



# Optimización en el proceso de búsqueda y análisis fundamental de compañías cotizadas en mercados de valores de EEUU

**Alex Rodriguez Just**

Grado en Ciencia de Datos Aplicada

Junio de 2023

## INDICE

1. Introducción .....	2
1.1. Contexto .....	2
1.2. Objetivos .....	3
1.3. Descripción del proyecto .....	3
1.4. Planificación .....	4
2. Viabilidad .....	5
2.1. Viabilidad técnica .....	5
2.2. Viabilidad económica .....	6
2.3. Viabilidad legal .....	7
3. Arquitectura del sistema .....	8
4. Analítica cualitativa .....	9
4.1. Datos, extracción e ingeniería prompt .....	9
5. Analítica cuantitativa .....	12
5.1. Datos, extracción y plantilla de valoración .....	12
6. Modelos Machine learning y Deep Learning .....	13
7. Aspectos éticos / sociales .....	14
8. Análisis y gestión de riesgo .....	14
9. Documentos adjuntos .....	15
9.1 Aviso de Exención de Responsabilidad (simplificado) .....	15
9.2 Ejemplo plantilla "Tesis de inversión por fundamentales" .....	16

# 1. Introducción

En este capítulo, se define el contexto en el cual se enmarca el proyecto, los objetivos que se desean alcanzar con su realización, la idea principal del proyecto y la planificación del trabajo que se ha establecido para su consecución.

## 1.1. Contexto

En el panorama económico de los Estados Unidos, más de 9500 empresas cotizan en los mercados de valores. Esta cifra es dinámica y puede fluctuar debido a diversos factores como las nuevas Ofertas Públicas de Venta (IPOs), fusiones, adquisiciones o incluso quiebras. La inversión centrada en valorar acciones cotizadas de una compañía exige un enfoque riguroso y analítico de los fundamentos de la empresa. En este escenario, es esencial distinguir entre el precio que paga el mercado por acción y su valor real intrínseco. Identificar este valor intrínseco en los fundamentos de una compañía es el objetivo primordial para aquellos que buscan un valor significativo a un precio óptimo.

Para analizar estos fundamentos, recurrimos a los datos públicos proporcionados a las autoridades reguladoras. Las compañías deben presentar formularios como el 10-K con parámetros cuantitativos cruciales como los datos de la cuenta de resultados, el balance de situación, el estado de flujo de caja, entre otros. Además deben justificar variables cualitativas que incluyen, entre otras, el posicionamiento competitivo de la empresa, el cumplimiento de las proyecciones, factores de riesgo, etc.

Un inversor que se concentra en el valor de una acción debe fundamentar su tesis de inversión en aspectos tanto cuantitativos como cualitativos, creando un marco que sea flexible y adaptable a las variaciones del tiempo y el mercado. Este enfoque sirve de brújula para la toma de decisiones, proporcionando un camino claro y directo en la compleja y dinámica arena de la inversión. El desafío es que por cada tesis de inversión, el coste de tiempo en investigación es elevado y si al multiplicar por el número de compañías cotizadas le sumamos sus dinámicas, nos encontramos con una tarea que exige recursos considerables.

En este sentido, nuestra labor se centra en proporcionar una tesis de inversión definida y de alto valor, ahondando en las capas de los informes financieros y de las dinámicas del mercado para desentrañar oportunidades de inversión potencialmente rentables. Nuestro objetivo primordial será reducir el tiempo y los recursos requeridos para cada tesis de inversión mejorando la calidad de la misma, permitiendo así a los inversores centrarse en la toma de decisiones estratégicas en lugar de en el análisis detallado de cada empresa.

La finalidad de estas tesis automatizadas será producir un primer borrador o proporcionar sugerencias y puntos de partida que luego pueden ser refinados por expertos humanos. El objetivo no es reemplazar completamente al humano, sino más bien liberarlo de tareas tediosas y repetitivas para que pueda concentrarse en la interpretación de los resultados y la toma de decisiones estratégicas.

## 1.2. Objetivos

El propósito principal de este proyecto radica en el desarrollo de un sistema automatizado, diseñado para producir una tesis de inversión de alto valor para cada compañía cotizada en EEUU que resulte de interés para un usuario. Así pues, a lo largo del desarrollo de este proyecto es necesario conseguir unos objetivos más específicos:

- Obtención histórica de informes financieros por cada compañía cotizada.
- Actualización trimestral de la presentación de informes financieros por compañía.
- Extracción de datos de los informes financieros.
- Desarrollo de una base de datos NoSQL robusta y eficiente.
- Envío y almacenaje de los datos anteriores a la base de datos del servidor.
- Transformación de datos cuantitativos en indicadores financieros clave.
- Transformación de datos cualitativos para alimentar la IA generativa.
- Ingeniería prompt para generación de texto en análisis financiero.
- Selección y análisis exploratorio de características relevantes (EDA).
- Desarrollo y aplicación de modelos de machine learning avanzados.
- Presentación de resultados en forma de tesis y formato HTML.

Teniendo presente esto, los requisitos necesarios para cumplir con los objetivos de este proyecto son:

- La creación de un archivo HTML individual para cada empresa investigada. Es decir, cada vez que un usuario introduzca un nombre de empresa o ticker (la denominación bursátil de la compañía), el sistema generará un documento HTML correspondiente, presentando una tesis de inversión detallada para dicha empresa.

## 1.3. Descripción del proyecto

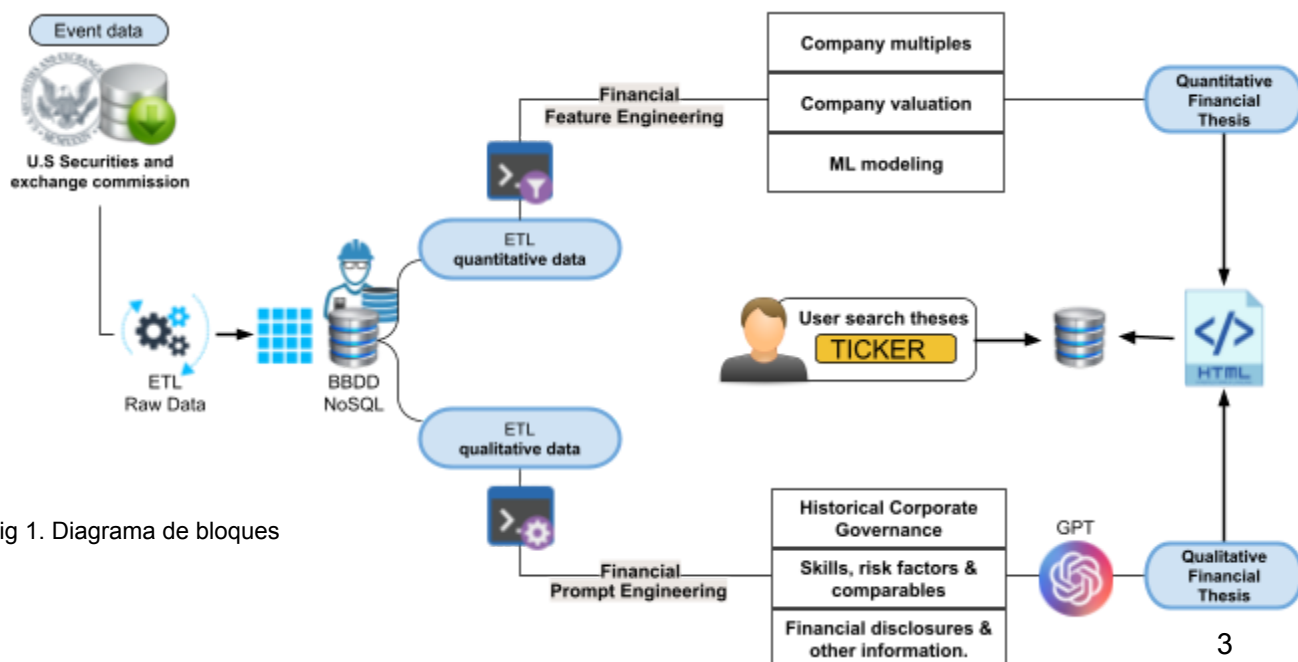


Fig 1. Diagrama de bloques

El proyecto es una iniciativa enfocada en simplificar y mejorar el proceso de inversión en las empresas cotizadas en los Estados Unidos. El objetivo principal del proyecto es desarrollar un sistema automatizado capaz de generar una tesis de inversión de alto valor para cualquier empresa cotizada que resulte de interés para un usuario.

Para alcanzar este objetivo, el proyecto ha definido diversas metas específicas, entre las que destacan la obtención histórica y actualización trimestral de informes financieros, la extracción de datos de estos informes, el desarrollo de una base de datos NoSQL eficiente, la transformación de datos cuantitativos en indicadores financieros clave y de datos cualitativos para alimentar un sistema de inteligencia artificial. Además, se utilizarán técnicas avanzadas de machine learning para la selección y análisis de características relevantes.

Como resultado, cada vez que un usuario introduzca el nombre o la denominación bursátil de una empresa, el sistema generará un archivo HTML individual con una tesis de inversión detallada para dicha empresa. En conclusión, este proyecto se presenta como una herramienta en el ámbito de la inversión financiera.

## 1.4. Planificación

CALENDARIO SEP 2024 – FEB 2025

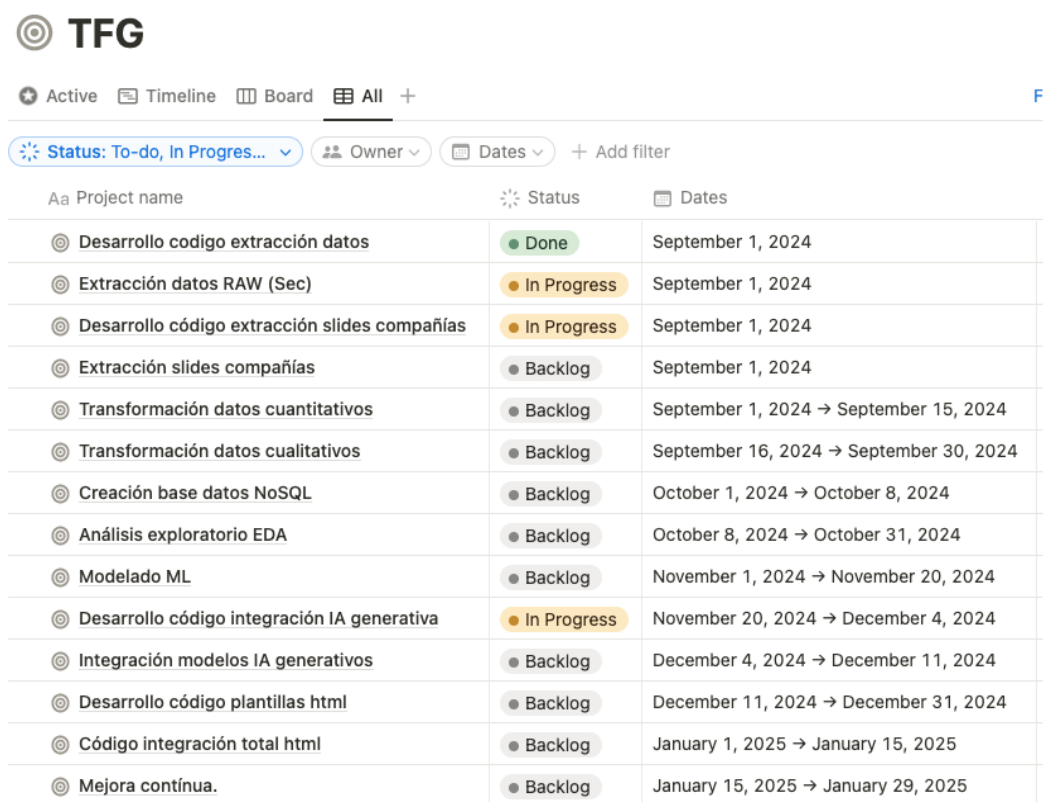


Figura 2. Planificació del projecte amb diagrama de Gantt

## 2. Viabilidad

En este capítulo se abordan las cuestiones relacionadas con la viabilidad técnica, económica y legal del proyecto a desarrollar.

### 2.1. Viabilidad técnica

Extracción - Todos los datos financieros cuantitativos y cualitativos serán recolectados de forma centralizada desde la Comisión de Bolsa y Valores de Estados Unidos (SEC) mediante su API Api SEC [1]. Esta recolección abarca informes de los últimos 10 años. Se estima que el volumen total de la descarga será inferior a 20 TB. También se considera la regularidad de actualización de estos datos automatizando este proceso.

Transformación - implicaría limpiar los datos y convertirlos en un formato que pueda ser fácilmente almacenado en la base de datos.

Carga - se realizará a una base de datos local; una base de datos orientada a documentos como MongoDB podría ser una buena opción debido a la naturaleza heterogénea de los datos. El desarrollo de una base de datos NoSQL robusta y eficiente es técnicamente factible. Los documentos por compañía podrían ser : colección de compañías, colección de informes financieros, colección de textos financieros.

Financial feature engineering - se programarán scripts para la automatización de la creación de nuevas características financieras a partir de fórmulas matemáticas para calcular las evaluaciones y estimaciones de valor de la compañía.

Financial Prompt engineering - se desarrollarán scripts estructurando prompts a partir de la colección de textos financieros, lo que facilitará la generación automatizada de contenido financiero de alto valor con GPT. [2]

EDA (Exploratory Data Analysis) y aplicación de modelos de Machine Learning - Automatizar el proceso de EDA y la aplicación de modelos ML para un gran número de compañías es un desafío considerable, pero factible con la tecnología actual. Se deben tomar decisiones estratégicas con respecto a cómo manejar ciertas situaciones particulares. Por ejemplo la presencia de outliers, la visión inicial de las características de los datos que permitan identificar patrones, tendencias y posibles problemas como correlaciones muy fuertes entre las características, etc. En última instancia, lo que estas decisiones aportan es eficiencia. Existen librerías como SHAP [3] y LIME [4] que pueden ayudar a entender la contribución de cada característica a las predicciones de un modelo.

Generación automática de informes - se desarrollarán scripts estructurando informes HTML con los resultados del análisis y las predicciones de los modelos para cada compañía en composición con los análisis cualitativos históricos y de tendencias generadas con GPT.

## 2.2. Viabilidad económica

En esta sección se verifica la viabilidad económica del sistema mediante el cálculo de los costos para la realización del prototipo y los costos para su industrialización (referido como al proceso de escalar el prototipo para poder manejar un mayor volumen de datos y usuarios) , con el fin de determinar en qué momento se amortizará la inversión inicial y se obtendrán beneficios.

En primer lugar, para la realización de un **prototipo** que abarca la generación de tesis de inversión para una muestra reducida de la población de compañías, los costos son menores. Para calcular la **industrialización** de una primera versión donde se recoge la población de 9500 compañías, se debería tener en cuenta los costos asociados al hardware y software.

Componente	Prototipo	Industrialización
Computadora de alto rendimiento	0	10239 €
Memoria externa	0	350 €
<b>TOTAL</b>	<b>0</b>	<b>10589 €</b>

Tabla 1. Coste hardware

En segundo lugar, el software utilizado en general es Open Source lo que reduce el costo simplemente al uso del software de OpenIA, el cual se hará una aproximación del costo. [5]  
Los precios del servicio GPT son en función de tokens de entrada y tokens de salida. Los precios actuales por tokens y modelo son:

Model	Input	Output
8K context	\$0.03 / 1K tokens	\$0.06 / 1K tokens
32K context	\$0.06 / 1K tokens	\$0.12 / 1K tokens

Para nuestro proyecto continuamos calculando para coste de prototipo y conste de industrialización. Donde para cada una de las tesis generadas se calcula [3 páginas de entrada, aproximadamente 1500 tokens ] + [ 4 páginas de salida, aproximadamente 2000 tokens ]. Aproximadamente se calcula que 1 página contiene unos 500 tokens. Si usamos un modelo 32K context obtenemos una primera aproximación del costo por tesis:

Modelo 32k	tokens	Fórmula	Coste 1 tesis	9500 tesis
entrada de tokens	1500	1500 tokens / 1000 * \$ 0.06	0.09 €	855 €
salida de tokens	2000	2000 tokens / 1000 * \$ 0.12	0.24 €	2280 €
<b>TOTAL</b>	<b>3500</b>		<b>0.33 €</b>	<b>3135 €</b>

Tabla 2. Coste software

Para el primer año, podemos sumar que el coste de industrializar el proyecto sería, sin tener en cuenta la mano de obra, a falta de oficializar precios y ajustar todo, sumaría un coste de 13315 €. Los cuales 3135 € son recurrentes y 10588 € una sólo inversión inicial. Esto supondría la

necesidad de 260 suscriptores anuales con una cuota de 50 €/año para recuperar la inversión inicial.

Y para finalizar, respecto a la industrialización, se tiene que valorar la posibilidad de trabajar en la nube. Este enfoque puede reducir tanto los costos iniciales como los costos operativos, especialmente si el volumen de datos o el número de usuarios varía con el tiempo. También proporciona una mayor flexibilidad y podría facilitar la expansión a nuevos mercados en el futuro. En este caso, los usuarios podrían pagar por lo que realmente utilizan, o bien podríamos establecer precios de suscripción basados en diferentes niveles de servicio.

### **2.3. Viabilidad legal**

Este proyecto se fundamenta en el análisis de datos financieros de acceso público suministrados por la SEC, lo que nos exime de enfrentar problemas éticos vinculados a la privacidad de datos. No obstante, es imprescindible asegurar la precisión e integridad de los datos, esquivando cualquier sesgo o manipulación. Es crucial garantizar un uso ético y equitativo de la inteligencia artificial y del aprendizaje automático para prevenir la emergencia de sesgos involuntarios en las predicciones y resultados.

En el contexto de un sitio web que recopila y proyecta precios de cotizaciones de acciones de diversas empresas, resulta esencial delimitar de manera explícita las limitaciones y el alcance de los servicios que dicho sitio web ofrece. Esto implica clarificar el papel y las responsabilidades del dueño del sitio web y su equipo administrativo. En línea con esto, proponemos la implementación de un "Aviso de Exención de Responsabilidad" (documento adjunto), donde se especifica que el propósito principal es proporcionar educación financiera y análisis no personalizados, sin ofrecer servicios de inversión regulados.



### 3. Arquitectura del sistema

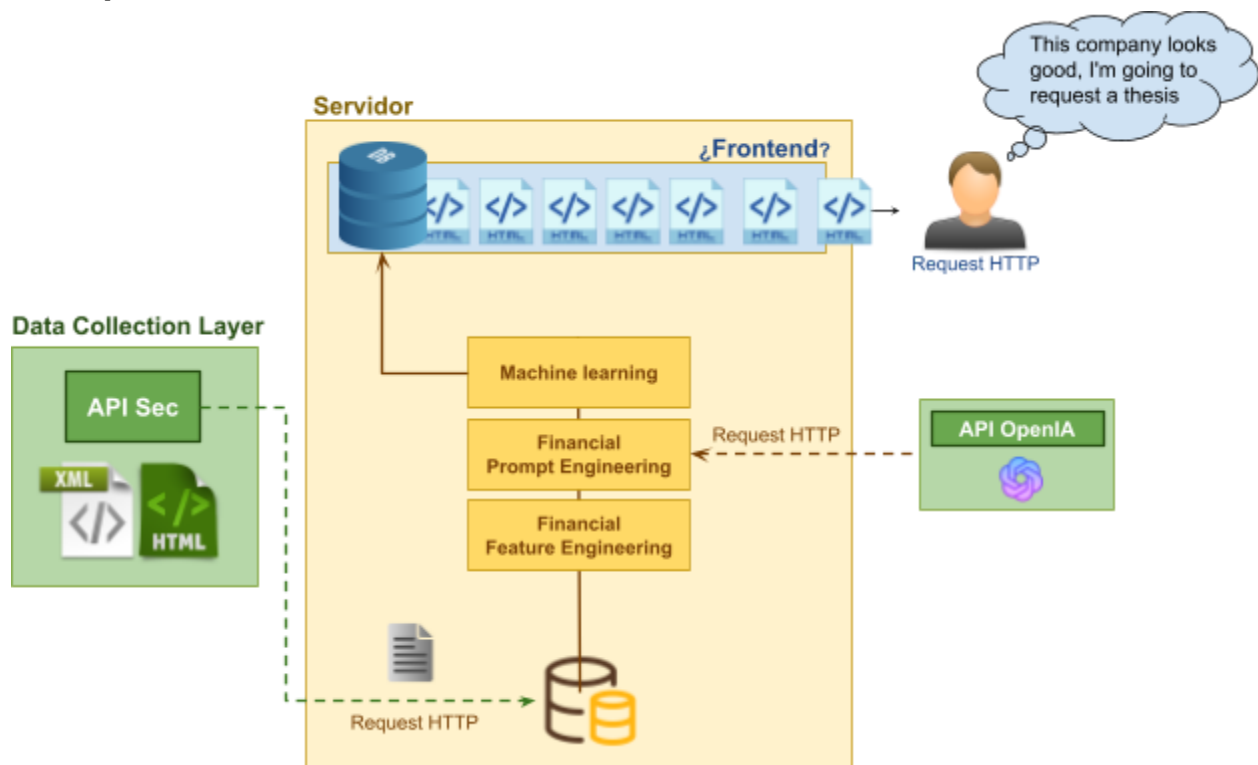


Figura 3. Arquitectura del sistema

#### Capa de Recopilación de Datos

- Bibliotecas de extracción : lxml [6], requests, BeautifulSoup
- Extracción datos **cuantitativos** estandarizados : taxonomía en esquema XML (.xsd)[7]
- Extracción datos **cualitativos** estandarizados : taxonomía en esquema HTML (.html)
- Formato datos limpios y ya transformados : BSON, una variante binaria de JSON.

#### Capa de Procesamiento de Datos

La carga se realizará a una base de datos orientada a documentos como MongoDB debido a la naturaleza heterogénea de los datos.

- Colección Compañías: Cada documento es una empresa única, con datos y metadatos.
- Colección Informes Financieros: Cada documento sería un informe financiero único. Con campos para la fecha, el tipo (10-K, 10-Q, etc.), ticker a la que pertenece el informe y campos para todos los datos financieros cuantitativos del informe.
- Colección Textos Financieros: representaría un texto financiero único (transcripción de la llamada de ganancias, presentación a inversores, etc.) y el contenido del texto.

Ingeniería de Características Financieras.

Ingeniería de Prompts para textos financieros : conexión vía http con OpenIA.

Aplicación de Modelos de Machine Learning.

#### Capa de Presentación de Resultados

Generación de Informes HTML ----> Tesis de Inversión

## 4. Analítica cualitativa

### 4.1. Datos, extracción e ingeniería prompt

#### Uso de EDGAR para investigar compañías

La base de datos EDGAR brinda acceso público gratuito a la información corporativa, lo que permite investigar la información financiera y las operaciones de una empresa pública mediante la revisión de las presentaciones que la empresa realiza ante la SEC.

#### Tipos de formularios

EDGAR presenta los resultados de la búsqueda en orden cronológico e identifica las presentaciones por tipos de formularios, que se muestran en la primera columna de los resultados de la búsqueda. Por ejemplo, el tipo de formulario para el informe anual en el Formulario 10-K presentado por empresas públicas nacionales es "10-K".

**Información Financiera y Resultados de Operaciones** - Las empresas públicas nacionales de EEUU presentan informes anuales y trimestrales , así como informes de actualidad cuando ocurren determinados hechos que requieren su pronta divulgación. Los informes anuales y trimestrales incluyen los estados financieros del período correspondiente.

Tipo de formulario	Descripción
--------------------	-------------

<b>10-K</b>	<b>Informe anual</b> : proporciona estados financieros anuales auditados, una discusión de los factores de riesgo importantes para la empresa y su negocio, y una discusión y análisis de la administración de los resultados de las operaciones de la empresa para el año fiscal anterior.
<b>10-Q</b>	<b>Informe trimestral</b> : proporciona estados financieros trimestrales no auditados, actualizaciones sobre los riesgos importantes a los que se enfrenta la empresa y la discusión y el análisis de la administración de los resultados de las operaciones de la empresa para el trimestre fiscal anterior.
<b>8-K</b>	<b>Informe actual</b> : divulga información o eventos importantes que la empresa elige o debe hacer pública antes de su próximo informe trimestral o anual programado.

#### Ejemplo modelo 10-K : Obteniendo información

El modelo documento ['10-K'] contiene de forma estructurada la información financiera deseada para extraer tanto datos cuantitativos como cualitativos. A continuación se muestra el índice:

Alphabet Inc. Form 10-K For the Fiscal Year Ended December 31, 2019		
TABLE OF CONTENTS		
	<a href="#">Note About Forward-Looking Statements</a>	Page <a href="#">3</a>
<b>PART I</b>		
Item 1.	<a href="#">Business</a>	<a href="#">5</a>
Item 1A.	<a href="#">Risk Factors</a>	<a href="#">9</a>
Item 1B.	<a href="#">Unresolved Staff Comments</a>	<a href="#">22</a>
Item 2.	<a href="#">Properties</a>	<a href="#">22</a>
Item 3.	<a href="#">Legal Proceedings</a>	<a href="#">22</a>
Item 4.	<a href="#">Mine Safety Disclosures</a>	<a href="#">22</a>
<b>PART II</b>		
Item 5.	<a href="#">Market for Registrant's Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities</a>	<a href="#">23</a>
Item 6.	<a href="#">Selected Financial Data</a>	<a href="#">26</a>
Item 7.	<a href="#">Management's Discussion and Analysis of Financial Condition and Results of Operations</a>	<a href="#">27</a>
Item 7A.	<a href="#">Quantitative and Qualitative Disclosures About Market Risk</a>	<a href="#">43</a>
Item 8.	<a href="#">Financial Statements and Supplementary Data</a>	<a href="#">46</a>
Item 9.	<a href="#">Changes in and Disagreements With Accountants on Accounting and Financial Disclosure</a>	<a href="#">90</a>
Item 9A.	<a href="#">Controls and Procedures</a>	<a href="#">90</a>
Item 9B.	<a href="#">Other Information</a>	<a href="#">90</a>
<b>PART III</b>		
Item 10.	<a href="#">Directors, Executive Officers and Corporate Governance</a>	<a href="#">91</a>
Item 11.	<a href="#">Executive Compensation</a>	<a href="#">91</a>
Item 12.	<a href="#">Security Ownership of Certain Beneficial Owners and Management and Related Stockholder Matters</a>	<a href="#">91</a>
Item 13.	<a href="#">Certain Relationships and Related Transactions, and Director Independence</a>	<a href="#">91</a>
Item 14.	<a href="#">Principal Accountant Fees and Services</a>	<a href="#">91</a>

Fig 4. Índice contenido por bloque formulario 10-K

Esta estructura es genérica para todos los documentos [10-K] es decir que la extracción se realiza por secciones. Cada sección va destinada a presentar la información establecida para ese apartado. Por ejemplo en la **PART I, Item 1**, una vez hemos hecho la extracción desde la API de la Sec y habiendo realizado transformaciones y limpiezas de datos, podemos aislar el texto cualitativo que nos interesa para la ingeniería de prompts destinada a GPT.

```
# Define el patrón regex
pattern = r'(?:\n\d+\n\nTable of Contents\n\nAlphabet Inc.\n\n)'
```

```
# Elimina las líneas no deseadas usando re.sub()
item_1 = re.sub(pattern, '', item_1)
```

```
print(item_1[:5000])
```

>ITEM 1.

BUSINESS

Overview

As our founders Larry and Sergey wrote in the original founders' letter, "Google is not a conventional company. We do not intend to become one." That unconventional spirit has been a driving force throughout our history – inspiring us to do things like tackling deep computer science problems, such as our investments in artificial intelligence (AI) and quantum computing.

Alphabet is a collection of businesses – the largest of which is Google. We report all non-Google businesses collectively as Other Bets. Our Other Bets include earlier stage technologies that are further afield from our core Google business. We take a long term view and manage the portfolio of Other Bets with the discipline and rigor needed to deliver long-term returns. Each of our businesses are designed to prosper through strong leaders and independence.

Access and technology for everyone

Fig 5. Extracto de código para descargar data RAW de un informe 10-K

## Ejemplo Ingeniería Prompt para GPT

Se muestra un simple ejemplo de cómo interactuar con GPT a partir de nuestros datos de texto de **PART I, Item 1**. En este caso las preguntas están diseñadas para obtener una visión más profunda de la estrategia de negocio de Alphabet, sus prioridades de inversión y sus perspectivas a largo plazo.

```
import openai
import os

#from dotenv import load_dotenv, find_dotenv
#_ = load_dotenv(find_dotenv())
openai.api_key = 'OPENAI_API_KEY'

def get_completion(prompt, model="gpt-3.5-turbo"):
    messages = [{"role": "user",
                  "content": prompt}]
    response = openai.ChatCompletion.create(
        model=model,
        messages=messages,
        temperature=0, # this is the degree of randomness of the model's output
    )
    return response.choices[0].message["content"]

text = f"""
BUSINESS 2023
As our founders Larry and Sergey wrote in the original founders' letter, "Google is not a conventional \
company. We do not intend to become one." That unconventional spirit has been a driving force throughout \
our history – inspiring us to do things like tackling deep computer science problems, such as our \
investments in artificial intelligence (AI) and quantum computing.

Alphabet is a collection of businesses – the largest of which is Google. We report all non-Google \
businesses collectively as Other Bets. Our Other Bets include earlier stage technologies that are \
further afield from our core Google business. We take a long term view and manage the portfolio of \
Other Bets with the discipline and rigor needed to deliver long-term returns. Each of our businesses \
are designed to prosper through strong leaders and independence...
"""

prompt = f"""
You will be provided with the {company}'s 2023 and 2022 business model, How does {company} manage the \
risk associated with long-term investments in early-stage technologies? What percentage of {company}'s \
total revenue comes from {principal business} compared to the "Other Bets"?,
re-write those instructions in the following format:

Output JSON

{company} : "question1" : "answer1",
            "question2" : "answer2",

\\\\"{text}\\\"\\\"\\\"
"""

response = get_completion(prompt)
print("Completion for Text 2:")
print(response)
```

Fig 6. Extracto de código para interactuar con API OpenIA

La calidad y precisión de las respuestas son derivadas de unas buenas prácticas, preguntas clave y concisas. Las respuestas van destinadas exclusivamente a lugares clave de nuestra plantilla. Todo el universo de cada compañía de los últimos 10 años estará en la BBDD bien compartimentado por año / trimestre y secciones. Una buena ingeniería de características nos facilitará la precisión y calidad de las preguntas. Temas clave entre otros:

- resumen ejecutivo,
- descripción del negocio,

- segmentación del negocio por áreas de ventas,
- tendencia del negocio y del sector,
- características positivas y negativas del negocio,
- cómo gana el dinero la compañía,
- El manejo y la participación de los insiders, con su descripción completa
- banderas rojas de la compañía a tener en cuenta,

## **5. Analítica cuantitativa**

### **5.1. Datos, extracción y plantilla de valoración.**

Procederemos a extraer los datos brutos (RAW) de los tres estados financieros de cada compañía, específicamente de:

- Cuenta de Resultados (Income Statement): Se extraerán 30 elementos, entre los cuales se incluyen Ingresos, Beneficio Bruto, etc.
- Balance (Balance Sheet): Se obtendrán 46 elementos, entre ellos Efectivo y Equivalentes, Inversiones a Corto Plazo, etc.
- Flujo de Caja (Cash Flow): Se adquirirán 34 elementos, como Ingresos Netos, Depreciación y Amortización, entre otros.

A partir de estos datos, implementaremos técnicas de ingeniería de características y generación de datos financieros claves que permitirán calcular las evaluaciones y estimaciones de valor de la compañía.

Esto incluirá la determinación de los múltiplos a los que cotiza la compañía, tales como:

- Forward Multiples: 7 datos (NTM Total Enterprise Value / Revenues, ...)
- Trailing Multiples: 11 datos (LTM Total Enterprise Value / Revenues, ...)
- Price Factors: 3 datos (Price, Total Enterprise Value (MM), Market Cap (MM), ...)
- Forward Factors: 5 datos (NTM Revenues, NTM EBITDA, ...)
- Trailing Factors: 10 datos (LTM Revenues, LTM Gross Profit, ... )


Estimaciones:

- Revenue, EBITDA, %EBITDA Margins, EBIT, % EBIT Margins, ...

#### **Plantilla de valoración**

Esta plantilla se compone con los datos anteriores y tiene la función de darnos un precio justo de la acción de forma compuesta para los siguientes 3 y 5 años. Para efectuar un cálculo compuesto de la proyección del valor de una acción, es indispensable contar con un mínimo de siete años de datos históricos para cada empresa.

Podemos emplear modelos de regresión para realizar una proyección temporal de valores en la propia plantilla.

												
	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027
Valuation ( millions ,except EP\$ )												
Market cap						660	660	660	660	660	660	660
Net DEBT (-) si es caja neta	0	0	6	21	19	24	20	10	2	-10	-24	-24
Deuda neta /EBITDA	#DIV/0!	#DIV/0!	0,32	0,83	0,57	0,67	0,41	0,15	0,02	-0,10	-0,20	-0,17
Deuda neta/equity	#DIV/0!	#DIV/0!	0,19	0,62	0,39	0,40	0,24	0,10	0,02	-0,07	-0,14	-0,11
Enterprise Value ( EV )						684	680	670	662	650	636	636
EBITDA	0	0	18	25	33	36	49	69	82	99	119	142
EBIT			15	22	26	28	40	58	69	83	100	120
Net income	0	0	11	15	17	20	27	42	50	61	73	87
FCF	0	0	6	2	11	16	-4	39	50	62	77	96
	3 YEARS	5 YEARS										
CAGR by PER	32%	27%	Price target PER+caja \$ 3					\$ 5	\$ 6	\$ 8	\$ 9	\$ 11
CAGR by P/FCF	33%	30%	Price target P/FCF +caja -\$ 0					\$ 5	\$ 6	\$ 8	\$ 10	\$ 12
CAGR by EV/EBITDA	34%	28%	Price target EV/EBITDA \$ 4					\$ 5	\$ 7	\$ 8	\$ 10	\$ 12
CAGR by EV/EBIT	34%	28%	Price target EV/EBIT \$ 4					\$ 5	\$ 7	\$ 8	\$ 10	\$ 12

Cotización actual 3,3

Multiple PER 25  
Multiple P/FCF 25  
Multiple EV/EBITDA 16  
Multiple EV/EBIT 19

Figura 7. Página última de la plantilla ejemplo en la valoración compuesta de acción por sus fundamentales

## 6. Modelos Machine learning y Deep Learning

El primer reto técnico es confiar en modelos que tengan sentido económicamente y la única suposición que hacemos en este trabajo es que los rendimientos futuros dependen de las características de la empresa. La relación entre estas características y el rendimiento se desconoce en gran medida y probablemente varíe con el tiempo. Es por eso que ML puede ser útil: para detectar algunos patrones ocultos más allá de las anomalías documentadas de precios de activos en el mercado vs el valor de los activos que le hemos dado.

Estas técnicas están destinadas a darnos información relativa a “Value investing”, “Factor investing”, “Dividend investing”. Alguno de los modelos a estudiar podrían ser:

### Máquinas de vectores de soporte (SVM) o las redes neuronales

pueden ser capaces de capturar relaciones más complejas en los datos. Estos métodos pueden ser especialmente útiles si tienes una gran cantidad de datos y muchas variables diferentes para considerar.

### Modelos de simulación Monte Carlo

Este tipo de modelo utiliza la simulación para generar una amplia gama de posibles resultados futuros basados en los datos históricos y la variabilidad de esos datos.

### Deep learning en valoración de activos

Utilizamos redes neuronales profundas para estimar un modelo de fijación de precios de activos para rendimientos de acciones individuales que aprovecha la gran cantidad de información condicionante, mantiene una forma totalmente flexible y tiene en cuenta la variación temporal. Las innovaciones clave consisten en utilizar la condición fundamental de no arbitraje como función de criterio para construir los activos de prueba más informativos con un

enfoque contradictorio y para extraer los estados de la economía de muchas series temporales macroeconómicas.

### **Gradient Boosting Machine o Máquina de Potenciación del Gradiente**

- para la selección de empresas pares con fundamentos similares
- para predecir el valor fundamental del precio de una acción.

“¿Nuestro enfoque de aprendizaje automático construye modelos que son congruentes con la teoría financiera subyacente? Esta es una pregunta importante porque los GBM adoptan un enfoque basado en datos. Las variables de entrada importantes elegidas por los GBM se alinean ampliamente con los determinantes teóricos del valor fundamental, lo que sugiere que los GBM, en cierto sentido, "aprenden" la teoría de valoración de los datos.”

## **7. Aspectos éticos / sociales**

La bolsa de valores desempeña un papel crucial en la economía y, por ende, en la sociedad, al valorar las empresas de manera eficaz. Este sistema permite a dichas empresas recaudar capital adicional a través de la emisión de acciones, brindando a los inversores la oportunidad de contribuir al crecimiento económico mediante la compra de estas acciones.

La ausencia de conocimiento popular en finanzas lleva a muchos (y yo me incluí alguna vez) a comparar erróneamente el mercado de valores con un casino, asumiéndolo como pura especulación. Sin embargo, contrariamente a esa percepción, subyace una ciencia cuantificable y comprensible que rige estos mercados; mercados que además cumplen un papel fundamental para la sociedad.

Nuestro propósito en la creación y puesta en marcha de este proyecto es simplificar el complejo proceso inherente en entender cuándo una compañía incrementa o disminuye su valor. Nuestra intención es guiar al usuario de manera pedagógica a través de las circunstancias cuantitativas y cualitativas de la compañía en tiempo real, convirtiendo la tarea en una experiencia más accesible y educativa..

## **8. Análisis y gestión de riesgo**

Se busca identificar, evaluar y priorizar posibles obstáculos o problemas que podrían surgir durante el desarrollo y la ejecución del proyecto, además mitigar los efectos adversos que estos riesgos podrían tener en los objetivos del proyecto.

**Riesgos técnicos: ¿Qué problemas podrían surgir con respecto a la recolección de datos, el procesamiento de datos, la implementación de algoritmos, etc.?**

Los algoritmos de IA son open source vía API. Este proyecto estará sujeto a posibles cambios que OpenIA pudieran aplicar o restringir. Gestión de riesgo : Para mitigar estos riesgos, se establecerán protocolos de seguimiento regular de las actualizaciones y cambios realizados en las APIs utilizadas. Además, se buscarán alternativas viables para reducir la dependencia de

un solo proveedor de API.

**Riesgos financieros: ¿Qué sucedería si el proyecto se va por encima del presupuesto o si hay un cambio en el financiamiento?**

Se ha calculado que el peso de los datos pueden estar alrededor de 20 TB, esto supone gestionar el almacenamiento, ya sea propio o en la nube, que en cualquier caso tendrá un gasto. Gestión de riesgo : Se explorarán diferentes opciones de almacenamiento y se seleccionará la más costo-efectiva sin comprometer la calidad y la eficacia de los procesos del proyecto.

**Riesgos de mercado: ¿Cómo podría afectar una variación en las condiciones del mercado al proyecto, especialmente en términos tecnológicos y utilidad de tu sistema?**

Existen plataformas web de análisis financiero con poder económico y tecnológico muy conocidas como Bloomberg, Factset, etc. Estas plataformas están en la vanguardia y pueden dejar fuera de mercado cualquier propuesta competitiva. Gestión de riesgo : Para combatir estos riesgos, se mantendrá una vigilancia constante de las tendencias del mercado y las innovaciones tecnológicas. Nuestra propuesta se centrará en ofrecer un enfoque diferencial, personalizado y educativo.

## **9. Documentos adjuntos**

### **9.1 Aviso de Exención de Responsabilidad (simplificado):**

"Este sitio web no ofrece servicios de inversión ni servicios auxiliares de inversión que podrían requerir autorización o registro por entidades reguladoras financieras en ninguna jurisdicción internacional. Por lo tanto, nuestras actividades operan fuera de la supervisión de dichas entidades. Instamos a los usuarios de este sitio web a que busquen asesoramiento legal y financiero a la hora de invertir en acciones.

Nuestro objetivo principal es promover la educación financiera y la transparencia. Los análisis y opiniones generados por nuestros algoritmos no son asesoramiento financiero personalizado ni garantizan la compra o venta de activos financieros. La información proporcionada se extrae de fuentes de acceso público. Intentamos asegurar su precisión, pero los mercados financieros volátiles pueden afectar su exactitud. Los precios objetivos y valoraciones presentados son estimaciones algorítmicas y están sujetos a volatilidad y riesgos de inversión. Es esencial recordar que los precios de las acciones dependen de factores externos como la economía general, las tasas de interés y la inflación. El rendimiento de una empresa no siempre se refleja en su cotización a corto plazo.

Finalmente, estamos comprometidos con la ética, la legalidad y la responsabilidad social. Creemos en la educación financiera y el poder de la tecnología para democratizar el acceso a la información financiera, pero siempre enfatizamos la importancia de tomar decisiones de inversión informadas y responsables."



## 9.2 Ejemplo plantilla “Tesis de inversión por fundamentales”

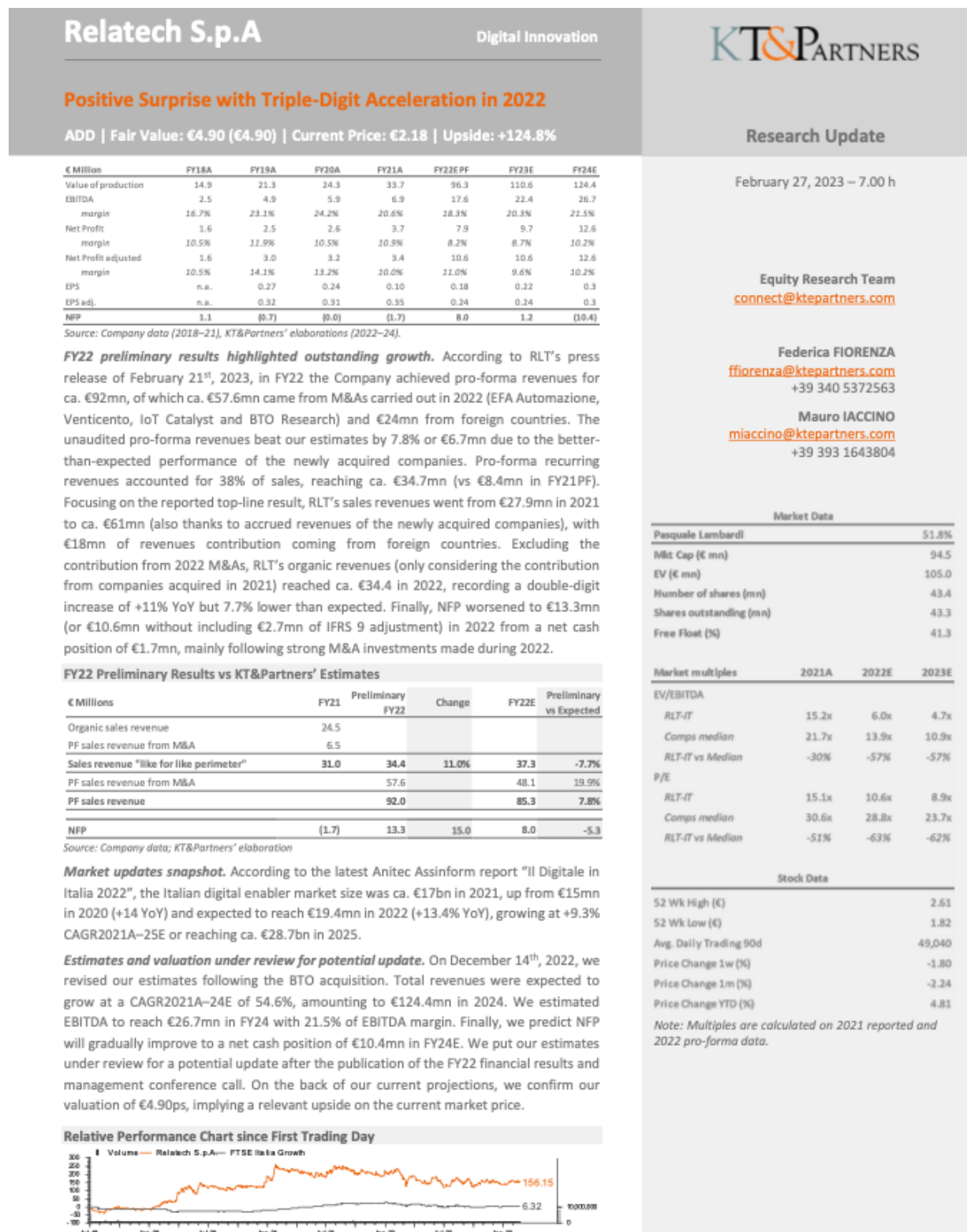


Figura 8. Página 1 la de tesis de inversión de KT&Partners respecto a la compañía Relatech S.p.A

## Referencias

- [1] «SEC.gov | EDGAR Application Programming Interfaces». <https://www.sec.gov/edgar/sec-api-documentation> (accedido 1 de julio de 2023).
- [2] «Morgan Stanley». <https://openai.com/customer-stories/morgan-stanley> (accedido 30 de junio de 2023).
- [3] «Welcome to the SHAP documentation — SHAP latest documentation». <https://shap.readthedocs.io/en/latest/index.html> (accedido 30 de junio de 2023).
- [4] «Papers with Code - LIME Explained». <https://paperswithcode.com/method/lime> (accedido 30 de junio de 2023).
- [5] «Pricing». <https://openai.com/pricing> (accedido 30 de junio de 2023).
- [6] «lxml - Processing XML and HTML with Python». <https://lxml.de/> (accedido 30 de junio de 2023).
- [7] «SEC.gov | EDGAR ABS XML Technical Specification (Version 3.0)». <https://www.sec.gov/info/edgar/specifications/absxml> (accedido 30 de junio de 2023).