



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

GRADO EN INGENIERÍA EN TECNOLOGÍAS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

ANÁLISIS DE EMPRESAS MEDIANTE ALGORITMOS
DE REGRESIÓN, CLUSTERING Y FORECASTING

Autor: Fernando Urrutia Gómez - Menor

Director: Celso Otero García

Co-Director: Diego Santo Domingo Vicario

Madrid

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
Análisis de empresas mediante algoritmos de regresión, clustering y forecasting
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el
curso académico 2024/25 es de mi autoría, original e inédito y
no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido
tomada de otros documentos está debidamente referenciada.

Fdo.: Fernando Urrutia Gómez-Menor

Fecha: 27/ 06/ 2025

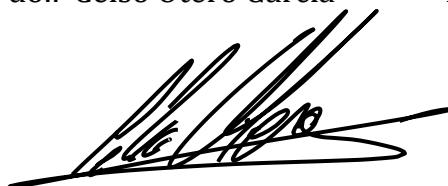


Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Celso Otero García

Fecha: 27..../ 06..../ 2025





COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

**GRADO EN INGENIERÍA EN TECNOLOGÍAS DE
TELECOMUNICACIÓN**

TRABAJO FIN DE GRADO

**ANÁLISIS DE EMPRESAS MEDIANTE ALGORITMOS
DE REGRESIÓN, CLUSTERING Y FORECASTING**

Autor: Fernando Urrutia Gómez - Menor

Director: Celso Otero García

Co-Director: Diego Santo Domingo Vicario

Madrid

Agradecimientos

A mis padres, hermano, familia y amigos,
por la ayuda y presencia en los momentos más difíciles

ANÁLISIS DE EMPRESAS MEDIANTE ALGORITMOS DE REGRESIÓN, CLUSTERING Y FORECASTING

Autor: Urrutia Gómez-Menor, Fernando.

Director: Otero García, Celso.

Entidad Colaboradora: Renta 4

RESUMEN DEL PROYECTO

El presente trabajo propone un marco metodológico que combina el estudio de indicadores relativos a los estados financieros de un conjunto de empresas del sector tecnológico junto a la aplicación de algoritmos de *Machine Learning* de tres campos distintos como son la regresión, *clustering* y predicción o *forecasting*.

Palabras clave: Machine Learning, Clustering no supervisado, Facebook Prophet, Regresión Lineal Múltiple, Análisis fundamental, Clasificación empresas, Predicción del valor de la acción.

1. Introducción

En un entorno financiero cada vez más complejo y saturado de datos, la toma de decisiones de inversión requiere de una integración conjunta del análisis financiero clásico basado en ratios y métricas junto con técnicas y algoritmos de ciencia de datos más modernos como Machine Learning. Frente a las limitaciones evidentes de herramientas convencionales como hojas de cálculo, el uso de modelos avanzados permite un análisis más preciso y eficiente.

A pesar del creciente interés en el uso de estos algoritmos, muchas de las gestoras aún dependen principalmente del análisis cuantitativo clásico. Además, aunque existe mucha literatura sobre el uso de estas técnicas en finanzas, se suele tender a aplicaciones de análisis técnico y objetivos irreales de predicción de acciones, difícilmente extrapolables a la realidad, dada la cantidad de variables externas que influyen en los precios y en el mercado [1] [2].

La aplicación de Machine Learning no solo mejora la precisión en los resultados y, por ende, en las decisiones, sino que optimiza los procesos analíticos y permite una base más eficiente sobre la que tomar decisiones.

2. Definición del proyecto

Este proyecto busca integrar, como se ha mencionado en la introducción, el análisis financiero tradicional junto con la capacidad predictiva y clasificatoria de técnicas modernas de Machine Learning aplicado a un conjunto de empresas seleccionadas, que en su mayoría pertenecen al sector tecnológico o eléctrico. Su valor diferencial reside en aplicar estos métodos a sectores menos cubiertos como el de las empresas tecnológicas recién salidas a bolsa, donde la incertidumbre es mayor, y a obtener resultados derivados exclusivamente de estos algoritmos, los cuales no son

presentados en herramientas y aplicaciones tradicionales utilizadas en el sector como *Bloomberg*.

Los principales objetivos de este trabajo son:

1. Realizar un análisis exploratorio de los datos y preprocesarlos para unificar la divisa y la fecha, entre otras cosas.
2. Aplicar técnicas de *forecasting* financiero como LSTM o *Prophet* a datos de series temporales históricas.
3. Utilizar estos resultados para proyectar estados financieros y calcular ratios de valoración aproximados que puedan dar una idea de la ponderación de las empresas respecto al mercado.
4. Implementar algoritmos de clasificación no supervisado como *K-means* y segmentar empresas en grupos con características comunes para buscar asociaciones.
5. Desarrollar modelos de regresión para explicar la dependencia de ciertos indicadores con el desempeño bursátil de las acciones.
6. Interpretar los resultados.

Con este marco metodológico se busca facilitar un análisis completo y estratégico para el conjunto de empresas seleccionadas.

3. Descripción del modelo/sistema/herramienta

El proyecto sigue un ciclo o proceso de tratamiento de datos que va desde su recopilación y descarga desde una base de datos ya creada, con información de *Factset*, hasta su presentación final a modo de resultados.

En primer lugar, se realiza un preprocesamiento que incluye limpieza, estandarización de formatos de fechas y divisas y normalización de variables para asegurar coherencia en el análisis. Después se seleccionan las variables más relevantes eliminando redundancia y problemas de multicolinealidad, es decir, dependencia lineal entre variables predictoras que es uno de los mayores problemas cuando se trabaja con datos financieros.

A continuación, se construyen modelos de predicción temporal como *Prophet* y LSTM, cuya precisión se evalúa con métricas como MAE y MSE. Con estas proyecciones, se estiman valores fundamentales de las empresas mediante métodos como el flujo de caja descontado (DCF).

Posteriormente, se aplican técnicas de regresión para identificar qué factores influyen más en indicadores clave como el rendimiento bursátil. En paralelo, se agrupan las empresas mediante algoritmos de clustering (*K-means*), identificando patrones comunes y segmentándolas según su perfil financiero en distintos plazos temporales (uno, tres y cinco años) para obtener, primero su estrategia a largo plazo y posteriormente sus resultados recientes más concretos.

Finalmente, todos los resultados se sintetizan en gráficos y tablas interpretables, mencionando los puntos fuertes y débiles encontrados.

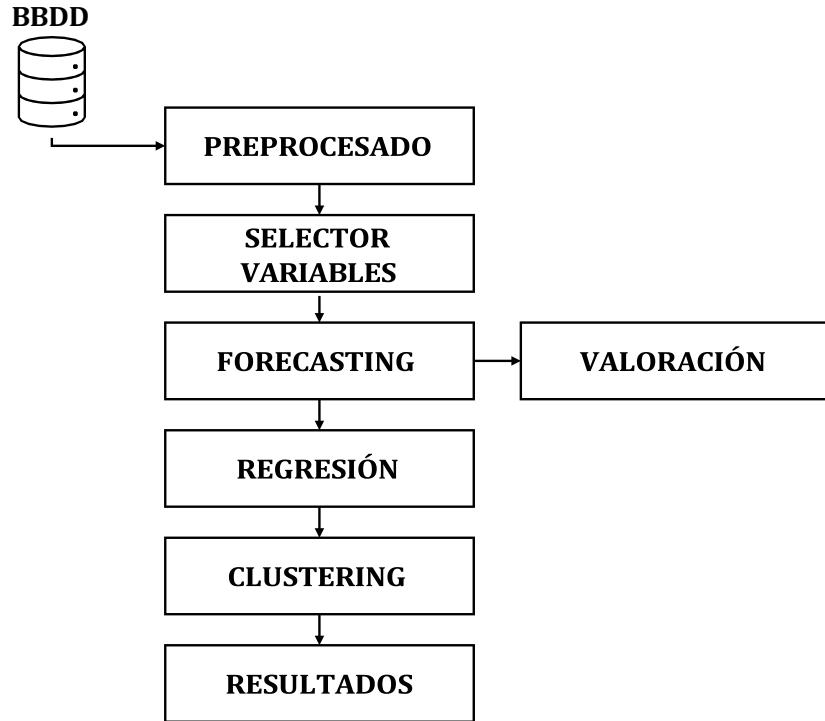


Ilustración a. Ciclo de trabajo

4. Resultados

- *Forecasting*

Los resultados muestran, en primer lugar, la escasa idoneidad de utilizar redes neuronales recurrentes como LSTM para la predicción del valor de la acción a futuro ya que, al trabajar con series temporales del estilo *random walk* [3], las predicciones derivan en un *naïve forecast*, en el que se toma como valor de predicción el anterior sumando un gradiente.

Por otro lado, *Prophet* logra ajustarse a la estacionalidad de los datos y realizar predicciones de ventas y otros indicadores de forma adecuada que posteriormente pueden utilizarse para valoraciones.

- Regresión

Las distintas regresiones indican que predecir el valor futuro de las acciones con únicamente datos de estados financieros es complejo debido a la gran cantidad de

variables externas y aleatoriedad que se acumula en la variación de precio con el paso del tiempo (explican, en el mejor de los casos, alrededor de un 6% de ese movimiento a un año)

Por otro lado, las características más correlacionadas tienen que ver con factores estructurales como márgenes positivos, eficiencia operativa y generación de caja. Por su parte, la liquidez o el crecimiento pasado no garantizan mayor rentabilidad de la acción en un futuro.

- *Clustering*

La segmentación a tres y cinco años produce cuatro grupos con características muy similares en ambos casos. En cambio, la clasificación del último año genera perfiles más concretos, en este caso seis.

A partir de estos grupos se identifican empresas en crisis operacional, empresas maduras con márgenes y crecimientos estables y, lo más importante, empresas con crecimientos elevados durante los últimos años que resultan más interesantes para una potencial inversión o, al menos, control de sus actividades a futuro.

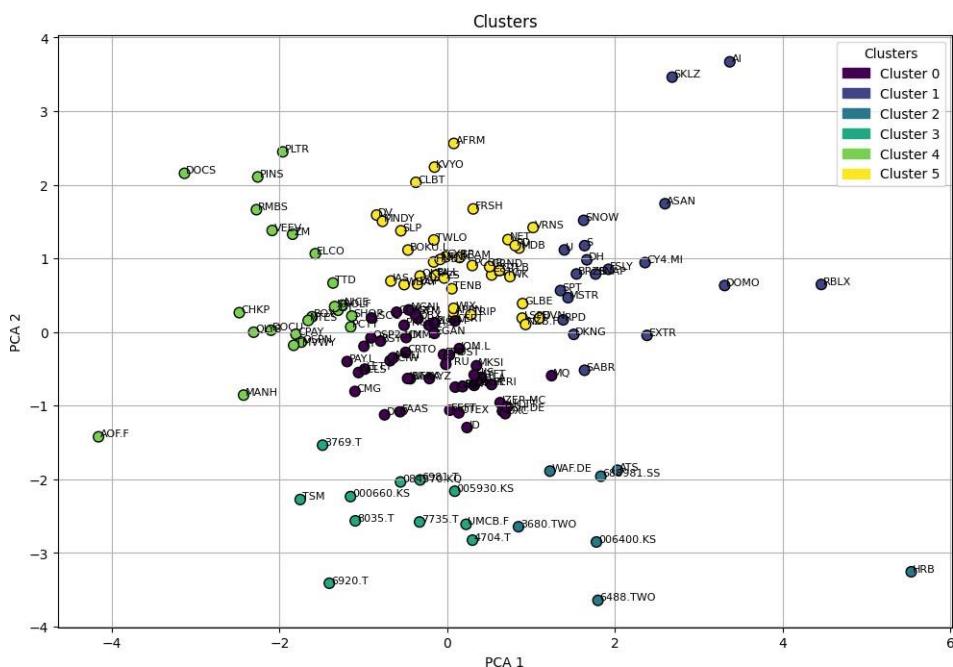


Ilustración b. Resultados Clustering 1 año

5. Conclusiones

La aplicación de técnicas de ML al análisis financiero representa un cambio de paradigma en la forma de interpretar y predecir el comportamiento de las empresas en el contexto económico actual. Si bien es cierto que todavía existen carencias significativas, como la correcta incorporación de shocks económicos a los modelos, las ventajas obtenidas son ya diferenciales.

Por último, es necesario comentar que este tipo de enfoques no pretende sustituir la experiencia humana y el juicio financiero tradicional de los expertos (o al menos, no debería), sino ser un complemento que presente evidencias empíricas robustas reforzando la toma de decisiones estratégicas en un entorno que es, cada vez más, volátil y globalizado.

6. Referencias

- [1] C.G, L.L, J.M & S.S. McKinsey & Co (2023). *Been there, doing that: How corporate and investment banks are tackling gen AI*
<https://www.mckinsey.com/industries/financial-services/our-insights/been-there-doing-that-how-corporate-and-investment-banks-are-tackling-gen-ai>
- [2] European Securities and Markets Authority (2025). *Artificial intelligence in EU investment funds: adoption, strategies and portfolio exposures*
https://www.esma.europa.eu/sites/default/files/2025-02/ESMA50-43599798-9923_TRV_Article_Artificial_intelligence_in_EU_investment_funds.pdf
- [3] Malkiel, B.G. (1973). *A Random Walk Down Wall Street*

REGRESSION, CLUSTERING AND FORECASTING ALGORITHMS APPLIED TO FINANCIAL DATA

Author: Urrutia Gómez-Menor, Fernando.

Supervisor: Otero García, Celso.

Collaborating Entity: Renta 4

ABSTRACT

The following project proposes a methodological scheme that combines the study of data from the financial statements of a selected group of technological companies with the application of Machine Learning algorithms from three different fields such as regression, clustering and forecasting.

Keywords: Machine Learning, Unsupervised Clustering, Facebook Prophet, Multiple Linear Regression, Fundamental Analysis, Company Segmentation, Stock Value Forecasting.

1. Introduction

Within a financial context that every day is more complex and overfilled of data, the financial investment decisions require from an ensemble of traditional financial analysis based on ratios and statistics with modern technics and algorithms from data science such as Machine Learning. Facing the evident limitations of conventional tools such as Excel files, the use of advanced models allows more accurate and efficient analysis.

Despite of the uprising interest in the use of these algorithms, lots of banks are still dependent on quantitative classical analysis. In addition, even if there is plenty of scientific literature about the implementation of these tools in finance, they tend to more technical analysis and irreal objectives of stock forecasting which are quite difficult to extrapolate to real world given the bunch of exogenous variables that have influence on prices and markets [1] [2]

Machine Learning approach not only upgrades accuracy on results and decisions but also processes and allows a strong foundation for decision making.

2. Project definition

The present project aims to combine, as mentioned before, classical financial analysis with the predictive and classificatory ability modern ML techniques. Its differential value lies on applying these models to less covered sectors such as technological startups with recent IPO, where uncertainty is bigger; and also, on obtaining results exclusively from these algorithms which are not usually presented in the most used applications and webs like Bloomberg.

The main objectives of the project are:

1. Make an Exploratory Data Analysis and a preprocessing to unify currency and time format among other things.
2. Applying financial forecasting algorithms such as LSTM or Prophet to historical time series
3. Using these results to project financial statements and make ratios of fundamental valuation which can give an idea of the current value of the company in comparison with the market.
4. Implementing clustering unsupervised algorithms like K-means and carry out a segmentation in groups with common characteristics to search for comps.
5. Developing regression models to explain correlation and dependence of a set of financial indicators with the stock behaviour.
6. Interpretation of the results

With this scheme it is aimed to give a complete and strategical analysis for the selected companies and measure the viability of the models.

3. Model

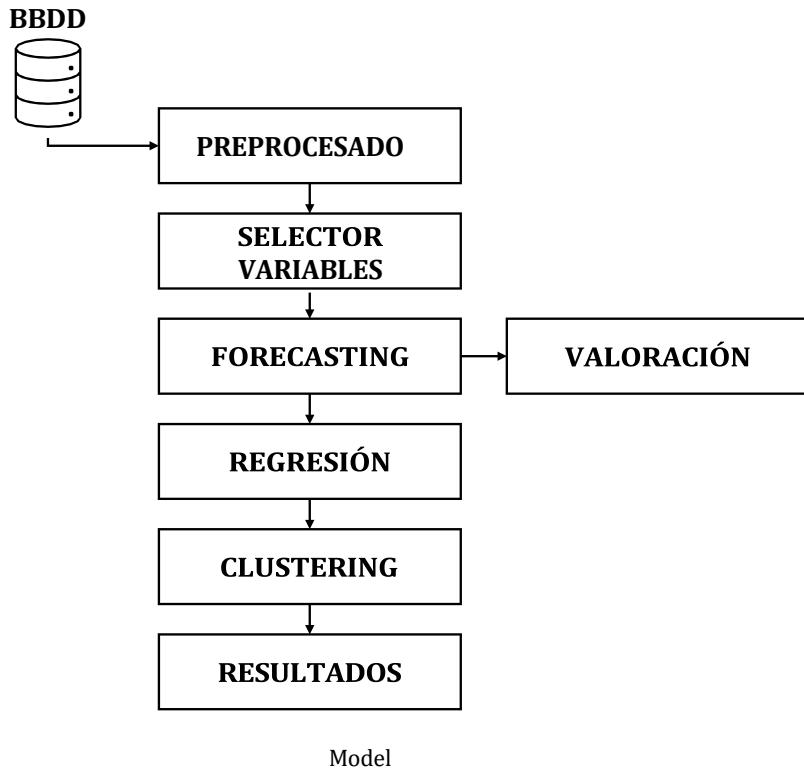
This work follows a process of data treatment going from its collection and download from a database with information from Facset, to its final presentation in a results way.

First, a preprocessing process is performed, including cleansing, standardizing date and currency formats, and normalizing variables to ensure consistency in the analysis. The most relevant variables are selected, eliminating redundancy and multicollinearity issues that is, linear dependence between predictor variables, which is one of the biggest issues when working with financial data.

Then, time based forecasting models such as Prophet and LSTM are built and implemented, measuring its accuracy with metrics such as MAE or MSE. With these forecasts, a fundamental valuation is done with *Discounted Cash Flow* methods.

Regression techniques are then applied to identify which factors most influence key indicators such as stock market performance. In parallel, companies are grouped using clustering algorithms (K-means), identifying common patterns and segmenting them according to their financial profile over different time horizons (one, three, and five years) to obtain, first, their long-term strategy and then their more specific recent results.

Finally, all the results are synthesized in graphics and tables and their strengths and weaknesses are mentioned.



4. Results

- *Forecasting*

Results in this chapter show, firstly, the impossibility of using recurrent neuronal networks as LSTM applied to the prediction of future stock value. This is because using time series with a random walk form, predictions tend to a naïve forecast, in which the value predicted is the previous one plus a small gradient, meaning useless interpretation and forecasting value.

On the other hand, Facebook Prophet model achieves good forecasts adjusting the model to data stationarity and it can project sales and other financial indicators in a good manner.

- Regression

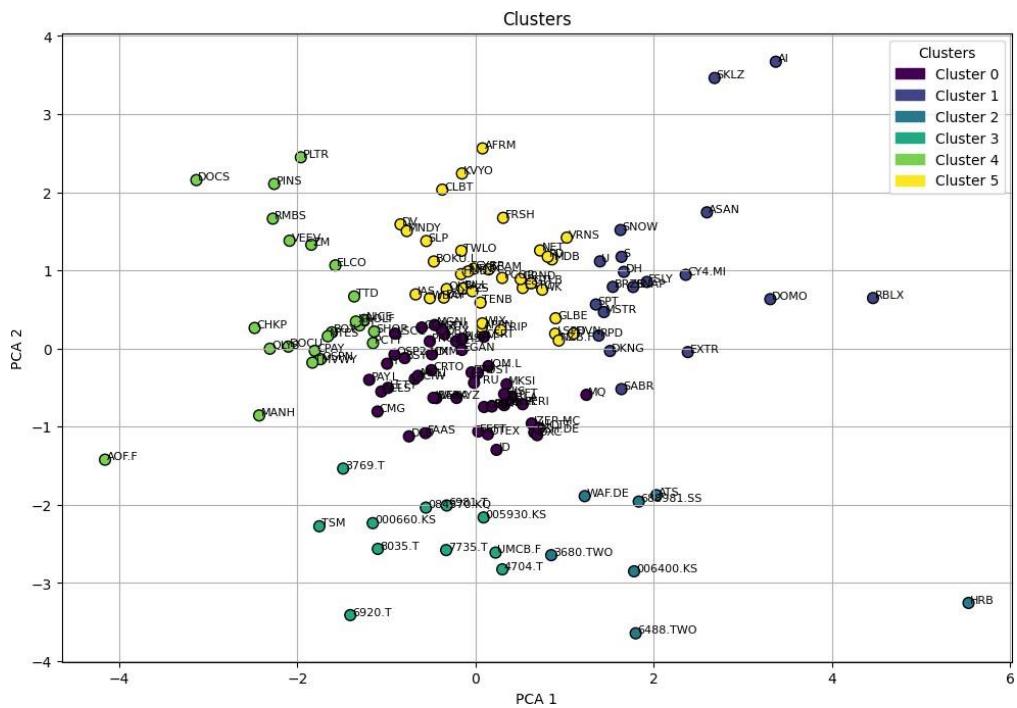
Different regressions at three temporal horizons show that trying to predict future profitability with only strict data from financials is complex due to the randomness accumulated during the time in price variation (they explain, in the best result, approximately a 6% of that movement)

On the other hand, most correlated characteristics are those related to structural factors such as positive margins, operative efficiency and cash generation. On its way, liquidity and past growth do not guarantee better profitability in future.

- *Clustering*

The three and five-year segmentation produces four groups with very similar characteristics in both cases. In contrast, the classification based on the most recent year generates more specific profiles, in this case, six.

From these groups, it is possible to identify companies in operational crisis, mature companies with stable margins and growth, and, most importantly, companies with high growth over recent years, which are more attractive for potential investment or, at the very least, for monitoring their future activities.



Clustering 1 year

5. Conclusions

Application of ML techniques to financial analysis represents an important change in the current paradigm about how to interpret and predict companies' behaviour within actual economic context. Even if it is also true that there are significant lacks, for example on the way of introducing shocks to models, the obtained advantages are already differential.

Finally, it is necessary to explain that these type of approaches are not aimed at replacing human experience and traditional expertise judgment (at least, it should not), but rather to be a complement which presents empirical evidence strengthening strategical decision-making in an environment that is increasingly volatile and globalized.

6. References

- [1] C.G, L.L, J.M & S.S. McKinsey & Co (2023). *Been there, doing that: How corporate and investment banks are tackling gen AI*
<https://www.mckinsey.com/industries/financial-services/our-insights/been-there-doing-that-how-corporate-and-investment-banks-are-tackling-gen-ai>
- [2] European Securities and Markets Authority (2025). *Artificial intelligence in EU investment funds: adoption, strategies and portfolio exposures*
https://www.esma.europa.eu/sites/default/files/2025-02/ESMA50-43599798-9923_TRV_Article_Artificial_intelligence_in_EU_investment_funds.pdf
- [3] Malkiel, B.G. (1973). *A Random Walk Down Wall Street*

Índice de la memoria

Capítulo 1. Introducción	7
Capítulo 2. Descripción de las Tecnologías	9
2.1 Extracción y Obtención de Datos	9
2.1.1 PostgreSQL	9
2.1.2 FactSet	11
2.1.3 Otros	11
2.2 Procesamiento de datos	12
2.3 Visualización de datos	13
Capítulo 3. Estado de la Cuestión	15
Capítulo 4. Definición del Trabajo	17
4.1 Justificación	17
4.2 Objetivos	19
4.3 Metodología	20
4.3.1 Esquema DIKW	20
4.3.2 Flujo de Trabajo	21
Capítulo 5. Análisis Exploratorio de Datos y Preprocesado	25
5.1 Datasets	25
5.1.1 Empresas	25
5.1.2 Estados Financieros	28
5.1.3 Stocks	32
5.1.4 Currency Exchange	33
5.1.5 Nasdaq	33
5.2 Preprocesado	34
Capítulo 6. Predicción y Forecasting	35
6.1 Random walk	36
6.1.1 Modelización de un mercado financiero como sistema caótico	37
6.2 Prophet	41

<i>6.2.1 Serie Temporal.....</i>	41
<i>6.2.2 Algoritmo de Forecast Facebook Prophet.....</i>	42
<i>6.3 Métodos de valoración.....</i>	45
<i>6.3.1 DCF – Discounted Cash Flows.....</i>	46
<i>6.3.2 DFE – Discounted Future Earnings.....</i>	47
<i>6.3.3 Múltiplos</i>	47
<i>6.4 Conclusiones</i>	49
<i>Capítulo 7. Regresión</i>	51
<i>7.1 Marco Teórico.....</i>	51
<i>7.2 Correlación y Multicolinealidad.....</i>	52
<i>7.3 Regresión</i>	54
<i>7.4 Conclusiones</i>	55
<i>Capítulo 8. Clustering</i>	57
<i>8.1 Marco Teórico.....</i>	57
<i>8.2 Kmeans</i>	58
<i>8.2.1 Agrupación.....</i>	58
<i>8.2.2 Normalización.....</i>	59
<i>8.2.3 PCA.....</i>	59
<i>8.2.4 Número Óptimo de Clusters.....</i>	59
<i>8.2.5 Resultados</i>	62
<i>8.3 Conclusiones</i>	69
<i>Capítulo 9. Conclusiones y Trabajos Futuros</i>	71
<i>9.1 Conclusiones</i>	71
<i>9.2 Líneas de trabajo futuro.....</i>	72
<i>9.2.1 Análisis de Sentimiento y Lenguaje Natural</i>	72
<i>9.2.2 Análisis Fundamental no lineal con Machine Learning.....</i>	72
<i>9.2.3 Modelos Híbridos de Datos Estructurados, No Estructurados y Alternativos</i>	73
<i>Capítulo 10. Bibliografía</i>	75
<i>ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS.....</i>	77
<i>ANEXO II: LIBRERÍAS utilizadas en PYTHON.....</i>	79

Índice de figuras

Figura I. Postgres Fuente: Wikipedia	9
Figura II. FactSet Fuente: Wikipedia.....	11
Figura III. Informe ESMA50	17
Figura IV. Pirámide DIKW	21
Figura V. Flujo de trabajo y procesos aplicados	23
Figura VI. Distribución de empresas según su año de fundación	26
Figura VII. Distribución de empresas por industria y sector	28
Figura VIII. Comparativa CAGR vs Crecimiento Simple.....	31
Figura IX. Selector de variables	32
Figura X. Ejemplo del Dataset de conversión de monedas	33
Figura XI. Distribución de empresas por industria en la bolsa Nasdaq Fuente: Nasdaq Composite	33
Figura XII. Ejemplo de Random Walk	36
Figura XIII. Serie histórica de la acción de Activision	38
Figura XIV. Test y forecast de la serie histórica de la acción de Activision.....	39
Figura XV. Test y LSTM forecast en escala temporal diaria	39
Figura XVI. Test, LSTM forecast y Naive forecast.....	41
Figura XVII. Descomposición de serie temporal.....	42
Figura XVIII. Forecast de ventas con Prophet	43
Figura XIX. Descomposición serie temporal con Prophet.....	44
Figura XX. Forecast FCF con Prophet.....	44
Figura XXI. Forecast Net Income con Prophet.....	45
Figura XXII. Secuencia de disminución de VIF	52
Figura XXIII. Elbow Method.....	60
Figura XXIV. Silhouette Method	61

Figura XXV. Combinación de Elbow y Silhouette.....	62
Figura XXVI. Clustering 5 años	64
Figura XXVII. Clustering 3 años.....	66
Figura XXVIII. Selección K para Clustering 1 año	67
Figura XXIX. Clustering 1 año	67

Índice de tablas

Tabla 1. N° de empresas por bolsa.....	27
Tabla 2. Errores LSTM vs Naive.....	40
Tabla 3. Correlación [%] de las variables con rentabilidad de la acción en el horizonte temporal	53
Tabla 4. Resultados de la regresión (R2).....	54
Tabla 5. Clustering 5 años.....	63
Tabla 6. Características notables Clustering 5 años	63
Tabla 7. Clustering 3 años.....	65
Tabla 8. Características notables Clustering 3 años	65
Tabla 9. Clustering 1 año.....	68
Tabla 10. Características notables Clustering 1 año.....	68

Capítulo 1. INTRODUCCIÓN

En un contexto de mercados financieros cada vez más dinámicos y volátiles, y en un mundo donde la cantidad de datos aumenta exponencialmente año tras año, la toma de decisiones de inversión exige de procesos y metodologías que integren tanto el análisis financiero-contable tradicional basado en ratios y estados financieros de las empresas, como las técnicas y algoritmos más avanzados de ciencia de datos, que permitan procesar volúmenes grandes y sintetizar la información para obtener conclusiones más sencillas y fáciles de interpretar.

El presente trabajo tiene como objetivo el desarrollo y evaluación de un marco metodológico que combine precisamente dichas mediciones financieras clásicas con algoritmos modernos de *Machine Learning* para estimar el funcionamiento y viabilidad presentes y futuros de un conjunto diversificado de compañías pertenecientes al sector tecnológico.

La inversión es fundamental en el ámbito económico, no solo para ajustar las operaciones cotidianas, sino también para planificar un crecimiento a largo plazo o mitigar riesgos en un entorno global cada vez más incierto y competitivo. Además, las decisiones tomadas pueden estar basadas en suposiciones erróneas o correctas y tienen repercusiones significativas, tanto para las instituciones financieras, como, de manera significativa, para sus clientes.

Dicho esto, predecir con precisión las tendencias del mercado, los movimientos de los activos financieros y las posibles situaciones económicas se está convirtiendo en una necesidad para maximizar la rentabilidad de los capitales invertidos inicialmente, lo que al final es el objetivo más primordial de cualquier inversión.

Históricamente, el desarrollo de estos modelos en el ámbito empresarial y financiero ha dependido en gran medida de herramientas convencionales, como las hojas de cálculo, que permitían realizar aproximaciones y simulaciones para predecir futuros comportamientos. No obstante, la predicción financiera es una ciencia compleja y la cantidad de variables involucradas en cada activo sobrepasa los límites de un modelo sencillo.

El rápido avance tecnológico ha transformado este campo como tantos otros, permitiendo la adopción de herramientas de *Machine Learning*. Estas herramientas ahora posibilitan el desarrollo de algoritmos más complejos y modelos predictivos más precisos, capaces de analizar datos de forma rápida y eficiente.

Por ello, y para hacer frente a los problemas anteriores, el trabajo combina un análisis exploratorio de datos junto a técnicas de predicción o *forecasting financiero* (en la medida de lo posible), regresión, en una búsqueda de relaciones entre variables y elementos del conjunto de datos, y modelos no supervisados de *clustering* para segmentar el conjunto de empresas en distintos grupos con características o patrones comunes que puedan indicar un grado de desempeño concreto.

El uso de *Machine Learning* en la predicción financiera y la clasificación de empresas utilizando este campo de la inteligencia artificial, no solo mejora la precisión en la toma de decisiones, sino que también optimiza los procesos de análisis y proporciona una base sólida sobre la que las instituciones, particulares o empresas pueden desarrollar su estrategia.

Capítulo 2. DESCRIPCIÓN DE LAS TECNOLOGÍAS

A continuación, se describen las herramientas y tecnologías utilizadas durante el proyecto.

2.1 EXTRACCIÓN Y OBTENCIÓN DE DATOS

2.1.1 POSTGRESQL

Conocido coloquialmente como “*Postgres*”, se trata de un sistema de gestión de bases de datos relacionales orientado a objetos, de código libre y extensible. Por su fiabilidad y compatibilidad con otras plataformas es ampliamente utilizado en el sector de la ciencia de datos. Ofrece, además, soporte al estándar SQL, lo cual simplifica mucho la interacción con los datos alojados.

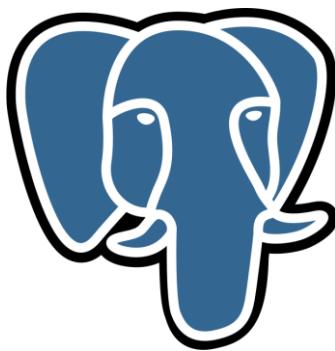


Figura I. Postgres / Fuente: Wikipedia

Para la interacción con este sistema se utiliza la herramienta oficial, *pgAdmin*. Ésta proporciona una interfaz gráfica intuitiva que permite administrar las bases de datos y ejecutar consultas.

En cualquier caso, para mayor comodidad, se ha creado además la siguiente función de *Python* para ser reutilizada desde el propio entorno y poder conectarnos a la base de datos sin tener que abrir la aplicación.

```
import psycopg2

def connection_postgress():

    host = "****"
    port = "****"
    database = "****"
    user = "****"
    password = "****"

    try:
        # Connect to the database
        connection = psycopg2.connect(
            host=host,
            port=port,
            database=database,
            user=user,
            password=password
        )
        # Create a cursor
        cursor = connection.cursor()

    except psycopg2.Error as e:
        print(f"Error: {e}")

    return connection, cursor
```

(los credenciales originales aparecen censurados *** debido a privacidad)

Un ejemplo de uso sería el siguiente

```
import pandas as pd
from postgres import connection_postgress

conn, cur = connection_postgress()

SECURITIES = cur.execute(
    """
    SELECT ...
    FROM ...
    WHERE ...
    """
)
```

```
SECURITIES = pd.DataFrame(cur.fetchall(), columns=[x[0] for x in
cur.description])
SECURITIES.to_excel('SECURITIES.xlsx', index=False)

#Closing
cur.close()
conn.close()
print("Database connection closed.")
```

Así, de este modo, obtenemos la información en formato *xlsx (Excel)*

2.1.2 FACTSET

Se trata de una de las plataformas financieras más ampliamente utilizada por los profesionales del sector para la obtención de: información referente a empresas de todo el mundo, tanto financiera como no financiera; métricas de macroeconomía y recomendaciones y análisis de profesionales.

Se trata de una herramienta de pago, aunque su uso para este trabajo está facilitado tanto por la Universidad como empresas privadas para este tipo de investigaciones o análisis profesionales.



Figura II. FactSet / Fuente: Wikipedia

2.1.3 OTROS

Se obtiene información de otras webs alternativas como, por ejemplo, la página oficial de *Nasdaq* para información sobre el índice con el mismo nombre o *Yahoo Finance* para información rápida acerca de las empresas.

2.2 PROCESAMIENTO DE DATOS

Para la manipulación y el procesamiento de datos se ha utilizado el lenguaje de programación *Python* implementado a través de un entorno de trabajo gestionado por *Visual Studio Code*.

La elección de este lenguaje deriva principalmente de su increíble flexibilidad y disponibilidad de contenido en librerías especializadas en *Machine Learning* y análisis de datos. Por ejemplo, `scikit-learn` o `keras` entre otras.

Como se ha mencionado anteriormente, una de estas librerías, `psycopg2`, permite importar datos a *Python* directamente desde la base de datos de *Postgres* sin requerir pasos intermedios.

Para la manipulación ordinaria y operaciones comunes se han utilizado `pandas` y `numpy`, con el objetivo de tratar los datos según estructuras tipo *DataFrame*.

Otras librerías que se han utilizado y que merecen especial mención por su importancia son las siguientes:

`yfinance`, permite descargar datos históricos y actuales del mercado financiero desde *Yahoo Finance*. Se utiliza en el proyecto para completar otros datos no accesibles de las empresas estudiadas o bien incluir nueva información de las mismas o métricas económicas generales.

`prophet`, es una librería desarrollada por *Meta (Facebook)* para la predicción y tratamiento de series temporales. Está diseñada para ser utilizada con datos que presentan estacionalidad o tendencias, por lo que es especialmente útil a la hora de tratar series temporales de datos financieros.

(Todas las librerías junto a la versión específica utilizada son presentadas al final del documento en el Anexo II. Librerías utilizadas en Python)

2.3 VISUALIZACIÓN DE DATOS

La representación de resultados de forma sintética y clara es una de las partes fundamentales de todo proyecto que incluya gran cantidad de información.

Para cumplir con este propósito se han utilizado principalmente las librerías `matplotlib.pyplot` y `seaborn`.

Capítulo 3. ESTADO DE LA CUESTIÓN

Ya en 1971, Robert E. Jensen [1] abordaba el uso del *clustering* para estudiar el desempeño de un conjunto de empresas seleccionadas. En este trabajo, *Jensen* utiliza técnicas de análisis multivariante para agrupar empresas según sus características financieras, como ingresos, rentabilidad y otras métricas.

Por aquel entonces las técnicas de *Machine Learning* estaban naciendo. Se desarrollaron las primeras contribuciones como el perceptrón (preursor de la red neuronal) y los primeros algoritmos de clasificación. No sería hasta el año 2010 en adelante cuando su uso empezó a ser masivo en una multitud de sectores y, gracias, sobre todo, a las librerías de código abierto e información de Internet, que se hicieron fácilmente accesibles también para particulares, lo que propició un desarrollo muy veloz de las implementaciones de estos algoritmos en multitud de campos distintos.

Por todo ello y debido a la vertiginosa velocidad de cambio y mejora, no tendría demasiado sentido hacer este repaso del estado del arte con más referencias anteriores a los últimos diez años.

El *Machine Learning* está impactando en numerosos sectores dentro de la economía, desde proyecciones futuras en cadenas de suministros [2] hasta el descubrimiento de fraude en transacciones o predicción de la volatilidad del mercado financiero [3].

Grandes consultoras como McKinsey [4] también analizan la acogida que está teniendo la IA generativa -y por consiguiente su derivado, el *Machine Learning*- en las operaciones diarias de los bancos de inversión. También se mencionan algunos de los principales obstáculos a superar en el futuro, relacionados principalmente con propiedad intelectual y costo computacional, además de la dificultad que ven algunos

bancos en implementar estas tecnologías en *front, middle o back office* (contacto con el cliente, procesos intermedios y administración, respectivamente).

Otro buen ejemplo lo podemos encontrar en *J.P. Morgan Chase*, banco conocido por ser pionero en adoptar tecnologías emergentes y en la implementación de técnicas de machine learning y lenguajes de procesado natural NLP, en sus operaciones de *equity research*.

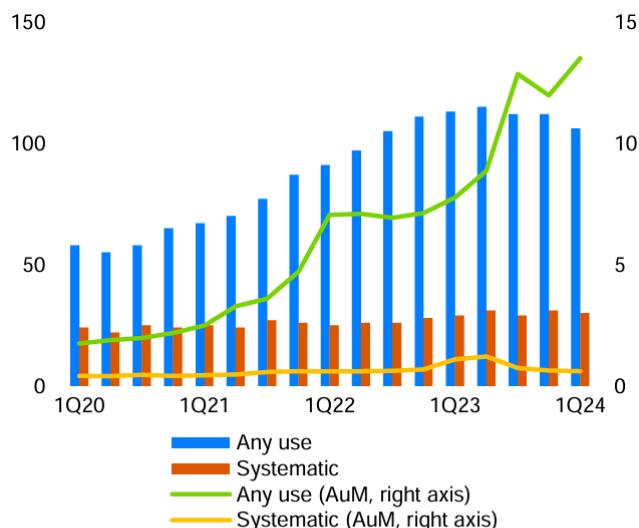
Por todo esto, queda de manifiesto la importancia de la inteligencia artificial en la mecánica de análisis y predicción financieras. Sin embargo, los tutoriales y ayudas de programación para aplicar estos algoritmos a datos financieros no son muy habituales, como pudieran serlo en otros campos y en la gran mayoría de los casos proceden de particulares. Además, muchas de estas guías están enfocados a la optimización de portfolios y profundizan más en análisis técnicos.

Capítulo 4. DEFINICIÓN DEL TRABAJO

Este capítulo presenta, en primer lugar, una serie de argumentos que justifican la realización del trabajo. En segundo lugar, los objetivos que se buscan cumplimentar con éste. Por último, se explica la metodología llevada a cabo: por un lado, la estrategia de trabajo general y por otro el flujo específico de algoritmos y procesos implementados.

4.1 JUSTIFICACIÓN

En el mundo financiero, pese al incremento en los últimos años del uso de algoritmos de *Machine Learning* o inteligencia artificial en el estudio de potenciales inversiones en empresas, la mayoría de las gestoras sigue anclada en un análisis cuantitativo clásico o basándose exclusivamente en datos de precios y métricas derivadas.



Note: Number (left axis) and AuM (right axis, in EUR bn) of EU investment funds that state using AI or ML in the investment process. 'Systematic' indicates funds for which the investment strategy is primarily determined by an AI-based algorithm.

Sources: Morningstar, ESMA

Figura III. Informe ESMA50

En este informe de la ESMA (Autoridad Europea de Valores y Mercados), Figura III, se recoge que solo 145 fondos de inversión (0,3% del total del estudio) mencionan IA o ML en su documentación. En apenas un 30% de esos fondos, la IA es el motor sistemático de su estrategia, es decir, se utiliza principalmente en tareas analíticas pero la decisión final sigue siendo humana [5].

Por otro lado, en Internet existe mucha literatura científica sobre la implementación de ciertos algoritmos e inteligencia artificial a datos financieros, así como guías de *Python* y otros lenguajes de programación. Sin embargo, mucha de esta información está enfocada al análisis técnico, cuya fiabilidad y metodología es objeto de debate entre los profesionales del sector; y a un intento de predicción del precio de las acciones, que resulta también de dudosa utilidad, debido a la cantidad de factores externos que influyen en la variación de los precios. Prueba fehaciente de esto último es la reciente situación económica, en la que unas palabras o declaraciones tienen casi más importancia que los propios números.

Además, los primeros resultados de búsqueda aluden a artículos que simplemente presentan la idea, pero no profundizan en cómo llevarla a cabo.

El aporte principal del proyecto radica, por tanto, en la integración de esta interpretación financiera clásica, que tan exitosa ha resultado ser, junto a la capacidad analítica y predictiva de algoritmos más modernos de *Machine Learning*.

El estudio aspira, además, a contribuir al estado de literatura académica existente sobre *Machine Learning* aplicado a datos financieros. Además, presenta una serie de ventajas como son:

- Conclusiones derivadas de aplicación directa de *Machine Learning* que no están disponibles en las aplicaciones financieras más utilizadas como *Bloomberg* o *FactSet*.

- Se cubre el sector de empresas tecnológicas de reciente IPO (Oferta Pública de Acciones, es decir, que han salido a bolsa recientemente) donde los gestores necesitan de más ayuda cuantitativa y donde la distribución de retornos es más dispar.

El análisis de los estados financieros con *Machine Learning* es el área donde todavía hay poca metodología abierta y fuertes barreras de acceso debido a la dificultad de obtención inicial de datos. Aprovecharlo ahora supondrá, para las empresas, consultoras o particulares que lo hagan, posicionarse un paso por delante de la competencia.

4.2 OBJETIVOS

El presente trabajo trata de cumplir los siguientes objetivos:

1. Realizar un **análisis exploratorio de los datos**, así como un flujo de preprocessado en el que se homogeneice el formato de los datos (misma moneda), las fechas y se explore la presencia de *outliers* y la importancia de su detección.
2. Utilizar técnicas de **predicción y *forecasting*** aplicadas a distintos datos financieros y aportar claridad sobre su relativa utilidad: partiendo de la información disponible se definirán conjuntos de *time series* (ingresos, márgenes, cotizaciones) para aplicar modelos de *forecasting*. En esta línea se utilizarán técnicas avanzadas como redes neuronales recurrentes *LSTM* o *Facebook Prophet*.
3. Realizar un **análisis fundamental** de las compañías basándose en las predicciones anteriores: a partir de las predicciones de variables clave se elaborarán estados financieros proyectados con los que calcular ratios fundamentales (PER, DCF...) y estimar el valor intrínseco de las compañías del índice para su comparación.

4. Realizar una segmentación/**clustering** de empresas mediante técnicas como *K-means*. Se evaluará el número óptimo de grupos con distintos criterios y finalmente se describirán características comunes promedio intragrupales.
5. Realizar una **regresión** para determinar la influencia de ciertos predictores en variables objetivo (desempeño en bolsa, retorno anual de las acciones). Partiendo de regresiones lineales multivariantes a modelos más avanzados que eviten multicolinealidad.
6. **Interpretar** los **resultados** obtenidos señalando puntos fuertes y débiles.

4.3 METODOLOGÍA

4.3.1 ESQUEMA DIKW

Como estrategia general se aplica el esquema DIKW al proyecto de forma que se consigue avanzar desde datos en bruto hasta llegar a decisiones informadas y juicios estratégicos [6].

En esencia, la pirámide está formado por cuatro escalones: en primer lugar, los datos, (en este caso información financiera en bruto). Al otorgar un significado, transformamos esos datos en proyecciones, ratios y mediciones que responden a preguntas concretas (información). Añadiendo un contexto (financiero, sector tecnológico...) y aplicando los distintos algoritmos de *Machine Learning* se transforma esa información en conocimiento. Por último, se interpretan esos resultados para generar decisiones informadas llegando al escalón de la sabiduría.

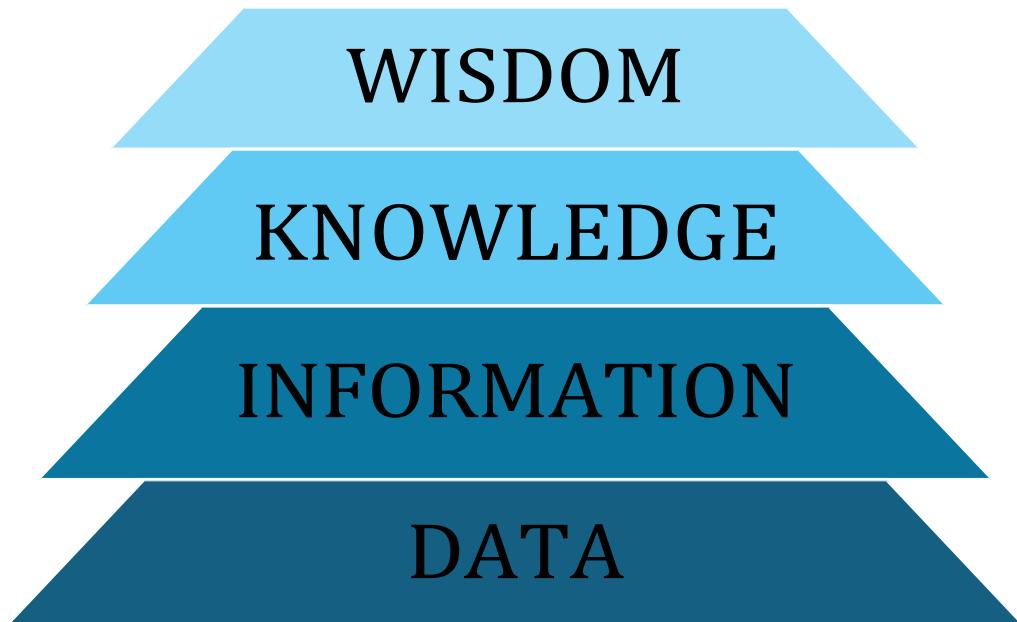


Figura IV. Pirámide DIKW

4.3.2 FLUJO DE TRABAJO

De forma más particular se presenta en la Figura V el ciclo o proceso de los datos desde su recolección en la base de datos a la visualización final de resultados.

4.3.2.1 Preprocesado

Se lleva a cabo la limpieza y homogeneización de datos, unificación de formatos (fechas, divisas) y un escalado de variables para garantizar consistencia y resultados realistas.

4.3.2.2 Selección de variables

Identificación y elección de las características más relevantes para evitar multicolinealidad.

4.3.2.3 *Forecasting*

Construcción y ajuste de modelos de series temporales como *Prophet* y de *Deep Learning* como LSTM, evaluando su capacidad predictiva mediante métricas como MAE o MSE.

4.3.2.4 *Valoración*

Elaboración de variables proyectadas y cálculo del valor intrínseco mediante métodos de análisis fundamental como DCF o DFE.

4.3.2.5 *Regresión*

Modelado de la relación entre las variables predictoras y variable objetivo, como el retorno de la acción en el trimestre, mediante regresión lineal múltiple, cuantificando la dependencia lineal y peso de cada factor.

4.3.2.6 *Clustering*

Agrupamiento de empresas mediante algoritmos de aprendizaje no supervisado (*K-means*) según sus perfiles financieros y operativos normalizados; selección del número óptimo de *clusters* con métodos como *Elbow* o *Silhouette* y caracterización final de cada grupo a tres horizontes temporales distintos.

4.3.2.7 *Resultados*

Síntesis de resultados en tablas y visualizaciones además de recomendaciones estratégicas basadas en estos.

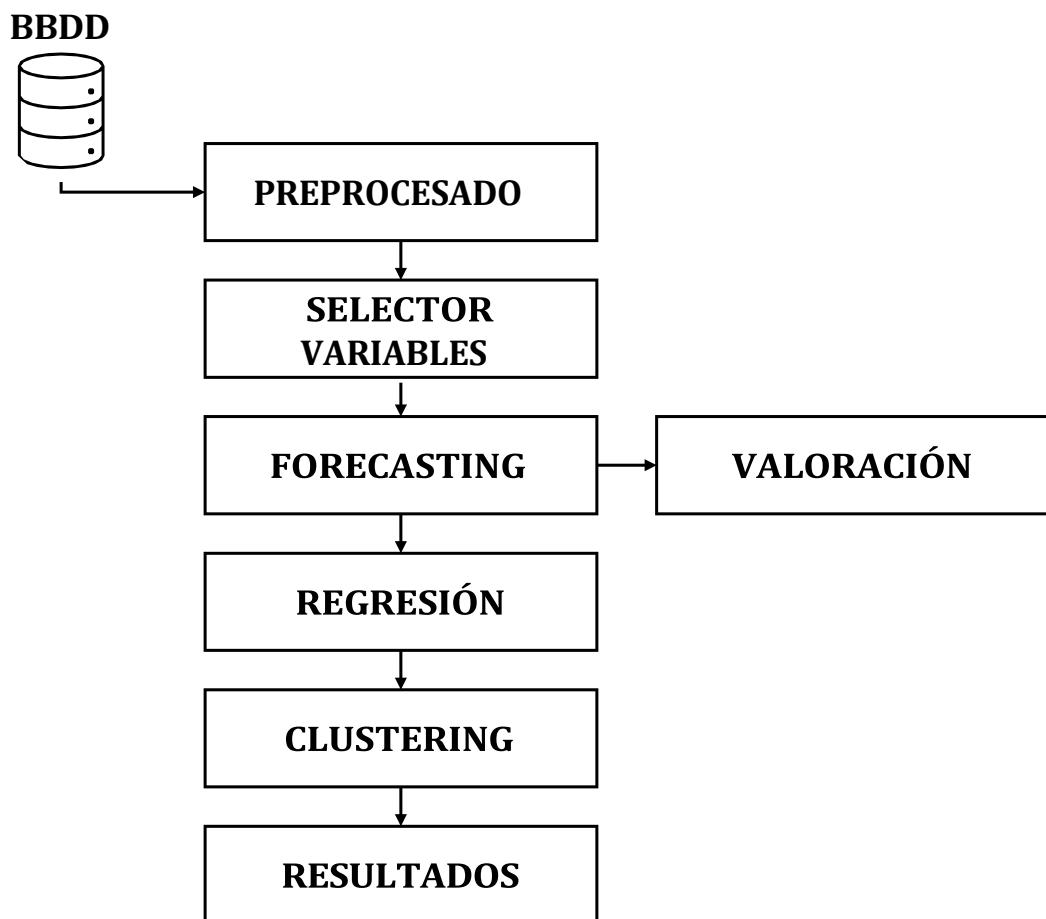


Figura V. Flujo de trabajo y procesos aplicados

Capítulo 5. ANÁLISIS EXPLORATORIO DE DATOS Y PREPROCESADO

En este capítulo se realiza un análisis exploratorio de los datos (*EDA*) con el objetivo de proporcionar una base sólida a partir de la cual sea posible comprender y visualizar la información utilizada en el proyecto previo a la aplicación de estadística y algoritmos avanzados.

En primer lugar, se describen los datos disponibles destacando su importancia relativa para el proyecto, sus características generales y potenciales limitaciones. Posteriormente, se lleva a cabo una selección de las variables más relevantes de entre todas las disponibles.

Se presenta además una serie de técnicas de preprocesado con el fin de depurar y normalizar en la medida de lo posible estos datos.

5.1 DATASETS

5.1.1 EMPRESAS

Para el análisis se han seleccionado un total de **164 empresas** a modo de índice sintético con el objetivo de estudiar parte de sus datos financieros mediante los algoritmos de Machine Learning descritos en la sección “4.2. Objetivos”.

Como se muestra en la Figura VI, la mayoría de las empresas seleccionadas son de reciente fundación; durante la segunda década de siglo XXI.

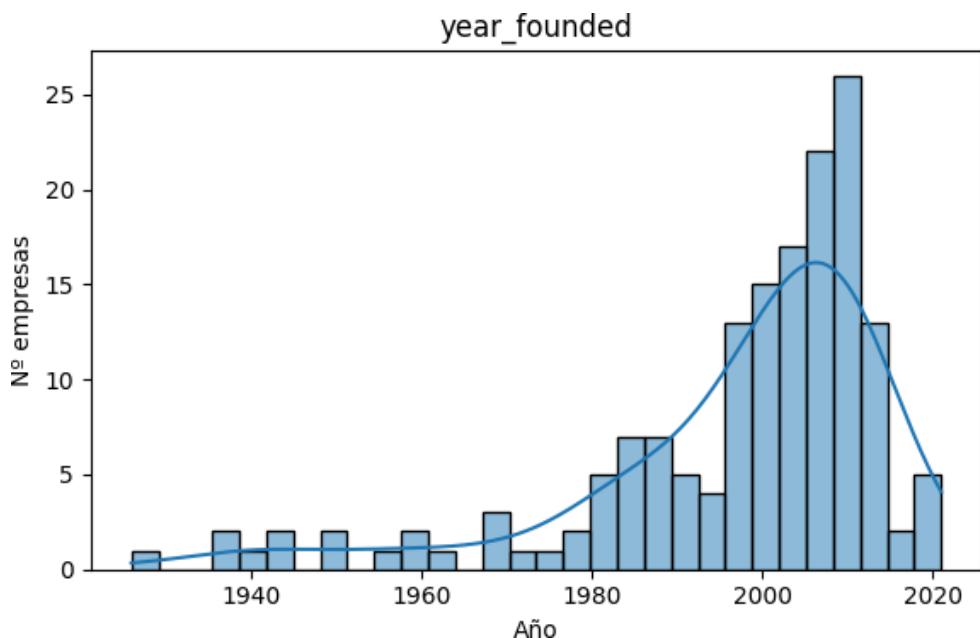


Figura VI. Distribución de empresas según su año de fundación

La mayoría de las empresas analizadas cotizan, de una manera u otra, en *Nasdaq*, que es un índice bursátil estadounidense que mide el rendimiento conjunto de una serie de empresas principalmente del sector eléctrico y tecnológico. Al estar ponderado por capitalización bursátil, las grandes empresas como *Apple* o *Microsoft* tienen más peso en esta cotización (Tabla 1)

La Figura VII muestra la distribución de empresas según la industria a la que pertenecen sus actividades comerciales y más concretamente al sector específico dentro de esta. Se puede observar que la gran mayoría pertenecen a la industria tecnológica, que se divide en tres grandes sectores: *software*, *hardware* y *semiconductores*. El segundo puesto sería para la industria de Consumo Cíclico, seguido de Telecomunicaciones. Otras industrias son Servicios Financieros o Salud.

<i>Bolsa</i>	<i>Nº de empresas</i>
NMS – Nasdaq Global Select	69
NYQ – New York	44
JPX - Japón	10
NGM – Nasdaq Global Market	8
FRA - Francia	5
NCM – Nasdaq Capital Market	5
LSE - Londres	3
PNK – Pink Sheets/OTC Markets	3
GER - Alemania	3
KSC – Corea del Sur	3

Tabla 1. N° de empresas por bolsa

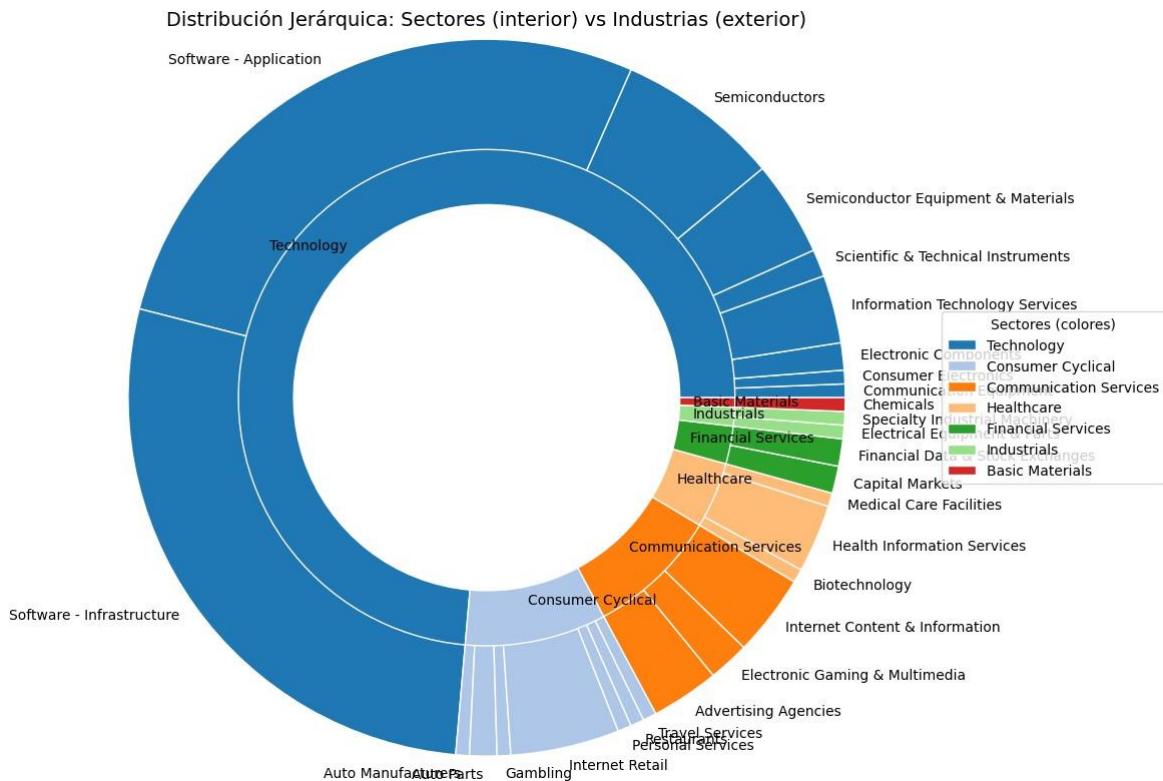


Figura VII. Distribución de empresas por industria y sector

5.1.2 ESTADOS FINANCIEROS

Los estados financieros de una empresa son una serie de informes contables que reflejan la situación económica y financiera de la misma en un periodo concreto. Suelen hacerse públicos al final de cada trimestre o de cada año de forma obligatoria para todas las empresas que cotizan en bolsa.

En total, la base de datos contiene información sobre 502 columnas. Sin embargo, muchas de las columnas tienen muchos datos vacíos que se tienen que eliminar. Para mayor simplicidad, se seleccionan variables con el objetivo de que representen distintos sectores contables de la empresa.

5.1.2.1 Cuenta de Pérdidas y Ganancias / Income Statement

Del *Income Statement* se extrae información relativa a los ingresos generados y los costes soportados durante el ejercicio.

- Ingresos (*Sales*)
- Gastos operacionales (*Opex*)
- EBITDA (Beneficios previos al pago de intereses, ingresos, depreciación y amortización)
- Ingresos netos (*Net Income*)
- Ingresos brutos (*Gross Income*)

5.1.2.2 Balance General / Balance Sheet

Del *Balance Sheet* se extrae información relacionada con el activo, pasivo y patrimonio de la empresa.

- Activo (*Assets*)
- Pasivo corriente (*Current Liabilities*)
- Deuda (*Debt*)
- Patrimonio (*Equity*)

5.1.2.3 Estado de Flujo de Efectivo / Cash Flow Statement

Del *Cash Flow Statement* se extrae información relacionada con el movimiento de efectivo en distintas actividades

- Flujo de caja libre (*Free Cash Flow, FCF*): la cantidad de efectivo que una empresa genera después de cubrir todos sus gastos operativos y las inversiones necesarias para mantener y expandir sus operaciones.

5.1.2.4 Ratios y Métricas derivadas

A partir de los datos anteriores se generan distintos ratios y métricas derivadas que presentan mayor interpretabilidad y permiten una comparación directa entre empresas con distinto tamaño. Otros ratios también directamente son extraídos de la base de datos.

$$\text{Acid Test (Quick Ratio)} = \frac{\text{Cash} + \text{Accounts Receivable}}{\text{Current Liabilities}}$$

$$\text{Current Ratio} = \frac{\text{Current Assets}}{\text{Current Liabilities}}$$

- FCF/Sales
- Opex/Sales
- *Earnings Per Share*

$$EPS = \frac{\text{Net Income} - \text{Preferred Dividends}}{\text{Nº shares outstanding}}$$

- Debt/Assets
- *Return Over Equity* - ROE
- *Return Over Assets* - ROA
- *Compound Annual Growth Rate* - CAGR

$$CAGR = \left(\frac{v_f}{v_i} \right)^{\frac{1}{n}} - 1$$

Usar el CAGR es generalmente mejor que usar tasas de crecimiento simples ya que ofrece una visión más realista y representativa del crecimiento a lo largo del tiempo ya que asume que las ganancias se reinvierten y componen a lo largo

del tiempo, además de suavizar la volatilidad, útil en empresas con ingresos estacionales. Se puede visualizar un ejemplo en la Figura VIII.

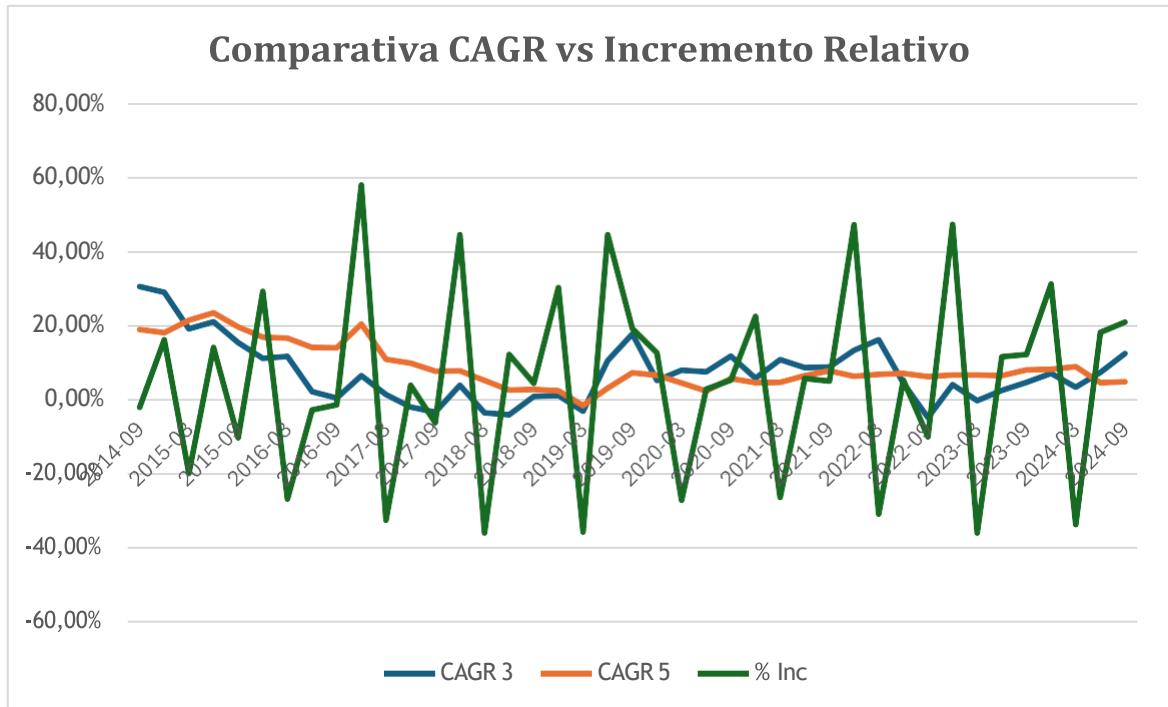


Figura VIII. Comparativa CAGR vs Crecimiento Simple

Finalmente, se crea un selector de variables para navegar por todas las columnas y filtrar el *dataset* con las variables seleccionadas. Un ejemplo de su funcionamiento se presenta en la Figura IX.

Columnas:

```

ff_net_inc_basic_aft_xord
ff_net_inc_basic_bef_xord
ff_net_inc_dil
ff_net_inc_dil_aft_xord
ff_net_income
ff_non_oper_inc
ff_oper_exp
ff_oper_inc
ff_oper_inc_aft_unusual
ff_oper_inc_bef_dep
ff_oper_lease_exp
ff_rd_exp
ff_restruct_exp
ff_sales
ff_stk_opt_exp
ff_stk_opt_exp_cogs
ff_stk_opt_exp_rd
ff_actg_chg_ps
ff_ebit_oper_ps
ff_eps
ff_ana_ban

```

	long_company_name	ticker	fp_end_ym	ff_ebitda_oper	ff_gross_inc	ff_net_income	ff_oper_exp	ff_sales
0	ACI Worldwide, Inc.	ACIW	2004-Q2	14.497	52.862	18.662	59.557	72.532
1	ACI Worldwide, Inc.	ACIW	2004-Q3	13.785	49.548	10.006	57.443	69.708
2	ACI Worldwide, Inc.	ACIW	2004-Q4	-14.335	60.864	12.923	96.221	80.606
3	ACI Worldwide, Inc.	ACIW	2005-Q1	17.149	56.083	11.193	59.583	75.626
4	ACI Worldwide, Inc.	ACIW	2005-Q2	16.383	57.362	9.995	62.817	78.003

Figura IX. Selector de variables

5.1.3 STOCKS

La variación del precio de la acción de una compañía es el elemento más significativo dentro de la inversión financiera, pues es, a fin de cuentas, el responsable directo de obtener beneficio o no.

Para el análisis se hará uso del precio de cierre de la acción en determinadas fechas, por ejemplo, para calcular el EPS o establecer los retornos trimestrales como variable objetivo en el apartado de regresión.

5.1.4 CURRENCY EXCHANGE

Debido a que existen algunas empresas cuyos datos están en monedas locales (tanto europeas o del resto del mundo), es necesaria una conversión para homogeneizar las unidades (divisa) en dólares americanos y que la comparación pueda realizarse de forma justa.

Para ello se ha utilizado una base de datos con la relación directa de conversión entre distintas monedas. Se ha utilizado la fecha más cercana al final de cada trimestre como referencia para la conversión debido a que no hay datos del 100% de todas las sesiones.

2024-07-01 00:00:00	EURUSD	1,07295
2024-08-01 00:00:00	EURUSD	1,07925
2024-09-30 00:00:00	EURUSD	1,11605
2024-10-01 00:00:00	EURUSD	1,10765
2024-11-01 00:00:00	EURUSD	1,0849
2024-12-31 00:00:00	EURUSD	1,0355
2025-01-01 00:00:00	EURUSD	1,0354693
2025-01-31 00:00:00	EURUSD	1,03955

Figura X. Ejemplo del Dataset de conversión de monedas

5.1.5 NASDAQ

La información relativa al índice Nasdaq se ha obtenido haciendo uso de *Yahoo Finance*.



Figura XI. Distribución de empresas por industria en la bolsa Nasdaq / Fuente: Nasdaq Composite

Como se muestra en la Figura XI, la mayoría de las empresas que cotizan en este índice presentan una distribución similar a la que se presenta en el conjunto seleccionado para este trabajo (Figura VII).

A partir de estos datos, ampliados con los datos históricos de las acciones de las empresas, se ha calculado la rentabilidad trimestral empleando retornos logarítmicos.

$$R_i = \log\left(\frac{P_f}{P_i}\right)$$

Los motivos de utilizar la rentabilidad logarítmica son los siguientes:

- Sumables/Aditivos en el tiempo
- Estabilidad numérica y log normalidad
- Mayor velocidad de procesamiento computacional

Además, cuando se trata de cifras no muy altas el cálculo es bastante similar a una rentabilidad simple.

5.2 PREPROCESADO

Finalmente, en adición a todo lo anterior, se han homogeneizado las fechas de resultados ya que algunas empresas presentan resultados en meses alternativos a la norma contable general, al finalizar los meses de marzo, junio, septiembre y diciembre.

Capítulo 6. **PREDICCIÓN Y FORECASTING**

Uno de los objetivos principales en el ámbito de la inversión y del mundo financiero en general consiste en tratar de predecir el valor futuro de las actividades comerciales de cualquier compañía.

La predicción y el *forecasting* financiero representan herramientas clave para anticipar comportamientos futuros en los mercados, facilitando la toma de decisiones estratégicas fundamentadas en datos históricos y tendencias observadas. Este capítulo aborda el desarrollo e implementación de modelos predictivos avanzados, específicamente orientados al análisis de indicadores financieros relevantes.

En primer lugar, se hace una revisión rápida de la teoría del *Random Walk* y su relación con las redes neuronales recurrentes LSTM cuando se aplican a la predicción del valor de la acción.

Posteriormente, se presenta la herramienta de *forecasting* elegida, *Prophet*, y los resultados que se obtienen tras su aplicación al estudio de ventas y otros indicadores financieros.

Finalmente, se utilizan las métricas proyectadas para una valoración intrínseca simple de las compañías estudiadas.

Además, se destacan las consideraciones metodológicas necesarias para garantizar la fiabilidad de las predicciones, evaluando la capacidad predictiva mediante métricas adecuadas. El objetivo último de este capítulo es proporcionar un marco riguroso que permita comprender mejor el comportamiento futuro de las variables financieras clave.

6.1 RANDOM WALK

Burton G. Malkiel en *A Random walk down Wall Street* [7] fundamenta la teoría del paseo aleatorio (en adelante RW por sus siglas en inglés *Random Walk*).

La teoría sostiene que la información pasada de la variación del precio de las acciones no contiene utilidad suficiente que permita al inversor tener más rendimiento en la gestión de una cartera de valores que con una estrategia en la que simplemente se compren y mantengan los títulos a largo plazo.

¿A qué se deben estas conclusiones tan pesimistas? En su forma más básica, los RW son cualquier proceso aleatorio donde la posición de un elemento en cierto instante de tiempo depende exclusivamente de su posición en un instante anterior y de una o varias variables aleatorias que varían con respecto al tiempo.

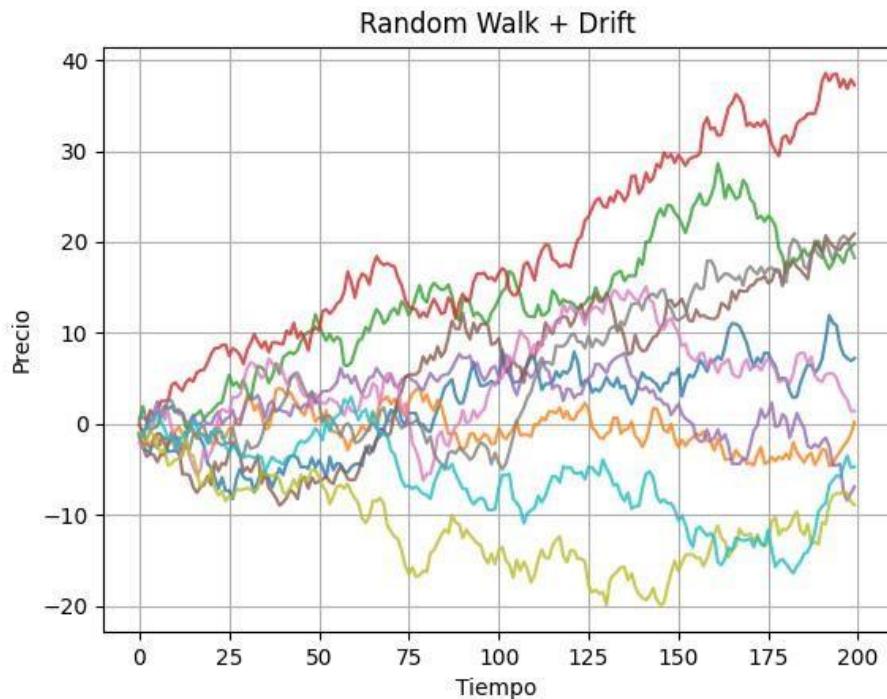


Figura XII. Ejemplo de Random Walk

En la Figura XII se presenta un ejemplo de RW y sus posibles caminos a lo largo del tiempo (diferentes colores). Se puede apreciar que siguen líneas parecidas a las que mantendría una acción de una empresa cualquiera.

Hay muchos factores que son necesarios tener en cuenta y que afectan al precio de las acciones, por lo que es prácticamente imposible crear un modelo que tenga en cuenta todas esas variables.

Otro de los motivos principales es que muchos de estos factores no son conocidos de antemano ni pueden llegar a conocerse hasta el momento en el que tienen lugar. Estos son los denominados *shocks* económicos: guerras, pandemias, casos de corrupción internos, discursos políticos, etc.

Por otro lado, si una sola persona pudiera predecir con exactitud valores futuros, ello implicaría que otras personas podrían hacerlo también. El simple hecho de la posibilidad de predicción afecta también a los movimientos del mercado ajustando los precios.

6.1.1 MODELIZACIÓN DE UN MERCADO FINANCIERO COMO SISTEMA CAÓTICO

Los mercados financieros se pueden modelar como sistemas caóticos de nivel 2 [8], es decir, aquellos sistemas que como se ha explicado antes, son capaces de reaccionar a las predicciones hechas sobre ellos mismos:

- Paradoja de la predicción perfecta (información perfecta): se imagina que un particular o grupo realizan una predicción con una precisión del 100%. En el momento en que dicha predicción se hiciese pública, los participantes en el mercado reaccionarían a esta información alterando el precio, invalidando la predicción inicial.
- Las decisiones de los inversores cambian según sus expectativas de las expectativas de los demás.

Una forma empírica de comprobar esto es tratar de utilizar una red neuronal recurrente para intentar predecir el precio de las acciones.

En este caso se demostrará con una red neuronal recurrente (RNN) llamada LSTM sobre el precio de las acciones de *Activision Blizzard*, empresa que fue comprada por Microsoft en 2023 y que, pese a pertenecer inicialmente al conjunto de datos, se ha descartado para el estudio por dicho motivo.



Figura XIII. Serie histórica de la acción de Activision

Si se hace *zoom* en el tramo de Train y Test (Figura XIV) se puede apreciar que la predicción tiene una precisión muy alta, y que incluso parece ser capaz de adaptarse a cambios más bruscos en el tiempo.



Figura XIV. Test y forecast de la serie histórica de la acción de Activision

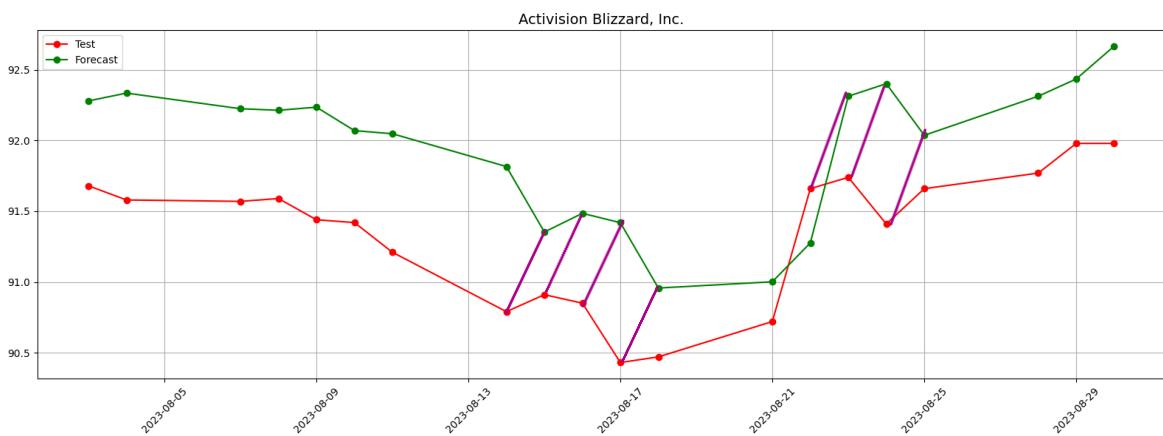


Figura XV. Test y LSTM forecast en escala temporal diaria

Con más “zoom” aún se puede finalmente apreciar qué está pasando en realidad (Figura XV). La red neuronal utiliza como valor de *forecast* el valor del día anterior junto con un incremento constante (líneas moradas).

Para explicar esto se puede añadir al modelo una predicción realizada a través del **método Naïve**.

Para este tipo de predicciones, el valor predicho siguiente es simplemente el valor real anterior, siguiendo la siguiente fórmula:

$$\hat{y}_{t+1} = y_t$$

Además, *naïve forecast* es óptimo cuando los datos siguen una estructura de RW. Es por ello por lo que a menudo se le denomina también *random walk forecast* [9]

En la Figura XVI se puede apreciar ya, a simple vista, que la predicción siguiendo esta fórmula mejora la realizada con LSTM. Para comprobarlo numéricamente se puede observar la Tabla 2 con los errores obtenidos.

<i>Modelo</i>	<i>MSE</i>	<i>MAE</i>
LSTM	2.3405	0.9927
Naive Forecast	2.1103	0.8911

Tabla 2. Errores LSTM vs Naive

Por un lado, **MSE** (*Mean Squared Error*) hace hincapié en los errores grandes y es útil para controlar la exposición a esos mismos errores. Por su parte, **MAE** (*Mean Absolute Error*) mide el promedio de errores absolutos por lo que es más sencillo de interpretar.



Figura XVI. Test, LSTM forecast y Naive forecast

6.2 PROPHET

La capacidad de estimar de forma sencilla, rápida y fiable la evolución futura de indicadores como las ventas, el flujo de caja libre o el beneficio neto resulta especialmente importante para determinar el desempeño de una empresa a lo largo del tiempo.

Frente a técnicas tradicionales como *ARIMA* o la aplicación directa de la regresión lineal, *Facebook Prophet* ofrece un enfoque robusto, muy adecuado para datos financieros debido a su adaptación a estacionalidades y tendencias, así como por su baja sensibilidad a *outliers*.

6.2.1 SERIE TEMPORAL

Antes de probar el modelo es necesario una breve explicación de los componentes de una serie temporal:

- Tendencia (t_t): comportamiento a largo plazo.
- Estacionalidad (s_t): fluctuaciones regulares que se repiten anualmente aproximadamente en el mismo espacio temporal y con intensidad similar.

- Ciclicidad u oscilaciones cíclicas (c_t): comportamiento a medio plazo. Asociado con ciclos económicos.
- Ruido (ε_t): comportamiento a corto plazo no explicado.

De modo que una serie temporal y_t puede definirse como la suma de las componentes anteriores (en el caso de un modelo aditivo):

$$y_t = t_t + s_t + c_t + \varepsilon_t$$

A continuación, se presenta la descomposición de la serie temporal de ventas de una de las empresas analizadas como ejemplo del modelo:

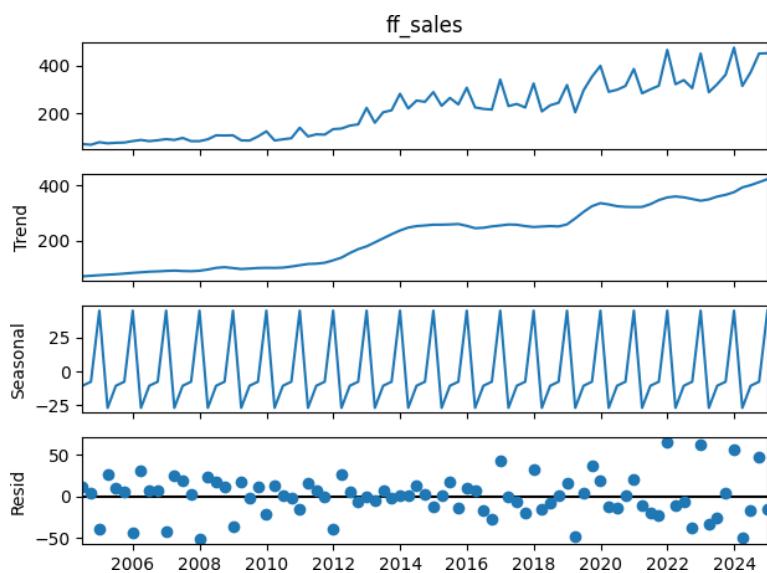


Figura XVII. Descomposición de serie temporal

6.2.2 ALGORITMO DE FORECAST FACEBOOK PROPHET

A continuación, se aplica el modelo *Prophet* [10] para predecir los tres indicadores financieros ya mencionados. Las figuras presentadas pertenecen únicamente a una empresa, en concreto *ACI Worldwide*, y sirven como ejemplo para visualizar gráficamente el comportamiento del modelo ya aplicado.

6.2.2.1 Ventas

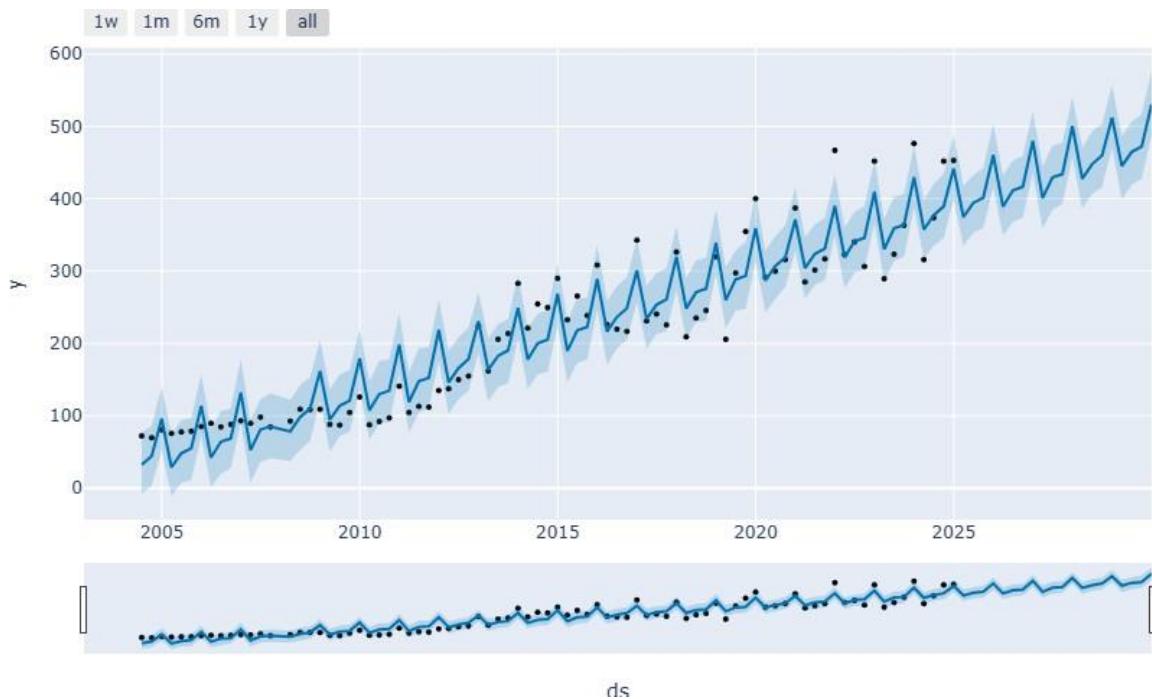


Figura XVIII. Forecast de ventas con Prophet

En la figura se encuentran los datos disponibles (puntos negros) junto con la predicción que realiza el modelo en promedio (línea azul) y, además, una predicción del posible rango en el que puede fluctuar. En el eje Y el valor de la variable y en el eje X la línea temporal.

Prophet también permite la descomposición en componentes. En la figura siguiente se presenta la tendencia de la serie anterior a lo largo del tiempo, así como la estacionalidad anual.

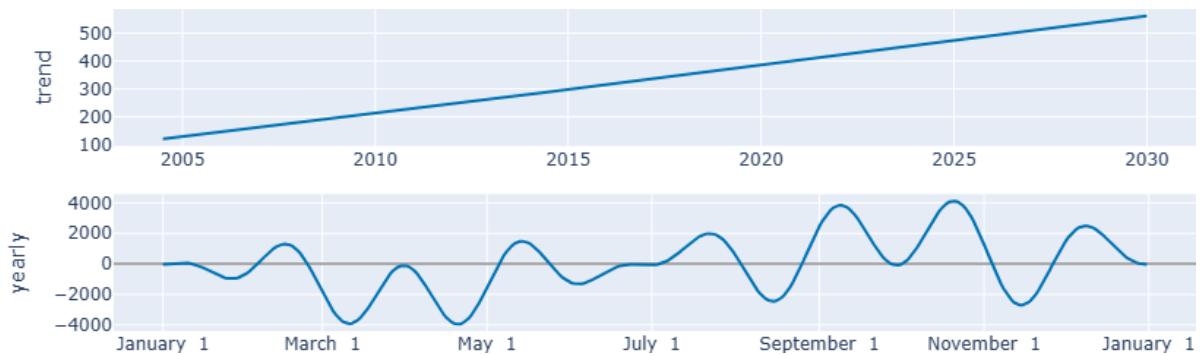


Figura XIX. Descomposición serie temporal con Prophet

6.2.2.2 FCF

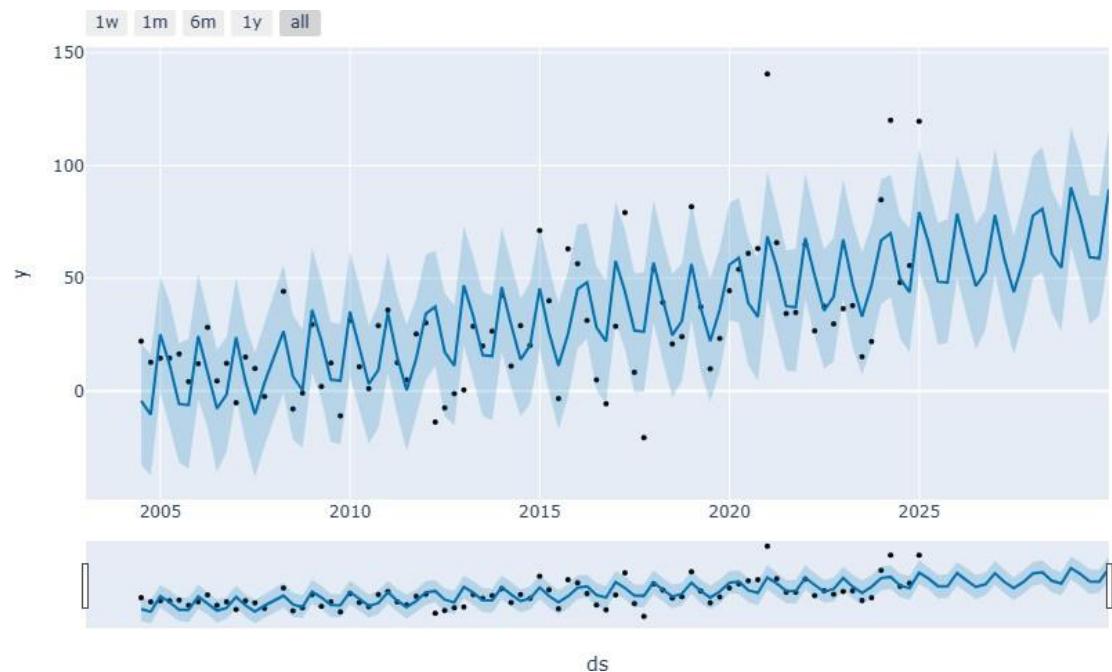


Figura XX. Forecast FCF con Prophet

6.2.2.3 *Net Income*

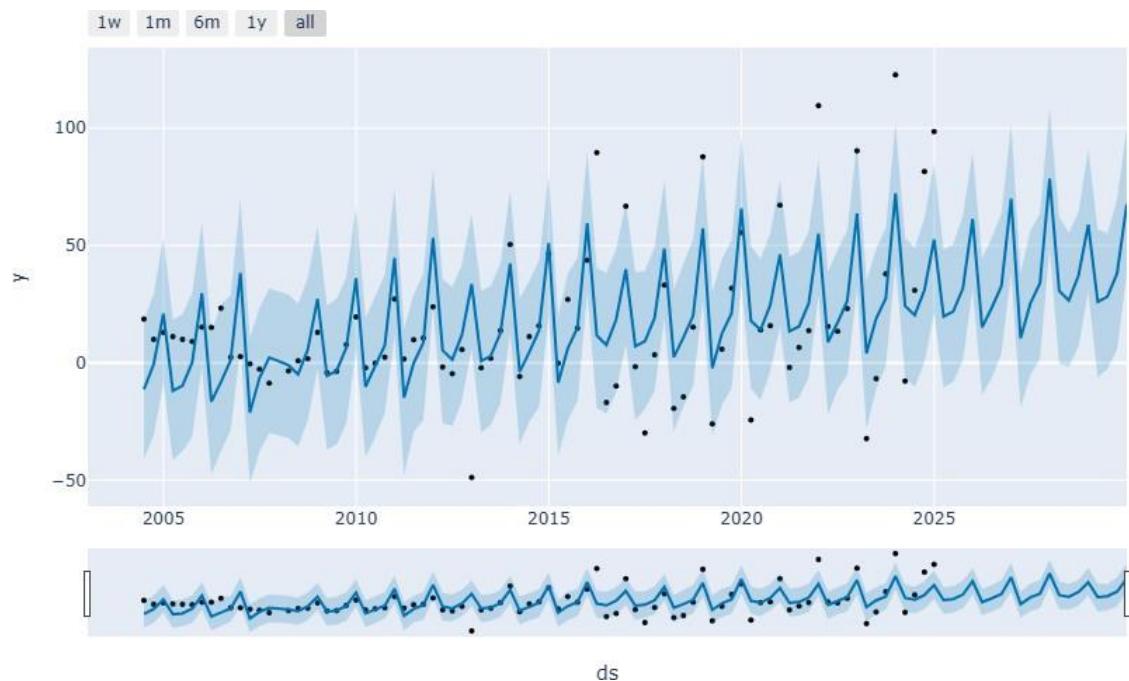


Figura XXI. Forecast Net Income con Prophet

6.3 MÉTODOS DE VALORACIÓN

La valoración de empresas es un proceso fundamental en el sector financiero que consiste en determinar el valor real o intrínseco de una compañía o elementos de ésta. Constituye, por tanto, una de las herramientas fundamentales para la toma de decisiones estratégicas, principalmente en operaciones de inversión.

Para este trabajo en concreto se utilizarán algunos modelos de valoración muy reconocidos por su eficacia y relevancia práctica. Sin embargo, debido a la complejidad inherente de estos modelos, se realizan ciertas simplificaciones respecto a una valoración personalizada y exhaustiva, típica cuando se analizan compañías de forma individualizada.

Además, los valores proyectados futuros son los obtenidos tras la aplicación de *Prophet* en la sección anterior.

6.3.1 DCF – *DISCOUNTED CASH FLOWS*

Este mecanismo estima el valor intrínseco de la acción de una empresa realizando primero una proyección de los flujos de caja (FCF) y posteriormente descontándolos a través de una tasa que refleja el coste de oportunidad y el riesgo natural del negocio, el coste medio ponderado de capital o WACC.

Este método es especialmente útil para inversores y analistas porque permite descomponer el valor en componentes temporales, evaluar el impacto de distintas hipótesis de crecimiento o margen, y comparar el precio de mercado con el valor fundamental, contribuyendo así a decisiones de inversión, fusiones y adquisiciones, o reestructuraciones de capital.

$$V = \sum_{t=1}^T \frac{FCF_{t+1}}{(1 + WACC)^t}$$

Para este trabajo se utiliza un horizonte temporal de $T = 5$ años y un WACC de 10%, que surge de una aproximación de las empresas estudiadas:

- Para grandes compañías tecnológicas consolidadas (hardware, semiconductores, software corporativo), el WACC se encuentra en un rango de 8%-10%.
- Para empresas más jóvenes o de nicho de alto crecimiento, es razonable asumir un WACC de 10 %-12 % dependiendo de su perfil de riesgo.

Calcular el WACC personalizado para cada empresa requeriría de información adicional y una complejidad añadida. De hecho, existen modelos exclusivamente orientados a esta actividad, por lo que excede a los objetivos de este proyecto [11].

Además de esto, se multiplica por un valor terminal al último año que capta los flujos existentes más allá del periodo explícito (suele ser un múltiplo utilizado como *benchmark*). En este caso, el valor terminal es 20.

Las conclusiones, por tanto, serán:

- DCF < Precio, existe una sobre-ponderación de la empresa por parte del mercado.
- DCF = Precio, la valoración respalda el mercado.
- DCF > Precio, existe una infra-ponderación de la empresa por parte del mercado. Este caso es el más llamativo porque representaría una posible oportunidad de compra o, al menos, una investigación más exhaustiva sobre la empresa para determinar el porqué de este resultado.

6.3.2 DFE – DISCOUNTED FUTURE EARNINGS

Este mecanismo estima el valor intrínseco de una empresa realizando primero una proyección de los beneficios por acción estimados (EPS) y posteriormente descontándolos de forma similar al anterior caso.

6.3.3 MÚLTIPLOS

Las dos ratios de valoración siguientes incorporan a la valoración una cierta medida del componente de sentimiento existente.

6.3.3.1 Media P/E a 5 años

Esta métrica, también conocido en el sector como PER, busca indicar cuánto está dispuesto a pagar un potencial inversor por cada unidad monetaria de beneficio generada por la compañía.

Un PER alto puede reflejar expectativas de crecimiento elevado o una sobrevaloración, mientras que un PER bajo podría indicar infravaloración o escaso potencial de

crecimiento. Es ampliamente utilizado para comparar la valoración de empresas dentro de un mismo sector o industria, sirviendo como referencia de lo que el mercado considera un nivel “razonable” de precio en función de los beneficios históricos o esperados. En este proyecto se utilizará la media del P/E de los últimos cinco años para suavizar la volatilidad de resultados puntuales y obtener una estimación más estable del valor de mercado relativo de cada empresa.

$$PER = \frac{Price/Share}{EPS}$$

6.3.3.2 *Media P/S a 5 años*

P/S mide la relación existente entre la capitalización bursátil de una empresa y sus ingresos totales. Permite valorar cuánto está dispuesto a pagar un potencial inversor por cada unidad monetaria de venta generada por la compañía. De forma alternativa se puede calcular de la siguiente forma:

$$\frac{P}{S} = \frac{\frac{Price}{Share} * Shares}{Net Income}$$

Esta medida es especialmente útil en empresas que aún no generan beneficios como ocurre en la fase más temprana del desarrollo de la mayoría de las empresas o en aquellas cuyo crecimiento es especialmente rápido.

Utilizar *Prophet* para la previsión de variables financieras con un DCF/DFE resulta un marco coherente y replicable de valoración ya que sobresale en la simplicidad y en el manejo de estacionalidades, aunque la fiabilidad del valor intrínseco depende finalmente de la solidez de los supuestos.

6.4 CONCLUSIONES

C1. *Prophet* parece ser un modelo muy eficaz a la hora de proyectar flujos futuros como se ha mostrado en el ejemplo. Sobre todo, con empresas con observaciones desde hace más años.

C2. El método de valoración DCF es útil para un gran porcentaje de las empresas estudiadas. Algunos ejemplos de los resultados obtenidos son:

- *ATOSS Software* recibe una valoración con *Prophet* de 122 USD frente a 115 USD que le otorgaba el mercado en diciembre de 2024. En este caso el mercado estaba infra ponderando la empresa.

En el mes de junio de 2025 el valor de la acción alcanza los 140 USD confirmando que la empresa tenía margen de crecimiento.

- *Bentley Systems* recibe una valoración con *Prophet* de 42 USD frente a 47 USD que le otorgaba el mercado en diciembre de 2024. En este caso el mercado respalda la valoración.

En el mes de junio de 2025 el valor de la acción alcanza los 51 USD confirmando que la empresa se encuentra en un valor estable.

C3. Los resultados muestran la escasa idoneidad de utilizar redes neuronales recurrentes como LSTM para la predicción del valor de la acción a futuro ya que, al trabajar con series temporales del estilo *random walk*, las predicciones derivan en un *naïve forecast*, en el que se toma como valor de predicción el anterior sumando un gradiente.

Capítulo 7. REGRESIÓN

Este capítulo se centra en la aplicación de una regresión lineal múltiple (en adelante MLR) con el propósito de cuantificar la capacidad explicativa de los indicadores financieros elegidos sobre el rendimiento de la acción en el mercado en cuatro horizontes temporales distintos: uno, tres, cinco y diez años.

En primer lugar, se explican brevemente los fundamentos de un MLR y qué dificultades presenta cuando se aplica a datos financieros de empresas.

Posteriormente se aplica el algoritmo y se explican los resultados obtenidos generando conclusiones significativas que, además, se utilizarán en el *Capítulo 8. Clustering*, donde se finalizará el proyecto.

7.1 MARCO TEÓRICO

Una MLR modela la relación lineal existente entre una variable dependiente y (objetivo) y un conjunto de variables independientes o explicativas x_n de forma que:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon$$

Los coeficientes β_n representan ese grado de explicación de la variable respecto al objetivo. Por su parte, ε recoge los efectos no observados o “ruido” existente.

El modelo general emplea el algoritmo de Mínimos Cuadrados Ordinarios (en adelante OLS), el cual busca minimizar la suma cuadrática de los residuos. Bajo supuestos de linealidad, variables independientes exógenas, homocedasticidad (varianza de errores constante en todos los niveles de las variables independientes) y ausencia de multicolinealidad, el estimador es insesgado, eficiente y consistente.

7.2 CORRELACIÓN Y MULTICOLINEALIDAD

En esta sección se hace patente que no basta con elegir variables diferentes si no que tienen que medir parámetros distintos y no ser versiones escaladas o derivadas unas de otras.

El **VIF** (*Variance Inflation Factor*) es una métrica utilizada para detectar multicolinealidad entre variables independientes. Esto ocurre cuando una variable explicativa puede predecirse linealmente a partir de otras, algo que ocurre con gran frecuencia con los indicadores financieros.

En la secuencia de tres imágenes siguientes se observa como al eliminar variables con relación lineal se consigue llegar a un $VIF < 5$ que no distorsione la estimación y fiabilidad del modelo.

Variable	VIF	Variable	VIF	Variable	VIF
ff_curr_ratio	218.038148	ff_curr_ratio	216.335559	ff_quick_ratio	2.120803
ff_quick_ratio	236.191085	ff_quick_ratio	234.041319	ff_gross_mgn	4.218292
ff_gross_mgn	2.550608	ff_gross_mgn	4.689138	net_margin	6.057508
ff_net_mgn	1.317746	net_margin	6.057508	ff_cash_ratio	16.002397
ff_entrepr_val	34.676504	ff_cash_ratio	16.002397	ROE	1.028058
ff_cash_ratio	16.472622	ROE	1.028820	ROA	2.229258
ff_cogs	435.739386	ROA	2.229258	fcf/sales	2.622767
ff_ebitda_oper	123.187863	fcf/sales	2.937564	oper_sales	4.401632
ff_net_income	6.586824	oper_sales	7.800081	CAGR_1y	1.137765
ff_oper_exp	2758.242011	CAGR_1y	1.148192	debt_assets	2.406485
ff_sales	2458.796824	debt_assets	2.634517	ev_ebitda	1.014897
ff_eps	1.671013	ev_ebitda	1.015844		
ff_capex	52.389594				
ff_free_cf	9.898332				
ff_assets	3489.556981				
ff_liabs_curr	146.744123				
ff_eq_tot	1830.728926				
ff_debt	117.661531				
CAGR_1y	1.199867				

Figura XXII. Secuencia de disminución de VIF

Por otra parte, se comprueba la **correlación** inicial existente entre las variables y el objetivo: retornos de la acción de las empresas a los horizontes temporales descritos.

<i>Variable</i>	<i>Retorno 1 año</i>	<i>Retorno 3 años</i>	<i>Retorno 5 años</i>	<i>Retorno 10 años</i>
Current ratio	-7	-6	-4	-3
Quick ratio	-7	-6	-4	-3
Gross margin	15	3	1	0
Net margin	22	7	6	5
Cash ratio	-7	-7	-1	-1
ROE	4	3	3	3
ROA	12	5	2	0
FCF/Sales	21	8	6	4
Opex/Sales	-17	-7	-6	-5
CAGR	-8	-1	-1	-1
Debt/Assets	-6	-3	-4	-5
EV/EBITDA	8	5	5	3

Tabla 3. Correlación [%] de las variables con rentabilidad de la acción en el horizonte temporal

7.3 REGRESIÓN

A continuación (Tabla 4) se presentan los resultados obtenidos tras la aplicación del modelo de regresión con las variables seleccionadas tras el estudio del VIF. Para ello se ha realizado un estudio *cross-sectional* en el que se seleccionan datos en puntos específicos del tiempo y de todas las empresas disponibles en ese periodo.

R^2 es el coeficiente de determinación. Representa la proporción de la varianza de la variable dependiente o, dicho de otro modo, cuánto del objetivo son capaces de explicar las variables seleccionadas.

<i>Modelo</i>	$R^2[\%]$
1 año	6.1
3 años	1.2
5 años	1.0
10 años	0.9

Tabla 4. Resultados de la regresión (R2)

7.4 CONCLUSIONES

C1. Los modelos explican muy poco de la varianza de la rentabilidad de la acción en los horizontes temporales seleccionados. Este resultado no es necesariamente malo. De hecho, confirma que en el contexto financiero el retorno está enormemente influido por multitud de factores impredecibles: macroeconómicos, noticias, sentimiento del comprador, sector privado...

Por otro lado, esta línea de resultados es también acorde a la teoría del paseo aleatorio explicada en la sección “6.1. *Random Walk*”, en la que se explica la importancia de la aleatoriedad de las acciones.

C2. Al ampliar el horizonte de predicción, el poder explicativo de los modelos de MLR disminuye progresivamente hasta acercarse a ser 0. Este resultado también es coherente con la literatura financiera existente, que destaca la dificultad de anticipar retornos a largo plazo únicamente a partir de indicadores fundamentales. Con el paso del tiempo la acumulación de más eventos exógenos y la aleatoriedad hacen que los modelos pierdan precisión.

C3. Los márgenes, eficiencia operativa y generación de caja son elementos claves para el retorno de la acción a largo plazo, por su gran correlación.

C4. La liquidez o el crecimiento pasado no garantizan mayor rentabilidad de la acción en un futuro. De hecho, tener demasiada liquidez puede indicar ineficiencia en el uso de capital al ser dinero que no está generando retornos.

C5. Al eliminar variables con alto VIF se ha tenido que elegir entre indicadores que son muy parecidos, pero representan características distintas. Se ha llegado a la conclusión que, para eliminar la colinealidad, es necesario elegir variables que midan sectores y aspectos distintos del negocio dentro de la misma compañía.

Estas variables son las que posteriormente se utilizan a la hora de realizar el *clustering* de empresas. Se muestran a continuación:

- Márgenes → Entre *Gross Margin* y *Net Margin* se selecciona *Gross Margin* pues muchas de las empresas no son rentables de forma neta durante el análisis. Con un margen bruto es más sencillo evaluar si el negocio tiene potencial.
- Conversión de ventas a caja → FCF/Sales.
- Rentabilidad → ROE y ROA.
- Eficiencia operativa → Opex/Sales.
- Liquidez → Entre *Current Ratio*, *Quick Ratio (Acid Test)* y *Cash Ratio* se ha optado por seleccionar el *Acid Test*, pues mide la liquidez de forma estricta, sin contar con inventarios que, en el sector de las tecnológicas, no son tan importantes.
- Crecimiento → CAGR.
- Apalancamiento → Debt/Assets.
- Retornos bursátiles.

Capítulo 8. **CLUSTERING**

Este capítulo se centra en la aplicación de algoritmos de aprendizaje no supervisado al conjunto de datos disponible. En primer lugar, se explican los fundamentos teóricos del *clustering* en general y qué objetivos principales derivan de la implementación de esta técnica; posteriormente se explica y aplica el algoritmo *K-means* para clasificación.

También se justifica la elección del número de grupos mediante dos criterios cuya interpretación también es explicada.

Finalmente se interpretan los resultados para generar grupos de interés.

8.1 MARCO TEÓRICO

El objetivo principal del *clustering* es la división del conjunto de observaciones en grupos (*clusters*) de manera que la distancia entre los elementos de un mismo grupo sea mínima y que la distancia entre grupos sea máxima. Esta distancia se puede calcular mediante diversas fórmulas, aunque la más conocida es la Euclídea.

El hecho de ser no supervisado implica que el método encuentra la estructura latente en los datos sin necesidad de etiquetas predefinidas. El objetivo principal es, por tanto, revelar patrones ocultos, segmentar datos, detectar anomalías o fraude y ayudar a reducir la complejidad inicial.

Más concretamente, aplicado a datos financieros, se añaden los siguientes objetivos:

- Encontrar pares comparables o “*comps*” con características similares evitando comparaciones sesgadas por el sector formal o el tamaño.
- Analizar salud financiera y riesgo crediticio para identificar posibles *defaults*, (empresas incapaces de pagar su deuda).

- Alertar de *outliers* que merezcan una vigilancia especial
- Dentro de los grupos, las empresas pueden realizar un *benchmarking* interno y fijar objetivos realistas para conseguir la media del clúster o incluso mejorarlo.

8.2 KMEANS

K-means es el algoritmo de *clustering* más utilizado debido a su simplicidad a la hora de aplicarse y la eficiencia en la segmentación. Sigue un proceso iterativo a la hora de calcular los grupos que se puede resumir de la siguiente forma:

1. Se eligen, *a priori*, k clústeres deseados. Es importante que este número respete la cantidad de observaciones de las que se dispone. Por ejemplo, no tendría demasiado sentido elegir $k=80$ en este proyecto ya que sólo disponemos de 160 empresas y cada grupo estaría formado por 2 observaciones.
2. Se inicializan los centroides (puntos representativos).
3. Se calcula la distancia de cada punto sobre los centroides y se le asigna al más cercano.
4. Se recalcula el centroide en el centro de las observaciones asignadas al mismo.
5. Se repite hasta que los centros no cambien o se llegue a un número fijo de iteraciones.

8.2.1 AGRUPACIÓN

Debido a que los datos inicialmente se presentan en trimestres no tendría mucho sentido realizar el *clustering* de este modo. Por ello se agrupan los datos para obtener observaciones anuales sumando o calculando la media dependiendo de la variable en cuestión.

8.2.2 NORMALIZACIÓN

Se lleva a cabo una normalización de los datos para conseguir una media cero y desviación típica de uno, evitando así problemas de escala y un sesgo por tamaño en la clasificación.

8.2.3 PCA

El paso siguiente consiste en un Análisis de Componentes Principales (*Principal Component Analysis*) con el objetivo de reducir la dimensionalidad a dos dimensiones para ofrecer un mayor rendimiento al algoritmo posteriormente aplicado y poder visualizar los resultados.

Consigue sustituir variables correlacionadas por unos pocos factores clave que capturan la máxima varianza posible, evitando así la colinealidad y la presencia de ruido en el modelo.

8.2.4 NÚMERO ÓPTIMO DE *CLUSTERS*

Determinar correctamente el número adecuado de *clusters*, es, como se ha mencionado anteriormente, una de las tareas más importantes a la hora de utilizar algoritmos como *K-means*. Para realizar esta tarea existen varios métodos o criterios de los cuales los dos siguientes son los más utilizados.

8.2.4.1 *ELBOW METHOD*

El método *Elbow* parte de la idea de que, si los *clusters* realmente representan estructuras naturales, añadir un nuevo centroide *k* debería reducir notablemente la dispersión interna de los puntos hasta un momento en el que dicha mejora se vuelve marginal.

Para concretar esto, se entrena los datos con *K-means* para *k* valores (generalmente de 1 a 9) y se calcula la inercia o Suma de Cuadrados de los Errores (en adelante SSE).

La gráfica suele presentar una caída brusca que cambia su pendiente en un punto, denominado codo o *elbow*. Es ese punto el número k que indica que la incorporación de más complejidad (más *clusters*) no compensa la ganancia del modelo.

Pese a que es un mecanismo fuerte debido a su sencillez y velocidad, puede resultar algo subjetivo determinar dónde se encuentra realmente el codo. Es por lo que se suele complementar con otros criterios para llegar a la solución final.

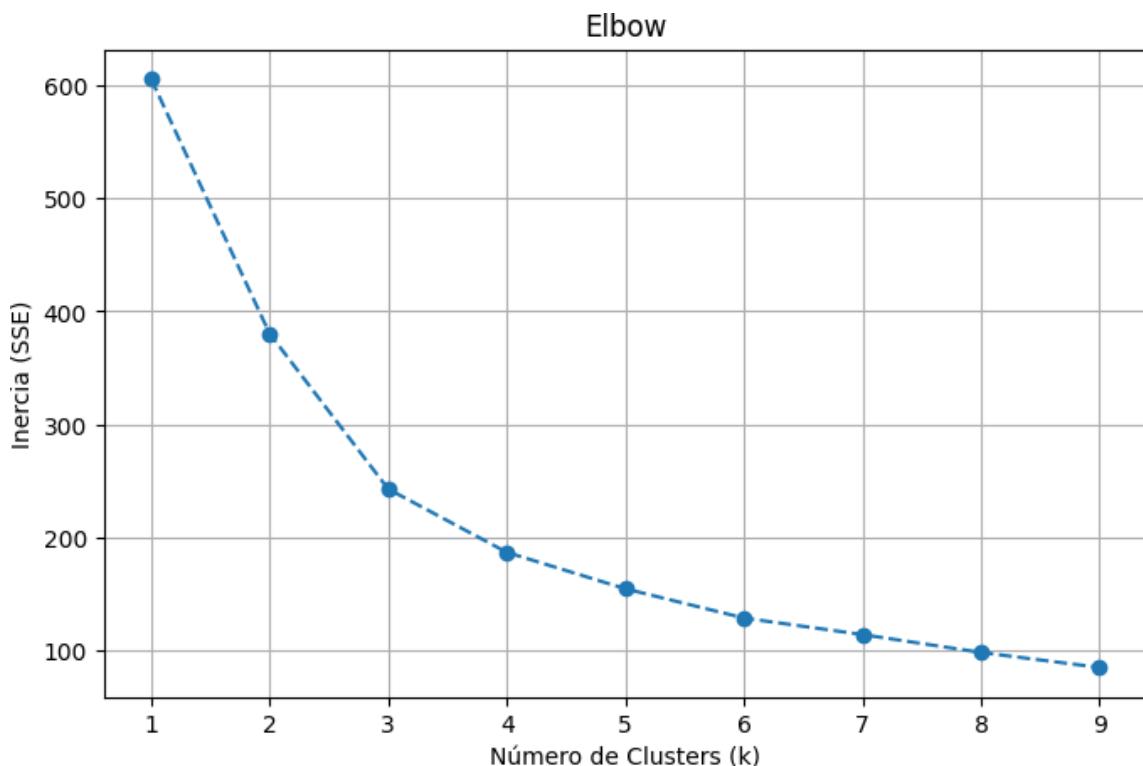


Figura XXIII. Elbow Method

En este caso el número óptimo según el criterio aparentemente sería $k=3$. Pero pueden surgir dudas respecto a $k=4$ o incluso $k=5$.

8.2.4.2 SILHOUETTE METHOD

El método *Silhouette* ofrece una medida del número óptimo de *clusters* menos sujeta a interpretación visual que el método anterior.

El algoritmo calcula para cada punto dos promedios de distancia: distancia media a los puntos de este *cluster* y distancia media al *cluster* vecino más cercano. A partir de esto define un coeficiente $\epsilon[-1,1]$. El coeficiente global es el promedio de los demás.

Este cálculo se repite para distintos valores de k y se genera una gráfica.

La ventaja evidente en este caso es la objetividad y su capacidad de identificar la calidad global, así como número de *clusters* problemáticos.



Figura XXIV. Silhouette Method

En este caso con el criterio *Silhouette* el número de *clusters* que mejor se adapta a los datos es $k=4$.

Elbow se suele utilizar como filtro rápido para acotar un rango razonable de k mientras que *Silhouette* permite afinar el número. En la Figura XXV se muestra un ejemplo de la complementación de estos dos criterios para seleccionar el número k final.

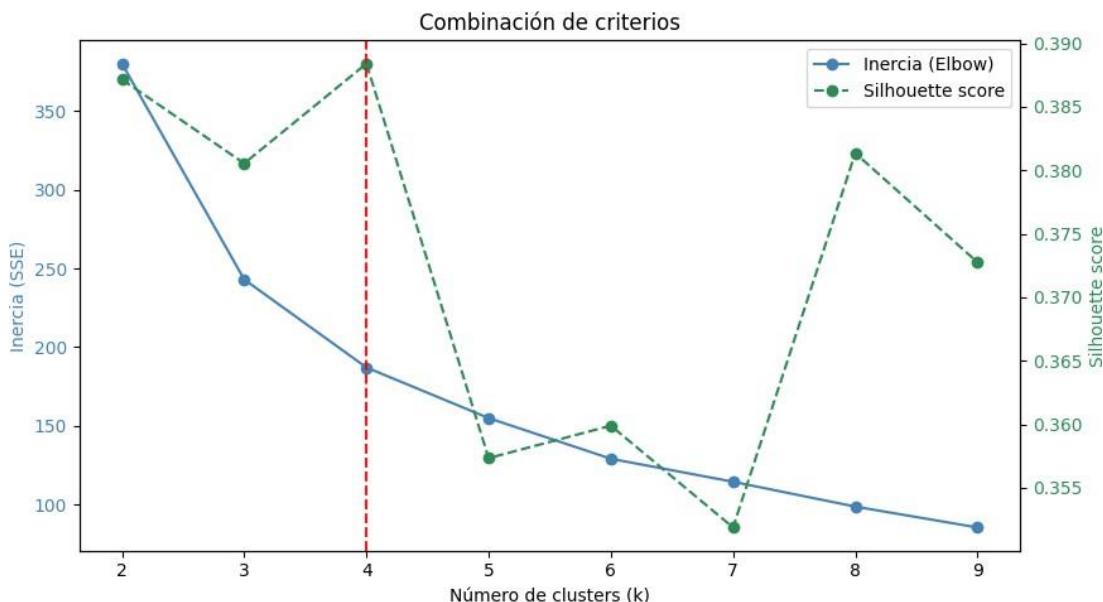


Figura XXV. Combinación de Elbow y Silhouette

8.2.5 RESULTADOS

El análisis a tres horizontes temporales enriquece el diagnóstico de cada empresa, atendiendo tanto a la coyuntura inmediata como a la trayectoria estratégica y enfoque a largo plazo durante los últimos años.

Los horizontes son:

- Cinco años: 2020-2024
- Tres años: 2022-2024
- Un año: 2024

8.2.5.1 Cinco años

El *clustering* a cinco años sirve para clasificar las empresas según su estrategia a largo plazo y resiliencia ante un ciclo económico completo (y en este caso a la salida de la crisis generada por la pandemia del Covid-19)

En este caso en concreto el número de clusters es k=4.

<i>Grupo</i>	<i>Nº empresas</i>	<i>Descripción</i>
0	30	Grandes tecnológicas, maduras y rentables, con fuerte conversión de ventas en caja.
1	35	Compañías que sacrifican beneficios actuales para ganar crecimiento y aumentar ventas.
2	18	“Startups” en fase temprana con fuertes inversiones en I+D. Poca rentabilidad o negativa
3	44	Empresas con crecimiento y rentabilidad discretos. Márgenes contenidos y deuda moderada. Gran proporción de empresas asiáticas.

Tabla 5. Clustering 5 años

Características numéricas notables por orden de importancia en el cluster:

<i>Grupo</i>	<i>1</i>	<i>2</i>	<i>3</i>
0	FCF/Sales 25%	Acid Test 3x	Gross Margin 78%
1	CAGR_5y 27%	Opex/Sales 118%	ROE -14%
2	FCF/Sales 17%	Acid Test 1.8x	Debt/Assets 20%
3	FCF/Sales 5%	Acid Test 1x	CAGR_5y 8%

Tabla 6. Características notables Clustering 5 años

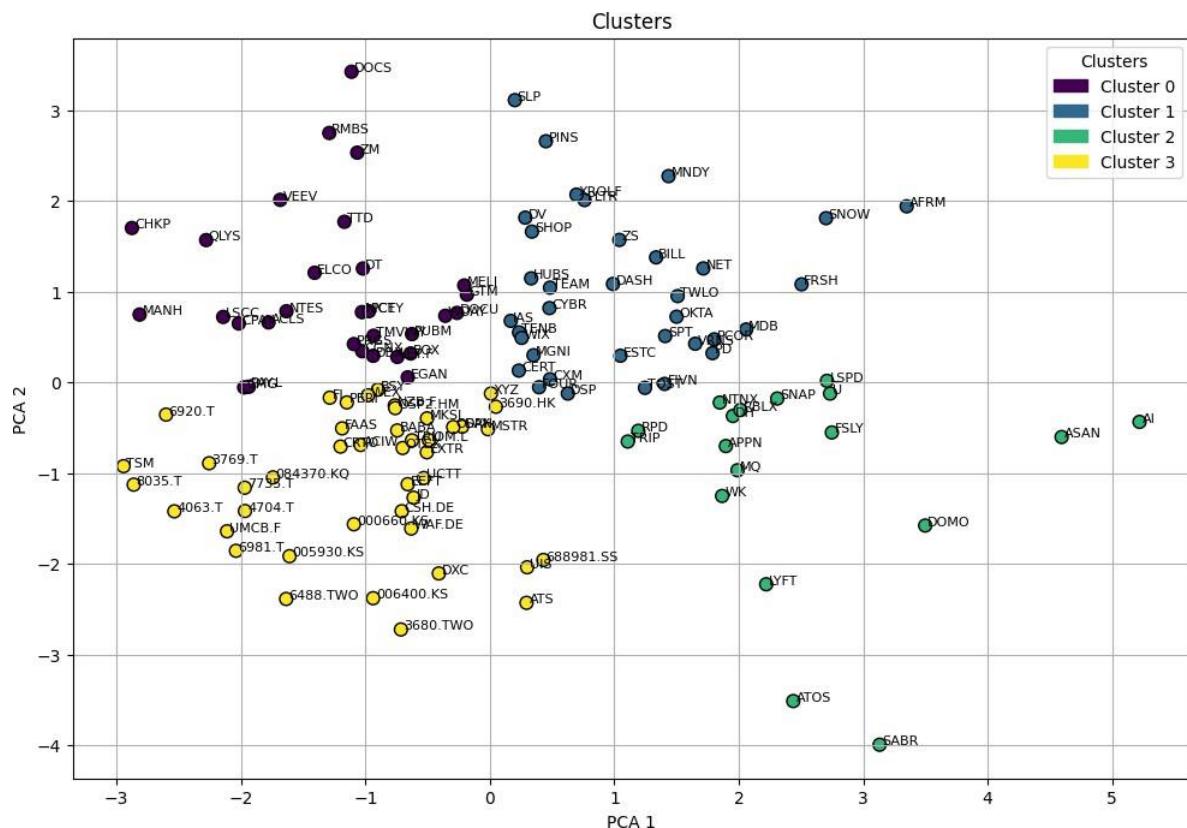


Figura XXVI. Clustering 5 años

8.2.5.2 Tres años

El *clustering* a tres años busca identificar tendencias estructurales en la evolución del modelo de negocio en un intervalo de tiempo algo menor al anterior.

En este caso el número idóneo de *clusters* es el mismo que en el anterior, $k=4$.

<i>Grupo</i>	<i>Nº empresas</i>	<i>Descripción</i>
0	37	Grandes tecnológicas, maduras y rentables, con fuerte conversión de ventas en caja.
1	49	Compañías consolidadas en nichos competitivos, con baja generación de caja y un crecimiento casi plano (periodo de madurez). De nuevo gran porcentaje de empresas asiáticas.
2	36	Empresas en estado de crecimiento con equilibrio de caja y alta inversión en gastos operacionales.
3	17	Startups tecnológicas en fase de fuerte expansión que sacrifican rentabilidad para intentar captar cuota de mercado.

Tabla 7. Clustering 3 años

<i>Grupo</i>	<i>1</i>	<i>2</i>	<i>3</i>
0	FCF/Sales 12%	Acid Test 3x	Gross Margin 70%
1	CAGR_3y 5%	Gross Margin 30%	FCF/Sales ≈ 0%
2	FCF/Sales 24%	Gross Margin 75%	Debt/Assets 50%
3	CAGR_3y 20%	Opex/Sales 140%	FCF/Sales -8%

Tabla 8. Características notables Clustering 3 años

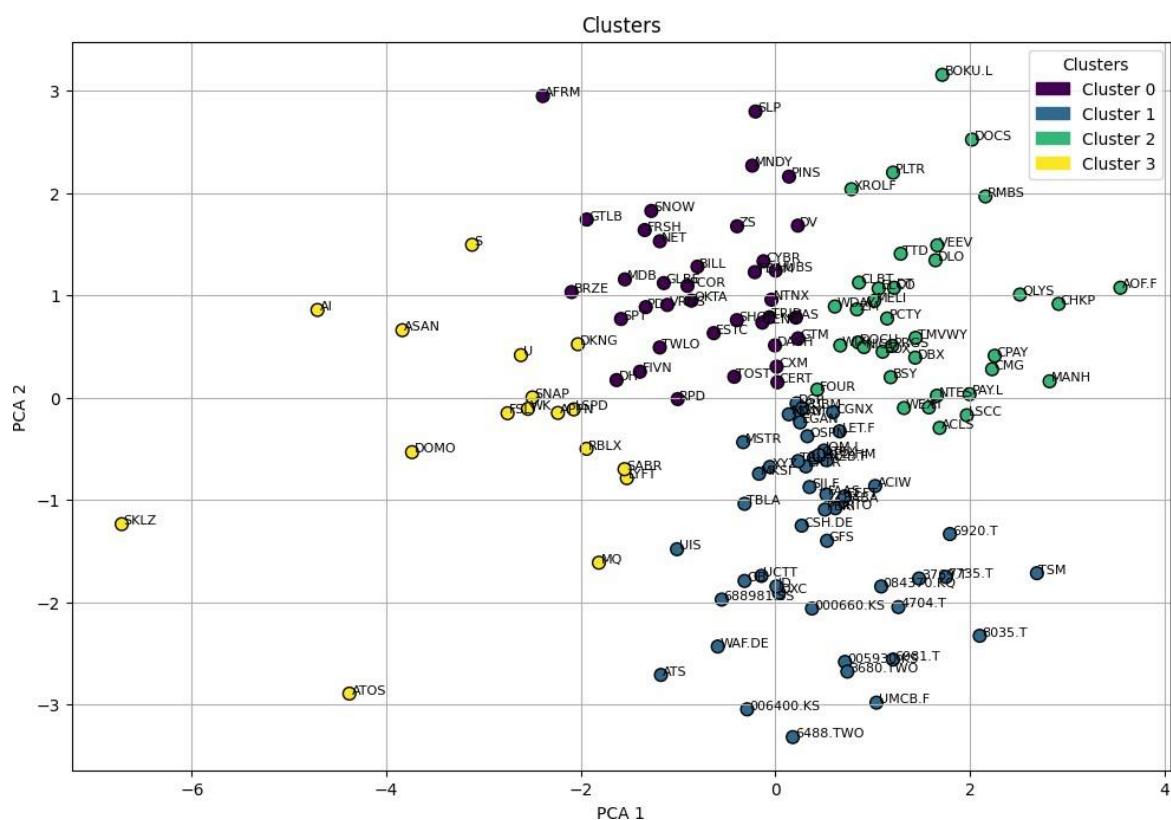


Figura XXVII. Clustering 3 años

8.2.5.3 Un año

El *clustering* a un año busca capturar una idea general del estado de las empresas durante el último año de estudio, que es 2024. Al existir un mayor número de observaciones, determinar el número de *clusters* idóneo ha resultado ser más complicado (Figura XXVIII).

Elbow sugiere un k aproximadamente igual a 4 mientras que *Silhouette* mejora a partir de $k=4$. Finalmente se elige $k=6$ como punto intermedio entre ambos criterios.

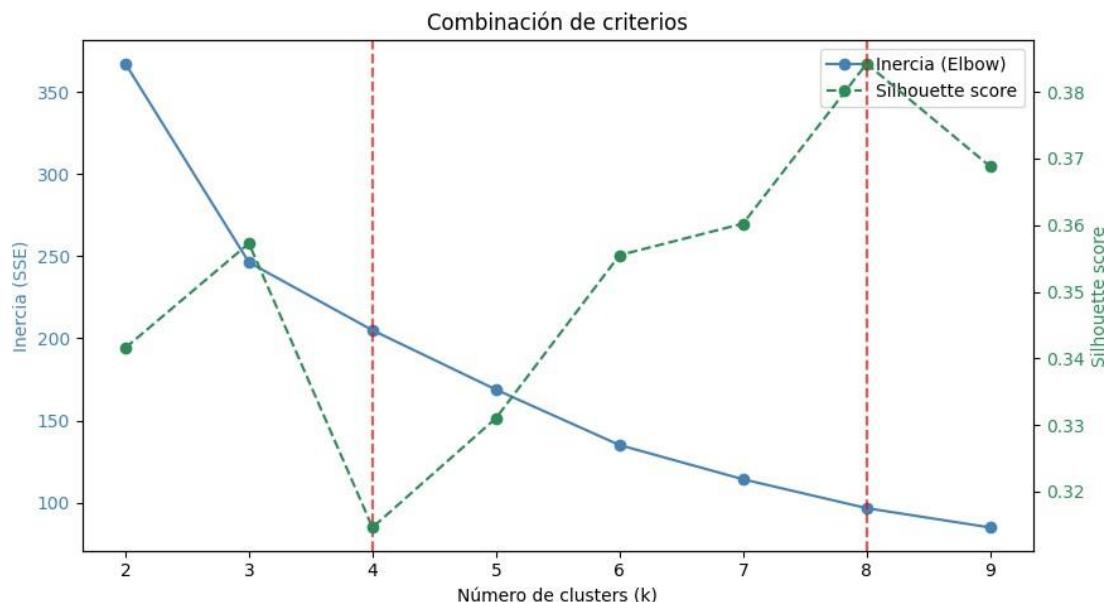


Figura XXVIII. Selección K para Clustering 1 año

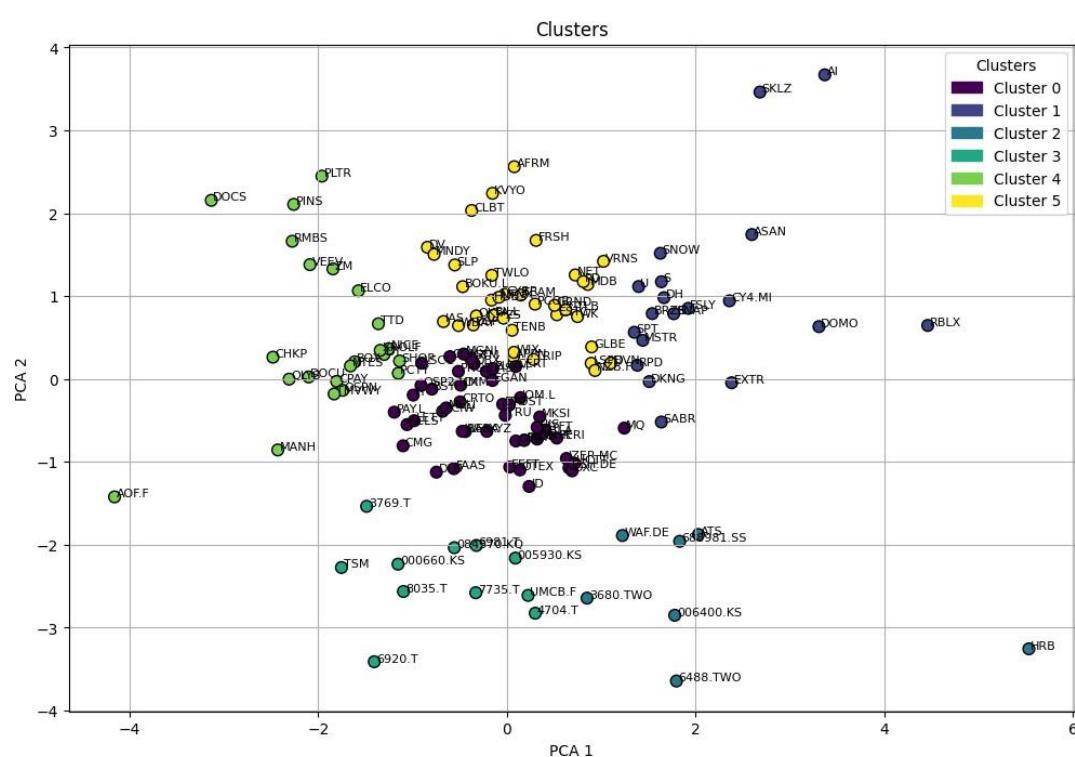


Figura XXIX. Clustering 1 año

Grupo	Nº empresas	Descripción
0	47	Crecimiento sólido y caja estable
1	19	Bajo crecimiento con altos gastos operacionales
2	7	Crisis operativa
3	11	Hipercrecimiento con gran <i>burn</i> de caja
4	23	Buenos márgenes y caja
5	36	Altos márgenes con ineficiencia de activos

Tabla 9. Clustering 1 año

Grupo	1	2	3
0	FCF/Sales 13%	CAGR_1y 12%	Gross Margin 43%
1	Opex/Sales 127%	CAGR_1y 1%	ROA -3.9%
2	FCF/Sales -52%	Gross Margin 7%	CAGR_1y -4.7%
3	CAGR_1y 23%	Gross Margin 0.4%	Acid Test 0.02x
4	FCF/Sales 31%	Gross Margin 75%	Return 4.7%
5	CAGR_1y 22%	Gross Margin 73%	ROA ≈ -1%

Tabla 10. Características notables Clustering 1 año

8.3 CONCLUSIONES

C1. La clasificación parece ser más sencilla en horizontes de tiempo más elevados debido a que las empresas tienden a parecerse más en el largo plazo que en desempeños anuales.

C2. **Clustering a 1 año:** los grupos con buen crecimiento en el último año son los más interesantes, pues son los que potencialmente aún tienen margen de mejora. Estos son los clusters 3 y 4. El primero tiene la desventaja de que quema tesorería (efectivo, ingresos) a cambio de buscar ese incremento. Dentro de este grupo se encuentran algunas empresas como:

- TSMC
- Palantir

Por su parte, el cluster 2 agrupa empresas con serios problemas debido a gastos operacionales que superan con creces sus ingresos.

- H&R Block
- ATS
- Siltronic AG

C3. Por un lado, los outliers eliminados previo al clustering nos indican ya estados financieros con problemas. Algunas empresas localizadas en este sentido son:

- Rivian Automotive
- Wolfspeed Inc
- Luminar Technologies

Capítulo 9. CONCLUSIONES Y TRABAJOS

FUTUROS

En este último capítulo se llevan a cabo dos funciones principales:

Un resumen de las conclusiones del proyecto, determinando qué aportaciones se han realizado al estado del arte y qué objetivos se han cubierto.

Una descripción de los objetivos que no se han cubierto o lo han hecho en menor medida, así como una posible línea de estudio futuro para afrontar dichos problemas.

9.1 CONCLUSIONES

Como se ha demostrado en el presente proyecto, la aplicación de técnicas de *Machine Learning* en el análisis financiero supone un avance evidente y una necesaria implementación para obtener una radiografía lo más precisa y eficiente posible de una compañía o empresa.

Los *insights* obtenidos deben ser utilizados para redefinir estrategias, sistemas, procesos o futuras inversiones.

Si bien es cierto que todavía existen carencias significativas, como la búsqueda de la forma de incorporar shocks económicos a los modelos, las ventajas obtenidas son ya diferenciales.

Por último, es necesario comentar que este tipo de enfoques no pretende sustituir la experiencia humana y el juicio financiero tradicional de los expertos (o al menos, no debería), sino ser un complemento que presente evidencias empíricas robustas

reforzando la toma de decisiones estratégicas en un entorno que es, cada vez más, volátil y globalizado.

9.2 LÍNEAS DE TRABAJO FUTURO

9.2.1 ANÁLISIS DE SENTIMIENTO Y LENGUAJE NATURAL

Quizás el objetivo futuro más claro que se desprende del trabajo realizado es la incorporación de un análisis de lenguaje natural. Como se observó en las distintas regresiones, la gran parte de la explicación del mundo financiero no viene dada directamente por los resultados históricos de las empresas. Por este motivo, cabe pensar que las palabras cada vez tienen mayor importancia e influencia en este sector. El volumen de información no estructurada ha crecido exponencialmente superando con creces la capacidad humana de procesarla de forma eficiente. Ante esta realidad el *Natural Language Processing* se vuelve prácticamente necesario en cualquier banco o fondo de inversión.

Algunas de las técnicas a desarrollar e implementar serían:

- Análisis de sentimiento mediante algoritmos de lenguaje como *VADER* o *FinBERT*
- *Topic Analysis* para determinar los temas de los que trata un texto o párrafo concreto con el fin de ahorrar tiempo en la búsqueda de información.
- Modelos de clasificación supervisados para detectar tonos optimistas o pesimistas respecto a distintas áreas de negocio: ventas, riesgos, innovación, desarrollo...

9.2.2 ANÁLISIS FUNDAMENTAL NO LINEAL CON MACHINE LEARNING

Como se ha visto sobre todo en el “Capítulo 7. Regresión”, los modelos clásicos de regresión como OLS parecen indicar una falta evidente de relación lineal entre

indicadores financieros y el desempeño bursátil futuro. En esta línea, explorar relaciones no lineales mediante algoritmos de *Machine Learning* podría permitir hallar relaciones no evidentes cuantitativamente (no lineales) o interacciones entre variables que arrojen mejores resultados.

Algunos ejemplos de algoritmos más complejos son *Support Vector Regression* basado en las conocidas SVM que funciona bien con *datasets* pequeños, regresión polinómica buscando interacciones entre indicadores o regresiones basadas en procesos gaussianos que modelan también la incertidumbre existente a la hora de realizar la predicción.

9.2.3 MODELOS HÍBRIDOS DE DATOS ESTRUCTURADOS, NO ESTRUCTURADOS Y ALTERNATIVOS

En relación con la sección 9.2.1. Análisis de Sentimiento y Lenguaje Natural, una línea de trabajo a futuro extra sería la creación de *datasets* que contengan información de tres tipos distintos:

- Datos estructurados → ratios financieros como los utilizados en este proyecto
- Datos no estructurados → textos de informes, entrevistas, notas de prensa, *earnings calls...*
- Datos alternativos → redes sociales, opiniones de consumidores o compradores, sensores o datos ESG (*Environmental, Social and Governance*)

Para la fusión de estos datos se podrían crear modelos basados en grafos que representen relaciones entre empresa, documento y sector aplicando posteriormente redes neuronales basadas, precisamente, en grafos (*Graphs Neural Networks*)

Capítulo 10. BIBLIOGRAFÍA

NOTA: todas las figuras que aparecen en el presente documento en las que NO se indica la fuente son de elaboración propia.

- [1] Jensen, R. E. (1971). *A Cluster Analysis Study of Financial Performance of Selected Business Firms*. *The Accounting Review*, 46(1), 36–56. <http://www.jstor.org/stable/243887>
- [2] Álvarez, A. d. (2025). *Impacto del Machine Learning en la cadena de suministro global: análisis de casos recientes y proyecciones futuras*. [Monografía]. Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/68418>
- [3] Jeaab, K., Ahmed, N., Falloul, M.E.M., Saoudi, Y., Yassine, A. (2025). Forecasting Volatility with Machine Learning: Case Study of the Dow Jones Industrial Average Index. In: Farhaoui, Y., Herawan, T., Lucky Imoize, A., Allaoui, A.E. (eds) *Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment*. ICAISE 2024. Lecture Notes in Networks and Systems, vol 1397. Springer, Cham. <https://doi.org/10.1007/978-3-031-90921-4>
- [4] C.G, L.L, J.M & S.S. McKinsey & Co (2023). *Been there, doing that: How corporate and investment banks are tackling gen AI*
<https://www.mckinsey.com/industries/financial-services/our-insights/been-there-doing-that-how-corporate-and-investment-banks-are-tackling-gen-ai>
- [5] European Securities and Markets Authority (2025). *Artificial intelligence in EU investment funds: adoption, strategies and portfolio exposures*
https://www.esma.europa.eu/sites/default/files/2025-02/ESMA50-43599798-9923_TRV_Article_Artificial_intelligence_in_EU_investment_funds.pdf
- [6] Wikipedia. *DIKW Pyramid*. https://en.wikipedia.org/wiki/DIKW_pyramid

- [7] Malkiel, B.G. (1973). *A Random Walk Down Wall Street*
- [8] D. B. (2024). *Financial Markets as Level 2 Chaotic Systems*
<https://www.daytrading.com/level-2-chaotic-systems>
- [9] Ellis, C.M. (2022). *LSTM time series + stock price prediction = FAIL on Kaggle*
- [10] Facebook Prophet by Meta, <https://facebook.github.io/prophet/>
- [11] Evdokimov, I., Kampuridis, M., Papastylianou, T. INNS DLIA, (2023). *Application of Machine Learning Algorithms to Free Cash Flows Growth Rate Estimation*

ANEXO I: ALINEACIÓN DEL PROYECTO CON LOS ODS

El presente proyecto se alinea con la Organización de las Naciones Unidas en su contribución y vinculación con los Objetivos de Desarrollo Sostenible siguientes:

ODS 8. Trabajo decente y crecimiento económico

El Objetivo número 8 trata de promover el crecimiento económico sostenido, inclusivo y sostenible, el empleo pleno y productivo y el trabajo decente para todos.

Al mejorar la calidad y disponibilidad de información financiera se favorece la creación de un entorno económico más estable.

Además, los análisis más precisos pueden ayudar a emprendedores o pequeñas empresas a introducirse en un mercado más justo, desarrollarse y crecer.

La parte predictiva, por su lado, ayuda a la planificación a largo plazo y a la resiliencia a posibles crisis o shocks económicos, y esto es muy importante porque este objetivo –el del crecimiento económico- la erradicación de la pobreza solo es posible mediante empleos estables y bien remunerados.

ODS 9. Industria, innovación e infraestructura

El Objetivo número 9 trata de construir infraestructuras resilientes, promover la industrialización inclusiva y sostenible y fomentar la innovación.

La gran cantidad de modelos utilizados suponen un ejemplo en la innovación tecnológica al acercar la ciencia de datos al mundo empresarial, concretamente al sector financiero.

Por otro lado, se sientan bases para el establecimiento de infraestructuras analíticas y posibles plataformas de datos que integren modelos clásicos con modelos más vanguardistas.

ODS 17. Alianzas para lograr objetivos

El proyecto fomenta la colaboración entre distintas entidades en materia de datos. Se impulsa la creación de una comunidad donde se compartan mejores prácticas de modelado financiero favoreciendo responsabilidad y responsabilidad conjunta.

El Objetivo número 17 versa sobre la necesidad de crear alianzas para revitalizar el Desarrollo Sostenible.

El proyecto fomenta la colaboración entre distintas entidades en materia de datos. Se impulsa la creación de una comunidad donde se compartan mejores prácticas de modelado financiero favoreciendo responsabilidad y responsabilidad conjunta.

Con carácter general, se puede asegurar también que la tecnología aplicada al desarrollo sostenible mediante la implementación de diferentes campos de inteligencia artificial como machine learning son necesarias para la integración de estrategias de sostenibilidad en los modelos de negocio de una empresa.

Además, la experta en ética y desarrollo legislativo de la Inteligencia Artificial, Idoia Salazar, sostiene que la IA es la herramienta perfecta para abordar desafíos como el cambio climático, la pobreza o la salud.

ANEXO II: LIBRERÍAS UTILIZADAS EN PYTHON

	Package	Version				
			[27]	importlib_resources	6.5.2	
			[28]	ipykernel	6.29.5	
[1]	absl-py	2.2.0	[29]	ipython	9.0.2	
[2]	asttokens	3.0.0	[30]	ipython_pygments_lexers	1.1.1	
[3]	astunparse	1.6.3	[31]	ipywidgets	8.1.7	
[4]	beautifulsoup4	4.13.4	[32]	jedi	0.19.2	
[5]	certifi	2025.1.31	[33]	joblib	1.4.2	
[6]	cffi	1.17.1	[34]	jupyter_client	8.6.3	
[7]	charset-normalizer	3.4.1	[35]	jupyter_core	5.7.2	
[8]	colorama	0.4.6	[36]	jupyterlab_widgets	3.0.15	
[9]	comm	0.2.2	[37]	keras	3.9.0	
[10]	contourpy	1.3.1	[38]	kiwisolver	1.4.8	
[11]	curl_cffi	0.11.1	[39]	libclang	18.1.1	
[12]	cycler	0.12.1	[40]	lxml	5.4.0	
[13]	Cython	3.1.2	[41]	Markdown	3.7	
[14]	debugpy	1.8.13	[42]	markdown-it-py	3.0.0	
[15]	decorator	5.2.1	[43]	MarkupSafe	3.0.2	
[16]	et_xmlfile	2.0.0	[44]	matplotlib	3.10.1	
[17]	executing	2.2.0	[45]	matplotlib-inline	0.1.7	
[18]	flatbuffers	25.2.10	[46]	mdurl	0.1.2	
[19]	fonttools	4.56.0	[47]	missingno	0.5.2	
[20]	frozendict	2.4.6	[48]	ml_dtypes	0.5.1	
[21]	gast	0.6.0	[49]	multitasking	0.0.11	
[22]	google-pasta	0.2.0	[50]	namex	0.0.8	
[23]	grpcio	1.71.0	[51]	nest-asyncio	1.6.0	
[24]	h5py	3.13.0	[52]	numpy	2.1.3	
[25]	holidays	0.73	[53]	openpyxl	3.1.5	
[26]	idna	3.10	[54]	opt_einsum	3.4.0	

[55]	optree	0.14.1	[85]	stanio	0.5.1
[56]	packaging	24.2	[86]	statsmodels	0.14.4
[57]	pandas	2.2.3	[87]	tabulate	0.9.0
[58]	parso	0.8.4	[88]	tensorboard	2.19.0
[59]	patsy	1.0.1	[89]	tensorboard-data-server	0.7.2
[60]	peewee	3.18.1	[90]	tensorflow	2.19.0
[61]	pillow	11.1.0	[91]	tensorflow-io-gcs-filesystem	0.31.0
[62]	pip	25.0.1	[92]	termcolor	2.5.0
[63]	platformdirs	4.3.6	[93]	threadpoolctl	3.6.0
[64]	pmdarima	2.0.4	[94]	tornado	6.4.2
[65]	prompt_toolkit	3.0.50	[95]	tqdm	4.67.1
[66]	protobuf	5.29.4	[96]	traitlets	5.14.3
[67]	psutil	7.0.0	[97]	typing_extensions	4.12.2
[68]	pure_eval	0.2.3	[98]	tzdata	2025.1
[69]	pycparser	2.22	[99]	urllib3	2.3.0
[70]	Pygments	2.19.1	[100]	wcwidth	0.2.13
[71]	pyparsing	3.2.1	[101]	websockets	15.0.1
[72]	python-dateutil	2.9.0.post0	[102]	Werkzeug	3.1.3
[73]	pytz	2025.1	[103]	wheel	0.45.1
[74]	pywin32	309	[104]	widgetsnbextension	4.0.14
[75]	pyzmq	26.3.0	[105]	wrapt	1.17.2
[76]	requests	2.32.3	[106]	XlsxWriter	3.2.3
[77]	rich	13.9.4	[111]	yfinance	0.2.61
[78]	scikit-learn	1.6.1			
[79]	scipy	1.15.2			
[80]	seaborn	0.13.2			
[81]	setuptools	65.5.0			
[82]	six	1.17.0			
[83]	soupsieve	2.7			
[84]	stack-data	0.6.3			

ANEXO III: EMPRESAS ANALIZADAS

long_company_name	domicile_region	ticker	sector
ACI Worldwide, Inc.	North America	ACIW	Technology
ARM Holdings PLC ADR	Asia	ARM	Technology
AT & S Austria Technologie & SA	West Europe	ATS	Industrials
ATOSS Software SE	West Europe	AOF.F	Technology
Affirm Holdings, Inc. Class A	North America	AFRM	Technology
Alibaba Group Holding Limited	Asia	BABA	Consumer Cyclical
Appian Corporation Class A	North America	APPN	Technology
Asana, Inc. Class A	North America	ASAN	Technology
Atlassian Corp Class A	North America	TEAM	Technology
Atos SE	West Europe	ATOS	Healthcare
Axcelis Technologies, Inc.	North America	ACLS	Technology
BILL Holdings, Inc.	North America	BILL	Technology
BOKU, Inc.	West Europe	BOKU.L	Technology
Bentley Systems, Incorporated Class B	North America	BSY	Technology
Block, Inc. Class A	North America	XYZ	Technology
Box, Inc. Class A	North America	BOX	Technology
Braze, Inc. Class A	North America	BRZE	Technology
C3.ai, Inc. Class A	North America	AI	Technology
CY4Gate SpA	West Europe	CY4.MI	Technology
Cellebrite DI Ltd.	North America	CLBT	Technology
Cenit AG	West Europe	CSH.DE	Technology
Certara, Inc.	North America	CERT	Healthcare
Check Point Software Technologies Ltd.	North America	CHKP	Technology
Cloudflare Inc Class A	North America	NET	Technology
Cognex Corporation	North America	CGNX	Technology
Coinbase Global, Inc. Class A	North America	COIN	Financial Services
Computer Modelling Group Ltd.	North America	CMG	Consumer Cyclical
Confluent Inc Class A	North America	CFLT	Technology
Corpay, Inc.	North America	CPAY	Technology
Criteo SA Sponsored ADR Repr 1 Sh	West Europe	CRTO	Communications
CyberArk Software Ltd.	North America	CYBR	Technology
DLocal Limited Class A	North America	DLO	Technology
DXC Technology Co.	North America	DXC	Technology
Datadog Inc Class A	North America	DDOG	Technology
Dayforce, Inc.	North America	DAY	Technology

Definitive Healthcare Corp. Class A	North America	DH	Healthcare
Disco Corporation	Asia	6146.T	Technology
DocuSign, Inc.	North America	DOCU	Technology
Domo, Inc. Class B	North America	DOMO	Technology
DoorDash, Inc. Class A	North America	DASH	Consumer Cyclical
DoubleVerify Holdings, Inc.	North America	DV	Technology
Doximity, Inc. Class A	North America	DOCS	Healthcare
DraftKings, Inc. Class A	North America	DKNG	Consumer Cyclical
Dropbox, Inc. Class A	North America	DBX	Technology
Dynatrace, Inc.	North America	DT	Technology
Elastic NV	North America	ESTC	Technology
Eleco Plc	West Europe	ELCO	Technology
Eugene Technology Co., Ltd.	Asia	084370.KQ	
Euronet Worldwide, Inc.	North America	EEFT	Technology
Extreme Networks, Inc.	North America	EXTR	Technology
Fabasoft AG	West Europe	FAAS	Technology
Fastly, Inc. Class A	North America	FSLY	Technology
Fiserv, Inc.	North America	FI	Technology
Five9, Inc.	North America	FIVN	Technology
Freshworks, Inc. Class A	North America	FRSH	Technology
GMO Payment Gateway, Inc.	Asia	3769.T	Technology
Gitlab, Inc. Class A	North America	GTLB	Technology
Global-e Online Ltd.	North America	GLBE	Consumer Cyclical
GlobalFoundries Inc.	North America	GFS	Technology
GlobalWafers Co., Ltd.	Asia	6488.TWO	Technology
Grindr Inc.	North America	GRND	Technology
Gudeng Precision Industrial Co., Ltd.	Asia	3680.TWO	Technology
H&R Block, Inc.	North America	HRB	Consumer Cyclical
HubSpot, Inc.	North America	HUBS	Technology
Integral Ad Science Holding Corp	North America	IAS	Communications
Izertis SA	West Europe	IZER.MC	Technology
JD.com, Inc. Sponsored ADR Class A	Asia	JD	Consumer Cyclical
Klaviyo, Inc. Class A	North America	KVYO	Technology
Lasertec Corp.	Asia	6920.T	Technology
Lattice Semiconductor Corporation	North America	LSCC	Technology
Lectra SA	West Europe	LET.F	Technology
Lightspeed Commerce Inc	North America	LSPD	Technology
Luminar Technologies, Inc. Class A	North America	LAZR	Consumer Cyclical
Lyft, Inc. Class A	North America	LYFT	Technology
MKS Instruments, Inc.	North America	MKSI	Technology
Magnite, Inc.	North America	MGNI	Communications
Manhattan Associates, Inc.	North America	MANH	Technology

Marqeta, Inc. Class A	North America	MQ	Technology
Meituan Class B	Asia	3690.HK	Consumer Cyclical
MercadoLibre, Inc.	North America	MELI	Consumer Cyclical
MicroStrategy Incorporated Class A	North America	MSTR	Technology
MongoDB, Inc. Class A	North America	MDB	Technology
Murata Manufacturing Co., Ltd.	Asia	6981.T	Technology
NCC Group plc	West Europe	NZB.F	Technology
NICE Ltd. Sponsored ADR	Mid East	NICE	Technology
Netease Inc Sponsored ADR	Asia	NTES	Communications
Nexxen International Ltd.	West Europe	NEXN	Communications
Nutanix, Inc. Class A	North America	NTNX	Technology
Okta, Inc. Class A	North America	OKTA	Technology
OneSpan Inc.	North America	OSPN	Technology
Open Text Corporation	North America	OTEX	Technology
Outbrain, Inc.	North America	OB	Communications
PDD Holdings Inc. Sponsored ADR Class A	West Europe	PDD	Consumer Cyclical
PagerDuty, Inc.	North America	PD	Technology
Palantir Technologies Inc. Class A	North America	PLTR	Technology
PayPoint plc	West Europe	PAY.L	Technology
Paylocity Holding Corp.	North America	PCTY	Technology
Perion Network Ltd	North America	PERI	Communications
Pinterest, Inc. Class A	North America	PINS	Communications
Procore Technologies Inc	North America	PCOR	Technology
Progress Software Corporation	North America	PRGS	Technology
PubMatic, Inc. Class A	North America	PUBM	Technology
Qualys, Inc.	North America	QLYS	Technology
QuantumScape Corporation Class A	North America	QS	Consumer Cyclical
Rambus Inc.	North America	RMBS	Technology
Rapid7 Inc.	North America	RPD	Technology
Rivian Automotive, Inc. Class A	North America	RIVN	Consumer Cyclical
Robinhood Markets, Inc. Class A	North America	HOOD	Financial Services
Roblox Corp. Class A	North America	RBLX	Communications
SCREEN Holdings Co., Ltd	Asia	7735.T	Technology
SK hynix Inc.	Asia	000660.KS	Technology
SUMCO Corporation	Asia	3436.T	Technology
Sabre Corporation	North America	SABR	Technology
Samsung Electronics Co., Ltd.	Asia	005930.KS	Technology
Samsung SDI Co., Ltd	Asia	006400.KS	Industrials
Semiconductor Manufacturing Int	Asia	688981.SS	Technology
SentinelOne, Inc. Class A	North America	S	Technology
Serviceware SE	West Europe	SJJ.F	Technology
Shift4 Payments, Inc. Class A	North America	FOUR	Technology

Shin-Etsu Chemical Co Ltd	Asia	4063.T	Basic Materials
Shopify, Inc. Class A	North America	SHOP	Technology
Siltronic AG	West Europe	WAF.DE	Technology
Simulations Plus, Inc.	North America	SLP	Healthcare
Skillz Inc. Class A	North America	SKLZ	Communications
Snap, Inc. Class A	North America	SNAP	Communications
Snowflake, Inc. Class A	North America	SNOW	Technology
Sprinklr, Inc. Class A	North America	CXM	Technology
Sprout Social, Inc. Class A	North America	SPT	Technology
Taboola.com Ltd.	North America	TBLA	Communications
Taiwan Semiconductor Manufacturing Co.	Asia	TSM	Technology
TeamViewer SE	West Europe	TMVWY	Technology
Tenable Holdings, Inc.	North America	TENB	Technology
Tencent Holdings Ltd.	Asia	0700.HK	Communications
Toast, Inc. Class A	North America	TOST	Technology
Tokyo Electron Ltd.	Asia	8035.T	Technology
Tokyo Ohka Kogyo Co., Ltd.	Asia	4186.T	Technology
Trade Desk, Inc. Class A	North America	TTD	Communications
TransUnion	North America	TRU	Financial Services
Trend Micro Incorporated	Asia	4704.T	Technology
TripAdvisor, Inc.	North America	TRIP	Consumer Cyclical
Twilio, Inc. Class A	North America	TWLO	Technology
USU Software AG	West Europe	OSP2.HM	Technology
Ultra Clean Holdings, Inc.	North America	UCTT	Technology
Unisys Corporation	North America	UIS	Technology
United Microelectronics Corp.	Asia	UMCB.F	Technology
Unity Software, Inc.	North America	U	Technology
Varonis Systems, Inc.	North America	VRNS	Technology
Veeva Systems Inc Class A	North America	VEEV	Healthcare
Viant Technology, Inc. Class A	North America	DSP	Technology
WEX Inc.	North America	WEX	Technology
Wix.com Ltd.	North America	WIX	Technology
Wolfspeed Inc	North America	WOLF	Technology
Workday, Inc. Class A	North America	WDAY	Technology
Workiva Inc. Class A	North America	WK	Technology
Xero Limited	Pacific	XROLF	Technology
Zomato Ltd.	Asia	ETERNAL.NS	Consumer Cyclical
Zoom Communications, Inc. Class A	North America	ZM	Technology
ZoomInfo Technologies Inc	North America	GTM	Technology
Zscaler, Inc.	North America	ZS	Technology
ams-OSRAM AG	West Europe	AMS	Healthcare
eGain Corporation	North America	EGAN	Technology

flatexDEGIRO AG	West Europe	FTK.DE	Financial Services
iomart Group plc	West Europe	IOM.L	Technology
monday.com Ltd.	North America	MNDY	Technology