



Facultad de Ciencias Económicas y Empresariales
ICADE

ESTUDIO DE LAS APLICACIONES DEL MACHINE LEARNING EN EL CAPITAL RIESGO

Autor: Alberto Santos-García Rico
Director: Susana Josefa Gago Rodríguez

MADRID | Junio 2024

ÍNDICE DE CONTENIDOS

1	INTRODUCCIÓN	3
1.1	Motivación y objetivos	5
1.2	Metodología y estructura del trabajo	5
2	FUNDAMENTOS DEL MACHINE LEARNING	6
2.1	Estado actual y evolución histórica del machine learning	6
2.2	Conceptos y definiciones clave	7
2.3	Tipos de aprendizaje: supervisado, no supervisado, reforzado	8
2.4	Principales algoritmos de Machine Learning en Finanzas	9
3	EL CAPITAL RIESGO	15
3.1	Descripción del capital riesgo y su importancia en los mercados financieros.....	15
3.2	Estructura y fases de inversión en capital riesgo.....	16
3.3	Medición y gestión de riesgos en capital riesgo.....	17
3.3.1	Tipos de Riesgos en Capital Riesgo	17
3.3.2	Estrategias de Medición de Riesgos	18
3.3.3	Estrategias de Gestión de Riesgos	18
4	MACHINE LEARNING EN FINANZAS	19
4.1	Estudios previos sobre la aplicación de machine learning en inversiones y finanzas ...	19
4.2	Análisis de resultados, interpretación y relevancia de los hallazgos.....	23
5	MACHINE LEARNING EN CAPITAL RIESGO	24
5.1	Potenciales áreas de aplicación.....	24
5.2	Metodología de creación y aplicación de modelos	26
6	Desafíos, Limitaciones y Perspectivas Futuras	28
6.1	Desafíos éticos, regulatorios y técnicos en la implementación de machine learning en finanzas.....	28
6.1.1	Desafíos Éticos:	28
6.1.2	Desafíos Regulatorios.....	29
6.1.3	Desafíos Técnicos.....	29
6.2	Limitaciones actuales	29
6.3	Tendencias emergentes, avances tecnológicos y su impacto en estrategias de inversión	
	30	
7	Conclusiones y Recomendaciones	30
7.1	Resumen de hallazgos clave.....	30
7.2	Sugerencias para futuras investigaciones y mejoras en la aplicación de machine learning en el capital riesgo	31

8	Declaración de uso de herramientas de inteligencia artificial generativa en Trabajos de Fin de Grado.....	32
9	Bibliografía	33

ÍNDICE DE FIGURAS

Ilustración 1: Ejemplo gráfico de Árbol de Decisión.	10
Ilustración 2: Ejemplo gráfico de Bosque Aleatorio.	11
Ilustración 3: Ejemplo gráfico de Boosting.....	12
Ilustración 4: Ejemplo gráfico de regresión lineal.	13
Ilustración 5: Ejemplo gráfico de regresión logística.....	13
Ilustración 6:Ejemplo gráfico de SVM.....	14
Ilustración 7: Ejemplo gráfico de red neuronal simple..	15
Ilustración 8: Figura tomada de "Machine learning methods in finance: Recent applications and prospects".	21
Ilustración 9: Cronología de una inversión en Capital Riesgo.....	24

Resumen

El uso del Machine Learning (ML) en el capital riesgo está emergiendo como una herramienta revolucionaria en el ámbito financiero. Este trabajo de fin de grado tiene como objetivo explorar las posibles aplicaciones del ML en el capital riesgo, evaluando cómo estas tecnologías pueden optimizar las decisiones de inversión y gestión de riesgos.

Inicialmente, se presenta un análisis del estado actual del ML y su evolución histórica, destacando los principales algoritmos utilizados actualmente en el ámbito de finanzas. Se examina el capital riesgo, sus fases de inversión y sus riesgos asociados, proporcionando un marco teórico para entender la industria.

A lo largo del estudio, se identifican áreas específicas donde el ML puede ser aplicado en el capital riesgo. Se proponen modelos de ML para la búsqueda de inversiones, análisis de due diligence, gestión de riesgos y optimización de la fase de venta. Estos modelos incluyen el uso de redes neuronales, máquinas de soporte vectorial y algoritmos de clustering, entre otros, para mejorar la precisión en la evaluación de startups y la predicción de tendencias de mercado.

El trabajo concluye que la integración del ML en el capital riesgo es no solo posible, sino también beneficiosa, permitiendo a los inversores tomar decisiones más informadas, mitigar riesgos y maximizar retornos.

Abstract

The use of Machine Learning (ML) in venture capital is emerging as a revolutionary tool in the financial sector. This study aims to explore the potential applications of ML in venture capital, evaluating how these technologies can optimize investment decisions and risk management.

Initially, an analysis of the current state of ML and its historical evolution is presented, highlighting the main algorithms currently used in the financial sector. Venture capital is examined, along with its investment phases and associated risks, providing a theoretical framework to understand the industry.

Throughout the study, specific areas where ML can be applied in venture capital are identified. ML models are proposed for investment search, due diligence analysis, risk management, and optimization of the exit phase. These models include the use of neural networks, support vector machines, and clustering algorithms, among others, to improve the accuracy of startup evaluations and market trend predictions.

The thesis concludes that the integration of ML in venture capital is not only possible but also beneficial, enabling investors to make more informed decisions, mitigate risks, and maximize returns.

Palabras Clave:

Capital riesgo, machine learning, big data, modelos predictivos, due diligence, medición y gestión de riesgos, estrategias de inversión, optimización de carteras, start-ups, transformación digital.

Key words:

Venture capital, machine learning, big data, predictive models, due diligence, risk measurement and management, investment strategies, portfolio optimization, start-ups, digital transformation.

1 INTRODUCCIÓN

1.1 Motivación y objetivos

La motivación para realizar este estudio nace del interés y expansión que está adquiriendo el machine learning y la inteligencia artificial en el ámbito de las finanzas en los últimos años. La implementación de modelos y algoritmos se está convirtiendo en un estándar para los jugadores más competitivos de la industria, surgiendo la pregunta de si estas tecnologías son aplicables y escalables en el ámbito del capital riesgo, una rama de este sector más nicho en España. Actualmente existe muy poca información pública acerca de modelos de machine learning aplicados al sector del capital riesgo, lo que nos lleva al objetivo de este estudio.

El objetivo de este TFG consiste en tratar de identificar las posibles aplicaciones que el machine learning puede aportar en el ámbito financiero del capital riesgo. Se pretende comenzar realizando un estudio generalista de ambas industrias en la actualidad y determinar áreas potenciales de aplicación, profundizando en el nexo de unión de ambas materias. Una vez obtenidas las conclusiones sobre los potenciales nexos de unión en las materias se pretenden identificar potenciales algoritmos de machine learning aplicables a las distintas áreas de la inversión en start-ups y activos de capital riesgo que permitan optimizar estos procesos e identificar factores de éxito útiles para los trabajadores e inversores de esta industria.

1.2 Metodología y estructura del trabajo

- a) Revisión bibliográfica: revisión exhaustiva de estudios y literatura existente en las áreas de machine learning y capital riesgo. Identificación de estudios previos que hayan explorado la aplicación de modelos y técnicas de machine learning en el ámbito del capital riesgo y similares. Por último, analizar las tendencias actuales en el uso de algoritmos de machine learning en la toma de decisiones financieras y de inversión.
- b) Análisis del Estado del Arte: evaluación de casos de estudio relevantes donde se haya implementado machine learning en las finanzas, (a ser posible lo más relacionado posible con capital riesgo), elaborando un análisis de aplicaciones, éxitos y desafíos.
- c) Identificación de nexos de unión: una vez completado el estudio del estado del arte y de estudios previos, tratar de identificar qué algoritmos y elementos del machine learning pueden ser de aplicación válida y realista en la industria del capital riesgo.
- d) Conclusiones: interpretación de resultados sobre los ámbitos anteriormente mencionados, identificando además limitaciones, recomendaciones y desafíos futuros.

2 FUNDAMENTOS DEL MACHINE LEARNING

Para comprender este documento, resulta esencial conocer primero una serie de términos y conceptos relacionados con la materia, por lo que conviene aportar una breve definición de estos términos relacionados con el machine learning.

2.1 Estado actual y evolución histórica del machine learning

El Machine Learning (ML), es definido por la Autoridad Bancaria Europea como “un proceso que utiliza algoritmos en lugar de codificación procedural que permite aprender de los datos existentes para predecir resultados futuros. En términos generales, es un campo de la informática que se ocupa del desarrollo de modelos cuyos parámetros se estimulan automáticamente a partir de datos con una intervención humana limitada o nula” (Autoridad Bancaria Europea, 2021, p.9).

En la actualidad, el Machine Learning está integrado en numerosos sectores, como la salud, marketing o finanzas. Las compañías utilizan el ML para analizar grandes volúmenes de datos, identificar patrones y a partir de estos realizar predicciones precisas. En el ámbito de las finanzas y más concretamente en capital riesgo, el ML se utiliza principalmente para evaluar riesgos, prever tendencias de mercado y optimizar carteras de inversión entre otros usos. Las principales diferencias entre un modelo de Machine Learning y un modelo informático tradicional, es que mientras los modelos informáticos tradicionales tienen una programación explícita, determinista y necesitan actualizaciones automáticas, los modelos de Machine Learning aprenden automáticamente a través de un conjunto preestablecido de datos o “training set” en inglés. A partir de este, el modelo busca patrones y relaciones dentro de esos datos y utiliza ese conocimiento para hacer predicciones o tomar decisiones. Además, se adaptan y actualizan automáticamente si se alimentan con nuevos datos, por lo que no es siempre necesaria la intervención humana para que mejore el modelo. Además, gracias al avance de las tecnologías e internet, actualmente existen grandes bibliotecas de datos y marcos como TensorFlow, PyTorch o Scikit-learn, y grandes bases de datos como Crunchbase.com, que facilitan la creación y ejecución de modelos de ML de manera eficiente y escalable.

En cuanto al contexto histórico, el Machine Learning ha recorrido un largo camino desde sus inicios hasta convertirse en lo que es hoy, una de las tecnologías más influyentes de este siglo. Los orígenes del Machine Learning pueden rastrearse hasta la década de 1950, una época de gran actividad en la investigación sobre inteligencia artificial (IA). Alan Turing, uno de los pioneros de la computación, introdujo el concepto de máquinas capaces de simular el aprendizaje humano en su famoso artículo “Computing Machinery and Intelligence”. En este documento, Turing plantea la pregunta fundamental de si las máquinas pueden pensar, introduciendo su conocido “Test de Turing”, utilizado para determinar si el comportamiento de una máquina es indistinguible con el de un ser humano a través de una interfaz de texto.

Uno de los primeros programas de ML fue el “perceptrón”, desarrollado por Frank Rosenblatt en 1957. Se trataba de un modelo simple de una red neuronal que podía aprender a clasificar entradas linealmente separables. En 1959, Arthur Samuel, uno de los pioneros en el campo de la inteligencia artificial, desarrolló uno de los primeros programas de aprendizaje autónomo, consiguiendo que un ordenador IBM 701 ganase a un ser humano en el juego de las damas aprendiendo a partir de jugar contra sí mismo. Samuel acuñó el término “Machine Learning y

lo definió como “el campo de estudio que da a las computadoras la capacidad de aprender sin ser explícitamente programadas”. (Samuel, 1959).

A lo largo de las décadas de los 70 y los 80, el campo del ML experimentó grandes avances gracias a la investigación en algoritmos y teorías de la computación. Se desarrollaron importantes técnicas como los árboles de decisión (“decision trees” en inglés), introducidos por J. Ross Quinlan en 1986 a través de su algoritmo ID3, los cuales explicaré más adelante en esta sección. En la década de los 90, aparece por primera vez el reconocimiento del ML como campo separado dentro de la inteligencia artificial y la informática, caracterizado por la introducción de nuevas técnicas, como las máquinas de soporte vectorial (“SVM”) por Vladimir Vapnik, y el redescubrimiento y mejora de las redes neuronales, que impulsaron el campo hacia aplicaciones más prácticas.

A partir de la década de los 2000, el ML experimenta un gran renacimiento, gracias al aumento de la capacidad computacional y la disponibilidad de grandes volúmenes de datos (“Big Data”). Los avances en hardware permitieron el entrenamiento de modelos más complejos y profundos, sobretodo mediante el uso de redes neuronales profundas o “deep learning”. Un hito significativo de esta era fue el desarrollo de AlexNet, una red neuronal que demostró la eficacia del Deep learning para el reconocimiento de imágenes. Este éxito fue un catalizador de las actuales aplicaciones del ML en diversas áreas, desde los modelos de procesamiento de lenguaje hasta la conducción autónoma.

2.2 Conceptos y definiciones clave

En los últimos años, el uso de ML en las finanzas ha crecido exponencialmente debido a su capacidad de analizar grandes volúmenes de datos, identificar patrones y realizar predicciones precisas. En esta sección se describen los conceptos principales y algoritmos más utilizados en este ámbito en el contexto actual.

1. Data Mining (Minería de datos): se trata del proceso de recaudar grandes conjuntos de datos con características en las cuales se prevé encontrar algún tipo de correlación, patrón o tendencia significativa. En el ámbito de las finanzas, la minería de datos se utiliza principalmente para identificar comportamientos de mercado, evaluar riesgos de varios tipos y detectar fraudes.
2. Modelos predictivos: consisten en un conjunto de técnicas y algoritmos de Machine Learning que tienen el objetivo de predecir eventos futuros a partir de datos históricos. Estos modelos se utilizan sobretodo en la predicción de precios de distintos activos, análisis de riesgos crediticios y estimación de rentabilidad.
3. Procesamiento de Lenguaje Natural (“NLP”): se trata de un campo en las ciencias de la computación que estudia las interacciones entre los ordenadores y el lenguaje humano corriente, principalmente enfocado a interpretar la carga semántica y emocional de los mensajes.
4. Análisis de sentimiento: consiste en el uso de técnicas de procesamiento de lenguaje natural (“NLP”), una tecnología que procesa e interpreta la carga sentimental de cualquier texto humano, otorgando una puntuación numérica en función del sentimiento. Este análisis es utilizado por instituciones financieras, sobretodo en textos de noticias y en redes sociales para integrar el sentimiento del mercado en su toma de decisiones.

5. Big Data: generalmente este término se utiliza para referirse a un conjunto de datos tan grande y/o complejo que no puede ser procesado utilizando métodos tradicionales de procesamiento de datos. En el ámbito de las finanzas el análisis de Big Data permite a inversores obtener insights profundos sobre el mercado. En el mundo de la inversión, el análisis de Big Data permite comprender el escenario macroeconómico con mucha más precisión que con un análisis tradicional.
6. Datos de entrenamiento (training data): se refiere al conjunto de datos o base de datos inicial que se utiliza e introduce en un modelo de machine learning. Estos datos son la base de todo algoritmo.
7. “Over / Underfitting”: en el contexto de los modelos de machine learning, se refiere al fallo de sobreajuste o subajuste de nuestro modelo al tratar de encajar o interpretar la información que se le proporciona. Existen varias maneras de medir y ajustar este tipo de error.
8. Dataset: término que significa conjunto de datos. Se utiliza para referirse a los distintos grupos de datos que se analizan en programas, modelos y aplicaciones de distinto tipo:
9. Valor atípico (“outlier”): dentro de un conjunto de datos, se refiere a una observación concreta que es numéricamente distante del resto de datos. En el contexto de bases de datos y machine learning, este término se suele utilizar para referirse a datos que se intentan identificar y eliminar por ser erróneos o excepciones, y para que no perjudiquen a la totalidad de la base de datos y por consiguiente al modelo.
10. Funciones de kernel: técnicas matemáticas utilizadas en las Máquinas de Soporte Vectorial (SVM) que permiten transformar datos no linealmente separables en su espacio original a un espacio de mayor dimensión donde es más probable que sean separables.
11. “Due diligence”: en el contexto de las finanzas, se refiere al estudio intensivo de una materia, sector, mercado o activo para realizar decisiones informadas a la hora de realizar una inversión o movimiento financiero.
12. Validación cruzada: método que permite probar el rendimiento de un modelo predictivo de machine learning a través de la comparación de muestras tomadas de datos de prueba y datos de entrenamiento. Permite predecir el ajuste de un modelo. Existen muchas variaciones de esta técnica.
13. Redes LSTM: tipo concreto de red neuronal caracterizada por aprender dependencias a largo plazo entre unidades de tiempo de datos secuenciales.

2.3 Tipos de aprendizaje: supervisado, no supervisado, reforzado

Previo a la explicación de los principales algoritmos de Machine Learning más utilizados en finanzas, es necesario comprender los paradigmas de aprendizaje para los distintos tipos de modelos de ML. Estos paradigmas se refieren a los distintos métodos y técnicas que se aplican en los algoritmos para que aprendan a partir de los datos introducidos. Son tres, y dependiendo del objetivo del modelo y de los datos necesarios/utilizados se utilizará uno u otro:

1. Aprendizaje supervisado: se trata de un tipo de aprendizaje en el que el modelo es entrenado utilizando un conjunto de datos etiquetados, es decir, datos que incluyen las respuestas correctas. Durante el entrenamiento del modelo, el algoritmo hace predicciones y el usuario proporciona correcciones en función de los errores

cometidos, guiando y amoldando el modelo para que este mejore sus predicciones futuras. Este aprendizaje se suele utilizar con el objetivo de aprender una función que asocie entradas con salidas. Es comúnmente utilizado en tareas de clasificación y regresión. Un ejemplo sencillo de uso de este tipo de aprendizaje es el de los modelos de predicción de precio de vivienda. A partir de un conjunto de datos de viviendas con características como el tamaño, el número de habitaciones y el precio de venta, se puede entrenar un modelo de regresión supervisado que prediga el precio de venta (“u otros valores”) de un inmueble en función del resto de sus características.

2. Aprendizaje no supervisado: al contrario que con el aprendizaje supervisado, este es un tipo de aprendizaje en el que los datos introducidos no están etiquetados. El objetivo suele ser identificar patrones y estructuras ocultas en los datos, pues comúnmente, al utilizar este tipo de aprendizaje se parte de un contexto en el cual no hay respuestas correctas conocidas y el modelo trata de aprender por sí mismo. Es comúnmente utilizado para tareas de agrupamiento y reducción de dimensionalidad. Un ejemplo de utilización de aprendizaje no supervisado en el ámbito de las finanzas es el de los modelos de detección de fraude. A partir de un conjunto de datos de transacciones bancarias sin etiquetas, el modelo agrupa las transacciones en diferentes categorías basadas en similitudes en los datos. Este proceso puede ayudar a identificar patrones de comportamiento inusual difíciles de apreciar a simple vista.
3. Aprendizaje por refuerzo: este tipo de aprendizaje se caracteriza por la manera de entrenar al software para que tome decisiones a fin de lograr los mejores resultados. Este método imita el proceso de aprendizaje por ensayo y error utilizado por los humanos para lograr objetivos. Primero se establece una recompensa, siendo esta el objetivo final del modelo. Las acciones del modelo que se acercan a la resolución del objetivo se refuerzan positivamente, mientras que las acciones que se alejan del objetivo final se descartan. De esta manera, a través del ensayo y error, el modelo va aprendiendo y aplicando las acciones que conducen a la mayor recompensa. Uno de los ejemplos más conocidos de utilización de aprendizaje por refuerzo es el software de recomendación de contenido que utiliza “Netflix” y otras plataformas similares. El software recomienda al usuario películas y series, analizando cuales son seleccionadas por el usuario, valorando positiva o negativamente en función de cuánto tiempo se visualiza cada contenido, y generando un perfil con las preferencias. Las series y películas que más cumplen con el objetivo (permanecer en la plataforma el máximo tiempo posible) serán recomendadas a usuarios con perfiles similares.

2.4 Principales algoritmos de Machine Learning en Finanzas

A continuación, se describen algunos de los algoritmos de Machine Learning más utilizados en el ámbito de las finanzas, junto con sus mayores pros, contras y aplicaciones (Jha, 2021).

1. Árbol de Decisión (Decision Tree): este formato de Machine Learning aparece por primera vez en 1959, cuando William Belson publica un documento llamado “Matching and Prediction on the Principle of Biological Classification”. Se trata de un modelo de aprendizaje supervisado en el cual se utilizan un conjunto de reglas de decisión derivadas de los datos, que se representan gráficamente en forma de árbol. Cada nodo interno del árbol representa una característica del conjunto de datos, cada

rama representa un resultado posible y cada nodo hoja representa una etiqueta de clase o valor. El algoritmo toma un camino de entre todos los posibles en función de las restricciones/decisiones que se le impongan, llegando a una conclusión específica.

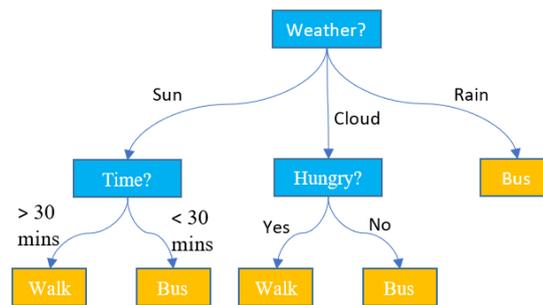


Ilustración 1: Ejemplo gráfico de Árbol de Decisión. Figura tomada de "ML Algorithms – Pros & Cons and Suitable Usages, por Rajneesh Jha, Analytics Vidhya, 2021, (<https://medium.com/analytics-vidhya/ml-algorithms-pros-cons-and-suitable-usages-b377c3c09f1b>). Copyright por el autor(es).

Uno de los principales puntos fuertes de este algoritmo es la facilidad que tiene de entenderse y de interpretarse, además de la facilidad que presentan de representarse gráficamente. Además, estos algoritmos pueden manejar información tanto numérica como categórica o cualitativa, por lo que no se requiere gran preparación previa de las bases de datos que se vayan a utilizar. También este algoritmo destaca por la similitud de razonamiento que presenta frente al razonamiento humano.

Por el contrario, los árboles de decisión no presentan demasiada robustez. Un pequeño cambio en los datos de entrenamiento puede alterar bastante la totalidad del árbol, y consecuentemente cambiar las predicciones finales. Al mismo tiempo, este modelo es muy susceptible al overfitting, y se pueden dar casos en los que se crea un árbol con demasiada complejidad el cual no es capaz de generalizar bien los datos de entrenamiento.

Una de las principales aplicaciones de este algoritmo en finanzas es para la evaluación de riesgos crediticios. A través de la creación de reglas con características como los ingresos, el historial de crédito y la deuda actual se pueden predecir las probabilidades de incumplimiento de un solicitante de crédito.

2. **Bosque Aleatorio (Random Forest):** un bosque aleatorio es un conjunto de árboles de decisión que trabajan que trabajan en conjunto. Cada árbol se entrena con un subconjunto aleatorio de los datos de entrenamiento, y las predicciones se realizan promediando las predicciones individuales de todos los árboles. (Breiman, 2001).

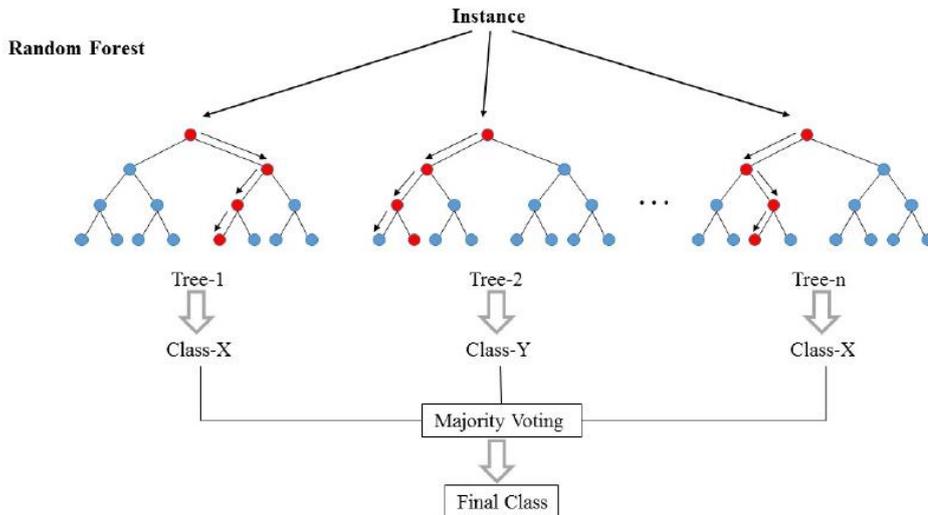


Ilustración 2: Ejemplo gráfico de Bosque Aleatorio. Figura tomada de "ML Algorithms – Pros & Cons and Suitable Usages, por Rajneesh Jha, Analytics Vidhya, 2021, (<https://medium.com/analytics-vidhya/ml-algorithms-pros-cons-and-suitable-usages-b377c3c09f1b>). Copyright por el autor(es).

Se utiliza principalmente para reducir el sobreajuste de los árboles de decisión, utilizando unos parámetros preestablecidos que son bastante sencillos de comprender. Se suele utilizar para tareas de regresión y clasificación, y es útil pues no es necesario normalizar los datos.

Sin embargo, se puede considerar como limitación la complejidad de este modelo. Al crear varios árboles (en la librería Scikit Learn de Python se crean 100 por defecto), se requiere más capacidad computacional y tiempo. Esto también significa que para entrenar un bosque aleatorio se requiere mucho más tiempo que con los árboles de decisión simples, y también comparte el defecto de que un pequeño cambio en los datos de entrenamiento puede crear un gran cambio en las conclusiones.

En el ámbito financiero, los bosques aleatorios se pueden utilizar por ejemplo para la predicción de precios de activos. A través de características como los precios históricos, el volumen de transacciones y otros indicadores se puede hacer una predicción de futuros precios de ese mismo activo u otros similares.

3. Algoritmos de Boosting (XGBoost y LightGBM): los algoritmos de boosting son un tipo de aprendizaje supervisado que combina múltiples modelos débiles (como los árboles de decisión) para crear un modelo más robusto. Cada modelo subsiguiente trata de corregir los errores de su predecesor, y se le más peso a las observaciones que los modelos anteriores predijeron incorrectamente. Intuitivamente, cada nueva iteración se centra en las observaciones más complejas que las iteraciones previas no fueron del todo capaces de predecir, obteniendo al final del proceso un algoritmo más potente con menos sesgo. XGBoost y LightGBM son implementaciones populares de boosting que son altamente eficientes y escalables. XGBoost es la abreviatura de "extreme gradient boosting", y se trata de un sistema de boosting de árboles que mejora el rendimiento de los modelos de aprendizaje automático mediante varias optimizaciones avanzadas. Estas incluyen un algoritmo consciente de la dispersión, un esquema de cuantiles ponderados para el aprendizaje de árboles aproximados, y técnicas de optimización del acceso a la caché y la compresión de datos. XGBoost es un algoritmo aplicable en Python (Chen & Guestrin, 2016).

De la misma manera, LightGBM es un sistema de boosting de árboles de decisión de gradiente que mejora el rendimiento de los modelos de aprendizaje automático mediante varias optimizaciones avanzadas. Estas incluyen Gradient-based One-Side Sampling (GOSS), que reduce el número de instancias de datos necesarias al enfocarse en aquellas con grandes gradientes, y Exclusive Feature Bundling (EFB), que agrupa características mutuamente exclusivas para reducir el número de características efectivas y mejorar la velocidad de entrenamiento. Estas técnicas permiten a LightGBM acelerar significativamente el proceso de entrenamiento mientras mantiene una alta precisión. Este algoritmo es de uso libre y fue originalmente desarrollado por Microsoft. (Ke et al., 2017).

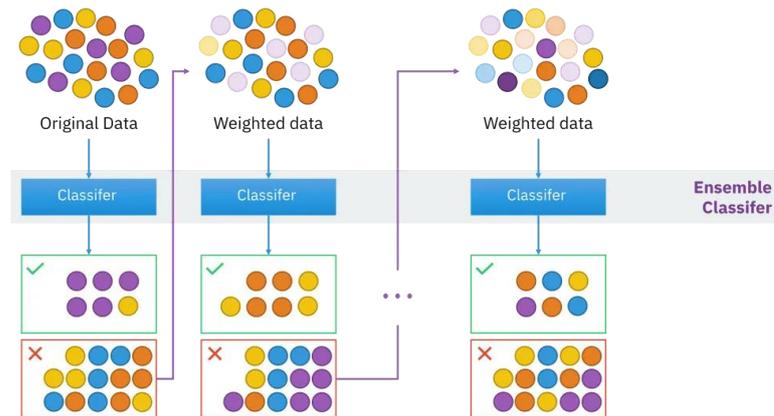


Ilustración 3: Ejemplo gráfico de Boosting. Figura tomada de "ML Algorithms – Pros & Cons and Suitable Usages, por Rajneesh Jha, Analytics Vidhya, 2021, (<https://medium.com/analytics-vidhya/ml-algorithms-pros-cons-and-suitable-usages-b377c3c09f1b>). Copyright por el autor(es).

El boosting ofrece una precisión predictiva que es difícil de superar. También, al no necesitarse pre-procesamiento y que este algoritmo se encarga de rellenar la información restante no incluida en los datos de entrenamiento, proporciona una gran flexibilidad para diferentes usos. Se suele utilizar para tareas complejas y de datasets muy extensos.

Al mismo tiempo, las técnicas de boosting generalmente ignoran el overfitting y las varianzas del dataset. La sensibilidad de los outliers es muy alta debido a que los estimadores basan el nivel de predicción correcta basándose en sus predecesores. La existencia de outliers hace que se extienda el error a lo largo de las iteraciones. También es una limitación la dificultad de escalar estos métodos, debido a que el proceso que utilizan es muy cerrado.

En el ámbito financiero, estos algoritmos tienen aplicaciones similares a los algoritmos previamente mencionados, porque en esencia el boosting es una optimización de los árboles de decisión.

4. Regresión Lineal: la regresión lineal está basada en el método de mínimos cuadrados o el método del error cuadrático medio mínimo. Las ideas en torno a los mínimos cuadrados o a el promedio de errores han evolucionado durante casi tres siglos. La primera publicación conocida de un “Método de Promedios” fue realizada por el astrónomo alemán Tobias Mayer en 1750. Es un método estadístico que modela la relación entre una variable dependiente y una o más variables independientes usando una línea recta. La fórmula general es $y=mx + b$, donde y es la variable dependiente, x

es la variable independiente, m es la pendiente y b es la intersección. La regresión lineal busca ajustar la línea que minimiza la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos.

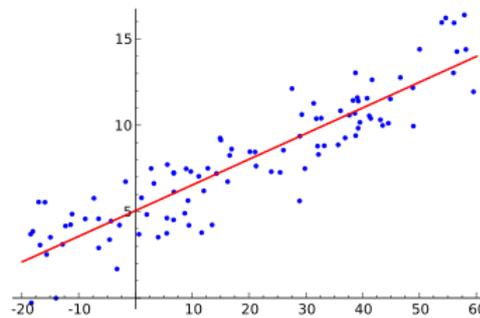


Ilustración 4: Ejemplo gráfico de regresión lineal. Figura tomada de "ML Algorithms – Pros & Cons and Suitable Usages, por Rajneesh Jha, Analytics Vidhya, 2021, (<https://medium.com/analytics-vidhya/ml-algorithms-pros-cons-and-suitable-usages-b377c3c09f1b>). Copyright por el autor(es).

En machine learning, se trata de un algoritmo de aprendizaje supervisado simple y fácil de interpretar a la par de que es bastante eficiente en términos computacionales. Sin embargo, asume relaciones lineales entre variables que no siempre tienen por qué estar relacionadas, y también es bastante sensible a valores atípicos y ruido en los datos.

Este método es muy utilizado en varios aspectos de las finanzas como en la predicción de precios de activos, pues es un método sencillo y aplicable de predecir tendencias futuras.

5. Regresión logística: es una técnica de clasificación que se utiliza para modelar la probabilidad de una variable binaria. Utiliza la función logística para restringir los valores predichos entre 0 y 1. Fue inventada por D.R. Cox en 1958.

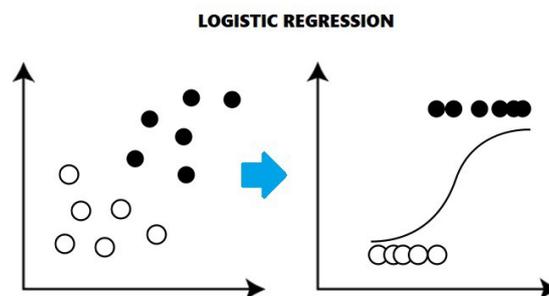


Ilustración 5: Ejemplo gráfico de regresión logística. Figura tomada de "ML Algorithms – Pros & Cons and Suitable Usages, por Rajneesh Jha, Analytics Vidhya, 2021, (<https://medium.com/analytics-vidhya/ml-algorithms-pros-cons-and-suitable-usages-b377c3c09f1b>). Copyright por el autor(es).

Al igual que con la regresión lineal, es un método sencillo de implementar y muy eficiente en términos de computación, a la par que proporciona probabilidades bien calibradas. En cambio, no destaca tanto si se trabaja con datos y relaciones no lineales, y requiere estrictamente que no haya ningún tipo de correlación entre las variables independientes.

Un ejemplo de su aplicación en el ámbito financiero es la clasificación de aprobaciones de crédito. A partir de unas características financieras, este algoritmo puede predecir la probabilidad de que un solicitante sea aprobado para un crédito y

responder a la pregunta de si el solicitante debería recibir el crédito o no a través de un valor entre 0 y 1.

6. Máquinas de Soporte Vectorial (SVM): las SVM son algoritmos de clasificación que buscan el hiperplano que mejor separa las diferentes clases en los datos. Utilizan funciones de kernel para transformar los datos y encontrar el margen máximo entre las clases.

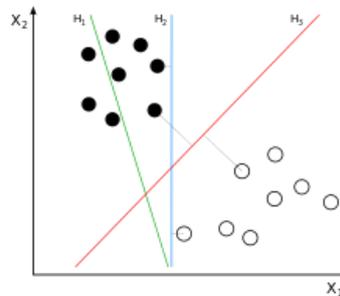


Ilustración 6: Ejemplo gráfico de SVM. Figura tomada de "ML Algorithms – Pros & Cons and Suitable Usages, por Rajneesh Jha, Analytics Vidhya, 2021, (<https://medium.com/analytics-vidhya/ml-algorithms-pros-cons-and-suitable-usages-b377c3c09f1b>). Copyright por el autor(es).

En este ejemplo gráfico se puede observar cómo las separaciones H_1 y H_2 no son capaces de separar las clases de datos mientras que H_3 las separa con el máximo margen. Este método es muy eficaz para clasificar datos que a priori no se sabe que significan, para clasificar datos relacionados en altas dimensiones y destaca también por la capacidad de este algoritmo para manejar datos no lineales mediante el uso de kernels. Por otro lado, este método puede resultar lento para computar conjuntos de datos grandes y al mismo tiempo, este algoritmo presenta la complejidad de interpretar y ajustar los hiperparámetros para realizar una predicción / clasificación precisa.

En el ámbito financiero, este algoritmo se utiliza para predecir quiebras empresariales utilizando características financieras y de rendimiento para clasificar empresas en categorías de alto o bajo riesgo de quiebra.

7. Redes Neuronales: es un programa, o modelo de machine learning que toma decisiones de forma similar al cerebro humano, utilizando procesos que imitan la forma en que las neuronas biológicas trabajan juntas para identificar fenómenos, sopesar opciones y llegar a conclusiones (IBM, s.f.). En concreto, las redes neuronales recurrentes ("RNN") son un tipo de red neuronal en el cual los resultados de las iteraciones previas se van incluyendo en las computaciones de las próximas iteraciones para llegar a un objetivo, obteniendo un algoritmo que aprende automáticamente. Son muy útiles para procesar datos de texto, audios, clasificación en problemas de predicción, regresión en problemas de predicción y modelos generativos. Este algoritmo es de los que contiene más potencial de crecimiento y de creatividad por su adaptabilidad a numerosos tipos de proyectos, con las limitaciones de que son muy complicados de entrenar correctamente.

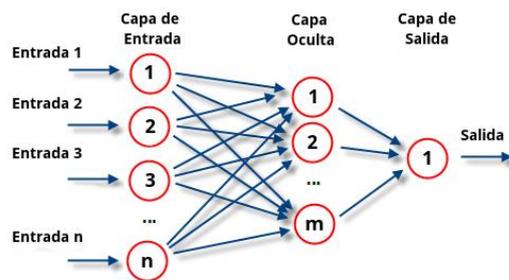


Ilustración 7: Ejemplo gráfico de red neuronal simple. Imagen obtenida de Wikipedia.

En resumen, cada algoritmo de Machine Learning ofrece un conjunto único de ventajas y desventajas, lo que hace que cada ejemplo sea más o menos adecuado dependiendo de la tarea que se requiere. Al comprender como funcionan estos algoritmos internamente y sus aplicaciones prácticas, se pueden crear modelos muy precisos y útiles para distintos ámbitos de las finanzas, lo que hace que el Machine Learning sea una herramienta poderosa a la hora de optimizar la correcta toma de decisiones ya sea en gestión de riesgos, posibles inversiones y eficiencia de operaciones entre otros.

3 EL CAPITAL RIESGO

3.1 Descripción del capital riesgo y su importancia en los mercados financieros

Según el diario Expansión, el capital riesgo es toda actividad financiera cuyo objeto principal es la toma de participaciones temporales, a medio o largo plazo, en el capital de empresas no cotizadas, para favorecer su nacimiento, crecimiento o expansión. El capital riesgo aporta a la empresa, además de recursos propios, apoyo y participación en la gestión empresarial. El objetivo final es obtener una rentabilidad para el inversor vía plusvalía, cuando se venda dicha participación empresarial. En España, el término capital riesgo engloba todas las fases de inversión, desde las iniciales (semilla y arranque), el capital expansión, o las operaciones apalancadas (“buy outs”). Es decir, engloba lo que en otros países llaman “venture capital” (inversión en primeras fases) y “private equity” (inversión en fases consolidadas) (Expansión, s.f.).

Por consolidar la definición, el capital riesgo es una forma de financiamiento proporcionada por inversores a empresas emergentes o no cotizadas, en las que el inversor adquiere participaciones en el capital de la empresa asumiendo un riesgo significativo con la esperanza de obtener un alto rendimiento.

Con el objetivo de aclarar posibles confusiones, de aquí en adelante cuando se refiera al término capital riesgo, se hace referencia al término inglés “venture capital” y todo lo relacionado con este, dejando de lado el “private equity”, que también se incluye como parte del capital riesgo en la definición española.

Esta actividad es de gran importancia en los mercados financieros por varias razones. En cuanto a las empresas emergentes, el término capital riesgo y el término innovación suelen ir de la mano pues generalmente las empresas más innovadoras son aquellas que cargan con un mayor riesgo pues suelen ser disruptivas. Empresas como Google, Facebook y Amazon recibieron capital riesgo en sus primeras etapas, lo que les permitió desarrollar y escalar sus

innovaciones. Gracias al capital riesgo, estas compañías han podido desarrollarse, contribuyendo al crecimiento económico, impulsando el desarrollo tecnológico y económico y generando miles de puestos de trabajo. Otro aspecto importante es el de la mitigación de riesgos; los inversores de capital riesgo asumen riesgos que otros inversores tradicionales no están dispuestos a tomar. Esto es crucial para la viabilidad de aquellas empresas prometedoras que presentan altos niveles de incertidumbre (Nexford University, 2024).

3.2 Estructura y fases de inversión en capital riesgo

La estructura del capital riesgo se puede desglosar en ocho fases clave que guían el proceso de inversión desde la identificación de oportunidades hasta la salida del inversor. Según Sahlman (1990), estas fases incluyen:

1. Inversión semilla (“Seed Investment”): se refiere a la provisión de una pequeña cantidad de capital a un inventor o emprendedor para determinar si una idea merece consideración y financiamiento adicional. Puede involucrar la construcción de un pequeño prototipo y no incluye producción para la venta (Sahlman, 1990, p. 7).
2. Inversión Startup: esta etapa va dirigida comúnmente a empresas que tienen menos de un año de antigüedad. Los fondos se utilizan para el desarrollo de productos, pruebas de prototipos y marketing experimental. También se enfocan en estudiar el potencial de penetración en el mercado, reunir un equipo de gestión y refinar el plan de negocios (Sahlman, 1990, p. 7).
3. Primera Etapa - Desarrollo Temprano (“First Stage - Early Development”): esta etapa procede solo si los prototipos son prometedores y los estudios de mercado indican viabilidad. La empresa configura un proceso de manufactura modesto y comienza a enviar productos en cantidades comerciales. Las empresas en esta etapa suelen no ser rentables (Sahlman, 1990, p. 7).
4. Segunda Etapa - Expansión (“Second Stage – Expansion”): en esta fase la empresa ha distribuido suficientes productos a suficientes clientes para recibir retroalimentación del mercado. Aunque es probable que aún no sea rentable o solo marginalmente rentable, la empresa necesita más capital para comprar equipo, inventarios y financiar cuentas por cobrar (Sahlman, 1990, p. 7).
5. Tercera Etapa - Rentable pero con Poca Liquidez (“Third Stage - Profitable but Cash Poor”): las empresas en esta etapa experimentan un crecimiento rápido en ventas y márgenes de beneficio positivos. Sin embargo, la rápida expansión requiere más capital de trabajo del que pueden generar internamente. El nuevo capital de VC puede ser utilizado para la expansión adicional de instalaciones, marketing y mejoras de productos (Sahlman, 1990, p. 7).
6. Cuarta Etapa - Crecimiento Rápido hacia el Punto de Liquidez (“Fourth Stage - Rapid Growth Towards Liquidity Point”): las empresas necesitan capital externo para sostener su crecimiento, pero son lo suficientemente exitosas y estables como para que el riesgo para los inversores externos sea reducido. La empresa puede preferir usar más financiamiento de deuda para limitar la dilución de capital (Sahlman, 1990, p. 7).
7. Etapa de Puente - Inversión Mezzanine (“Bridge Stage - Mezzanine Investment”): la empresa puede tener una idea de qué forma de salida es más probable (oferta pública de venta, adquisición o compra apalancada) y conoce el tiempo aproximado, pero necesita más capital para sostener el crecimiento mientras tanto. Esta financiación

puede corresponder a una liquidación limitada de inversores iniciales o una reestructuración de posiciones entre inversores de VC (Sahlman, 1990, p. 7).

8. Etapas de Liquidez - Salida (“Liquidity Stage - Cash-Out or Exit”): en esta última fase, los inversores de VC buscan obtener liquidez para una parte sustancial de sus participaciones. Esto puede venir en forma de una oferta pública inicial (IPO) o mediante una adquisición. La liquidez puede venir en forma de efectivo, acciones en una empresa que cotiza en bolsa, o deuda a corto plazo, situación en la que el inversor termina con una posición menos líquida que la anterior (Sahlman, 1990, p. 7).

3.3 Medición y gestión de riesgos en capital riesgo

La medición y gestión de riesgos son fundamentales para asegurar la estabilidad y el éxito a largo plazo de las empresas participadas. Este apartado combina los hallazgos de los estudios de Reid, Terry y Smith (1995) y Proksch, Stranz, Pinkwart y Schefczyk (2016) para proporcionar una visión completa de las estrategias de gestión de riesgos en la industria del capital riesgo. Se trata primero los tipos de riesgo que hay en este tipo de inversión, para luego abordar las estrategias de medición y gestión de riesgos.

3.3.1 Tipos de Riesgos en Capital Riesgo

1. Riesgo de Agencia: se refiere a la posibilidad de que la empresa se vuelva ilíquida o incluso quiebre. Está considerado como uno de los principales riesgos de capital riesgo debido a los potenciales problemas de selección adversa y de moral entre los inversores y los propietarios de la empresa en cuestión. Para evitar este posible conflicto de intereses, existen mecanismos de intervención como la contratación financiera o la provisión gradual de capital (Proksch et al., 2016, p. 7).
2. Riesgo Financiero / Liquidez: se refiere a la posibilidad de que la empresa se vuelva ilíquida o incluso quiebre. Las decisiones de inversión se ajustan para mitigar este riesgo, prefiriendo inversiones en tecnología avanzada y etapas iniciales (Proksch et al., 2016, p. 8).
3. Riesgo Tecnológico: en ocasiones sinónimo de riesgo de producto, evaluar el riesgo asociado con la tecnología o el producto es crucial antes de invertir, dado que muchas tecnologías aún no están listas para el mercado. Se realizan investigaciones (“due diligence”) tecnológicas y se consulta a expertos del sector para mitigar este riesgo lo máximo posible (Proksch et al., 2016, p. 8).
4. Riesgo de Mercado: este riesgo está relacionado con la comercialización de nuevas tecnologías. En ocasiones, las empresas de capital riesgo carecen de la capacidad de marketing necesaria para lanzar un producto nuevo al mercado. Los inversores aplican investigaciones (“due diligence”) de mercado y utilizan su red de contactos para facilitar la introducción en el mercado y mitigar posibles riesgos de tecnología y de mercado (Proksch et al., 2016, p. 8).
5. Riesgo de recursos humanos: asociado con la calidad y capacidades del equipo directivo de la empresa participada. Los inversores verifican el historial del equipo directivo y prefieren equipos conocidos previamente. Se observan distintos factores, como la personalidad del CEO, la experiencia previa de los empleados, historial de gestión de empresas, falta de personal cualificado para gestionar y otras capacidades humanas (Proksch et al., 2016, p. 8).

3.3.2 Estrategias de Medición de Riesgos

1. Rendimientos Esperados: son una medida clave en la evaluación del riesgo. Tanto inversores como participadas reconocen la exposición significativa a riesgos. Los inversores suelen manejar estos riesgos mediante un análisis detallado y continuo de las proyecciones de retorno y las probabilidades asociadas. La mayoría de los inversores no se involucran en proyectos con alta probabilidad de fracaso, prefiriendo aquellos con un equilibrio adecuado entre riesgo y retorno (Reid, Terry, & Smith, 1995, p. 10).
2. Balance de Cartera: es una estrategia crucial para la gestión de riesgos. Los inversores buscan equilibrar los buenos y malos resultados mediante la diversificación de sus inversiones, evitando concentraciones en un solo sector o tecnología. Esta diversificación ayuda a mitigar el riesgo sistémico y específico de cada inversión individual (Reid et al., 1995, p. 13).
3. Selección de Inversiones: la selección rigurosa de inversiones es esencial para reducir la selección adversa. Los inversores emplean análisis de riesgo-retorno y otros métodos de evaluación para seleccionar las inversiones que ofrezcan un potencial de alto retorno con riesgos manejables. Este proceso implica una revisión detallada del plan de negocio, la validación de proyecciones y la verificación de la experiencia del equipo directivo de la empresa participada (Reid et al., 1995, p. 16).

3.3.3 Estrategias de Gestión de Riesgos

1. Compartición de Riesgos: la compartición de riesgos es una estrategia clave en la relación entre inversores y empresas participadas. Los contratos de capital riesgo están diseñados para alinear los intereses de ambas partes, incentivando el esfuerzo y la gestión eficaz de la empresa. Los inversores buscan estructuras contractuales que motiven a los gestores a maximizar el valor de la empresa mientras comparten el riesgo asociado a la inversión (Reid et al., 1995, p. 20).
2. Monitoreo y Control: una vez realizada la inversión, los inversores implementan mecanismos de monitoreo y control para asegurar que la empresa participada sigue el rumbo previsto. Esto incluye la participación en juntas directivas, revisiones periódicas de desempeño y la implementación de sistemas de reporte financiero y operativo. Estos mecanismos permiten a los inversores intervenir y ajustar estrategias cuando sea necesario para proteger su inversión (Reid et al., 1995, p. 23).
3. Uso de Hitos: los hitos o etapas de logro son utilizados para gestionar el riesgo y asegurar el cumplimiento de objetivos críticos. Los desembolsos de capital suelen estar condicionados a la consecución de hitos específicos, lo que asegura que la empresa participada está progresando según lo planeado. Este enfoque escalonado permite a los inversores minimizar el riesgo de fracaso en etapas tempranas del desarrollo del negocio, y va de la mano de la gestión del área de gobernanza (Proksch et al., 2016, p. 15).

4 MACHINE LEARNING EN FINANZAS

Una vez entendido el contexto y el estado del arte actual de ambos machine learning y capital riesgo por separado, esta sección trata de arrojar luz sobre el estado del arte de la combinación de estas dos disciplinas, comenzando por las finanzas u inversión en activos más tradicionales. Al ser la aplicación del ML en el ámbito del capital riesgo un reto ciertamente novedoso pero con mucho potencial, este apartado explora la integración de técnicas de machine learning en el ámbito de las finanzas, para posteriormente focalizarse en el capital riesgo, destacando estudios previos que demuestran su eficacia, análisis de resultados obtenidos en aplicaciones específicas y las áreas potenciales de aplicación que pueden beneficiarse de su implementación. Además, se discuten los beneficios que el uso de ML puede aportar a la industria del capital riesgo, mejorando la precisión y eficiencia de las decisiones de inversión. Al examinar cómo el machine learning puede transformar las prácticas tradicionales de due diligence en finanzas y capital riesgo, se proporciona una comprensión profunda de cómo estas tecnologías avanzadas pueden ofrecer ventajas competitivas significativas. La adopción de ML no solo promete mejorar la precisión de las predicciones y la gestión de riesgos, sino que también abre nuevas oportunidades para innovar en la estrategia de inversión, asegurando que las empresas inversoras en capital riesgo puedan adaptarse y prosperar en un entorno financiero cada vez más complejo y dinámico.

4.1 Estudios previos sobre la aplicación de machine learning en inversiones y finanzas

Para entender el estado del arte actual, se han escogido de entre un gran número de documentos y estudios, los tres que se consideran más relevantes en cuanto a importancia de la materia, correlación con este estudio y fecha de publicación.

1. “AI and Big Data Investments: Handbook by CFA Institute” (2023): este manual proporciona una exploración exhaustiva de como la inteligencia artificial y el big data están siendo integrados en los procesos de inversión, ofreciendo una guía práctica para los profesionales del sector.

La relevancia de este documento para el presente trabajo radica en su enfoque detallado y práctico sobre la implementación de ML y big data en el ámbito de las inversiones. En el contexto del capital riesgo, donde la toma de decisiones se basa en la evaluación de riesgos y oportunidades en startups y empresas emergentes, las técnicas de ML pueden proporcionar ventajas significativas. El documento del CFA Institute aborda integralmente las diversas aplicaciones de ML en finanzas, desde la predicción de precios hasta la optimización de carteras, lo cual es directamente aplicable a la gestión de inversiones en capital riesgo.

Este manual destaca varias áreas clave donde el Machine Learning ha resultado ser efectivo:

- a) Predicción de Retornos de Acciones: utilizando datos históricos de precios y volúmenes de transacciones, estos se introducen para entrenar modelos de Machine Learning como redes neuronales y árboles de decisión para identificar patrones y tendencias. Estos modelos se validan con técnicas como la validación cruzada y se ajustan progresivamente para mejorar la precisión de las predicciones. Estos modelos demostraron ser más precisos en la predicción de los retornos de acciones en comparación con los métodos tradicionales. Este enfoque está siendo adoptado actualmente por numerosas firmas de inversión para mejorar sus

estrategias de trading y para monitorear las estrategias actuales. En el estudio del CFA Institute, se informa de la obtención de cinco resultados principales obtenidos, siendo los siguientes:

- 1) Los algoritmos de ML superan significativamente a los enfoques lineales tradicionales, como se muestra en el estudio donde una red neuronal de cuatro capas logra una ratio de Sharpe anualizada de 1.35 bajo ponderación por valor y de 2.45 bajo ponderación igualitaria, en comparación con 0.61 y 0.83 para predicciones basadas en mínimos cuadrados ordinarios (OLS).
 - 2) Los algoritmos de ML pueden formular dependencias funcionales de orden superior, como la no linealidad y la interacción, lo que mejora el rendimiento predictivo en comparación con los modelos lineales simples.
 - 3) Las características clave como las reversiones a corto plazo, el momentum de retorno de acciones y sectores, la volatilidad del retorno y el tamaño de la empresa son consistentemente importantes para los resultados de predicción, independientemente del algoritmo de ML utilizado.
 - 4) Los algoritmos de ML simples, debido a su menor propensión al sobreajuste, superan a los modelos más complicados en escenarios con datos limitados, como se confirma con el mejor rendimiento fuera de la muestra de una red neuronal con cuatro capas ocultas.
 - 5) La cantidad de datos es fundamental para el rendimiento de los algoritmos de ML, con mejores resultados predictivos cuanto mayor es el volumen de datos disponible, lo que ha sido confirmado en aplicaciones tanto generales como financieras de ML (Cao, 2023, p. 33).
- b) Optimización de Carteras: a través de la utilización de rendimientos históricos de activos, correlaciones entre activos y otros datos financieros relevantes se crean algoritmos de clustering para identificar grupos (o “clusters”) de activos que se comportan de una manera similar. A través de reducción de dimensionalidad se identifican los factores subyacentes que explican las variaciones en los rendimientos. Estos factores se utilizan para diversificar la cartera y minimizar los riesgos específicos del mercado.
A través de este enfoque, los gestores de carteras pueden optimizar la distribución de activos, maximizando el binomio rentabilidad – riesgo. Las simulaciones de escenarios y ajustes continuos son fundamentales para realizar este proceso correctamente.
- c) Análisis de Sentimiento del Mercado: mediante la recolección de textos de noticias financieras, publicaciones en redes sociales y transcripciones de conferencias, se utilizan modelos de procesamiento de lenguaje natural (NLP) para extraer de estos textos el sentimiento general del mercado (positivo, negativo, neutral). Estos datos obtenidos del análisis de sentimiento se integran en modelos predictivos de mercado para anticipar movimientos basados en las emociones y opiniones predominantes. El estudio indica que la inclusión del análisis de sentimiento mejora la precisión de las predicciones del mercado, proporcionando a los inversores una ventaja competitiva en la toma de decisiones, no solo en cuanto a retorno sino también en otras áreas de relevancia como la gobernanza o el ESG.

En resumen, este documento es altamente relevante para el presente trabajo pues ofrece una exploración exhaustiva de cómo las técnicas de ML y big data están siendo implementadas en el sector de inversiones, proporcionando una base sólida para comprender las aplicaciones prácticas de estas tecnologías. La capacidad del ML para analizar grandes volúmenes de datos, identificar patrones ocultos y realizar predicciones precisas es particularmente valiosa en el capital riesgo, donde las decisiones deben basarse en proyecciones precisas del rendimiento futuro de las inversiones. Los estudios revisados en este documento no solo ilustran el potencial de ML para transformar las prácticas tradicionales de inversión, sino que también proporcionan directrices prácticas y metodológicas que pueden ser aplicadas en el contexto del capital riesgo.

2. “Machine Learning Methods in Finance: Recent Applications and Prospects” (2022): este documento publicado en 2022 en *European Financial Management*, examina cómo se pueden aplicar los distintos métodos de ML en el ámbito financiero, ofreciendo casos prácticos. El estudio tiene como objetivo explorar aplicaciones actuales y futuras del machine learning en la investigación financiera, comparando estas con enfoques econométricos tradicionales. Los autores dividen las aplicaciones de ML en tres categorías principales: construcción de medidas superiores y noveladas, reducción del error de predicción, y extensión del conjunto de herramientas econométricas. Se proporcionan ejemplos concretos de cómo el ML mejora la precisión y eficacia en varios contextos financieros.

Por ejemplo, se muestra un caso de predicción de precios de bienes raíces en Alemania, se utilizan métodos de ML como árboles de regresión potenciados para predecir los precios utilizando un conjunto de datos con más de cuatro millones de listados de propiedades, duplicando la precisión en comparación con la regresión lineal (OLS) (Hoang & Wiegatz, 2022, p. 1659-1660).

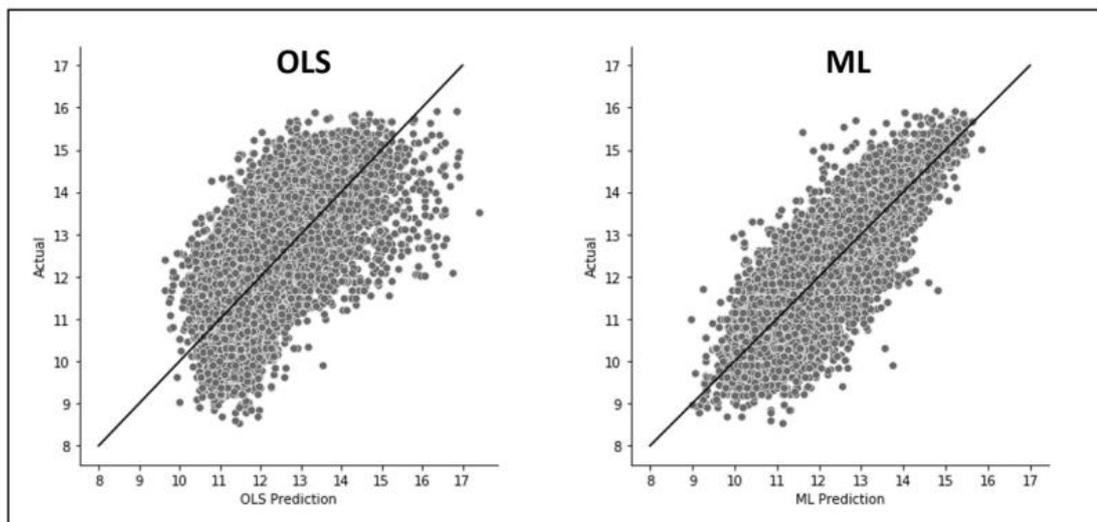


Ilustración 8: Figura tomada de "Machine learning methods in finance: Recent applications and prospects", por D. Hoang y K. Wiegatz, 2022, *European Financial Management*, 29(5), p. 1660 (<https://doi.org/10.1111/eufm.12408>). Copyright 2022 por el autor(es).

En la ilustración se puede observar el resultado de la desviación de precio entre el precio real de cada bien raíz frente a la predicción realizada por una regresión lineal de mínimos cuadrados (gráfico de la izquierda) frente a las predicciones realizadas por el algoritmo de ML utilizado en el estudio (gráfico de la derecha). Se aprecian mejores predicciones a través del uso de ML, especialmente en la parte superior derecha de la gráfica, la cual representa activos de precios más caros.

Otro ejemplo incluido es la construcción de medidas de sentimiento del mercado mediante técnicas de procesamiento de lenguaje natural (NLP) aplicadas a datos textuales de redes sociales, noticias y reportes de analistas, mejorando significativamente la precisión en la predicción de retornos de acciones y la volatilidad del mercado sobre los enfoques tradicionales basados en diccionarios. El documento destaca varios beneficios de usar ML en finanzas, como la mejora de la precisión de predicción, la reducción de errores de medición y la exploración de nuevas variables, proporcionando nuevas perspectivas y variables previamente inmedibles en la investigación financiera. Sin embargo, también se reconocen desafíos significativos, como la difícil interpretación de los modelos de ML, la necesidad de grandes conjuntos de datos y los altos costos computacionales asociados con el entrenamiento de modelos avanzados.

En conclusión, el uso de ML en finanzas ofrece una serie de ventajas que pueden revolucionar la precisión y la eficiencia de las investigaciones financieras, aunque no sin enfrentar ciertos desafíos técnicos y prácticos (Hoang & Wiegratz, 2022).

3. “Financial applications of machine learning: A literature review” (2023): este documento publicado en la revista Expert Systems with Applications revisa y examina la literatura reciente sobre la convergencia de dos campos diversos: el machine learning y sus aplicaciones en finanzas. A medida que el uso del machine learning en las finanzas crece constantemente, se revisan en detalle seis aplicaciones principales que demuestran el poder de esta tecnología: predicción del mercado de valores, gestión de carteras, criptomonedas, mercado de divisas, crisis financieras, y predicción de quiebras e insolvencias. Tras analizar 126 artículos publicados en 44 revistas reconocidas desde 2015, discutiendo características, conjuntos de datos, enfoques de validación, modelos y métricas de rendimiento utilizados en cada aplicación financiera el estudio concluye que la aplicación del machine learning en los mercados de valores ha demostrado un rendimiento predictivo efectivo que refuta la teoría de la hipótesis de los mercados eficientes. Los modelos como redes neuronales (ANN, FFNN, BPNN) y máquinas de soporte vectorial (SVM) son los más utilizados, incorporando precios históricos, indicadores técnicos y datos textuales de noticias, blogs y redes sociales.

En gestión de carteras, las aplicaciones de ML han superado los modelos de referencia al obtener mayores rendimientos. Recientemente, se ha utilizado el aprendizaje por refuerzo para crear agentes comerciales que operan racionalmente como los humanos. En el ámbito de las criptomonedas, aunque la aplicación de ML está en su infancia y falta supervisión de mercado, hay un creciente interés por predecir el comportamiento del mercado. La predicción de divisas, especialmente del par EUR/USD, ha sido exitosa gracias a la capacidad del ML para modelar series temporales no lineales.

Como conclusión, el estudio decreta que cada modelo tiene ventajas y desventajas dependiendo de las circunstancias. Aunque hay instancias donde los modelos estadísticos superan a los de ML, la tendencia general favorece a los modelos de ML, incluyendo deep learning, híbridos y ensamblajes. Hay un amplio margen para implementar ML y sus subgrupos en finanzas (Expert Systems with Applications, 2023).

4.2 Análisis de resultados, interpretación y relevancia de los hallazgos

El análisis de estos estudios relevantes sobre la aplicación de machine learning (ML) en las finanzas revela hallazgos que subrayan la transformación y las oportunidades significativas que estas tecnologías ofrecen en el sector financiero. Los resultados obtenidos destacan la mejora en la precisión y eficiencia en varias aplicaciones financieras, a la par que el potencial del ML para revolucionar el capital riesgo. A continuación, se interpretan estos resultados y se discute su importancia en el contexto del capital riesgo a partir de las principales secciones de las finanzas en las que se enfocan principalmente estos estudios.

Con respecto a la predicción de valores, los estudios muestran que los algoritmos de ML, como las redes neuronales y las máquinas de soporte vectorial (SVM), superan significativamente a los métodos tradicionales como la regresión lineal (OLS) en la labor de predecir correctamente distintos tipos de activos. Esta capacidad de ML para manejar grandes volúmenes de datos y capturar patrones complejos es crucial en el capital riesgo, donde la precisión en la evaluación de startups y empresas emergentes puede determinar el éxito o el fracaso de las inversiones.

Para la gestión de carteras y distribución de activos, utilizando rendimientos históricos y correlaciones entre activos, los algoritmos de clustering pueden identificar grupos de activos con comportamientos similares, permitiendo una diversificación efectiva y minimización de riesgos específicos del mercado. Para los inversores de capital riesgo, esta capacidad de optimizar la distribución de activos y maximizar el rendimiento ajustado por riesgo es fundamental para gestionar las inversiones realizadas.

Para el sentimiento del mercado, usando procesamiento de lenguaje natural (NLP) para analizar textos de noticias financieras, redes sociales y transcripciones de conferencias, los modelos de ML pueden extraer el sentimiento general del mercado. En el contexto del capital riesgo, el análisis de sentimiento puede proporcionar a los inversores una ventaja competitiva, permitiendo identificar oportunidades de inversión emergentes y anticipar cambios en el mercado basados en las percepciones del público y los expertos.

Tras analizar los resultados, la mayor relevancia de estos hallazgos nos indica que la adopción de ML en el capital riesgo tiene el potencial de transformar las prácticas tradicionales de inversión, proporcionando herramientas avanzadas para la evaluación de riesgos y oportunidades. La capacidad de ML para analizar grandes volúmenes de datos, identificar patrones ocultos y realizar predicciones precisas es especialmente valiosa en el contexto del capital riesgo, donde las decisiones deben basarse en proyecciones detalladas del rendimiento futuro de las inversiones. La implementación del ML en el sector del capital riesgo presenta el potencial de ayudar a los inversores a la hora de optimizar la selección de inversiones,

gestionar las carteras de inversión, anticipar movimientos del mercado y mitigar riesgos financieros entre otros.

5 MACHINE LEARNING EN CAPITAL RIESGO

En las secciones anteriores de este trabajo, se ha explorado en profundidad tanto el concepto y la importancia del capital riesgo como el impacto revolucionario del machine learning en el ámbito financiero. El capital riesgo, con su enfoque en financiar startups y empresas emergentes, puede beneficiarse enormemente de estas herramientas avanzadas que mejoran la toma de decisiones y la gestión de riesgos. Al mismo tiempo, hemos visto cómo el machine learning, con su capacidad para analizar grandes volúmenes de datos y hacer predicciones precisas, ofrece un potencial significativo para transformar estas prácticas de inversión. En este contexto, la siguiente sección del trabajo se centrará en las potenciales aplicaciones prácticas del ML en la industria del capital riesgo, cómo estas técnicas pueden implementarse en la realidad y ofreciendo ejemplos concretos y estudios de caso que ilustran el impacto de estas técnicas en el sector.

5.1 Potenciales áreas de aplicación

Como se ha visto en la sección de este trabajo que explica el capital riesgo, su estructura, fases de inversión y riesgos asociados, los inversores de capital riesgo pueden aplicar técnicas de machine learning en todas estas fases para optimizar el proceso de inversión y las probabilidades de éxito.

Observando todas las áreas y proceso que siguen los inversores, podemos asumir que toda inversión de capital riesgo sigue el siguiente curso:



Ilustración 9: Cronología de una inversión en Capital Riesgo

Uniendo el conocimiento de esta cronología con las técnicas de Machine Learning que ya se están aplicando en otros ámbitos de las finanzas, surgen distintas aplicaciones tangibles que pueden ayudar al inversor de capital riesgo, cada una asociada a cada una de estas fases:

1. **Búsqueda de Inversiones:** la primera fase de una inversión de capital riesgo consiste en encontrar empresas o startups con potencial de crecimiento. Para asistir a un inversor de capital riesgo a encontrar inversiones atractivas, se pueden utilizar técnicas de machine learning para facilitar al inversor en su búsqueda de inversiones que se ajusten a sus necesidades. Actualmente existen múltiples plataformas como AngelList, StartEngine o SeedInvest, que conectan a los emprendedores con los inversores. No obstante, ninguna de estas plataformas web proporciona un filtro o discernimiento avanzado que encuentre inversiones ajustadas para las necesidades específicas de cada inversor. Utilizando técnicas de ML como las redes neuronales, árboles de decisión / bosques aleatorios y algoritmos de clustering (todos algoritmos de ML que brillan por su capacidad para clasificar entre otras), se puede crear un radar de búsqueda de inversiones adaptable a las necesidades específicas de los inversores. Utilizando una

base de datos extensa de startups/inversiones históricas se pueden encontrar patrones de éxito y variables relevantes que faciliten este proceso.

2. Análisis o Due Diligence Individual: una vez que un inversor de capital riesgo ha identificado una startup o inversión potencialmente interesante, el siguiente paso crucial es llevar a cabo un análisis detallado o due diligence, proceso que implica evaluar exhaustivamente varios aspectos de la empresa para determinar su viabilidad y potencial de crecimiento. El machine learning puede desempeñar un papel vital en esta fase, proporcionando herramientas avanzadas para analizar grandes cantidades de datos y extraer información relevante que facilite la toma de decisiones.

Entre las posibles aplicaciones para esta sección, una propuesta atractiva es el análisis financiero automatizado, impulsado por modelos de machine learning como redes neuronales o máquinas de soporte vectorial, entrenados para analizar estados financieros, identificar tendencias y detectar anomalías. También se pueden analizar datos de mercado, industria y competidores para evaluar la posición de la startup en su sector a través del uso de algoritmos de clustering. Por último, para obtener contexto extra, se puede integrar un análisis de sentimiento y reputación utilizando un algoritmo de NLP que procese menciones en redes sociales, reseñas, y noticias para evaluar la percepción pública y la reputación de la empresa analizada, a la par que una evaluación del equipo fundador que analice el historial y el desempeño del equipo fundador utilizando datos disponibles públicamente y redes de profesionales.

3. Inversión acorde a la Etapa: como se explicó anteriormente en el apartado 3.2 Estructura y fases de inversión en Capital Riesgo, las inversiones pasan por diversas fases, desde la inversión semilla hasta la etapa de liquidez. Cada etapa presenta sus propios desafíos y oportunidades, y las decisiones acertadas en cada fase son cruciales para maximizar el retorno de la inversión y minimizar los riesgos. El machine learning puede jugar un papel fundamental en este proceso, proporcionando herramientas avanzadas para analizar datos y apoyar la toma de decisiones en cada una de estas etapas. En las fases iniciales (inversión semilla, etapa de crecimiento...) los inversores se encuentran con empresas con potencial disruptivo pero que aún no han probado su modelo de negocio en el mercado. Los modelos de ML como las redes neuronales, los modelos de procesamiento de lenguaje natural y algoritmos de clustering pueden analizar grandes cantidades de datos no estructurados, como propuestas de negocio, planes de producto y menciones en redes sociales, para identificar patrones de éxito temprano, utilizando como ejemplo datos históricos de empresas de capital riesgo que tuvieron éxito en su progreso. Para etapas más avanzadas, estos mismos algoritmos pueden arrojar luz sobre el procedimiento recomendado a seguir para el inversor en función de lo involucrado que esté en el negocio y el desarrollo de la situación de la empresa. La integración de técnicas de machine learning orientadas a las distintas fases de inversión dentro del capital riesgo puede permitir a los inversores tomar decisiones más informadas y estratégicas. Al aprovechar el poder del ML para analizar datos complejos y hacer predicciones precisas, los inversores pueden identificar las mejores oportunidades de inversión en cada etapa, optimizar la asignación de recursos y maximizar el retorno de la inversión. Esta capacidad es especialmente crucial en el

capital riesgo, donde las decisiones acertadas pueden significar la diferencia entre el éxito y el fracaso.

4. Gestión de Riesgos: como se describe en el punto 3.3 de este trabajo, la gestión de riesgos es una parte integral del proceso de inversión en capital riesgo, ya que los inversores deben evaluar y mitigar una variedad de riesgos que pueden afectar el éxito de sus inversiones. El machine learning puede ser una herramienta poderosa para gestionar estos riesgos, proporcionando análisis avanzados y predicciones precisas para apoyar la toma de decisiones informadas una vez la inversión ya está realizada. Para mitigar riesgos de mercado, riesgos financieros, riesgos operativos y riesgos de liquidez, se pueden aplicar diversas técnicas de machine learning. Por ejemplo, los modelos de redes neuronales y Long Short-Term Memory (LSTM) pueden predecir tendencias y volatilidades futuras del mercado, mientras que los modelos de procesamiento de lenguaje natural (NLP) pueden analizar el sentimiento del mercado a partir de noticias y redes sociales. Para evaluar la salud financiera de una startup, algoritmos como las máquinas de soporte vectorial (SVM) y los bosques aleatorios pueden identificar patrones en los datos financieros que sugieran problemas potenciales. Además, los algoritmos de clustering pueden monitorear y analizar procesos operativos para detectar ineficiencias y áreas de mejora, y los modelos de regresión pueden predecir necesidades futuras de liquidez basándose en el análisis de flujo de caja y otros indicadores financieros. Integrando estas técnicas, los inversores pueden anticipar y mitigar diversos tipos de riesgos, protegiendo sus inversiones y maximizando los retornos.
5. Venta: por último, en el ciclo de vida de una inversión de capital riesgo, la fase de venta es crucial para maximizar los retornos. Esta fase implica la salida de la inversión, ya sea a través de una oferta pública inicial (IPO), una venta a otra empresa o inversores, o mediante una recompra por parte de la startup. El machine learning puede desempeñar un papel vital en la optimización de este proceso, proporcionando análisis y predicciones para apoyar la toma de decisiones informadas. Por ejemplo, los modelos de regresión y las máquinas de soporte vectorial (SVM) pueden predecir la valoración futura de la startup en función de las tendencias de mercado y el desempeño financiero. Los algoritmos de clustering pueden segmentar potenciales compradores o mercados objetivo, ayudando a identificar las mejores oportunidades de venta. Además, las técnicas de procesamiento de lenguaje natural (NLP) pueden analizar el sentimiento del mercado y las percepciones públicas para determinar el momento óptimo para la venta. Los modelos de redes neuronales pueden simular diferentes escenarios de salida para evaluar el impacto potencial en los retornos. Al integrar estas técnicas, los inversores pueden optimizar el proceso de venta, asegurando una salida rentable y estratégica de sus inversiones.

5.2 Metodología de creación y aplicación de modelos

Una vez descritos todos estos escenarios de potenciales aplicaciones e integración de modelos de machine learning en las distintas etapas de una inversión de capital riesgo, se puede observar que existe cierta correlación entre estas potenciales aplicaciones y las aplicaciones ya existentes en otros ámbitos más generalizados de las finanzas. Al no existir disponibilidad pública de modelos reales en materia de capital riesgo, en esta sección se explica la metodología y proceso a seguir para crear un modelo realista como los explicados en el

anterior punto de este trabajo, ya que todos ellos requieren el mismo proceso a seguir para materializarlos y obtener resultados tangibles. Consiste en los siguientes pasos:

1. Entrada de Datos / Base de Datos: la primera fase en la creación de un modelo de machine learning es la recopilación de datos relevantes. Para el capital riesgo, esto incluye datos históricos y actuales sobre startups, indicadores financieros, métricas de rendimiento, información del mercado, y datos cualitativos como menciones en medios y redes sociales. Las fuentes de datos necesarias pueden incluir bases de datos financieras, registros de empresas, plataformas de inversión (como AngelList o Crunchbase), y fuentes de datos no estructurados como redes sociales. Una vez obtenidos todos los datos que consideremos relevantes y necesarios, es necesario organizar y almacenar los datos en un formato adecuado para su análisis. Esto puede incluir el uso de bases de datos relacionales, almacenes de datos en la nube, o sistemas de almacenamiento de big data.
2. Preprocesamiento: este paso es crucial para preparar los datos recopilados para el análisis y modelado. Implica limpiar, transformar y estructurar los datos para asegurarse de que sean adecuados para el uso en algoritmos de machine learning. Se comienza por la limpieza de datos, eliminando datos duplicados, corrigiendo errores, manejando valores faltantes y eliminando outliers que puedan sesgar el análisis. A continuación, si es necesario debemos convertir datos no estructurados en un formato estructurado adecuado para el análisis. Esto puede incluir la tokenización de texto, normalización de datos numéricos y codificación de variables categóricas. El siguiente paso es identificar y seleccionar las características más relevantes que influirán en el modelo. Esto puede implicar técnicas de reducción de dimensionalidad como PCA (Análisis de Componentes Principales). Por último, es necesario dividir los datos en conjuntos de entrenamiento, validación y prueba para evaluar la precisión del modelo. Dependiendo de lo rigurosa que sea nuestra base de datos, como preparemos los datos utilizados, y que variables incluyamos teniendo en cuenta su ponderación, la precisión de nuestro modelo variará sustancialmente. Un gran modelo sin buenos datos de input o con sesgo resulta inservible.
3. Modelado: implica seleccionar y entrenar algoritmos de machine learning para crear un modelo predictivo basado en los datos preprocesados. Dependiendo de la funcionalidad de nuestro modelo, se elegirán algoritmos de machine learning más adecuados para el problema en cuestión, como redes neuronales, máquinas de soporte vectorial, árboles de decisión, bosques aleatorios, etc. Una vez montado nuestro modelo, hay que entrenarlo, usando el conjunto de datos de entrenamiento que se ha preseleccionado, ajustando los parámetros del algoritmo para optimizar el rendimiento. Por último, hay que validar el modelo, utilizando el conjunto de validación para evaluar el rendimiento del modelo y ajustar los parámetros para mejorar la precisión y evitar el sobreajuste.
4. Evaluación: este paso es vital para medir el rendimiento y efectividad del modelo. Esto se realiza mediante la aplicación del modelo a los datos de prueba y la

comparación de las predicciones con los resultados reales. Para realizar este proceso correctamente, se deben utilizar métricas como la precisión o el error cuadrático medio para evaluar el rendimiento del modelo. Además, técnicas como la validación cruzada para asegurar que el modelo generaliza bien datos no vistos y no está sobreajustado ayudarán a asegurarnos de que el modelo es válido. También es importante analizar los resultados para interpretar las métricas de evaluación e identificar áreas de mejora para ajustar el modelo según sea necesario.

5. **Despliegue:** esta última fase consiste en poner en producción el modelo entrenado para que pueda ser utilizado en entornos reales para realizar predicciones y apoyar la toma de decisiones. Es importante integrar el modelo en las plataformas y sistemas utilizados por los inversores, asegurando su accesibilidad y eficiencia. También es necesario establecer sistemas de monitoreo para evaluar el rendimiento del modelo en producción y realizar ajustes continuos basados en nuevos datos y resultados obtenidos, actualizando el modelo regularmente para asegurar que se mantenga preciso y relevante con los cambios en los datos y las condiciones del mercado.

6 Desafíos, Limitaciones y Perspectivas Futuras

6.1 Desafíos éticos, regulatorios y técnicos en la implementación de machine learning en finanzas

La implementación de machine learning en el sector financiero presenta una serie de desafíos éticos, regulatorios y técnicos que deben ser abordados para asegurar su adopción efectiva y responsable. Estos desafíos son críticos, dado el impacto significativo que las decisiones automatizadas pueden tener en los mercados financieros y en la sociedad en general.

6.1.1 Desafíos Éticos:

Transparencia: los algoritmos de ML a menudo son complejos y opacos, conocidos como "cajas negras", que dificultan la comprensión de cómo se toman las decisiones. La falta de transparencia puede generar desconfianza entre los usuarios y reguladores, y plantea problemas éticos relacionados con la responsabilidad y la rendición de cuentas. Es crucial desarrollar técnicas que permitan explicar las decisiones de los modelos de ML de manera comprensible para los humanos, asegurando que los stakeholders puedan entender y confiar en los resultados.

Sesgos y Equidad: los modelos de machine learning pueden exacerbar los sesgos existentes en los datos, lo que puede llevar a decisiones injustas o discriminatorias. En el ámbito financiero, esto puede afectar la concesión de créditos, la fijación de precios de seguros o las decisiones de inversión. Es esencial implementar métodos de detección y mitigación de sesgos para garantizar que los modelos de ML operen de manera justa y equitativa.

Privacidad de los datos: el ML requiere grandes cantidades de datos para entrenar modelos efectivos. La recolección y uso de datos financieros y personales plantea importantes preocupaciones sobre la privacidad. Las instituciones financieras deben cumplir con las leyes y regulaciones de protección de datos, como el Reglamento General de Protección de Datos (GDPR) en Europa, y asegurarse de que los datos se manejen de manera segura y ética.

6.1.2 Desafíos Regulatorios

Cumplimiento Normativo: las instituciones financieras están sujetas a estrictas regulaciones que buscan proteger a los consumidores y asegurar la estabilidad del sistema financiero. La implementación de ML debe alinearse con estas regulaciones, lo que puede ser complejo debido a la naturaleza innovadora y en rápida evolución de estas tecnologías.

Supervisión y Auditoría: los modelos de ML deben ser auditables para garantizar que cumplan con las regulaciones y estándares éticos. Esto implica desarrollar procesos de supervisión que puedan evaluar la precisión, equidad y seguridad de los modelos.

Riesgo Sistémico: el uso generalizado de ML en finanzas puede aumentar el riesgo sistémico si muchos actores del mercado dependen de modelos similares. Un fallo en un modelo ampliamente utilizado podría tener consecuencias significativas para el sistema financiero en su conjunto. Los reguladores deben considerar estas implicaciones sistémicas y promover la diversidad en los enfoques de modelado.

6.1.3 Desafíos Técnicos

Calidad y Gestión de Datos: la calidad de los datos es crucial para el rendimiento de los modelos de ML. Datos incompletos, incorrectos o sesgados pueden llevar a resultados inexactos y decisiones erróneas.

Escalabilidad y Eficiencia Computacional: los modelos de ML pueden requerir recursos computacionales significativos para entrenarse y desplegarse a gran escala. Las instituciones deben invertir en infraestructuras adecuadas y tecnologías avanzadas para manejar estas demandas.

Mantenimiento y Actualización de Modelos: como se mencionó en la sección anterior, los modelos de ML deben mantenerse y actualizarse regularmente para seguir siendo precisos y relevantes en un entorno financiero dinámico. Esto implica monitorear continuamente el rendimiento del modelo y realizar ajustes a medida que cambian las condiciones del mercado.

6.2 Limitaciones actuales

A pesar del notable potencial y las múltiples aplicaciones del machine learning en el sector financiero, existen varias limitaciones actuales que deben considerarse:

1. Datos Insuficientes y Sesgados: los modelos de machine learning requieren grandes cantidades de datos de alta calidad para entrenarse de manera efectiva. En muchos casos, los datos disponibles pueden ser insuficientes, incompletos o sesgados, lo que afecta negativamente la precisión y fiabilidad de los modelos. También, en el contexto del mercado español, el mercado de capital riesgo tiene otras características y mucho menor volumen que en otros países, por lo que existen complicaciones a la hora de crear modelos uniformes.

2. Riesgo de Sobreajuste: el sobreajuste es un problema común en machine learning, donde un modelo se ajusta demasiado a los datos de entrenamiento y no generaliza bien a datos nuevos. Esto puede llevar a predicciones inexactas y decisiones erróneas, especialmente en un entorno financiero dinámico.

3. Costos Computacionales: el entrenamiento y la implementación de modelos avanzados de machine learning requieren recursos computacionales significativos. Esto puede ser costoso y limitar la capacidad de las instituciones más pequeñas para aprovechar completamente estas tecnologías

4. Regulaciones y Cumplimiento: la integración de machine learning en finanzas debe alinearse con las regulaciones existentes, lo que puede ser complicado y restrictivo. La evolución constante de la tecnología a menudo supera el ritmo de las actualizaciones regulatorias, creando incertidumbre y posibles conflictos (Federal Reserve Bank of Atlanta, 2023).

5. Dependencia de la Infraestructura Tecnológica: la eficacia de los modelos de machine learning depende en gran medida de la infraestructura tecnológica disponible. Las limitaciones en hardware, software y capacidades de red pueden restringir la implementación y operación eficiente de estos modelos (IMF, 2021).

6.3 Tendencias emergentes, avances tecnológicos y su impacto en estrategias de inversión

El sector financiero está experimentando una transformación significativa debido a las tendencias emergentes y los avances tecnológicos en machine learning y otras áreas de la inteligencia artificial. Estos desarrollos están redefiniendo las estrategias de inversión y ofreciendo nuevas oportunidades para los inversores.

Entre las tendencias emergentes que están transformando la manera en que los inversores toman decisiones, destaca el aprendizaje automático híbrido, que combina técnicas tradicionales con enfoques avanzados como el aprendizaje profundo (Deep learning) y el aprendizaje por refuerzo. Estos modelos híbridos son capaces de capturar relaciones complejas en los datos, lo que resulta en una mayor precisión en las predicciones financieras.

Otra tendencia significativa es la de las plataformas de datos integradas. Estas plataformas consolidan la información de distintas fuentes, facilitando el acceso a datos de alta calidad. Al estar los modelos de ML necesitados de grandes bases de datos con constantes actualizaciones, estas herramientas pueden mejorar significativamente la mejora de los modelos y la toma de decisiones en el ámbito financiero.

En conclusión, el avance de las tecnologías de inteligencia artificial y machine learning está avanzando de manera exponencial, con revoluciones constantes e implementaciones de nuevos modelos o actualizaciones de los existentes. Sin duda el mundo de las finanzas se dirige hacia un contexto cada vez más tecnológico para optimizar todas las áreas existentes del sector.

7 Conclusiones y Recomendaciones

7.1 Resumen de hallazgos clave

Este análisis de las posibles aplicaciones del machine learning en el ámbito del capital riesgo informa al usuario sobre el estado del arte de ambas industrias, cuáles son las corrientes principales del estado del machine learning en las finanzas actuales y cómo estas se pueden aplicar de forma realista en la industria del capital riesgo.

Tras investigar todos estos aspectos con el apoyo de documentos oficiales recientes y ejemplos de modelos en otros ámbitos de las finanzas, se observan varios indicios y evidencias que indican la posible aplicación de todos los algoritmos de machine learning mencionados en este trabajo para integrar en las distintas etapas de una inversión de capital riesgo. Todas distintas etapas en una inversión de capital riesgo (búsqueda de inversiones, análisis o due diligence individual, inversión acorde a la etapa, gestión de riesgos y venta) presentan unas características específicas con una gran correlación a las aplicaciones actuales del machine learning en las finanzas.

Esto lleva a la conclusión de que en el contexto actual es posible, factible y realista introducir varios posibles modelos de machine learning en el ámbito del capital riesgo, pues se disponen de los datos necesarios, herramientas de pre procesamiento y evaluación para desplegar algoritmos efectivos que ayuden a la predicción de factores relevantes y de éxito que ayuden a los inversores de capital riesgo a optimizar el proceso de recopilar información, tomar decisiones acertadas, mitigar riesgos durante la inversión y maximizar los retornos.

7.2 Sugerencias para futuras investigaciones y mejoras en la aplicación de machine learning en el capital riesgo

La sugerencia más natural que surge para futuras investigaciones tras concluir este estudio se trata de la implementación real de un modelo de machine learning aplicado en alguna de las áreas mencionadas de capital riesgo para estudiar su efectividad y rendimiento en inversiones reales. Como se menciona en el apartado “5.1 Potenciales áreas de aplicación” de este trabajo, desde un radar de posibles inversiones de capital riesgo hasta un simulador de escenarios de venta, se plantean varios modelos de machine learning atractivos que pueden resultar altamente efectivos a la hora de facilitar las tareas de un inversor de capital riesgo.

Por otro lado, hay muchos factores que giran alrededor del sector del capital riesgo y de las inversiones que son de gran importancia e interés donde también se pueden aplicar técnicas de machine learning, como el ámbito regulatorio o ESG. Cada vez se le está dando más importancia a las inversiones sostenibles y se están lanzando nuevas normativas europeas que los inversores tienen que adaptar para quedar dentro del marco regulatorio de la unión europea. Iniciativas como la taxonomía verde europea, las normas sobre fondos de inversión de artículo ocho, artículo nueve y artículo nueve “plus” invitan a estudiar las implicaciones que tienen y tendrán en las inversiones, tanto en los activos tradicionales como en el sector del capital riesgo.

8 Declaración de uso de herramientas de inteligencia artificial generativa en Trabajos de Fin de Grado

Por la presente, yo, Alberto Santos-García Rico, estudiante de Administración y Dirección de Empresas Bilingüe (E-2 Bilingüe) de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "[Título del trabajo]", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. Brainstorming de ideas de investigación: Utilizado para idear y esbozar posibles áreas de investigación.
2. Crítico: Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. Referencias: Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. Constructor de plantillas: Para diseñar formatos específicos para secciones del trabajo.
5. Revisor: Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
6. Traductor: Para traducir textos de un lenguaje a otro.

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

Fecha: 04/06/2024

Firma: ALBERTO SANTOS-GARCÍA RICO

9 Bibliografía

Autoridad Bancaria Europea. (2021). *Discussion Paper on Machine Learning for IRB Models*. https://www.eba.europa.eu/sites/default/files/document_library/Final%20Report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210-229. <https://doi.org/10.1147/rd.33.0210>

Jha, R. (2021, agosto 10). ML algorithms: Pros, cons and suitable usages. *Analytics Vidhya*. <https://medium.com/analytics-vidhya/ml-algorithms-pros-cons-and-suitable-usages-b377c3c09f1b>

Breiman, L. (2001). Random forests. *Statistics Department University of California Berkeley*. <https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. <https://arxiv.org/abs/1603.02754>

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. https://papers.nips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

IBM. (s.f.). Página web oficial de IBM. ¿Qué son las redes neuronales? <https://www.ibm.com/es-es/topics/neural-networks>

Expansión. (s.f.). Diccionario del diario Expansión. Capital riesgo. <https://www.expansion.com/diccionario-economico/capital-riesgo.html>

Nexford University. (2023). Advantages and Disadvantages of Venture Capital 2024. <https://www.nexford.edu/insights/pros-cons-venture-capital>

Sahlman, W. A. (1990). The structure and governance of venture-capital organizations. *Journal of Financial Economics*, 27(2), 473-521. https://www.edegan.com/pdfs/ERBC/Hellmann,%205.28.11/Sahlman_1990.pdf

Proksch, D., Stranz, W., Pinkwart, A., & Schefczyk, M. (2016). Risk management in the venture capital industry: Managing risk in portfolio companies. *The Journal of Entrepreneurial Finance*, 18(2), 1-33. <https://digitalcommons.pepperdine.edu/jef/vol18/iss2/1/>

Reid, G. C., Terry, N. G., & Smith, J. A. (1995). Risk Management in Venture Capital: Investor-Investee Relations. University of St Andrews, Centre for Research into Industry, Enterprise, Finance and the Firm (CRIEFF). <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6633bdbe899f0b6bf4c6dc69dd93efb96edd47b>

Cao, L. (Ed.). (2023). Handbook of artificial intelligence and big data applications in investments. CFA Institute Research Foundation. <https://www.cfainstitute.org/-/media/documents/article/rf-brief/ai-and-big-data-in-investments.pdf>

Hoang, D., & Wiegatz, K. (2022). Machine learning methods in finance: Recent applications and prospects. *European Financial Management*, 29(5), 1657-1701. DOI: [10.1111/eufm.12408](https://doi.org/10.1111/eufm.12408)

Expert Systems with Applications. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications*, 219, 119640. DOI: <https://doi.org/10.1016/j.eswa.2023.119640>

International Monetary Fund (IMF). (2021). Powering the Digital Economy: Opportunities and Risks of Artificial Intelligence in Finance. <https://www.imf.org/en/Publications/Departmental-Papers-Policy-Papers/Issues/2021/10/21/Powering-the-Digital-Economy-Opportunities-and-Risks-of-Artificial-Intelligence-in-Finance-494717>