fuente: Elaboración propia

Costes variables:

- Coste marketing
- Inventario inicial

Costes variables	Coste mensual
Marketing	2.500,00 €
Inventario inicial	205.000,00 €

Tabla 16: Costes variables.

Fuente: Elaboración propia

Proyección de ingresos:

Al analizar el mercado, se observa que los precios de los softwares de risk management varían considerablemente en función de las funcionalidades que ofrecen. Por ejemplo, según datos de Capterra, estos productos oscilan entre los 120 y los 800 dólares mensuales, dependiendo de su nivel de complejidad y prestaciones.[48].

Dado que el producto presentado en este proyecto busca posicionarse como una alternativa accesible y competitiva en términos de precio, se ha decidido establecer un coste mensual de 300 euros para la elaboración del plan financiero. Este importe se sitúa en el tramo inferior del rango de precios del mercado, lo que refuerza su atractivo comercial y su accesibilidad para pequeñas y medianas empresas.

Análisis del plan financiero:

A continuación, se presenta un resumen del plan financiero, en el cual se muestra el flujo de caja de la empresa durante el periodo de análisis definido, teniendo en cuenta los costes e ingresos previamente detallados. Una de las variables clave incluidas es el número de clientes, estimándose que, tras la campaña de marketing inicial, el producto se lanzará con una base inicial de 14 clientes, y que se mantendrá una tasa de crecimiento mensual del 5% (calculada sobre el número de clientes del mes anterior), similar a lo estipulado en el plan de marketing.

Esta tendencia se aplicará de forma constante durante todo el periodo de estudio, redondeando siempre al número entero inferior de clientes, para ofrecer una estimación conservadora y más realista. El plan financiero también refleja el flujo de caja mensual y el flujo de caja acumulado, lo que permite identificar el punto de equilibrio: es decir, el número de clientes y el mes en el que la empresa comenzaría a generar beneficios y alcanzar la solvencia económica.

Mes	Cientes crecimiento	Cientes	Ingreso por cliente	Ingresos	Costes marketing	Costes totales	Flujo de caja mensual	Flujo total
1	14	14	300,00 €	4.200,00 €	840,00 €	225.006,67 €	-220.806,67 €	-220.806,67 €
2	14,70	14	300,00 €	4.410,00 €	882,00 €	17.548,67 €	-13.138,67 €	-233.945,33 €
3	15,435	15	300,00 €	4.630,50 €	926,10 €	17.592,77 €	-12.962,27 €	-246.907,60 €
4	16,20675	16	300,00 €	4.862,03 €	972,41 €	17.639,07 €	-12.777,05 €	-259.684,65 €
5	17,0170875	17	300,00 €	5.105,13 €	1.021,03 €	17.687,69 €	-12.582,57 €	-272.267,21 €
6	17,86794188	17	300,00 €	5.360,38 €	1.072,08 €	17.738,74 €	-12.378,36 €	-284.645,57 €
7	18,76133897	18	300,00 €	5.628,40 €	1.125,68 €	17.792,35 €	-12.163,95 €	-296.809,52 €
8	19,69940592	19	300,00 €	5.909,82 €	1.181,96 €	17.848,63 €	-11.938,81 €	-308.748,33 €
9	20,68437621	20	300,00 €	6.205,31 €	1.241,06 €	17.907,73 €	-11.702,42 €	-320.450,74 €
10	21,71859502	21	300,00 €	6.515,58 €	1.303,12 €	17.969,78 €	-11.454,20 €	-331.904,95 €
11	22,80452477	22	300,00 €	6.841,36 €	1.368,27 €	18.034,94 €	-11.193,58 €	-343.098,53 €
12	23,94475101	23	300,00 €	7.183,43 €	1.436,69 €	18.103,35 €	-10.919,93 €	-354.018,45 €
13	25,14198856	25	300,00 €	7.542,60 €	1.508,52 €	18.175,19 €	-10.632,59 €	-364.651,04 €
14	26,39908799	26	300,00 €	7.919,73 €	1.583,95 €	18.250,61 €	-10.330,89 €	-374.981,93 €
15	27,71904239	27	300,00 €	8.315,71 €	1.663,14 €	18.329,81 €	-10.014,10 €	-384.996,03 €
16	29,10499451	29	300,00 €	8.731,50 €	1.746,30 €	18.412,97 €	-9.681,47 €	-394.677,49 €
17	30,56024424	30	300,00 €	9.168,07 €	1.833,61 €	18.500,28 €	-9.332,21 €	-404.009,70 €
18	32,08825645	32	300,00 €	9.626,48 €	1.925,30 €	18.591,96 €	-8.965,49 €	-412.975,19 €
19	33,69266927	33	300,00 €	10.107,80 €	2.021,56 €	18.688,23 €	-8.580,43 €	-421.555,61 €
20	35,37730274	35	300,00 €	10.613,19 €	2.122,64 €	18.789,30 €	-8.176,11 €	-429.731,73 €
21	37,14616787	37	300,00 €	11.143,85 €	2.228,77 €	18.895,44 €	-7.751,59 €	-437.483,31 €
22	39,00347627	39	300,00 €	11.701,04 €	2.340,21 €	19.006,88 €	-7.305,83 €	-444.789,15 €
23	40,95365008	40	300,00 €	12.286,10 €	2.457,22 €	19.123,89 €	-6.837,79 €	-451.626,94 €
24	43,00133258	43	300,00 €	12.900,40 €	2.580,08 €	19.246,75 €	-6.346,35 €	-457.973,28 €
25	45,15139921	45	300,00 €	13.545,42 €	2.709,08 €	19.375,75 €	-5.830,33 €	-463.803,61 €
26	47,40896917	47	300,00 €	14.222,69 €	2.844,54 €	19.511,20 €	-5.288,51 €	-469.092,13 €

27	49,77941763	49	300,00 €	14.933,83 €	2.986,77 €	19.653,43 €	-4.719,61 €	-473.811,74 €
28	52,26838851	52	300,00 €	15.680,52 €	3.136,10 €	19.802,77 €	-4.122,25 €	-477.933,99 €
29	54,88180794	54	300,00 €	16.464,54 €	3.292,91 €	19.959,58 €	-3.495,03 €	-481.429,02 €
30	57,62589834	57	300,00 €	17.287,77 €	3.457,55 €	20.124,22 €	-2.836,45 €	-484.265,47 €
31	60,50719325	60	300,00 €	18.152,16 €	3.630,43 €	20.297,10 €	-2.144,94 €	-486.410,41 €
32	63,53255291	63	300,00 €	19.059,77 €	3.811,95 €	20.478,62 €	-1.418,85 €	-487.829,27 €
33	66,70918056	66	300,00 €	20.012,75 €	4.002,55 €	20.669,22 €	-656,46 €	-488.485,73 €
34	70,04463959	70	300,00 €	21.013,39 €	4.202,68 €	20.869,35 €	144,05 €	-488.341,68 €
35	73,54687157	73	300,00 €	22.064,06 €	4.412,81 €	21.079,48 €	984,58 €	-487.357,10 €
36	77,22421515	77	300,00 €	23.167,26 €	4.633,45 €	21.300,12 €	1.867,14 €	-485.489,96 €
37	81,0854259	81	300,00 €	24.325,63 €	4.865,13 €	21.531,79 €	2.793,84 €	-482.696,12 €
38	85,1396972	85	300,00 €	25.541,91 €	5.108,38 €	21.775,05 €	3.766,86 €	-478.929,26 €
39	89,39668206	89	300,00 €	26.819,00 €	5.363,80 €	22.030,47 €	4.788,54 €	-474.140,72 €
40	93,86651616	93	300,00 €	28.159,95 €	5.631,99 €	22.298,66 €	5.861,30 €	-468.279,43 €
41	98,55984197	98	300,00 €	29.567,95 €	5.913,59 €	22.580,26 €	6.987,70 €	-461.291,73 €
42	103,4878341	103	300,00 €	31.046,35 €	6.209,27 €	22.875,94 €	8.170,41 €	-453.121,32 €
43	108,6622258	108	300,00 €	32.598,67 €	6.519,73 €	23.186,40 €	9.412,27 €	-443.709,05 €
44	114,0953371	114	300,00 €	34.228,60 €	6.845,72 €	23.512,39 €	10.716,21 €	-432.992,83 €
45	119,8001039	119	300,00 €	35.940,03 €	7.188,01 €	23.854,67 €	12.085,36 €	-420.907,48 €
46	125,7901091	125	300,00 €	37.737,03 €	7.547,41 €	24.214,07 €	13.522,96 €	-407.384,52 €
47	132,0796146	132	300,00 €	39.623,88 €	7.924,78 €	24.591,44 €	15.032,44 €	-392.352,08 €
48	138,6835953	138	300,00 €	41.605,08 €	8.321,02 €	24.987,68 €	16.617,40 €	-375.734,68 €
49	145,6177751	145	300,00 €	43.685,33 €	8.737,07 €	25.403,73 €	18.281,60 €	-357.453,08 €
50	152,8986638	152	300,00 €	45.869,60 €	9.173,92 €	25.840,59 €	20.029,01 €	-337.424,07 €
51	160,543597	160	300,00 €	48.163,08 €	9.632,62 €	26.299,28 €	21.863,80 €	-315.560,27 €
52	168,5707769	168	300,00 €	50.571,23 €	10.114,25 €	26.780,91 €	23.790,32 €	-291.769,95 €
53	176,9993157	176	300,00 €	53.099,79 €	10.619,96 €	27.286,63 €	25.813,17 €	-265.956,78 €
54	185,8492815	185	300,00 €	55.754,78 €	11.150,96 €	27.817,62 €	27.937,16 €	-238.019,62 €

55	195,1417456	195	300,00 €	58.542,52 €	11.708,50 €	28.375,17 €	30.167,35 €	-207.852,27 €
56	204,8988328	204	300,00 €	61.469,65 €	12.293,93 €	28.960,60 €	32.509,05 €	-175.343,22 €
57	215,1437745	215	300,00 €	64.543,13 €	12.908,63 €	29.575,29 €	34.967,84 €	-140.375,38 €
58	225,9009632	225	300,00 €	67.770,29 €	13.554,06 €	30.220,72 €	37.549,56 €	-102.825,81 €
59	237,1960114	237	300,00 €	71.158,80 €	14.231,76 €	30.898,43 €	40.260,38 €	-62.565,44 €
60	249,0558119	249	300,00 €	74.716,74 €	14.943,35 €	31.610,02 €	43.106,73 €	-19.458,71 €

Tabla 17: Plan financiero.

Fuente: Elaboración propia

Se comprueba que, en cuanto al número de clientes, se parte con 14 usuarios iniciales (similar al dato estimado en el plan de marketing), alcanzando una cifra alrededor de 250 clientes en el quinto año. Tal como se mencionó anteriormente, se ha mantenido una tasa de crecimiento mensual constante, lo que permite proyectar esta evolución. No obstante, esta estimación puede resultar algo elevada, considerando que durante estos años el producto estaría enfocado principalmente en el mercado español. Al tratarse de un sector de nicho, con una competencia existente y un número limitado de empresas potencialmente interesadas, alcanzar dicha cifra podría representar un reto significativo.

A medida que crece la base de clientes, también se observa un incremento progresivo de los ingresos. Sin embargo, este crecimiento va acompañado de un aumento en los costes, especialmente en lo relacionado con marketing y los costes variables asociados al crecimiento.

En lo que respecta al flujo de caja mensual, se puede observar que durante los primeros 33 meses no se generan beneficios. Las pérdidas alcanzan su punto máximo en el primer mes, debido a la concentración de gastos iniciales, como los costes de marketing y de infraestructura. El punto de equilibrio se alcanza en el mes 34, momento en el que se cuenta con 70 clientes. A partir de ese punto, la empresa comienza a generar beneficios mensuales.

No obstante, el beneficio neto acumulado no se alcanza hasta el mes 61, un mes después del tiempo de estudio del proyecto, cuando los ingresos totales superan finalmente la suma de todos los costes en los que se ha incurrido desde el inicio del proyecto. Es decir, aunque el flujo mensual ya es positivo desde el mes 34, la rentabilidad total no se consigue hasta el mes 61.

En resumen, se estima que durante los primeros cinco años la empresa no alcanzará la rentabilidad, siendo a partir del mes 61 cuando comenzaría a generar beneficios reales de forma sostenida. Esta proyección se basa en el supuesto de que se mantenga la tasa de crecimiento de clientes estimada, que el precio del software sea lo suficientemente competitivo para atraer a pequeñas y medianas empresas, y que los costes considerados sean representativos y realistas dentro del mercado en el que se posiciona el producto.

8. Conclusiones

Conclusiones extraídas del estudio y del prototipo inicial desarrollado:

- El conjunto de datos utilizado está claramente desbalanceado, con una clase mayoritaria, las empresas solventes, que tiene un peso significativamente mayor que la clase minoritaria, empresas en quiebra. Esta desproporción influye considerablemente en los resultados finales del modelo.
- Al comparar los dos enfoques analizados (datos tratados con SMOTE y datos mejorados mediante la incorporación de ratios financieros), se observa que los datos mejorados con ratios ofrecen un rendimiento superior en términos de métricas generales, evaluando los distintos algoritmos empleados.
- Para el caso de los datos tratados con SMOTE, la clasificación del rendimiento de los algoritmos es la siguiente, empleando la media de las métricas empleadas:
 - Bagging random forest
 - Bagging GB
 - Random forest
 - Gradient Boosting
 - Redes neuronales 5 capas
- Para el caso de los datos mejorados, la clasificación del rendimiento de los algoritmos es la siguiente, empleando la media de las métricas empleadas:
 - Redes neuronales (3,5 y 10 capas respectivamente)
 - Random forest
 - SVM, Bagging RF, Bagging SVM, Bagging RN
- Se comprueba que Random Forest se posiciona como la técnica con mejor rendimiento general, destacándose especialmente en combinación con Bagging, situándose entre las mejores opciones en ambos escenarios analizados.
En el caso de SMOTE, al contar con un conjunto de datos más equilibrado, Gradient Boosting también demuestra ser una técnica eficaz, mientras que SVM y redes neuronales quedan rezagadas en cuanto a rendimiento.
Por otro lado, en el caso de los datos originales con ratios añadidos, las redes neuronales se posicionan como la técnica más destacada, seguidas nuevamente por Random Forest, tanto con Bagging como sin él.

Por tanto, para futuras mejoras y en la evolución del producto, sería recomendable enfocar los esfuerzos en estas técnicas, ya que han demostrado ofrecer los mejores resultados según los diferentes escenarios analizados.

- No obstante, aunque los datos mejorados presentan un mejor rendimiento global, los modelos entrenados con datos SMOTE obtienen mejores resultados en la predicción de casos de quiebra. Esto se debe a que el oversampling equilibra parcialmente las clases, otorgando mayor peso a la clase minoritaria. Como resultado, aunque el rendimiento global sea menor, la capacidad del modelo para identificar quiebras mejora notablemente.
- Por el contrario, los datos originales (incluso tras la adición de ratios) siguen mostrando un fuerte desbalance. Esto provoca que los casos de quiebra tengan poco peso en el conjunto, lo que puede llevar a que el modelo obtenga buenos resultados en métricas generales, pero no logre detectar correctamente las quiebras. De hecho, en algunos casos, la tasa de predicción correcta de quiebras es prácticamente nula, algo crítico dado el objetivo del proyecto.
- Sería de gran valor desarrollar un estudio específico centrado en el análisis de las variables más relevantes detectadas por los distintos modelos. Se han observado patrones comunes y diferencias entre algoritmos que podrían aportar información clave. Este análisis sería muy útil en una versión más profesional del producto, ya que permitiría perfeccionar el modelo, optimizar su rendimiento y aportar valor añadido a empresas interesadas en su implementación.
- En conclusión, los resultados obtenidos muestran que el prototipo es técnicamente viable, pero también evidencian la necesidad de seguir desarrollando y perfeccionando el modelo para que alcance un nivel óptimo para su uso en la industria. El uso de los datos mejorados muestra mejores métricas generales, pero no resuelve la problemática crítica de la predicción de quiebras. El enfoque SMOTE ofrece mejores resultados en ese aspecto, lo que sugiere que para evolucionar este producto será necesario:
- Ampliar el dataset, reduciendo el desbalance entre clases.
- Optimizar los parámetros en cada algoritmo.
- Incorporar nuevos algoritmos que ofrezcan perspectivas diferentes y mejores resultados.

El prototipo ha demostrado el gran potencial de esta herramienta para abordar la predicción de quiebras financieras, pero también ha puesto de manifiesto la complejidad técnica y el alto nivel de especialización necesario para transformar esta idea en un producto profesional y robusto.

Conclusiones extraídas del plan de negocio:

- Se trata de un producto altamente especializado, orientado a un mercado y a un perfil de cliente muy concreto.
- Los principales costes en los que se incurrirá corresponden al equipo humano y a la infraestructura tecnológica necesaria, tanto para el desarrollo del prototipo final como para su posterior mantenimiento y funcionamiento operativo.
- Los ingresos provendrán de un modelo de suscripción mensual o anual, a través del cual los clientes obtendrán una licencia de uso del producto.
- El producto está dirigido inicialmente a pequeñas y medianas empresas en España, con la intención de ampliar posteriormente su alcance a nivel de la Unión Europea. Por ello, el precio es un factor clave para garantizar la competitividad.
- El gran valor diferencial del producto frente a las herramientas tradicionales radica en la incorporación de inteligencia artificial, lo que permite un mayor rendimiento, así como un aprendizaje y evolución constantes que garantizan su actualización frente a las condiciones del mercado.
- Se trata de un sector altamente competitivo, con numerosos actores establecidos, donde la especialización y el elevado nivel técnico requerido constituyen barreras de entrada importantes.
- El plan financiero presentado muestra que los principales costes son fijos, mientras que el precio fijado permite posicionar el producto de forma competitiva y atractiva para el público objetivo.
- Se observa que, hasta el tercer año, con aproximadamente 70 clientes, no se comienza a generar beneficios, acumulando pérdidas durante los meses anteriores.
- A lo largo del periodo de estudio no se alcanza la solvencia completa del proyecto, manteniéndose el balance neto negativo. Sin embargo, en el mes 61 el balance se vuelve positivo, generando beneficios netos por primera vez. Esto refleja claramente los retos asociados a la comercialización del producto.
- La tasa de crecimiento de clientes asumida puede ser realista en las etapas iniciales, pero resulta difícil de mantener a largo plazo, dado que se trata de un mercado muy específico con un número limitado de potenciales clientes, especialmente si se circunscribe inicialmente a España.
- El producto no sería rentable hasta pasados cinco años y con una base de unos 250 clientes, lo que demuestra la complejidad del lanzamiento y las dificultades de viabilidad económica del proyecto.
- Sería conveniente investigar las subvenciones y ayudas disponibles para empresas tecnológicas y del sector financiero que fomenten la innovación, ya que podrían representar un importante apoyo económico y contribuir significativamente a la viabilidad del proyecto, especialmente aquellas ofrecidas por la Unión Europea y organismos nacionales en España.

En conclusión, se trata de un producto con gran potencial y valor añadido, que enfrenta importantes desafíos técnicos para alcanzar un nivel de perfección y sofisticación que garantice su utilidad real. En cuanto a su modelo de negocio, se encuentra en un sector de nicho en el que la competitividad exige un alto rendimiento técnico y una estrategia de precios ajustada. El plan financiero sugiere que la rentabilidad es posible a medio-largo plazo, siempre y cuando se mantenga un crecimiento sostenido de clientes en un mercado tan limitado y se controlen rigurosamente los costes e ingresos establecidos.

9. Alineamiento con los Objetivos de Desarrollo Sostenible (ODS)

Este proyecto, aunque en un primer momento pueda parecer ajeno a los compromisos y Objetivos de Desarrollo Sostenible (ODS), en realidad tiene un impacto significativo en esta materia. Si bien la relación puede no ser evidente de forma directa, existe una conexión sólida que convierte a este proyecto en una herramienta valiosa para el avance y cumplimiento de diversas metas sostenibles.

Los principales Objetivos de Desarrollo Sostenible con los que este proyecto se alinea son los siguientes:

- **Objetivo 8 – Trabajo decente y crecimiento económico:**
Dentro de las metas de este objetivo, destacan especialmente dos por su estrecha conexión con este proyecto. El producto presentado actúa como un canalizador clave y una herramienta de gran utilidad para facilitar el cumplimiento y avance de ambas metas en particular.
 - **8.2** “Lograr niveles más elevados de productividad económica mediante la diversificación, la modernización tecnológica y la innovación, entre otras cosas centrándose en los sectores con gran valor añadido y un uso intensivo de la mano de obra.”[49].
 - **8.3** “Promover políticas orientadas al desarrollo que apoyen las actividades productivas, la creación de puestos de trabajo decentes, el emprendimiento, la creatividad y la innovación, y fomentar la formalización y el crecimiento de las microempresas y las pequeñas y medianas empresas, incluso mediante el acceso a servicios financieros.”[49].

Es fácil comprobar la estrecha relación que este proyecto mantiene con el objetivo y las metas señaladas. Se trata de un producto que incorpora tecnología de última generación como machine learning, lo que impulsa la modernización tecnológica y su aplicación en un sector clave como el financiero, generando un alto valor para este sector y para la sociedad en general.

Además, este proyecto es una iniciativa emprendedora en la que la innovación ocupa un papel central, con el objetivo de mejorar el rendimiento y la eficiencia del sistema financiero. Por tanto, resulta evidente su alineación con las metas planteadas, contribuyendo activamente a su cumplimiento y a la promoción del desarrollo de este objetivo.

- **Objetivo 9 – Industria. Innovación e infraestructura:**
En este caso, también es importante destacar la alineación de este proyecto con el cumplimiento de las siguientes dos metas:
 - **9.3** “Aumentar el acceso de las pequeñas industrias y otras empresas, particularmente en los países en desarrollo, a los servicios financieros,

incluidos créditos asequibles, y su integración en las cadenas de valor y los mercados.”[49].

- **9.4** “De aquí a 2030, modernizar la infraestructura y reconvertir las industrias para que sean sostenibles, utilizando los recursos con mayor eficacia y promoviendo la adopción de tecnologías y procesos industriales limpios y ambientalmente racionales, y logrando que todos los países tomen medidas de acuerdo con sus capacidades respectivas.”[49].

El alineamiento del producto presentado con este Objetivo de Desarrollo Sostenible es evidente, y se refuerza al analizar las metas específicas asociadas. Se trata de una herramienta clave para el sector financiero, diseñada para facilitar el acceso a tecnologías avanzadas que permitan anticipar situaciones de quiebra empresarial. Esto proporciona a las empresas una visión más clara del entorno económico, fortaleciendo su capacidad de decisión y gestión de riesgos.

Además, el proyecto impulsa la modernización del sector financiero mediante la integración de tecnologías de última generación, como la inteligencia artificial, contribuyendo así a la transformación digital del sistema económico. Todo ello demuestra el compromiso firme del proyecto con este objetivo y su papel activo en el avance hacia un desarrollo más sostenible.

Cabe destacar que esto no implica que el proyecto no contribuya también al avance de otros Objetivos de Desarrollo Sostenible, como la reducción de las desigualdades (Objetivo 10) o el objetivo 16 “ paz, justicia e instituciones sólidas”[49].

Todo lo expuesto en este apartado permite evidenciar la importancia e impacto que puede tener el desarrollo de un producto como el que se presenta. Esta iniciativa emprendedora no solo genera beneficios económicos y sociales, sino que también se alinea y contribuye activamente a los objetivos establecidos por las Naciones Unidas. Así, el proyecto demuestra su compromiso con la sociedad y con el progreso de la humanidad, consolidándose como una propuesta coherente con los valores del desarrollo sostenible.

10. Referencias

- [1] ChatGPT. (2025). Introducción Trabajo Fin de Master.
- [2] Banco de España. (2017, mayo). Informe sobre la crisis financiera y bancaria en España, 2008 – 2014. BDE.
https://www.bde.es/f/webbde/GAP/Secciones/SalaPrensa/InformacionInteres/ReestructuracionSectorFinanciero/Arc/Fic/InformeCrisis_Completo_web.pdf
- [3] Álvarez, J. I. (2023, 27 febrero). España lidera el alza de quiebras de empresas en la UE con un 163% más. elEconomista.es.
<https://www.eleconomista.es/economia/noticias/12163038/02/23/Espana-lidera-el-alza-de-quiebras-de-empresas-en-la-UE-con-un-163-mas.html>
- [4] RAE. Definición de quiebra. (s.f). Real Academia Española, RAE.
<https://dle.rae.es/quiebra>
- [5] European Justice. (2025,3 marzo). Insolvencia y quiebra. Unión Europea.
https://e-justice.europa.eu/topics/money-monetary-claims/insolvency-bankruptcy/es_es
- [6] European Justice. (2023, 30 mayo). Insolvencia y quiebra. Unión Europea.
https://e-justice.europa.eu/topics/money-monetary-claims/insolvency-bankruptcy_es
- [7] Portal Europeo de e-Justicia. (2025, 3 marzo). Insolvencia y quiebra.
https://e-justice.europa.eu/topics/money-monetary-claims/insolvency-bankruptcy/es_es
- [8] INE. Instituto Nacional de Estadística.(s. f.). Glosario de conceptos. INE.
<https://www.ine.es/DEFIne/concepto.htm?c=5402>
- [9] Asociación Española de Afectados Por la Deuda. (s.f).Insolvencia Económica: Qué es, Causas, Consecuencias. Afectados Deuda. Asociación Española de Afectados Por la Deuda.
<https://asociacionafectadosporladeuda.org/blog/insolvencia-economica-que-es-causas-consecuencias-soluciones/>
- [10] IBM. (2021, 22 septiembre). What is Machine Learning?. IBM.
<https://www.ibm.com/es-es/topics/machine-learning>
- [11] SAP. (s. f.). ¿Qué es machine learning ?. SAP.
<http://sap.com/spain/products/artificial-intelligence/what-is-machine-learning.html>

- [12] Chen, M. (2024, 25 noviembre). *What is machine learning?*. Oracle
<https://www.oracle.com/es/artificial-intelligence/machine-learning/what-is-machine-learning/>
- [13] IBM. (s.f). ¿Qué es un algoritmo de Machine Learning?. IBM.
<https://www.ibm.com/es-es/think/topics/machine-learning-algorithms>
- [14] Iberdrola. (s. f.). Descubre los principales beneficios del Machine Learning. Iberdrola.
<https://www.iberdrola.com/conocenos/nuestro-modelo-innovacion/machine-learning-aprendizaje-automatico>
- [15] IBM. (2023, 16 octubre). 10 casos de uso cotidianos del Machine Learning. IBM.
<https://www.ibm.com/es-es/think/topics/machine-learning-use-cases>
- [16] Repsol. (2025, 29 abril). ¿Qué es el machine learning y que usos tiene?. REPSOL.
<https://www.repsol.com/es/energia-futuro/tecnologia-innovacion/machine-learning/index.cshtml>
- [17] Banco Santander Impulsa Empresas. (2024, 20 mayo). Machine learning: qué es, para qué sirve y cómo funciona. Banco Santander.
<https://www.impulsa-empresa.es/machine-learning/#APP>
- [18] Telefónica. (2025, 27 febrero). 5 ejemplos de machine learning. *Telefónica*.
<https://www.telefonica.com/es/sala-comunicacion/blog/machine-learning-ejemplos>
- [19] Amazon Web Services, Inc. (s.f). ¿Qué es el machine learning? - Explicación sobre el machine learning empresarial - AWS.
<https://aws.amazon.com/es/what-is/machine-learning/>
- [20] Francés. T. (2020, abril). Impacto del Machine Learning en el sistema financiero. Universidad Pontificia Comillas.
<https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/42692/Impacto%20del%20machine%20learning%20en%20el%20sistema%20financiero%20-%20Frances%20Monedero%2c%20Teresa.pdf?sequence=1&isAllowed=y>
- [21] Oracle.(s.f). Fusion Cloud EPM Trabajo con Planning. IQR (Rango intercuartílico). Page 389. Oracle.
https://docs.oracle.com/cloud/help/es/pbcs_common/PFUSU/insights_metrics_IQR.htm#PFUSU-GUID-CF37CAEA-730B-4346-801E-64612719FF6B

- [22] Garcia. J. (2021, mayo). Comparativa de técnicas de balanceo de datos. Aplicación aun caso real para la predicción de fuga de clientes .Universidad de Oviedo.
https://digibuo.uniovi.es/dspace/bitstream/handle/10651/60629/TFM_Joaqu%C3%ADnGarc%C3%ADaAbad.pdf?sequence=4
- [23] Google.(s.f). Clasificación: Exactitud, recuperación, precisión y métricas relacionadas. (s. f.). Google For Developers.Machine Learning.
<http://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=es-419>
- [24] Microsoft Azure. (2025, 6 junio).Métricas de evaluación.Micorosf Azure, AI Foundry.
<https://learn.microsoft.com/es-es/azure/ai-services/language-service/custom-text-classification/concepts/evaluation-metrics>
- [25] IBM. (s.f). ¿Qué es el bosque aleatorio?. IBM
<https://www.ibm.com/mx-es/think/topics/random-forest>
- [26] IBM.(2021,17 agosto).Funcionamiento de SVM.IBM
<https://www.ibm.com/docs/es/spss-modeler/saas?topic=models-how-svm-works>
- [27] Clark.B,lee.F.(2025,7 abril). What is Gradient Boosting?. IBM
<https://www.ibm.com/think/topics/gradient-boosting>
- [28] IBM.(s.f).¿Que son las redes neuronales?. IBM.
<https://www.ibm.com/es-es/think/topics/neural-networks>
- [29] IBM.(2021,23 septiembre).¿En qué consiste el bagging?. IBM.
<https://www.ibm.com/es-es/think/topics/bagging>
- [30] De los Santos, P. R. (2022, 24 enero). Datos de entrenamiento vs datos de test. Telefónica Tech..

<https://telefonicatech.com/blog/datos-entrenamiento-vs-datos-de-test>

- [31] Bembade.D. (2023).What is random state?.Kaggle.
<https://www.kaggle.com/discussions/general/422466>
- [32] Mathworks.(s.f).Introducción a Support Vector Machines (SVM). MATLAB & Simulink.
<https://es.mathworks.com/discovery/support-vector-machine.html>
- [33] Google. (s. f.). Redes neuronales: Funciones de activación. Google For Developers.
<https://developers.google.com/machine-learning/crash-course/neural-networks/activation-functions?hl=es-419>
- [34] CSIC. (2022,14 julio). ‘Machine learning’ para impulsar la medicina de precisión. Consejo Superior de Investigaciones Científicas, CSIC.
<https://www.csic.es/es/actualidad-del-csic/machine-learning-para-impulsar-la-medicina-de-precision>
- [35] Dhapte. A. (s.f). Descripcion general del mercado de software de insolvencia. Market Research Future (Summary).
<https://www.marketresearchfuture.com/es/reports/insolvency-software-market-32484>
- [36] IEP. Instituto Europeo de Posgrado. (2025, 13 enero). Herramientas y software para la gestión de riesgos . Instituto Europeo de Posgrado.
<https://iep.edu.es/herramientas-y-software-para-la-gestion-de-riesgos/>
- [37] IBM. (s. f.). IBM OpenPages. IBM
<https://www.ibm.com/es-es/products/openpages>
- [38] EMBAT. (s.f). Soluciones de Gestión de Tesorería - Software Tesorería Digital. EMBAT
https://www.embat.io/?gad_source=1&gad_campaignid=22221746839&gbraid=0AAAArBxOQcCM6uZTJZVA0G2m2S5JbpKG&gclid=Cj0KCQjwucDBBhDxARIsANqFdr3ep2z9Vv9IJ_16bf9LKI2nBi8AkI5T_ec4wM48_Jcmy-Ox-cTjxeEaAtypEALw_wcB
- [39] Morningstar Investor Affiliate Program. (s. f.). Morningstar, Inc.
https://www-morningstar-com.translate.goog/products/investor-affiliate-program?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sge&_x_tr_hist=true

- [40] Pardo.A . (2024, 8 octubre). ¿Cuánto gana un Machine Learning Engineer en 2025?. Deusto Formación. Deusto.
<https://www.deustoformacion.com/cursos/programacion-tecnologia/curso-machine-learning/sueldo>
- [41] Universidad Francisco de Vitoria. (2023, 25 agosto). ¿Cuánto gana un experto en finanzas en España?. UFV.
<https://www.ufv.es/cuanto-gana-un-finanzas-en-espana-preguntas-gradados/>
- [42] Bogusch, K. (2024, 24 enero). Cloud Computing Costs in 2024. Oracle.
<https://www.oracle.com/es/cloud/cloud-computing-cost/>
- [43] Veritas. (s.f). La guía definitiva sobre los precios del almacenamiento en la nube. Veritas.
<https://www.veritas.com/es/es/information-center/the-definitive-guide-to-cloud-storage-pricing>
- [44] CiberSafety.(s.f).¿cuanto cuesta la ciberseguridad para empresas?. CiberSafety.
<https://cibersafety.com/precio-ciberseguridad-para-empresas/#precio-de-ciberseguridad-en-sector-finanzas>
- [45] Business Plans. (2025, 4 abril). ¿Cuáles son los 9 costos operativos para las empresas de software de contabilidad? Business Plan Templates.
<https://businessplan-templates.com/es/blogs/running-costs/accounting-software#:~:text=Los%20costos%20de%20marketing%20t%C3%ADpicos,contenido%20y%20generaci%C3%B3n%20de%20leads.>
- [46] Afianza.QualityConta.(2025). Gastos de constitución de una empresa en España (2025). QualityConta.

<https://www.qualityconta.com/gastos-constitucion-empresa/#:~:text=Ejemplos%20de%20gastos%20de%20constituci%C3%B3n,registro%20mercantil%20y%20asesor%C3%ADa%20b%C3%A1sica.>

- [47] Vaqueriza.M.(2025, 15 enero). ¿Qué son los gastos de constitución?. PayFit.
<https://payfit.com/es/contenido-practico/gastos-de-constituci%C3%B3n/>
- [48] Capterra (s.f). Risk Management Software. How to find the right risk management software. Capterra.
<https://www.capterra.com/risk-management-software/>
- [49] Naciones Unidas (s.f). Objetivos de Desarrollo Sostenible. Naciones Unidas.
<https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>

11. Anexos

Información relevante a los procedimientos de insolvencia en España, extraído de la Unión Europea:

Condiciones para la apertura del procedimiento de insolvencia:

“La Ley exige como condiciones para declarar el concurso determinados requisitos subjetivos y objetivos.

A) Presupuesto subjetivo (artículo 1): cualquier deudor puede acudir a un procedimiento de insolvencia (concurso, restructuración, procedimiento especial de microempresas), sea persona física o jurídica, empresario o consumidor.

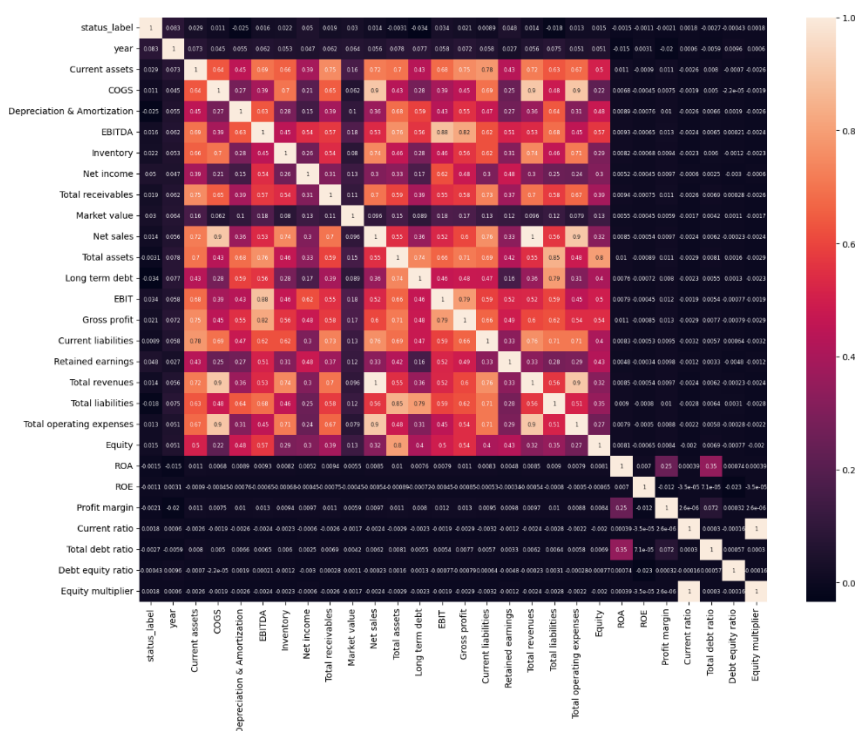
No pueden ser declaradas en concurso las entidades que integran la organización territorial del Estado, los organismos públicos y demás entes de derecho público.

B) Presupuesto objetivo: es la insolvencia del deudor, entendida como imposibilidad de pagar regularmente sus obligaciones exigibles, que puede ser actual o inminente, si la previsión del impago se sitúa dentro de tres meses siguientes (artículo 2). También puede acudirse al procedimiento de reestructuración en los casos de probabilidad de la insolvencia en los próximos dos años (artículo 585).”.

Link al anexo:

https://e-justice.europa.eu/topics/money-monetary-claims/insolvency-bankruptcy/es_es

Estudio datos sintéticos



Base de datos Kaggle:

<https://www.kaggle.com/datasets/utkarshx27/american-companies-bankruptcy-prediction-dataset/data>

Código desarrollado:

https://colab.research.google.com/drive/12Jz7_0392yWuLJU1NETTvZW6hrbP0jQd?usp=sharing