



MODELO INTEGRADO PARA LA PREDICCIÓN DE QUIEBRA Y LA VALORACIÓN DINÁMICA DE EMPRESAS MEDIANTE TÉCNICAS AVANZADAS DE MACHINE LEARNING Y SIMULACIÓN FINANCIERA ESTOCASTICA

Model for bankruptcy prediction and dynamic valuation of companies using advanced machine learning techniques and stochastic financial simulation

Sara Elizabeth Martínez Osorio

Proyecto de grado

Aesor, docente

Paula María Almonacid Hurtado

UNIVERSIDAD EAFIT
ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA
MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA
MEDELLÍN
2025

ABSTRACT	3
INTRODUCCIÓN	4
OBJETIVOS	5
General	5
Específicos	5
MARCO TEÓRICO y estado del arte	6
Valoración de empresas (Business valuation) mediante Machine Learning.....	7
Predicción de Riesgo de Quiebra (Bankruptcy Predictor) mediante Machine Learning.....	8
Modelo Integrado de Riesgo y Valor	10
Variables (Variables, Features, Predictors)	11
Técnicas de Machine Learning (MLDL)	12
Modelado de Riesgo y Simulación Estocástica	13
Resultados Obtenidos	13
Limitaciones	14
Diseño metodológico	16
Fases del desarrollo metodológico	16
RESULTADOS ESPERADOS	17
PRESUPUESTO PROYECTADO	19
CRONOGRAMA DE ACTIVIDADES	20
REFERENCIAS	21

RESUMEN

En un entorno empresarial cada vez más incierto y altamente competitivo, la capacidad de anticipar el riesgo financiero y valorar adecuadamente a las empresas se ha convertido en un aspecto crítico para la toma de decisiones estratégicas. A pesar de los avances en técnicas de análisis financiero tradicional, muchas organizaciones continúan utilizando métodos estadísticos que no incorporan el aprendizaje automático ni la simulación de escenarios para modelar la incertidumbre futura. Este trabajo de grado tiene como propósito desarrollar un modelo analítico basado en técnicas de aprendizaje automático (machine learning) que integre dos enfoques complementarios: la predicción de quiebra empresarial y la valoración financiera de empresas.

Para ello, se recopiló y procesó un conjunto de datos con información financiera histórica de empresas, enfocando en implementar técnicas de aprendizaje supervisado para hallar el valor apropiado de las compañías con base en sus características financieras y si es probable que quiebre o no.

El estudio comparó el desempeño de diferentes modelos de machine learning utilizando matrices específicas para cada tipo de problema. Asimismo, se analizará la importancia relativa de las variables empleadas, con el fin de identificar los factores financieros que inciden en mayor medida tanto en el riesgo de quiebra como en la generación de valor.

La aplicación conjunta de machine learning y simulación financiera estocástica (Monte Carlo) ofrece una oportunidad para desarrollar modelos más adaptativos y realistas, capaces de estimar simultáneamente el riesgo de insolvencia y la variabilidad del valor de una empresa bajo distintos escenarios.

Palabras clave: Valoración de empresas, predicción de quiebra, machine learning, indicadores financieros, simulación estocástica, riesgo.

ABSTRACT

In an increasingly uncertain and highly competitive business environment, the ability to anticipate financial risk and adequately value companies has become critical for strategic decision-making. Despite advances in traditional financial analysis techniques, many organizations continue to use static methods that do not

Incorporate machine learning or scenario simulation to model future uncertainty. The purpose of this thesis is to develop an analytical model based on machine learning techniques that integrates two complementary approaches: business bankruptcy prediction and financial valuation.

To this end, a dataset with historical financial information on companies will be collected and processed, focusing on implementing supervised learning techniques to determine the approximate value of companies based on their financial characteristics and their likelihood of bankruptcy.

The study will compare the performance of different machine learning models using specific metrics for each type of problem. Likewise, the relative importance of the variables used will be analyzed to identify the financial factors that have the greatest impact on both bankruptcy risk and value creation.

The combined application of machine learning and stochastic financial simulation (Monte Carlo) offers an opportunity to develop more adaptive and realistic models capable of simultaneously estimating the risk of insolvency and the variability of a company's value under different scenarios.

Keywords: Business valuation, bankruptcy prediction, machine learning, financial indicators, stochastic simulation, risks.

INTRODUCCIÓN

En el entorno actual, caracterizado por la volatilidad de los mercados y la creciente incertidumbre, las empresas se enfrentan a la necesidad de anticipar riesgos financieros y valorar de manera precisa su posición económica. (Barragán-Villaverde & Calderín-Calderín, 2024). Estos procesos son fundamentales no solo para la gestión interna, sino también para inversionistas, acreedores y demás partes que participan en la toma de decisiones estratégicas, se enfrenta a pesar de los avances en los métodos tradicionales de análisis, como el modelo de Altman Z-Score para la predicción de quiebras o el Price-to-Capex Descartado (PCD) para la valoración, dichas metodologías presentan limitaciones al ser enfoques estáticos que no capturan la dinámica de los mercados en la incertidumbre futura (Barragán, n.d.).

Este proyecto plantea un interrogante central: ¿es posible desarrollar un modelo analítico integrado, basado en técnicas de machine learning y simulación financiera estocástica, que permita evaluar la probabilidad de quiebra y, de forma complementaria, estimar el valor de una empresa bajo distintos escenarios de incertidumbre, superando así las limitaciones de los métodos tradicionales?

Este estudio se justifica por la necesidad crítica de dotar a los profesionales del ámbito financiero de herramientas analíticas de vanguardia que reflejen la realidad dinámica del mercado y permitan una toma de decisiones más informada y resiliente. El presente trabajo de grado busca dar respuesta a estas interrogantes mediante el análisis del efecto y valor de los factores de riesgo en la valoración de empresas, considerando el impacto de quiebra empresarial y la valoración financiera. Para ello, se recopiló y procesó datos financieros históricos, se entrenaron algoritmos supervisados de aprendizaje automático, y se implementó un módulo de simulación de Monte Carlo que permita capturar la variabilidad de los resultados del valor de una empresa bajo diferentes escenarios y a su vez establecer un punto de corte entre los métodos tradicionales versus estos nuevos enfoques metodológicos que han surgido.

El impacto esperado de esta investigación se refleja en dos niveles. En el ámbito académico, aporta al avance del conocimiento en analítica financiera aplicada, combinando modelos estadísticos clásicos con herramientas modernas de inteligencia artificial y simulación. En el ámbito práctico, ofrece a gestores empresariales e inversionistas un instrumento más robusto y adaptativo para evaluar tanto la estabilidad financiera de una empresa como su valor bajo incertidumbre, mejorando la toma de decisiones estratégicas en contextos altamente dinámicos.

General

Generar un modelo integrado que combine la predicción dinámica del riesgo de quiebra con la valoración empresarial, mediante técnicas de machine learning y simulación financiera estocástica, para generar métricas conjuntas de riesgo y valor.

1. Diseñar y entrenar con base en un marco teórico sólido un modelo de machine learning supervisado para estimar la probabilidad de quiebra empresarial utilizando información contable y financiera histórica.

Consentado [SM1]: Ortiz, F., 2025. El Impacto de la Inteligencia Artificial y el Machine Learning en la Valoración de Activos Financieros.

Comentario [G42]: Revisa el alcance de estos dos verbos:
Con la validación se incluye la implementación pero desarrollo es sobre algo que no existe

2. Entrenar bajo el modelo de machine learning supervisado para estimar la probabilidad de quiebra empresarial utilizando información contable y financiera histórica.
3. Establecer las variables financieras más influyentes en la predicción de quiebra mediante técnicas de interpretación de modelos.
4. Implementar un modelo de valoración empresarial basado en flujo de caja descontado (DCF) incorporando ajustes consistentes con la realidad del negocio.
5. Desarrollar un módulo de simulación de Monte Carlo para generar escenarios financieros y estimar la distribución probabilística del valor de la empresa.

MARCO TEÓRICO Y ESTADO DEL ARTE

En un entorno empresarial caracterizado por la incertidumbre y la volatilidad, anticipar el riesgo de quiebra y estimar de manera precisa el valor de una empresa son elementos esenciales para la toma de decisiones financieras y estratégicas. Según (Jaramillo et al., 2023), existen tres enfoques principales frente a la quiebra: la prevención, la alerta temprana y la superación, este último orientado a evitar la liquidación mediante la recuperación empresarial.

La valoración de empresas, por su parte, constituye la base para determinar un precio justo y ofrecer información clave a los grupos de interés (Morales & González, 2023). Sin embargo, en la práctica, los inversionistas deben analizar múltiples variables que interactúan de manera compleja y no lineal, lo que exige modelos analíticos más sofisticados. Los avances en machine learning (ML) ofrecen nuevas posibilidades para mejorar la precisión y reducir la subjetividad en los procesos de valoración (Deetsma & Lu, 2023).

Como lo argumenta Kiklik, (2023) las técnicas de machine learning tienen mejores desempeños que modelos tradicionales como la regresión lineal estimada por el método de mínimos cuadrados. El hecho especial destaca en métodos estadísticos

no paramétricos. Son capaces de incorporar relaciones no lineales y efectos de interacción, y pueden aproximar funciones de cualquier tipo y complejidad, la regularización para evitar el sobreentrenamiento, el ajuste de hiperparámetros, el uso de validación cruzada, y finalmente se enfoca en la predicción del valor y no de los coeficientes.

La inteligencia artificial cobra mayor relevancia en la estrategia empresarial y tiene el reto enfrentarse a la evolución de estas nuevas tecnologías. Como lo confirman Pfeiffer & Krietsch, (2023) la adopción con estas tecnologías es fundamental para obtener mejores resultados y fomentar una verdadera transformación digital, sin embargo, esto implica un esfuerzo significativo en múltiples ámbitos ya que se requiere una combinación de recursos técnicos, humanos y organizacionales para desarrollar una verdadera capacidad de IA.

Valoración de empresas (Business valuation) mediante Machine Learning

El uso de Machine Learning (ML) para la valoración de empresa tiene unas ventajas como brindar mejores herramientas a los inversionistas o grupos de interés, reducir la complejidad y subjetividad para valorar una compañía, y a destacar, la flexibilidad de los modelos de ML, para trabajar con diferentes tamaños, alta correlación y dimensionalidades de los datos (Koklev, 2023).

En línea con las tendencias y necesidades actuales en la era de la información resulta oportuno aprovechar los sistemas automáticos en la valoración de empresas para agilizar los procesos de toma de decisiones y aumentar la precisión en las evaluaciones. La literatura reciente ha explorado la aplicación de ML en la valoración, demostrando ser una herramienta eficaz y económica para predecir la capitalización de mercado utilizando datos financieros (Morales & González, 2023).

Los modelos de ML, en particular los Gradient Boosting Machines (GBMs), superan sustancialmente a los modelos tradicionales en la precisión de la valoración fuera de la muestra, y sus predicciones se comportan como valores fundamentalistas

(Geertsema & Lu, 2023). Un marco de regresión basado en características de la empresa (Regression with Company Characteristics) puede reducir el error de valoración en un 30% en comparación con los estándares de la industria, demostrando que la integración de características financieras y no financieras conduce a mejoras significativas en la precisión predictiva(Blanquet et al., 2025).

Otro trabajo para revisar encontramos el realizado por Geertsema & Lu, (2023) con relación a la valoración realista, el usa un amplio set de variables aproximadamente 97 de EE. UU. (1980-2019), que se agrupan en categorías de Profitability, Growth, Efficiency, Financial soundness, Industry, etc. priorizando información fundamental (con beta CAPM y clasificación FFM3 como complementos). El estudio implementa GBM (LogitGBM) sobre log-múltiples, con validación sub-muestra mensual y comparaciones frente a modelos tradicionales; además usa SHAP y una descomposición de pesos identificar qué variables eran más importantes (ajustabilidad y crecimiento) y qué empresas eran menos influyentes en cada predicción. Los errores de valoración del GBM (diferencia entre el valor estimado y el valor real de mercado) predicen rendimientos futuros. Si el modelo dice que una empresa está sobrevalorada, su acción tiende a caer; Si dice que está infravalorada, su acción tiende a subir.

Cabe resaltar que la literatura relacionada con la aplicación de modelos ML en la valoración de empresas es aún limitada. La falta de investigaciones suficientes en este campo representa una oportunidad significativa tanto para el desarrollo de este trabajo, como para futuras investigaciones que busquen profundizar en el tema.

Predicción de Riesgo de Quiebra (Bankruptcy Prediction) mediante Machine Learning

La predicción del riesgo de quiebra empresarial es fundamental para la gestión de riesgos (Jorgevski-Vilanova & Colantoni-Carlinovi, 2024) y para la toma de decisiones críticas, como la concesión de préstamos(Wei et al., 2024).

Existe un amplio consenso en que los modelos de Machine Learning (ML) son significativamente mejores que los modelos estadísticos tradicionales (como la Regresión Logística y el Análisis Discriminante Lineal) en términos de rendimiento estadístico y financiero para la predicción de quiebra (Radwansky & Hain, 2022). Los modelos de ML tienen la capacidad de modelar relaciones no lineales entre las variables predictoras y el resultado de la quiebra (Shetty et al., 2022).

En la revisión realizada por Morales & Gonzalez, (2022) evidencia que ciertos algoritmos de Machine Learning muestran una capacidad predictiva para la predicción de quiebra mucho mayor que las metodologías tradicionales, como la puntuación Z de Altman. Entre los modelos con mejor desempeño, se destacan algoritmos como XGBoost, SVM, Simult. Random Forest y Análisis de Decisión. Estos algoritmos no solo son eficientes para la predicción, sino que también pueden trabajar con conjuntos de datos de alta dimensionalidad.

Por otra parte, el trabajo realizado por Lombardo et al. (2022) implementa en un conjunto de 8262 empresas estadounidenses algoritmos como support vector machine, random forest, algoritmos de boosting, regresión logística y redes neuronales. Hace uso de 18 variables contables y financieras como activo circulante, ingresos, deudas, ganancias. Aquí se enfatiza el uso de la métrica del área bajo la curva (AUC) como el desempeño de los datos, dado que una empresa quiebra es un evento poco usual. Se hace énfasis en el costo de clasificar una empresa en quiebra como saludable por lo que es importante la sensibilización del error tipo II, se debe dar particular importancia a la reducción de los fallos negativos y a la maximización de la sensibilidad (recall). Lombardo et al. (2022) destacan también, los retos en esta línea: la privatización de datos, el desequilibrio de clases y la falta de métricas estandarizadas.

Lisaberto et al., (2023) realiza un estudio para la predicción de quiebras en donde utiliza 13 indicadores financieros provenientes del modelo Z-Score de Altman que

representar la solidez financiera de la empresa. En este trabajo se usó modelos de machine learning y métodos de balanceo de datos. Uno de los hallazgos más importantes es que en muestras pequeñas los modelos de bagging y Random Forest, combinados con New-Miss y Random under-sampling, mostraron los mejores resultados para identificar empresas en quiebra. Y para muestras grandes las redes neuronales artificiales y los métodos de árboles de decisión, junto con SMOTE y el muestreo aleatorio, funcionaron mejor.

Finalmente el trabajo de Alarín et al., (2022) compara el rendimiento de ocho modelos para la predicción de quiebra de empresas con datos de compañías estadounidenses. Usó modelos de regresión penalizada, bosque de árboles y redes neuronales, aquí se enfatiza que estos algoritmos pueden manejar un gran número de covariables y descubrir relaciones no lineales complejas. En este trabajo se destacan los modelos como XG Boost con un 92% de AUC, y en cuanto a las variables se destacan las variables de mercado ya que estas revelan información que puede ser relevante para el futuro de la empresa. Sin embargo, las variables económicas adquieren importancia en los periodos de crisis. También, el incluir variables textuales como documentos corporativos y de sentimiento hacia la compañía o variables macroeconómicas y de nivel de industria no contribuyen a una mejora significativa del desempeño del algoritmo.

Modelo Integrado de Riesgo y Valor

Aunque la predicción de quiebra y la valoración son campos bien investigados por separado, un enfoque integrado se justifica por las limitaciones en las metodologías existentes:

1. **Riesgo vs. Valor:** Es crucial que los inversionistas comprendan y se adapten a las limitaciones y políticas que afectan la valoración empresarial para gestionar una posición más sólida (Morales & González, 2023).

<div><div><div>UNIVERSIDAD</div><div>EAFIT</div></div></div>		
<p>2. Riesgo Interno y de Contagio: Tradicionalmente, la predicción de quiebra se ha centrado en el riesgo interno de la empresa (pérdidas), ignorando el riesgo de contagio y sus efectos combinados. Es una contribución novedosa considerar ambos riesgos simultáneamente para la inferencia de quiebra (Wei et al., 2024).</p> <p>3. Simulación Estadística: La modelización del desarrollo de los flujos de efectivo puede mejorarse utilizando un proceso estocástico generado por una simulación de Monte Carlo. La simulación de mercados es una herramienta eficaz que contribuye a optimizar el proceso de valoración empresarial. En contextos empresariales donde existe incertidumbre en la valoración, es crucial utilizar un análisis de riesgos basado en la simulación (Moralés & González, 2023).</p>		
Variables (Variables, Features, Predictors) La literatura identifica una amplia diversidad de variables cruciales para la predicción de riesgos y la valoración (Business Valuation Resources, 2025).		
Tipo Variables	de Ejemplos y Contexto	Citas
Financieras (Ratios)	Son las variables más comunes, representando más del 80% de las características utilizadas en la predicción de insolvencia (Ali et al., 2024). Incluyen rentabilidad, eficiencia operativa, solvencia, liquidez, endeudamiento, crecimiento y estabilidad (Wei et al., 2024). Los ratios de rentabilidad (ROCE-type) y los relacionados con el crecimiento tienen efectos marginales positivos esperados en todos los modelos de valoración, ya que impulsan los flujos de efectivo futuro (Grieshaber & Li, 2023).	
No Financieras (Calificativas)	Información no financiera, como el crecimiento, el desempeño operativo (Moralés & González, 2023), el capital intelectual, el potencial tecnológico, la escalabilidad, el riesgo, los indicadores de desempeño, y las características	

<div>UNIVERSIDAD EAFIT</div> <div>de los fundadores/estructura de propiedad (Blarquet et al., 2025).</div> <div>Valuación Específica</div> <div>Múltiplos basados en las ganancias futuras (muestran el rendimiento más destacado), seguidos por las medidas de ganancias históricas, el Ego de director y el valor en libros del patrimonio. La sostenibilidad económica, social y ambiental es un factor recurrente y crucial en la generación de valor (Morales & González, 2023).</div>		
Técnicas de Machine Learning (ML/DL)		
Categoría	Algoritmos Comunes	Aplicaciones Notables
Modelos Ensemble	XGBoost, Random Forest, LightGBM, Gradient Boosting Trees (GBDT), Stacking Classifier (SCC)	Predicción de quiebra con alta precisión. XGBoost es el mejor (RF). LightGBM: NOBrow, clasificador para datos tabulares. Gradient Boosting: Decision Tree ha sido el algoritmo con mejor desempeño en algunos estudios. Stacking (SCC) mejora la precisión de quiebra (87.83% accuracy) (Santaguida-Cabrera, 2024).
	Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Deep Learning (DL)	Predicción de quiebra y prioridades. Redes Neuronales (NN): Deep la detección de fraudes (92%) por Learning (DL) (Shetty et al., 2022).
Otros Superintendidos	Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression Classifier (LRC)	Predicción de quiebra y riesgo crediticio. Se utilizan para clasificar quiebras y riesgo crediticio. Fuentes & Macías Arias, 2025).
ML Interpretación	SHAP (Shapley Additive), LIME (Local Interpretable Model-agnostic Explanations)	Explicación de modelos "caja negra". SHAP: Permutación, Local. LIME: Permutación, Local.

Interpretable Model-Agnostic characteristic to the prediction final
Explanations (LIME) (Aho et al (Kokkeiv, 2023).
et, 2024)

Modelado de Riesgo y Simulación Estocástica

- Modelos de Intra-Riesgo y Contagio: Se ha propuesto un marco unificado (Corrigan) para capturar el riesgo interno (usando características significativas de la empresa) y el riesgo de contagio (usando Redes Neuronales Gráficas Heterogéneas, HYGNN, y Redes Jerárquicas de Transformación Gráfica, HDTN) basado en un grafo de conocimiento empresarial (EKG) (Wu et al., 2024).
- Simulación de Monte Carlo (SMC): Es una herramienta matemática utilizada para el análisis de sistemas complejos, facilitando la estimación y aproximación de escenarios a la realidad. En finanzas, se aplica para evaluaciones y cálculos de cobertura de opciones, buscando aproximar la esperanza a través de la media de una muestra de variables aleatorias. Un estudio lo usó para determinar la probabilidad de que un indicador financiero (Roi) fuera positivo (García, 2021).
- Modelo de DuPont: Utilizado para evaluar y pronosticar indicadores financieros, centrándose en el indicador de rentabilidad ROE (Return on Equity), que integra indicadores de actividad y rentabilidad (García, 2021).

Resultados Obtenidos

- Precisión de ML: Los modelos de ML basados en ensamble han mostrado una eficiencia superior, alcanzando precisiones consistentemente altas en la predicción de quiebras (Shetty et al., 2022). Por ejemplo, el algoritmo SCC alcanzó un accuracy del 97.63% en la predicción de quiebras en un estudio (Vernagima-Villaverde & Caballitas-Carbonell, 2024). En la

valoración, los modelos ML superan a los tradicionales, con mejoras significativas en la reducción de errores (Geertman & Lu, 2023).

- ML vs. modelos estadísticos: Los modelos de ML superan sistemáticamente a los modelos estadísticos tradicionales (como la Regresión Logística y el Análisis Discriminante Lineal) en los escenarios considerados (Radovanovic & Haas, 2023).

- Importancia de las variables: En la valoración, los ratios relacionados con la rentabilidad, el crecimiento, la solidez financiera y la eficiencia son más importantes que la clasificación industrial (Geertman & Lu, 2023).

- Impacto de la BMC: La BMC puede usarse para pronosticar la probabilidad de que un indicador como el ROE (rentabilidad) sea positivo.

Al comparar la BMC con una RNA entrenada para predecir el ROE, se demostró que la RNA tuvo una mayor precisión, aunque ambos modelos fueron suficientemente precisos (por debajo del 5%) (García, 2021).

- Métricas sensibles al costo: Los modelos de ML son exitosos en la reducción de los costos de quiebra incurridos por los inmenoristas, superando a los métodos estadísticos tradicionales. Pequeñas diferencias en el rendimiento estadístico (accuracy, AUC) pueden llevar a diferencias sustanciales en el rendimiento financiero y socioeconómico, destacando la necesidad de métricas sensibles al costo (Radovanovic & Haas, 2023).

Limitaciones

Para orientar su discusión sobre la investigación futura y la justificación de este trabajo, es vital incluir las limitaciones encontradas en la literatura:

1. Problema de la "Caja Negra" (Black Box): A pesar de la alta precisión, los modelos de ML basados en aprendizaje son a menudo "caja negra". Esto dificulta la interpretación de los modelos, un desafío que la investigación intenta superar con herramientas de IA explicable (XAI) como SHAP/KSHAP.

- 2023). La dificultad para interpretar los modelos puede obstaculizar su adopción y su integración en la práctica(Stafkova, 2024).
2. Falta de métricas sensibles al costo. Los modelos estadísticos a menudo se evalúan únicamente con métricas de rendimiento estadístico (precisión, AUC, F1-score), sin considerar los diferentes costos de las malas clasificaciones (errores Tipo I y Tipo II) o el impacto socioeconómico real (como el número de empleos perdidos) (Rabacovovic & Huan, 2023).
3. Sesgo y Limitación de Datos:
- Desequilibrio de clases: La predicción de quiebra es un suceso debido al fuerte desequilibrio de clases, ya que la quiebra es un evento raro en la economía real (Lombardo et al., 2022).
 - Sesgo de prospectación (Look-Ahead Bias): En la predicción del éxito empresarial, muchos estudios anteriores contienen sesgos porque utilizaban características que no habían estado disponibles en el momento de la decisión (Zelikowski & Antkowiak, 2021).
 - Tamaño muestral y alcance geográfico: Algunos estudios se ven limitados por conjuntos de datos relativamente pequeños(Chalvatzis, 2024). Por ejemplo, un estudio de valoración se basó principalmente en startups de la región nórdica, lo que puede limitar la generalización de los hallazgos (Blanchard et al., 2022).
4. Modelos estáticos vs. dinámicos: Existe la necesidad de modelos dinámicos que utilicen series de tiempo para explotar mejor la información temporal y predecir el estado de la empresa con mayor anticipación (Lombardo et al., 2022).
5. Exclusión de factores externos/complejos: Los modelos de valoración a menudo no tienen en cuenta factores de conexión con el mercado de forma suficiente, lo cual es necesario para obtener una estimación razonable del valor económico(Morales & Gonzalez, 2023).

DISEÑO METODOLÓGICO

El presente estudio adopta un enfoque cuantitativo de tipo no experimental, correlacional y explicativo, orientado a analizar y modelar la relación entre los indicadores financieros y la probabilidad de quiebra, así como a estimar el valor de las empresas bajo distintos escenarios de incertidumbre. La metodología integra técnicas de Machine Learning supervisado con simulación financiera estocástica, combinando el rigor analítico de los modelos estadísticos con la capacidad predictiva de los algoritmos inteligentes.

Fases del desarrollo metodológico con enfoque cuantitativo

1. Recolección y preparación de datos:
Se recopilarán datos financieros históricos de empresas provenientes de fuentes públicas y bases de datos financieras (por ejemplo, Yahoo Finance, Compustat, Bloomberg o EMB). Las variables incluirán indicadores de rentabilidad, liquidez, endeudamiento, eficiencia operativa y crecimiento. Se realizarán procesos de limpieza, normalización, detección de valores atípicos y balanceo de clases (por ejemplo, mediante SMOTE) para garantizar la calidad del conjunto de datos.
2. Selección y definición de variables:
Se identificarán las variables predictoras más relevantes a partir de la literatura revisada y se realizará un análisis exploratorio y correlacional para determinar su impacto en el riesgo de quiebra y el valor de la empresa.
3. Entrenamiento de modelos de Machine Learning:
Se implementarán diferentes algoritmos supervisados —como XGBoost, Random Forest, Support Vector Machine (SVM), Regresión Logística y *K-Nearest Neighbors (KNN)*— para la predicción de quiebras. Los modelos se evaluarán mediante técnicas de validación cruzada y métricas como Accuracy, Precision, Recall, F1-score y AUC-ROC.

4. Modelado de valoración financiera:

Se desarrollará un modelo de valoración basado en el Flujo de Caja Descuento (DCF) complementado con un componente de simulación estocástica. Para cada empresa, se estimará la distribución probabilística del valor mediante la Simulación Monte Carlo, generando miles de escenarios que consideren la volatilidad de los principales supuestos (crecimiento, tasa de descuento, márgenes operativos, entre otros).

5. Evaluación y validación del modelo integrado:

El desempeño del modelo será evaluado comparando sus resultados con los de metodologías tradicionales (como el modelo Z-Score de Altman y la valoración DCF convencional). Se aplicarán análisis de sensibilidad y robustez para validar la estabilidad de los resultados.

6. Visualización y análisis interpretativo:

Finalmente, se implementarán herramientas de interpretación de modelos (como SHAP y LIME) para identificar la importancia relativa de las variables en la predicción, garantizando la transparencia y explicabilidad del modelo.

RESULTADOS ESPERADOS

El desarrollo del modelo integrado permitirá generar resultados cuantitativos y cualitativos que contribuyan al avance en el campo de la analítica financiera. Se espera obtener un modelo predictivo robusto, interpretable y adaptable a distintos contextos empresariales.

Objetivo Específico	Producto Esperado	Resultado Cuantificable o Indicador
---------------------	-------------------	-------------------------------------

Diseñar y entrenar un Algoritmo entrenado
modelo de ML supervisado (XGBoost, SVM, RF, etc.)

Objetivo Especifico	Producto Esperado	Resultado Cuantificable o Indicador
para estimar la y sus métricas de probabilidad de quiebra.	desempeño.	
Identificar las variables más influyentes en la predicción de quiebras.	Reporte interpretativo basado en SHAP y LIME.	Ranking de importancia de variables y visualizaciones explicativas.
Implementar y calibrar un modelo de valoración DCF.	Modelo financiero parametrizable en Python con supuestos ajustables.	Diferencia media de error (MAPE) < 10%.
Desarrollar un módulo de simulación de Monte Carlo.	Simulador estocástico de Distribución probabilística de valores de salida valor esperado por escenarios generados.	
Integrar los módulos de predicción de quiebra y valoración.	Prototipo del modelo ajustado por riesgo, integrado riesgo-valor condicional.	Métricas conjuntas: Valor por riesgo, probabilidad de quiebra condicional.
Validar el modelo propuesto frente a métodos tradicionales.	Informe comparativo de desempeño.	de Mejora \geq 15% respecto a modelos base.

PRESUPUESTO PROYECTADO

El presupuesto del proyecto contempla recursos principalmente tecnológicos y de procesamiento de datos, dado que la investigación se desarrollará con herramientas de código abierto.

Recurso	Descripción	Costo estimado (COP)
Computador de rendimiento	Procesamiento de datos y almacenamiento de modelos (uso de GPU/CPU).	\$2.000.000
Licencias o servicios cloud (Google Cloud Pro, Azure ML, etc.)	Entrenamiento de modelos, simulación Monte Carlo.	\$400.000
Software de análisis y visualización (Python, Power BI, Excel)	Herramientas analíticas (versión gratuita o institucional).	\$0
Bibliografía y artículos académicos	Adquisición de documentos disponibles de acceso libre.	\$300.000
Asesoría y revisión técnica	Revisión metodológica conceptual del modelo.	\$500.000
Total estimado		\$3.200.000 COP

Nota: El proyecto se apoyará principalmente en software de libre acceso (Python, Solid-learn, Pandas, Matplotlib), lo que permite mantener los costos dentro de un rango académico razonable.

CRONOGRAMA DE ACTIVIDADES			
Actividad	Descripción	Duración estimada	Periodo
Revisión bibliográfica y consolidación del marco teórico	Identificación de modelos de marco variables y metodologías relevantes.	4 semanas	Febrero – Marzo 2025
Recolección, limpieza y estructuración de datos	y Preparación del dataset para entrenamiento.	3 semanas	Marzo – Abril 2025
Entrenamiento de modelos de Machine Learning	Construcción y calibración de modelos predictivos.	5 semanas	Abril – Mayo 2025
Implementación del modelo de valoración DCF y simulación Monte Carlo	Modelado del componente financiero y generación de escenarios.	7 semanas	Junio – Julio 2025
Integración del modelo de riesgo y valor	Desarrollo del prototipo integrado.	3 semanas	Julio – Agosto 2025
Validación, análisis de resultados y visualización	de Evaluación de desempeño.	4 semanas	Agosto – Septiembre 2025



Actividad	Descripción	Duración estimada	Periodo
	comparativa y análisis interpretativo.		
Redacción final y sustentación del trabajo de grado	Elaboración del documento final, revisión 4 semanas y presentación.		Octubre - Noviembre 2025

REFERENCIAS

Aparisi, E., Chava, S., & Shah, A. (2022). *Benchmarking Machine Learning Models to Predict Corporate Bankruptcy*. <http://taylor.org/abla/2212.12051>

Altamirano, H. Jr. (s.f.). *El Impacto de la Inteligencia Artificial y el Machine Learning en la Valoración de Activos Financieros*.

Alirol, J., Ait, I., & Nizam, K. (2024). Advancing financial resilience: A systematic review of default prediction models and future directions in credit risk management. *Heliyon*, 10(21). <https://doi.org/10.1016/j.heliyon.2024.e39770>

Blanchet, L. B., Perrin, M. A., & Petrov, S. (2025). An interpretable machine learning framework for explaining company valuation. *Decision Analytics Journal*, 16. <https://doi.org/10.1016/j.dajour.2025.100611>

Business Valuation Resources, L. (2025). *Business Valuation Year in Review 2024: The Rise of AI*.

21

García, I. (2021). Simulación y predicción de indicadores de gestión financieros en PYMES mediante el uso de Redes Neuronales Artificiales. *Universidad Politécnica de Valencia*. <https://ruwet.upv.es/handle/10251/172984>

Goehmann, P., & Lu, W. (2023). Relative Valuation with Machine Learning. *Journal of Accounting Research*, 61(1), 329–376. <https://doi.org/10.1111/1475-679X.12464>

Izarraguirre-Villanueva, O., & Cabanillas-Carbonell, M. (2024). Predicting business bankruptcy: A comparative analysis with machine learning models. *Journal of Open Innovation Technology, Market, and Complexity*, 10(3). <https://doi.org/10.1016/j.joitmc.2024.100375>

Kuklev, P. (2023). PREREQUISITES FOR THE USE OF MACHINE LEARNING FOR BUSINESS VALUATION. *Brazilian Journal of Law & International Relations*, 45(6), 1–12. <https://doi.org/10.4001/BJLIR.4504.7373>

Lisachenko, O., Kisilev, T., & Kostovetsky, V. (2023). Machine Learning and Data Balancing Methods for Bankruptcy Prediction. *Ekonomika*, 10(2), 28–46. <https://doi.org/10.15388/Ekon.2023.102.2.2>

Lombardi, G., Pellegrino, M., Adaloglou, G., Cagnoni, S., Pansino, P. M., & Poggi, A. (2022). Machine Learning for Bankruptcy Prediction in the American Stock Market: Dataset and Benchmarks. *Future Internet*, 14(8). <https://doi.org/10.3390/fi14080244>

Morales, T., & Gonzalez, J. (2023). Valoración de empresas: Mapeo científico, evolución y tendencias de investigación. *Revista de Ingeniería Interfaces*, 6(1), 1–20. <https://doi.org/10.6009-0004-3885-3663>

Nguyen, H. H., Viviani, J. L., & Ben Jabeur, S. (2025). Bankruptcy prediction using machine learning and Shapley additive explanations. *Review of Quantitative*

- Finance and Accounting, 69(1), 107–148. <https://doi.org/10.1007/s11156-023-01192-4>
- Pertturi, N. A., & Khoo, F. (2023). Investigating the Influence of Artificial Intelligence on Business Value in the Digital Era of Strategy: A Literature Review. In *Information (Switzerland)* (Vol. 14, Issue 2). MDPI. <https://doi.org/10.3390/info14020085>
- Radovanovic, J., & Hsiao, C. (2023). The evaluation of bankruptcy prediction models based on socio-economic costs. *Expert Systems with Applications*, 227. <https://doi.org/10.1016/j.eswa.2023.120275>
- Shetty, S., Musa, M., & Bickart, X. (2022). Bankruptcy Prediction Using Machine Learning Techniques. *Journal of Risk and Financial Management*, 15(1). <https://doi.org/10.3390/jrfm15010005>
- Šušteršič, V. (2024). OUTSMART THE PEER GROUP IN MARKET COMPARISON: BUILDING BUSINESS VALUATION MULTIPLES BY MACHINE LEARNING. *European Journal of Business Science and Technology*, 10(2), 156–172. <https://doi.org/10.11118/ejbsat.2024.011>
- Tuñón Fuentes, J. B., & Macías Arias, E. J. (2025). Modelos de machine learning para la detección de fraudes financieros: Una revisión de la literatura. *UNEDUM - Ciencias: Revista Científica Multidisciplinaria*, 9(2), 220–234. <https://doi.org/10.47250/unedum-ciencias.v9i2.02025.220-234>
- Wu, S., Li, J., Gao, Y., Yang, Q., Chen, X., Zhao, Y., Li, Q., Zhuang, F., & Kou, G. (2024). Combining intra-risk and contagion risk for enterprise bankruptcy prediction using graph neural networks. *Information Sciences*, 659. <https://doi.org/10.1016/j.ins.2023.120081>

