**Desarrollo de un Sistema de Inversión y Calificación ESG basado en Señales y Análisis de Sentimiento con NLP y Redes LSTM**

Luis Marti Ávila

Máster en Inteligencia Artificial y Computación cuántica aplicada a mercados financieros. BME

Mayo 2025

**ÍNDICE**

1. INTRODUCCIÓN
2. MARCO TEÓRICO
   1. La integración de los criterios ESG en los mercados financieros
   2. Limitaciones de las calificaciones esg tradicionales
   3. El tratamiento de datos no estructurados como alternativa ESG
   4. Procesamiento de lenguaje natural para calificación ESG
   5. Predicción de scores ESG con modelo LSTM
   6. Estrategias de inversión basadas en señales ESG
3. EXTRACCIÓN DE DATOS
   1. Selección de empresas y sectores
   2. Fuentes utilizadas y justificación
   3. Proceso de validación y homogeneización
4. PRIMER MODELO NLP
   1. Etiquetado de noticias
      1. Descripción del proceso de etiquetado automatizado
   2. Arquitectura y entrenamiento del modelo
      1. Conjunto de datos de entrenamiento
      2. Preprocesamiento y tokenización
      3. Arquitectura del modelo
      4. Procedimiento de entrenamiento
      5. Resultado y evaluación
      6. Conclusión
   3. Inferencia sobre el dataset completo
      1. Preparación del corpus para inferencia
      2. Procedimiento
      3. Validación y control de calidad
5. SEGUNDO MODELO NLP
   1. Arquitectura y entrenamiento del modelo
      1. Conjunto de datos de entrenamiento
      2. Procedimiento de entrenamiento
      3. Resultado y evaluación
   2. Inferencia sobre el dataset completo
6. TERCER MODELO NLP
   1. Arquitectura y entrenamiento del modelo
      1. Entrenamiento y aplicación de un modelo propio de análisis de sentimiento
      2. Arquitectura del modelo
      3. Resultados
7. PREPARACIÓN DE LOS DATOS
8. REPUTACIÓN MEDIÁTICA
   1. Fundamento teórico
   2. Desarrollo
9. EXPLORACIÓN DE DATOS
   1. Preparación de los datos
   2. Distribución de noticias por empresa y categoría ESG
   3. Análisis de clustering
10. LSTM
    1. Creación de embeddings
       1. Preparación de los datos
       2. Generación de embeddings
       3. Agrupación y reducción dimensional
       4. Enriquecimiento de embeddings
    2. Generación de ventanas
       1. Fundamento teórico
       2. Generación de ventanas con padding
       3. Estadísticas y validaciones
       4. Exportación de resultados
    3. Asociación de ventanas
       1. Fundamento teórico
       2. Preparación de los datos
       3. Algoritmo de asociación tolerante
       4. Validación y exportación
    4. Entrenamiento
       1. Carga y preparación de los datos
       2. Diseño de la arquitectura LSTM
       3. Configuración del entrenamiento
       4. Evaluación
11. CREACIÓN DE UN DATASET COMPLETO
    1. Fundamento teórico
    2. Carga de datos
    3. Generación de nuevas ventanas de 7 días
    4. Entrenamiento de un nuevo modelo LSTM
    5. Filtrado de empresas y categorías
    6. Interpolación de scores oficiales y predicciones
    7. Dataset finale generado
12. DESARROLLO DE UN ALGORITMO DE INVERSIÓN
    1. Carga y preparación de los datos
    2. Diseño modular del sistema de trading
    3. Estrategia de inversión por categorías ESG
    4. Backtest operativo
    5. Análisis de resultados
    6. Comparativa de la estrategia ESG frente al benchmark
    7. Interpretación y conclusiones
    8. Líneas de mejora y optimización de la exposición
    9. Resumen
13. STREAMLIT Y GOOGLE CLOUD
    1. ESG dashboard
       1. Objetivos
       2. Arquitectura general del proyecto
       3. Desarrollo de la aplicación
       4. Justificación de la arquitectura cloud
       5. Despliegue
    2. Simulador de estrategias ESG y generador de señales
       1. objetivo
       2. Desarrollo
    3. Entrenamiento en Vertex AI
       1. Escalabilidad
       2. Despliegue
       3. Conclusiones
14. DESCARTES REALIZADOS
    1. Uso de los sustainability reports de las compañías
    2. Información constante para complementar la red LSTM además del cluster
15. BIBLIOGRAFÍA

**1.INTRODUCCIÓN**

En los últimos años los factores medioambientales y la responsabilidad social han adquirido una gran importancia en los mercados financieros. La evolución regulatoria, los usos sociales, y el papel activo de los inversores institucionales han contribuido a la integración de los criterios ESG (Environmental, Social and Governance) en los procesos de análisis de inversión (Friede, Busch & Bassen, 2015). Existen estudios que demuestran la correlación entre el desempeño ESG y el crecimiento a largo plazo de las compañías.

Debido a esto, ha surgido una necesidad crítica: desarrollar herramientas capaces de medir la evolución de estos factores de forma eficiente, dinámica y cuantificable. Sin embargo, a pesar de su creciente relevancia, la medición del desempeño ESG, en gran medida, sigue siendo un reto no resuelto en gran medida. Existen diversas empresas que otorgan estas calificaciones, si bien no ha dado respuesta a varios problemas: la opacidad de las metodologías, la heterogeneidad de los criterios aplicados y la ausencia de estandarización. Todo ello se traduce en discrepancias notables entre puntuaciones de una misma empresa (Berg, Koelbel & amp; Rigobon, 2022).

A estos problemas se añade la baja frecuencia de actualización de las calificaciones ESG convencionales ya que suelen emitirse con carácter trimestral o incluso anual, lo que hace difícil que reflejen eventos recientes que puedan tener un impacto relevante en la reputación o el comportamiento corporativo.

En este contexto, el tratamiento de grandes volúmenes de información no estructurada, como pueden ser las noticias financieras, los comunicados de prensa o las menciones en medios digitales, se posiciona como una alternativa para el análisis ESG. La ventaja de este tipo de contenido es la inmediatez con la que se reflejan eventos importantes, lo que permite una visión más reactiva y sensible a los cambios en la conducta empresarial (Krueger, Sautner & Starks, 2020).

La irrupción de técnicas de procesamiento del Lenguaje Natural (NLP) ha facilitado enormemente la explotación sistemática de este tipo de datos. En particular, la aparición de modelos pre entrenados basados en arquitecturas Transformer, ha supuesto un salto de calidad en la capacidad de clasificar textos, extraer noticias concretas de compañías determinadas, evaluar sentimientos y generar representaciones semánticas robustas. Estas técnicas permiten capturar de forma más precisa matices relevantes dentro de los textos, lo que permite enriquecer el análisis ESG.

Complementariamente, en el campo de las finanzas cuantitativas, los modelos de redes neuronales recurrentes, especialmente las redes LSTM (Long Short-Term Memory) (Hochreiter & Schmidhuber, 1997), han demostrado una gran rendimiento en trabajos de predicción de series temporales debido a su capacidad para retener información a lo largo de periodos extensos.

Este trabajo se enmarca en la intersección entre sostenibilidad, finanzas cuantitativas e inteligencia artificial, y tiene como objetivo principal diseñar, implementar y evaluar un algoritmo de inversión basado en señales ESG generadas a partir de noticias, además de la emisión de calificaciones ESG también a partir de estas. La propuesta incluye 3 componentes claves:

1. Un módulo de clasificación ESG y análisis de sentimiento, utilizando modelos NLP para identificar noticias relevantes y clasificarlas según su dimensión (E, S o G) y polaridad.
2. Un modelo LSTM que, partiendo de dichas noticias, predice diariamente un score ESG dinámico por empresa, permitiendo captar su evolución a lo largo del tiempo.
3. Un algoritmo de inversión basado en señales, que utiliza estos scores como input para generar decisiones de compra o venta, incorporando además aspectos realistas de ejecución.

Desde un punto de vista financiero, este enfoque busca validar la hipótesis de que el uso de información ESG no estructurada permite una mejora tangible en la toma de decisiones de inversión. Se plantea que el modelo puede detectar con mayor antelación eventos importantes y, por tanto, generar señales más alineadas con la realidad de los mercados.

Para asegurar la representatividad y robustez del sistema desarrollado, el análisis se ha centrado en una muestra de 30 empresas pertenecientes al índice S&P 500, distribuidas equitativamente entre seis sectores económicos distintos: tecnología, salud, energía, servicios financieros, consumo discrecional y comunicación. Esta selección busca garantizar una cobertura amplia y diversificada del mercado estadounidense, permitiendo observar cómo varían las señales ESG y su impacto financiero en función del contexto sectorial. Además, facilita el estudio de patrones comunes y divergencias entre industrias, aportando una mayor profundidad al análisis y favoreciendo el diseño de estrategias adaptadas a cada sector.

Finalmente, este proyecto no solo pretende aportar valor desde un punto de vista académico, sino también contribuir a la práctica profesional, proponiendo un enfoque replicable y escalable para la integración de criterios ESG en procesos de inversión automatizada. El documento se estructura de la siguiente forma: en primer lugar, se presenta el marco teórico ESG y el estado del arte en técnicas NLP y LSTM; a continuación, se detallan los datos y la metodología utilizada; posteriormente, se expone el diseño del sistema completo y se presentan los resultados experimentales; y finalmente, se discuten las conclusiones y líneas futuras de investigación.

**2.MARCO TEÓRICO**

## **a.La integración de criterios ESG en los mercados financieros**

El enfoque ESG busca ampliar los tradicionales criterios financieros incorporando métricas no financieras relacionadas con el comportamiento ambiental de una empresa (por ejemplo, emisiones de carbono, eficiencia energética o gestión de residuos), su impacto social (condiciones laborales, diversidad, relaciones con la comunidad) y la calidad de su gobernanza corporativa (estructura del consejo, políticas anticorrupción, transparencia). A través de estos ejes, se pretende evaluar el riesgo reputacional, operacional y regulatorio que enfrenta una entidad, así como su potencial para generar valor sostenible en el tiempo.

La regulación internacional ha jugado un papel clave. Iniciativas como el *European Green Deal*, el *Sustainable Finance Disclosure Regulation* (SFDR) en Europa o los Principios de Inversión Responsable (PRI) promovidos por la ONU han establecido un marco normativo que exige mayor transparencia y responsabilidad en la incorporación de criterios sostenibles en los productos financieros.

En segundo lugar, los grandes inversores institucionales —como fondos de pensiones, aseguradoras y gestoras de activos— han comenzado a exigir estándares ESG como condición para asignar capital. Organizaciones como BlackRock han expresado públicamente su compromiso con el cambio hacia una economía más sostenible, alineando sus políticas de inversión con objetivos de largo plazo como los ODS (Objetivos de Desarrollo Sostenible) de Naciones Unidas.

Por otro lado, existe una motivación estrictamente financiera: diversos estudios empíricos han demostrado una correlación positiva entre el buen desempeño ESG y la rentabilidad ajustada al riesgo de las empresas (Friede, Busch & Bassen, 2015). Las compañías que gestionan adecuadamente sus riesgos sociales y ambientales tienden a ser más resilientes, más innovadoras y menos expuestas a escándalos o litigios, lo que se traduce en una menor volatilidad y una mayor estabilidad en sus resultados financieros.

Este contexto ha propiciado un crecimiento exponencial de los activos gestionados bajo criterios ESG. Según datos de la *Global Sustainable Investment Alliance (GSIA)*, en 2022 el volumen de activos sostenibles superó los 35 billones de dólares a nivel global, representando casi el 40% del total de activos bajo gestión en mercados desarrollados. Asimismo, ha aumentado el número de índices bursátiles especializados, productos financieros ESG (fondos, ETFs, bonos verdes) y agencias de rating que ofrecen calificaciones sostenibles.

No obstante, este rápido crecimiento también ha revelado importantes desafíos metodológicos, especialmente en lo que respecta a la medición, comparación e interpretación de los datos ESG. Estos desafíos serán abordados en detalle en la siguiente sección.

## **b.Limitaciones de las calificaciones ESG tradicionales**

A pesar del creciente interés por integrar factores ESG en la toma de decisiones de inversión, las herramientas disponibles para medir y comparar el desempeño ESG de las empresas presentan notables limitaciones. Las calificaciones ESG proporcionadas por agencias especializadas —como MSCI, Sustainalytics, S&P Global o Refinitiv— han sido ampliamente adoptadas por gestores de activos, analistas y entidades reguladoras. Sin embargo, diversos estudios han puesto de manifiesto su baja coherencia interna y externa, lo que dificulta su uso como indicadores objetivos y comparables.

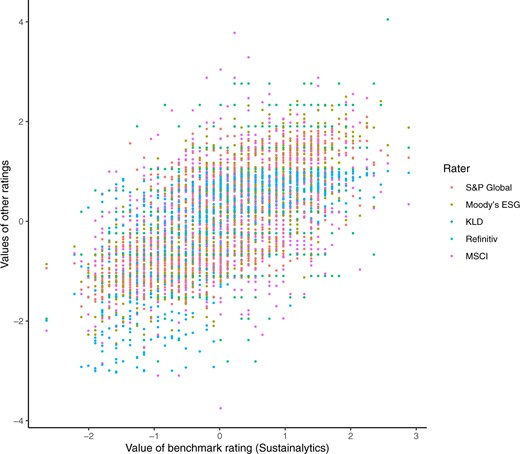
Una de las principales críticas es la falta de estandarización metodológica entre los distintos proveedores. Cada agencia utiliza su propio marco de evaluación, ponderando diferentes indicadores, criterios y fuentes de información. Como consecuencia, es común que una misma empresa reciba puntuaciones ESG significativamente distintas dependiendo de la agencia evaluadora. Según el estudio de Berg, Koelbel y Rigobon (2022), la correlación promedio entre ratings ESG de diferentes proveedores se sitúa en torno al 0,61, muy por debajo de la correlación habitual en otras métricas financieras como las calificaciones crediticias, donde suele superar el 0,90.

Esta divergencia en las calificaciones no es solo un problema técnico, sino que tiene implicaciones directas en la asignación de capital. Los gestores pueden tomar decisiones de compra o venta basadas en evaluaciones que no reflejan necesariamente un consenso sobre el comportamiento sostenible de una compañía.

Otro problema relevante es la escasa transparencia de las metodologías empleadas. Muchas agencias no publican en detalle los algoritmos, ponderaciones y fuentes de datos que utilizan, lo que impide a los usuarios finales comprender cómo se han construido las calificaciones. Esta opacidad complica la verificación externa, dificulta la replicabilidad de los resultados y limita la capacidad de auditar o mejorar los sistemas de puntuación.

Por último, cabe destacar la frecuencia limitada con la que se actualizan estas calificaciones. Al tratarse de informes exhaustivos, muchas agencias solo revisan sus ratings con periodicidad trimestral o anual. Esta falta de inmediatez provoca que eventos ESG relevantes —como conflictos laborales, derrames medioambientales o cambios regulatorios— no se reflejen en las calificaciones hasta semanas o meses después de haber ocurrido. En un mercado financiero cada vez más dependiente de la información en tiempo real, esta desincronización puede generar oportunidades perdidas o errores en la gestión del riesgo.

Frente a estas limitaciones, ha surgido la necesidad de explorar fuentes de información alternativas, más dinámicas y reactivas, capaces de capturar con mayor precisión y velocidad la evolución del comportamiento ESG de las empresas. En la siguiente sección se analizará el papel del tratamiento de datos no estructurados, como las noticias, en este nuevo paradigma de evaluación ESG.



*Esta figura muestra que, aunque las calificaciones ESG de diferentes agencias están positivamente correlacionadas, existe una dispersión considerable. Esto significa que una empresa puede recibir calificaciones significativamente diferentes según la agencia evaluadora, lo que dificulta identificar claramente a los líderes en desempeño ESG.*

## **c.El tratamiento de datos no estructurados como alternativa ESG**

Frente a las limitaciones asociadas a las calificaciones ESG tradicionales, ha ganado relevancia el uso de fuentes de información alternativas y más dinámicas, capaces de reflejar el comportamiento de las empresas en relación con factores ambientales, sociales y de gobernanza. Entre estas fuentes, los datos no estructurados, como las noticias, comunicados de prensa, publicaciones en redes sociales o transcripciones de conferencias corporativas, se presentan como una vía prometedora para complementar y enriquecer la evaluación ESG.

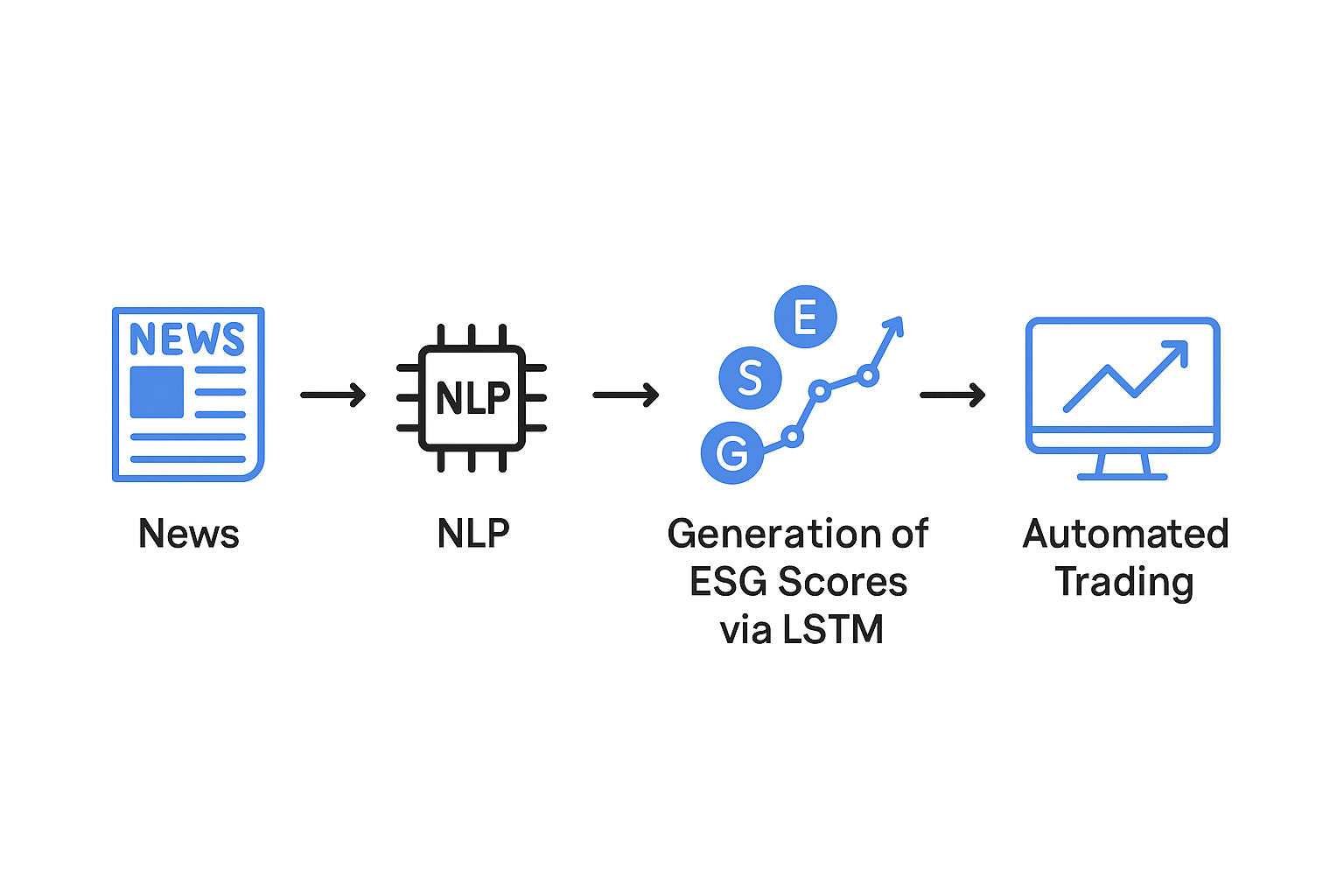
Los datos no estructurados tienen una característica distintiva: no siguen un formato tabular predefinido y, por tanto, requieren técnicas específicas para su procesamiento e interpretación. En el contexto ESG, este tipo de información puede capturar eventos relevantes con una inmediatez muy superior a la de los informes tradicionales. Por ejemplo, un escándalo de corrupción, un vertido tóxico o una protesta laboral pueden aparecer en titulares informativos horas o días antes de que afecten al precio de las acciones o sean recogidos por las agencias de rating.

Diversos trabajos han señalado el potencial de las noticias como indicadores tempranos de riesgo ESG (Krueger, Sautner & Starks, 2020). Las noticias financieras y sectoriales permiten monitorizar de forma continua las actividades de las empresas, su exposición reputacional y su relación con los distintos stakeholders. Además, ofrecen una gran variedad de perspectivas, ya que provienen de medios independientes, analistas especializados, periodistas o agencias gubernamentales.

El uso de esta información requiere superar varios desafíos técnicos. Por un lado, es necesario identificar si una noticia está efectivamente relacionada con una empresa concreta —para lo cual se aplican técnicas de reconocimiento de entidades (Named Entity Recognition, NER)— y, por otro, clasificar su contenido dentro de una de las tres dimensiones ESG. Asimismo, es fundamental determinar la carga emocional o el sentimiento de la noticia, ya que esto puede influir en la percepción pública del evento y, por ende, en su impacto financiero.

El tratamiento automatizado de grandes volúmenes de texto se ha visto enormemente potenciado por el desarrollo de técnicas de Procesamiento del Lenguaje Natural (NLP), especialmente con la aparición de modelos de lenguaje avanzados basados en redes neuronales profundas. Estas herramientas permiten convertir texto libre en representaciones estructuradas que pueden ser analizadas y utilizadas por sistemas de decisión o predicción.

La principal ventaja de utilizar datos no estructurados es su capacidad para generar una visión más reactiva y contextual del desempeño ESG, adaptada a los cambios en el entorno y al flujo continuo de información. Además, permiten una actualización diaria o incluso intradiaria de los indicadores ESG, lo que favorece que la toma de decisiones sea más ágil y alineada con la realidad de los mercados.



## **d.Procesamiento del Lenguaje Natural (NLP) para clasificación ESG**

El Procesamiento del Lenguaje Natural (NLP) es una rama de la inteligencia artificial que se centra en la interacción entre ordenadores y lenguaje humano. Su objetivo principal es dotar a las máquinas de la capacidad para comprender, interpretar, generar y clasificar textos escritos o hablados en lenguaje natural.

Tradicionalmente, las técnicas de NLP se apoyaban en enfoques basados en reglas o modelos estadísticos simples, como bolsas de palabras (bag-of-words), TF-IDF o Naive Bayes. Sin embargo, el panorama cambió radicalmente con la aparición de los modelos de lenguaje preentrenados basados en arquitecturas Transformer, como BERT (Devlin et al., 2018) y sus variantes más modernas como RoBERTa (Liu et al., 2019). Estos modelos permiten capturar de forma mucho más precisa el contexto y la semántica de las palabras, al considerar su significado en función del entorno en el que aparecen dentro de una frase.

La aplicación de modelos NLP en el ámbito ESG tiene múltiples dimensiones, como la calificación temática y el análisis de sentimiento.

En resumen, el NLP permite estructurar y transformar grandes cantidades de información textual en variables cuantificables y útiles para modelos predictivos. En este trabajo, esta tecnología constituye el primer paso del sistema, al encargarse de filtrar, clasificar y etiquetar todas las noticias que servirán como input para la generación de scores ESG dinámicos mediante modelos secuenciales.

## **e.Predicción de scores ESG con modelos LSTM**

El análisis del desempeño ESG de una empresa no solo requiere identificar y clasificar eventos relevantes, sino también modelizar su evolución en el tiempo. Para ello, es fundamental contar con herramientas capaces de procesar secuencias temporales y capturar patrones dinámicos en series de datos. En este contexto, los modelos de redes neuronales recurrentes (RNN) y, en particular, las Long Short-Term Memory (LSTM), se han consolidado como una de las metodologías más eficaces para tareas de predicción basadas en datos secuenciales.

Las LSTM fueron introducidas por Hochreiter y Schmidhuber (1997) como una solución a las limitaciones de las RNN tradicionales, en especial la dificultad para aprender dependencias a largo plazo debido al problema del desvanecimiento del gradiente. Las LSTM introducen una arquitectura basada en "celdas de memoria" y mecanismos de puertas (input, forget y output gates), que permiten controlar de forma selectiva qué información se almacena, olvida o transmite a través del tiempo. Esto las convierte en herramientas especialmente adecuadas para capturar tanto la persistencia como los cambios abruptos en la evolución de una variable.

En el contexto de este trabajo, las LSTM se emplean para generar un score ESG dinámico diario por empresa, a partir de las noticias previamente clasificadas mediante técnicas de NLP.

El modelo LSTM recibe como entrada una secuencia temporal de embeddings diarios de noticias. A partir de esta información, la red neuronal aprende a estimar la puntuación ESG más probable para cada día, replicando de forma aproximada el comportamiento de un rating real, pero con mayor frecuencia y sensibilidad al flujo de información.

Este enfoque no solo permite captar mejor la evolución ESG en el corto y medio plazo, sino que también ofrece la posibilidad de realizar predicciones anticipadas, en función de patrones históricos y señales recientes. Además, al tratarse de un modelo autorregresivo, puede integrarse de forma natural en sistemas de inversión cuantitativa, sirviendo como generador de señales para decisiones de compra o venta de activos financieros.

En conclusión, el uso de modelos LSTM va a construir una representación numérica continua y sensible de la sostenibilidad empresarial, superando el carácter estático y discreto de las métricas tradicionales. Esta representación es clave para alimentar el sistema de inversión desarrollado en este trabajo, cuya lógica se basa en la evolución diaria de los scores ESG generados.

## **f.Estrategias de inversión basadas en señales ESG**

El uso de señales en sistemas de inversión cuantitativa es una práctica ampliamente extendida en las finanzas modernas. Una señal puede definirse como una variable derivada de datos financieros o extrafinancieros que indica una posible oportunidad de inversión, ya sea para entrar o salir de una posición, ajustar el riesgo o modificar la exposición de una cartera. En el contexto de este trabajo, las señales ESG cumplen esa función, actuando como catalizadores para la toma de decisiones dentro de un sistema automatizado de trading.

En este trabajo, las señales ESG se construyen a partir de los scores dinámicos diarios generados mediante modelos LSTM, los cuales sintetizan la evolución reciente del comportamiento ESG de una empresa según las noticias clasificadas por el módulo de NLP. A partir de estos scores, se definen diferentes reglas de activación de señales, por ejemplo:

* Señal de compra (BUY): score ESG creciente y superior a cierto umbral.
* Señal de venta (SELL): caída abrupta del score o aparición de eventos negativos significativos.
* Señal de mantener (HOLD): score estable sin cambios relevantes.

Además, se pueden integrar otras señales complementarias basadas en desacoplamientos entre sentimiento y score, eventos extremos, o momentum ESG, generando una estrategia más sofisticada y adaptativa.

Estas señales alimentan un algoritmo de inversión automatizado que gestiona una cartera de activos reales (en este caso, acciones de 30 empresas del S&P 500). El sistema implementa reglas realistas de ejecución, incluyendo la consideración de comisiones, slippage y stop-loss, así como restricciones de exposición máxima y gestión del capital disponible.

Finalmente, cabe destacar que este enfoque permite una gran flexibilidad y escalabilidad. Las señales pueden adaptarse por sector, ajustarse en función del perfil de riesgo del inversor, o integrarse en modelos más complejos como redes neuronales multicapa o agentes de aprendizaje por refuerzo. De este modo, el sistema propuesto no solo es innovador en su concepción, sino también replicable en distintos contextos y adaptable a múltiples objetivos de inversión.

**3.EXTRACCIÓN DE DATOS**

La primera parte del trabajo consiste en la construcción de un dataset de noticias con contenido ESG, que sea capaz de alimentar los diferentes módulos del sistema propuesto.

Con el objetivo de maximizar la representatividad, diversidad y relevancia del conjunto de noticias, se ha seleccionado un periodo de 5 años , desde marzo de 2020 hasta marzo de 2025. Esta amplitud temporal permite capturar tanto eventos ESG excepcionales (por ejemplo, escándalos de gobernanza, incidentes ambientales o conflictos laborales), como también tendencias sostenidas (estrategias de sostenibilidad, planes de descarbonización, políticas de inclusión, etc.).

### **a.Selección de empresas y sectores**

El universo de análisis incluye 30 empresas del índice S&P 500, cuidadosamente seleccionadas para cubrir seis sectores económicos distintos: tecnología, energía, salud, consumo discrecional, servicios financieros y comunicación. Esta diversidad sectorial busca garantizar la generalización del modelo y evitar sesgos derivados de la concentración temática. La extracción de noticias se ha realizado de forma individualizada para cada una de estas compañías. Empresas seleccionadas:

* Salud: Johnson & Johnson, Pfizer, United Healthcare, CVS Health, Abbvie
* Tecnología: Apple, Microsoft, Google, Amazon, Nvidia
* Energía: ExxonMobil, Chevron, ConocoPhillips, Marathon Petroleum, Valero
* Finanzas: JP Morgan, Goldman Sachs, Morgan Stanley, American Express, Wells Fargo
* Comunicación y entretenimiento: Verizon, Comcast, Netflix, Disney
* Consumo Discrecional: McDonald's, Nike, Starbucks, The Home Depot, General Motors, Tesla

La elección de este conjunto de empresas se basa en que el índice S&P 500 es uno de los índices más representativos a nivel global, lo que puede asegurar una mayor cobertura mediática. Cabe destacar que las noticias seleccionadas cumplen el requisito de estar redactadas en inglés, lo que facilita todos los pasos posteriores.

### **b.Fuentes utilizadas y justificación**

Se han empleado tres fuentes principales de información, seleccionadas en función de su disponibilidad y cobertura histórica:

#### 1. Quantexa API News (2023–2025)

Esta plataforma es la fuente principal para los años más recientes, y destaca por su capacidad de aplicar reconocimiento automático de entidades (NER) en el proceso de extracción. Esta funcionalidad permite filtrar y descargar únicamente aquellas noticias que mencionan explícitamente el nombre de una empresa objetivo, con independencia del lenguaje utilizado, lo que reduce drásticamente la ambigüedad semántica.

Además, Quantexa incluye la opción de definir filtros temáticos adicionales, como “sostenibilidad”, “ética corporativa” o “energía renovable”, lo que garantiza que las noticias descargadas estén alineadas con los conceptos ESG.

2. NewsAPI (2020–2022)

Para los años anteriores se ha utilizado NewsAPI, dando acceso a noticias históricas a través de consultas por palabras clave. En este caso, se definió un conjunto de términos relevantes como “climate change”, “sustainability”, “diversity”, “ESG investing”, “human rights” o “governance issues”, entre otros. Aunque esta fuente ofrece un gran volumen y diversidad temática, no incluye funcionalidades de reconocimiento de entidades.

Para solucionar este problema, es necesario aplicar un proceso de filtrado posterior mediante técnicas de NER, utilizando modelos preentrenados de identificación de nombres de empresas en inglés. De esta manera, se garantiza que cada noticia esté correctamente asociada con al menos una de las 30 compañías seleccionadas. Esta doble capa de filtrado (por keywords y por entidades) ha conseguido reducir el ruido semántico de forma significativa.

Esta técnica se encuentra en el siguiente notebook: ‘02\_NER\_NEWSAPI’.

Se ha utilizado Spacy como herramienta de NER por su robustez a la hora de procesar textos en inglés. Dentro de la variedad de modelos, se ha optado por ‘en\_core\_web\_sm’ para equilibrar precisión con coste computacional, y el análisis se aplicó sobre el título y contenido de las noticias para dar un mayor contexto. Para validar los resultados, se ha realizado un muestreo manual.

#### 3. Financial Modeling Prep (FMP)

Como tercera fuente se incorporan noticias y comunicados oficiales extraídos de Financial Modeling Prep. Aunque el volumen obtenido ha sido mucho menor en comparación con las otras dos plataformas, su inclusión ha tenido un objetivo estratégico: equilibrar el sesgo mediático de los contenidos provenientes de fuentes externas. Al tratarse de notas de prensa emitidas directamente por las empresas, estas noticias reflejan la narrativa oficial, lo que permite contrastar posiciones y detectar posibles desacoplamientos entre el discurso interno y la cobertura mediática externa, aspecto relevante para las estrategias de inversión basada en señales.

De esta última fuente, no solo se han extraído algunas noticias sino que ha sido la única fuente de extracción de precios OHLC de cada compañía y de calificaciones trimestrales ESG. A su vez , FMP ha obtenido o generado dichas puntuaciones a partir de los informes 8-K y 10-K de la SEC (Securities and Exchange Commission). Este tipo de informes reflejan, desde eventos significativos relacionados con cuestiones ESG, hasta resultados financieros y eventos corporativos. Ejemplo:



### **c.Proceso de validación y homogeneización**

Uno de los principales desafíos del trabajo es integrar noticias procedentes de fuentes con formatos, estructuras y criterios de clasificación diferentes. Para garantizar la coherencia del dataset final se han implementado las siguientes etapas de procesamiento:

* Normalización estructural: unificación de campos clave como : title, content, publication\_date, source, ticker, category\_ESG, sentiment.
* Limpieza de datos: se han eliminado noticias sin cuerpo de texto, sin mención clara a la empresa, o con errores de codificación.
* Homogeneización del idioma: solo se conservaron noticias escritas en inglés, para evitar sesgos introducidos por traducciones automáticas.

El resultado final es un conjunto definido de más de 1 millón de noticias únicas (noticias\_totales.csv) , etiquetadas por empresa y con metadatos estandarizados, preparado para alimentar modelos de NLP y predicción ESG con una calidad aceptable para entornos de modelización robustos.

**4.PRIMER MODELO ROBERTA**

**a.Etiquetado de noticias**

Notebook: 01\_LONGCHAIN.ipynb

Uno de los principales desafíos al trabajar con grandes volúmenes de texto no estructurado es la necesidad de identificar y clasificar aquellas unidades de información que son realmente relevantes para los objetivos del análisis. En el presente trabajo, dicha necesidad se traduce en la distinción entre noticias que abordan cuestiones relacionadas con los criterios ESG (ambientales, sociales o de gobernanza) y aquellas que no lo hacen, a pesar de los filtrados previos en la extracción de noticias. Esta etapa es crítica, ya que actúa como filtro de entrada al sistema.

Para el entrenamiento de un modelo NLP, es necesario contar con una muestra de noticias etiquetadas. Para ello, se ha optado por una solución basada en modelos de lenguaje de gran escala (LLM), utilizando la librería LangChain como framework de orquestación y la API de OpenAI, específicamente el modelo GPT-4o, como motor de inferencia. A diferencia de los clasificadores tradicionales basados en aprendizaje supervisado, este enfoque permite aprovechar las capacidades semánticas avanzadas de los modelos de lenguaje, sin necesidad de contar con un conjunto previamente etiquetado por humanos.

En el contexto del presente trabajo, una noticia se considera “relacionada con ESG” si aborda directa o indirectamente alguna de estas tres dimensiones, incluso aunque no mencione explícitamente el acrónimo "ESG". Por ejemplo, una noticia sobre la imposición de multas por vertidos contaminantes o sobre un cambio en el consejo de administración debido a escándalos de corrupción es, indudablemente, ESG-relevante aunque no haga referencia directa a esa terminología.

### i.Descripción del proceso de etiquetado automatizado

La implementación técnica del sistema de clasificación se ha de realizar utilizando la combinación de LangChain y OpenAI, con una lógica de procesamiento secuencial:

1. Definición del prompt: Se ha diseñado un prompt detallado en el que se instruye al modelo para actuar como un analista experto en ESG. Prompt utilizado:

*‘’You are an expert in ESG (Environmental, Social, and Governance) analysis.*

*Analyze the following financial news article and determine whether it is related to ESG topics.*

*Please interpret ESG in a broad and inclusive way:*

*- Environmental includes sustainability, climate change, emissions, energy, biodiversity, etc.*

*- Social includes diversity, inclusion, employee well-being, human rights, education, community impact, etc.*

*- Governance includes ethical behavior, transparency, corporate governance, executive compensation, stakeholder rights, etc.*

*Even if the news is not explicitly labeled as ESG, it may still relate to ESG themes based on its content.*

*Always respond in English, even if the original article is in another language.*

*Return:*

*- `esg`: true or false*

*Text:*

*{input}’’*

1. Modelo utilizado: El modelo elegido ha sido gpt, un LLM de última generación optimizado para respuestas rápidas y contextualmente precisas.
2. Muestra clasificada: Se ha extraído una muestra aleatoria de 6.000 noticias del conjunto total. A cada noticia se le ha aplicado el prompt definido, y el modelo ha devuelto una salida estructurada con una etiqueta booleana (True si la noticia es ESG, False en caso contrario).

Este esquema permite construir un dataset etiquetado de forma rápida (sample\_esg\_or\_not.csv), consistente y con una base interpretativa sólida.

**b.Arquitectura y entrenamiento del modelo**

Notebook: 02\_ROBERTA\_ESG\_OR\_NOT.ipynb

### Una vez que contamos con el dataset de muestra etiquetado, el siguiente paso consiste en el entrenamiento de un modelo de procesamiento del lenguaje natural. En este proceso, el elegido ha sido RoBerta. La elección de este modelo deriva de la necesidad de interpretar no solo palabras clave aisladas, sino también el contexto semántico completo en el que se enmarcan las expresiones. Por esta razón, se ha optado por utilizar un modelo basado en la arquitectura RoBERTa (Robustly Optimized BERT Approach), desarrollado por Liu et al. (2019), que representa una evolución sobre el modelo BERT tradicional. RoBERTa es un modelo de lenguaje basado en Transformers, preentrenado a través del aprendizaje auto-supervisado en grandes cantidades de datos textuales.

La elección de RoBERTa responde a varios criterios estratégicos:

* Capacidad para manejar textos financieros y de noticias con matices complejos, ironías o relaciones implícitas, típicos del lenguaje empresarial.
* Transfer learning efectivo: dado que el modelo ya ha sido preentrenado en un corpus amplio, solo se requiere un ajuste fino ("fine-tuning") sobre el conjunto específico de noticias ESG.
* Robustez frente a ruido textual: esencial en noticias extraídas de diferentes fuentes con estilos de redacción variados.

En conjunto, RoBERTa proporciona una solución sólida y eficiente para afrontar la clasificación binaria ESG / no-ESG requerida en esta fase del sistema.

### i.Conjunto de datos de entrenamiento

El dataset utilizado para entrenar el modelo RoBERTa proviene del proceso de etiquetado inicial descrito en el apartado anterior. A partir del corpus completo de noticias extraídas entre 2020 y 2025, se ha seleccionado una muestra de 6.000 noticias etiquetadas de forma automática mediante un modelo de lenguaje, validándose manualmente en una proporción representativa. Para preparar los textos antes de ser procesados por los modelos NLP, se ha aplicado un proceso de limpieza y estructuración inicial. En primer lugar, se reemplazaron los valores nulos en las columnas ‘title’ y ‘content’ por cadenas vacías (''), garantizando así que no se produzcan errores durante la concatenación de campos. Posteriormente, se genera un nuevo campo denominado ‘input\_text’, compuesto por la concatenación del título de la noticia y los primeros 300 caracteres del contenido principal.

Esta decisión responde a la necesidad de mantener un balance entre contexto y eficiencia computacional: el título suele capturar la esencia de la noticia, mientras que un fragmento inicial del contenido aporta detalles adicionales sin sobrecargar el input con texto excesivamente largo.

Para evitar problemas de desbalanceo entre las clases, se ha implementado un proceso de resampleo:

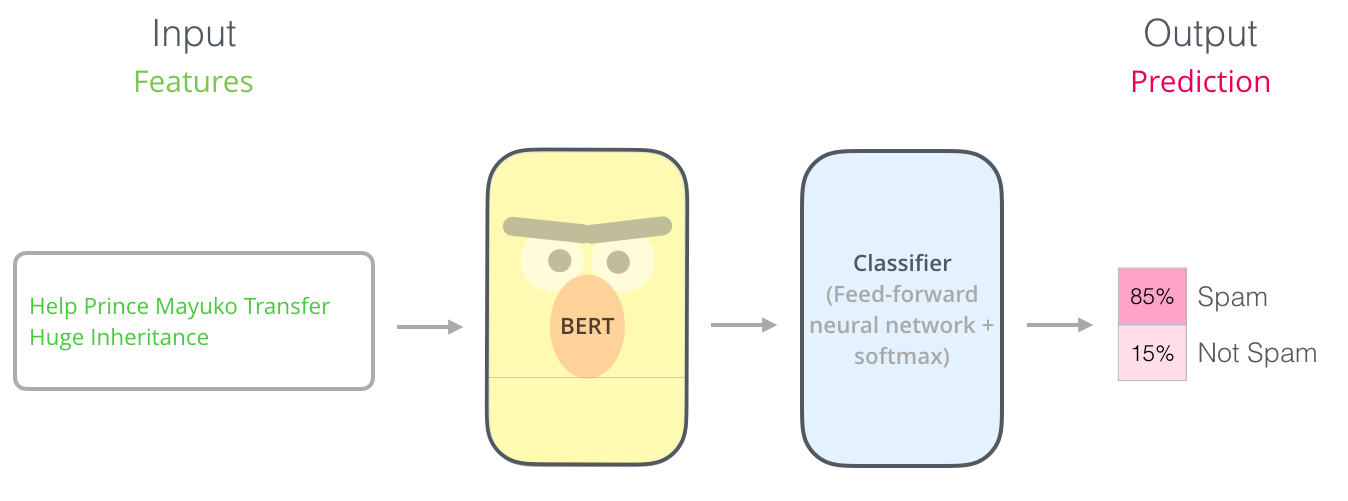
* La clase mayoritaria (noticias ESG) se reduce hasta igualar el número de ejemplos de la clase minoritaria (noticias no ESG).
* Este proceso permite entrenar un modelo equilibrado, minimizando el sesgo hacia la clase mayoritaria.

### ii.Preprocesamiento y tokenización

Antes de introducir los datos en el modelo RoBERTa, es necesario aplicar un proceso llamado tokenización. Este es un paso clave en cualquier modelo de procesamiento del lenguaje natural (NLP) porque convierte el texto crudo (por ejemplo, una frase) en una secuencia de números que el modelo puede interpretar. El tokenizador seleccionado es el oficial de ‘roberta-base’. Las principales operaciones realizadas son:

* Conversión de texto a tokens según el vocabulario preentrenado. Por ejemplo, la frase *‘The environment is critical’* puede transformarse en algo como:[0, 1332, 1234, 19, 4567, 2]
* Padding automático para ajustar todas las secuencias a la misma longitud (256 tokens) ya que estos modelos requieren que todas las secuencias tengan la misma longitud dentro de un batch.
* Truncamiento en los casos donde los textos superaban la longitud máxima admisible por el modelo.

El resultado de este proceso ha sido un conjunto de tensores de entrada (input\_ids) y máscaras de atención (attention\_mask). Los tensores de entrada representan las secuencias, y las máscaras de atención, las posiciones que son contenido real y las que son padding, todo ello necesario para alimentar adecuadamente el modelo RoBERTa.



*Esta imagen representa un ejemplo de clasificación de frases.*

Para poder alimentar adecuadamente un modelo NLP en Pytorch, es fundamental estructurar los datos en un formato que sea compatible con las herramientas internas de la biblioteca. PyTorch proporciona un módulo llamado torch.utils.data.Dataset, que es una clase base estándar para crear datasets personalizados. Nos va a permitir que los datos, aunque provengan de un archivo tabular (como un DataFrame de pandas), puedan transformarse en tensores y batches, que son las unidades de trabajo de los modelos de deep learning.

Para ello se ha diseñado una clase personalizada denominada ‘CustomDataset’. Esta clase permite llevar a cabo la transformación. En particular, el dataset contiene dos columnas clave: text (con el texto de la noticia) y label (la clase objetivo, codificada como entero).

La clase ‘CustomDataset’ ha implementado las siguientes funcionalidades principales:

* Inicialización: almacena la lista de textos y sus respectivas etiquetas, junto con el tokenizer y la longitud máxima (max\_len) permitida para las secuencias.
* Método \_\_len\_\_: devuelve la cantidad total de muestras, requisito estándar para la integración con PyTorch DataLoader.
* Método \_\_getitem\_\_: procesa cada texto aplicando el ‘tokenizer.encode\_plus’, que realiza la tokenización y devuelve los ‘input\_ids’ (tokens) y la ‘attention\_mask’ necesaria para los modelos basados en Transformers.

El método devuelve un diccionario con:

* input\_ids: secuencia tokenizada,
* attention\_mask: máscara de atención,
* targets: etiqueta real (convertida en tensor PyTorch).

Posteriormente, para la fase de entrenamiento y evaluación, se ha dividido el dataset balanceado en tres subconjuntos: train (80%), validación (10%) y test (10%).

Esta división se realiza utilizando la función ‘train\_test\_split’ de Scikit-learn, aplicando el argumento stratify para asegurar que la proporción de clases se mantuviera constante en todos los subconjuntos, lo cual resulta ser fundamental para evitar sesgos y garantizar una evaluación robusta.

Una vez definidos los datasets (train\_dataset, valid\_dataset, test\_dataset), se crean objetos llamados DataLoaders, que son herramientas fundamentales en PyTorch porque: dividen el dataset en mini-batches automáticos durante el entrenamiento (parámetro batch\_size), aleatorizan (shuffle) las muestras en cada época para mejorar la generalización y   
optimizan la carga de datos en memoria de forma eficiente.

### iii.Arquitectura del modelo

El modelo diseñado para la clasificación de noticias ESG se basa en un enfoque de fine-tuning (ajuste fino) de otro preentrenado, concretamente roberta-base. Con este enfoque se consigue aprovechar el conocimiento lingüístico adquirido previamente por RoBERTa , adaptándolo rápidamente a una tarea nueva con relativamente pocos datos adicionales.El esquema general del modelo es el siguiente:

1. Carga del modelo base (RoBerta Model): RobertaModel es la implementación del modelo RoBERTa disponible en la librería Hugging Face Transformers. Está basado en la arquitectura Transformer, una red neuronal profunda diseñada para trabajar con secuencias de texto y capturar relaciones complejas de contexto entre palabras. Transforma los textos tokenizados (los input\_ids y attention\_mask) en representaciones densas de alta dimensión, llamadas embeddings contextuales. Cada token de la entrada recibe un embedding que no solo representa su significado aislado, sino también su relación con los demás tokens en la secuencia.

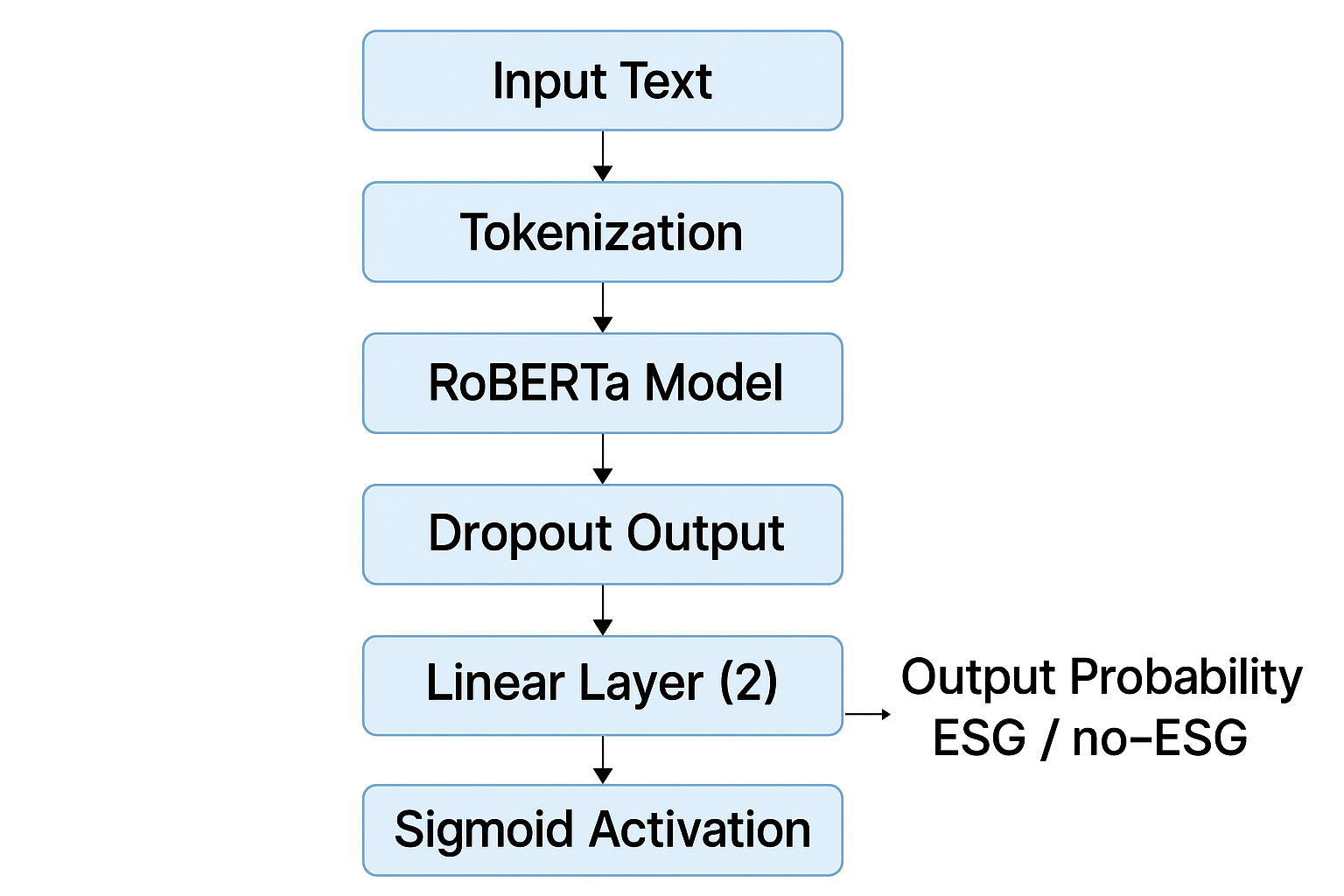
La salida principal usada en este proyecto es el pooler\_output, que corresponde a la representación final del token especial [CLS] (que siempre está al principio de la secuencia). Este vector es tratado como un resumen del significado global del texto y se usa típicamente para tareas de clasificación de secuencias completas.

1. Capa dropout: La capa Dropout (con probabilidad de 0.3 en este proyecto) es una técnica de regularización muy común en redes neuronales. Implementada mediante nn.Dropout en PyTorch, funciona desactivando aleatoriamente un porcentaje de las neuronas durante cada pasada de entrenamiento. Esto evita que el modelo memorice demasiado los datos de entrenamiento (overfitting) y mejora su capacidad para generalizar a datos nuevos. En este caso, al aplicar Dropout sobre el pooler\_output, se está regularizando justo la parte que más influye en la predicción final.
2. Capa de proyección lineal: Es una capa totalmente conectada (implementada como nn.Linear en PyTorch) que toma la salida del modelo RoBERTa (un vector de alta dimensión, por ejemplo 768 dimensiones en roberta-base) y la proyecta a una dimensión más baja. En este caso: 2 neuronas, una para cada clase (ESG / no-ESG).

Esta capa es la responsable de convertir la representación contextual de RoBERTa en una predicción final concreta. Funciona como un “puente” entre la parte del modelo que entiende lenguaje y la parte que toma decisiones.

1. Funcion de activación (sigmoid): La función sigmoide (Sigmoid) transforma la salida de la capa lineal en un número entre 0 y 1, que se interpreta como la probabilidad de que la noticia pertenezca a la clase ESG.

Esta ha sido la arquitectura escogida debido a su simplicidad y eficiencia ya que el modelo aprovecha la potencia de RoBERTa sin añadir capas complejas adicionales, lo que permite mantener la arquitectura ligera y eficiente. Además cabe recalcar su robustez al usar únicamente una capa lineal y Dropout encima del pooler\_output sigue una práctica recomendada y contrastada en la literatura para tareas de clasificación de texto con transformers (ver Liu et al., 2019).



### iv.Procedimiento de entrenamiento

El entrenamiento del modelo RoBERTa ajustado para la clasificación ESG se ha llevado a cabo mediante un procedimiento cuidadosamente diseñado para garantizar estabilidad numérica, buena convergencia y evitar problemas típicos como el sobreajuste. A continuación, se describen los componentes clave y su función:

1. Función de pérdida: Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss): Esta función combina de manera eficiente la entropía cruzada binaria (utilizada para problemas de clasificación binaria) con la función sigmoide, todo en un solo paso. Al integrar la función sigmoide directamente dentro de la pérdida, se evita la inestabilidad numérica que puede surgir si se aplican por separado. Es la elección estándar y recomendada para tareas de clasificación binaria en redes neuronales profundas como Transformers.
2. Optimizador: AdamW (Adam con Weight Decay): AdamW es una mejora del optimizador Adam tradicional. Introduce un decaimiento de peso (weight decay) de forma desacoplada, lo que mejora la capacidad del modelo para generalizar y evita que los pesos crezcan demasiado durante el entrenamiento. AdamW es altamente recomendado para modelos basados en Transformers, ya que mantiene una buena estabilidad durante la optimización y ayuda a evitar sobreajuste, algo crucial en tareas NLP donde los modelos suelen ser muy grandes.

En este proyecto no se ha llevado a cabo una optimización sistemática de hiperparámetros mediante métodos exhaustivos como grid search, búsqueda aleatoria o algoritmos bayesianos. En su lugar, se ha adoptado un enfoque basado en prueba y error (trial-and-error), ajustando los valores de hiperparámetros clave (como el batch size y la tasa de aprendizaje) en función de la experiencia previa y las mejores prácticas descritas en la literatura para modelos basados en Transformers.

Hiperparámetros principales:

* Batch size: 32 ejemplos por iteración.
* Número de épocas: 10 pasadas completas sobre el conjunto de entrenamiento.
* Learning rate inicial: ajustado mediante scheduler, comenzando en 1e-5.

Durante todo el proceso de entrenamiento se monitoriza la pérdida (loss) en el conjunto de entrenamiento y validación y la precisión (accuracy) para evaluar la calidad de las predicciones. Esta monitorización permite identificar posibles problemas de overfitting y aplicar early stopping (detección temprana) , en este caso, después de 5 épocas, cuando la mejora en el conjunto de validación se estabilizó o comenzó a deteriorarse.

Epoch 1/10, Train Loss: 0.4900, Train Acc: 0.7346, Val Loss: 0.3309, Val Acc: 0.8476

Epoch 2/10, Train Loss: 0.2619, Train Acc: 0.8908, Val Loss: 0.3018, Val Acc: 0.8614

Epoch 3/10, Train Loss: 0.1809, Train Acc: 0.9264, Val Loss: 0.3871, Val Acc: 0.8614

Epoch 4/10, Train Loss: 0.1051, Train Acc: 0.9619, Val Loss: 0.3742, Val Acc: 0.8568

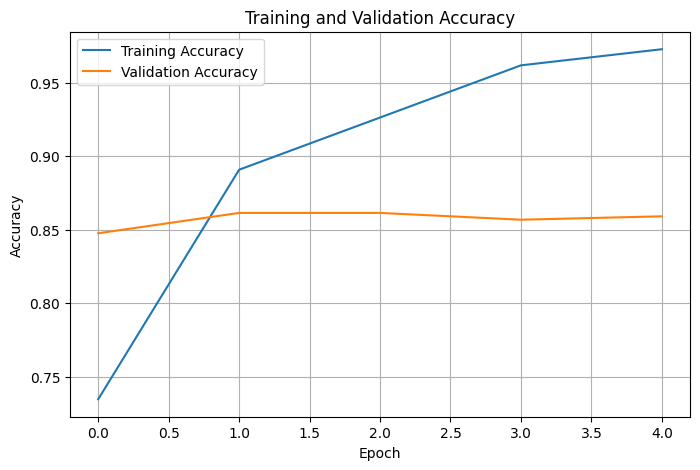
Epoch 5/10, Train Loss: 0.0808, Train Acc: 0.9729, Val Loss: 0.4182, Val Acc: 0.8591

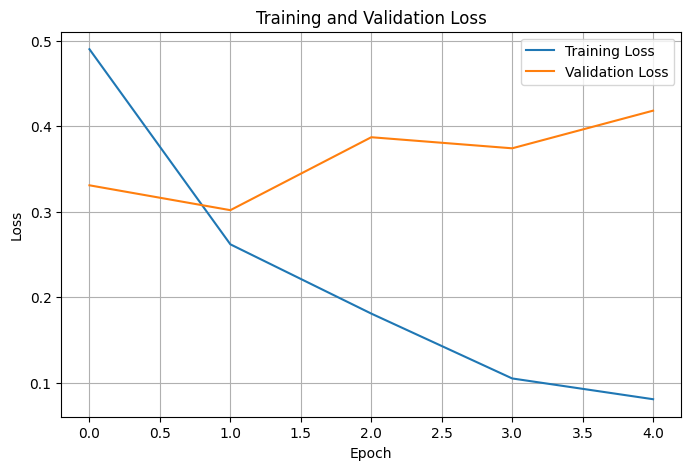
Early stopping triggered after 5 epochs.

El entrenamiento completo se ha realizado en un entorno de GPU, permitiendo tiempos razonables de ajuste incluso trabajando con el modelo RoBERTa.

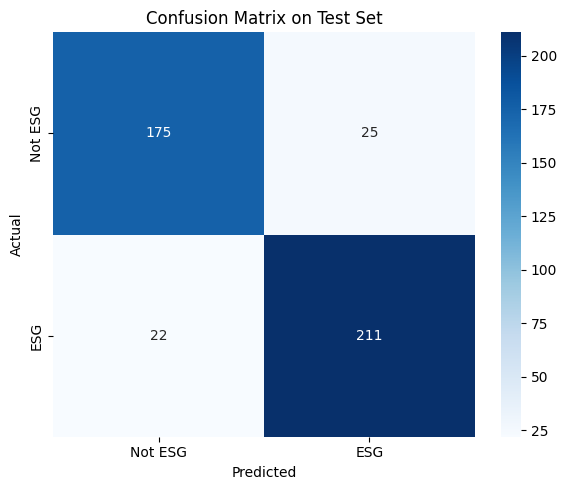
### v.Resultados y evaluación

Finalizado el proceso de entrenamiento, se ha evaluado el desempeño del modelo sobre el conjunto de validación. Los principales resultados obtenidos fueron:

* Accuracy (precisión global): 88,9%
* Precision (precisión en clase ESG): 89%
* Recall (sensibilidad para detectar noticias ESG): 91%
* F1-Score: 88%  
  



Además, se construyó una matriz de confusión que refleja el balance entre falsos positivos y falsos negativos. Estos resultados indican que el modelo es capaz de identificar de manera efectiva noticias relacionadas con ESG, manteniendo un equilibrio adecuado entre precisión y sensibilidad.



La tasa relativamente baja de falsos negativos (noticias ESG clasificadas erróneamente como no ESG) es especialmente importante, dado que la prioridad del sistema es maximizar la detección de eventos relevantes para la sostenibilidad empresarial.

### vi.Conclusión

El modelo RoBERTa ajustado mediante fine-tuning ha demostrado ser una herramienta eficaz para abordar el primer gran reto del sistema: la filtración precisa de noticias relacionadas con ESG.  
Gracias a su capacidad de interpretar el contexto de las noticias, superar el enfoque puramente basado en palabras clave y adaptarse rápidamente al dominio financiero, el modelo proporciona una base fiable sobre la cual construir los siguientes módulos de análisis.

Modelo: esg\_model\_weights.pt

**c.Inferencia sobre el dataset completo**

Notebook: 03\_INFERENCE\_ROBERTA\_1.ipynb

Tras la finalización del entrenamiento y validación del modelo RoBERTa especializado en la detección de noticias ESG (Environmental, Social, Governance), se procede a aplicar el modelo sobre el conjunto completo de noticias recopiladas en el marco temporal de marzo de 2020 a marzo de 2025. Esta fase ha permitido clasificar automáticamente y de manera eficiente cada noticia según su relevancia ESG, sirviendo como paso preliminar indispensable para los análisis temáticos y estratégicos posteriores.

### i.Preparación del corpus para inferencia

El conjunto de noticias utilizadas para la inferencia se ha almacenado en un único archivo ‘noticias\_totales.csv’. Sin embargo, dado que el modelo RoBERTa requiere entradas textuales bien estructuradas y de longitud controlada, ha sido necesario realizar el mismo procesamiento que se llevó a cabo en la fase de entrenamiento. Este paso garantiza que las predicciones sean consistentes y comparables con los resultados obtenidos durante la validación.

ii.Procedimiento de inferencia

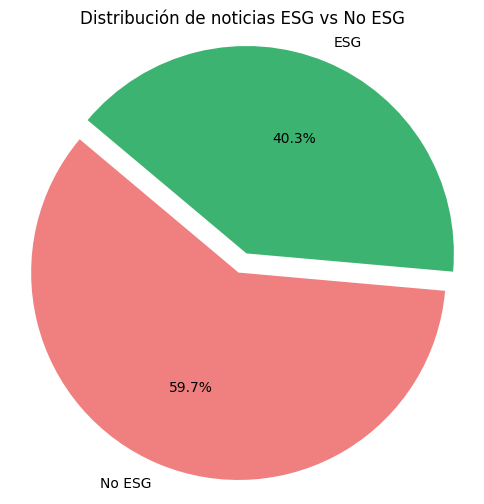
El modelo se aplica sobre el corpus completo de noticias utilizando un DataLoader, lo que permite procesar los datos en batches y optimizar la eficiencia computacional, especialmente en GPU. El flujo de operaciones ha sido el siguiente:

1. Predicción de logits: para cada batch, el modelo genera salidas brutas (logits) correspondientes a las clases ESG y no ESG. Los logits representan las puntuaciones antes de aplicar la función sigmoide.
2. Conversión a predicciones discretas: mediante la función torch.max, se selecciona la clase con mayor probabilidad para cada noticia. Esto convierte las predicciones en valores discretos: 0 → Noticia no relacionada con ESG y 1 → Noticia relacionada con ESG.
3. Almacenamiento de resultados: las predicciones se asignan a una nueva columna esg\_pred en el DataFrame original.

### iii.Validación y control de calidad

Aunque el modelo había sido previamente validado en el conjunto de test, se ha realizado una verificación básica post-inferencia (total de noticias clasificadas como ESG –400K):

* Revisión manual de una muestra: Se selecciona aleatoriamente un subconjunto de varias noticias clasificadas como ESG y no ESG para inspección manual..
* Coherencia temática: Las noticias etiquetadas como ESG incluían predominantemente eventos relevantes como iniciativas de sostenibilidad, conflictos laborales, mejoras en gobernanza, sanciones regulatorias o emisiones contaminantes.



Ejemplo de noticia con categoría ESG:

* Goldman Sachs commences latest round of layoffs, NY Times reports.
* Woman accused of attacking McDonald's coworker with headset ordered to stand trial.

**5.SEGUNDO MODELO NLP**

Tras identificar qué noticias del corpus están relacionadas de forma general con los factores ESG mediante el primer modelo RoBERTa, el siguiente paso consiste en clasificar de manera más granular dichas noticias en una de las tres dimensiones tradicionales del análisis ESG: Environmental (E), Social (S) o Governance (G).

Esta segunda clasificación es indispensable para poder construir posteriormente indicadores separados, analizando la evolución de cada categoría.

La tarea de asignar una noticia a una categoría ESG concreta presenta varias dificultades:

* Muchos artículos pueden tratar varios aspectos simultáneamente (por ejemplo, políticas de diversidad dentro de una estrategia de sostenibilidad ambiental).
* Los enfoques tradicionales de clasificación basados en palabras clave son insuficientes, dado que los conceptos ESG suelen aparecer de forma implícita o entrelazada en los textos.
* La precisión y la consistencia en la clasificación son críticas para evitar sesgos en los indicadores posteriores.

Notebook: 04\_LONGCHAIN:2.ipynb

La forma de crear la muestra etiquetada es la misma que en el primer modelo: LangChain + API de OpenAI (gpt-4o).Se filtran todas las noticias previamente etiquetadas como ‘esg\_pred == 1’ (es decir, relevantes ESG) en el proceso de inferencia inicial, conformando un conjunto inicial de trabajo.

Posteriormente, se extrae una muestra aleatoria de 8.000 noticias, buscando un tamaño de muestra suficiente para capturar la diversidad temática y garantizar una base sólida para el entrenamiento y validación de modelos posteriores. Se ha elaborado un prompt estructurado que instruye al modelo para actuar como analista ESG, proporcionando definiciones claras de cada categoría:

*‘You are an ESG analyst.*

*Classify the following news article into one of the ESG categories:*

*- Environmental (e.g., emissions, energy, pollution, biodiversity, sustainability)*

*- Social (e.g., labor practices, community, diversity, equality, employee well-being)*

*- Governance (e.g., corporate ethics, transparency, board structure, regulation)*

*Always choose one category only: "Environmental", "Social", or "Governance".*

*Text:*

*{input}’*

Cada noticia ha sido evaluada de forma individual, obteniendo como resultado una etiqueta categórica (Environmental, Social, Governance) que se ha añadido como nueva columna en el DataFrame.

**a.Arquitectura y entrenamiento del modelo**

Notebook: 05\_ROBERTA\_E\_S\_G.ipynb

Para esta segunda tarea de clasificación, se ha mantenido la estrategia de utilizar un modelo basado en la arquitectura RoBERTa (Robustly Optimized BERT Approach), por las mismas razones que motivaron su elección inicial.

### i.Conjunto de datos de entrenamiento

El dataset utilizado en esta fase procede del archivo ‘sample\_e\_s\_g.csv’, generado en la fase de etiquetado temático descrita en el apartado anterior. Cada noticia incluye: el campo del texto, que contiene el título concatenado con el cuerpo abreviado (primeros 300 caracteres) del artículo y el campo esg\_category, que contiene la etiqueta asignada (Environmental, Social o Governance).  
Para preparar el conjunto de datos de entrenamiento: se mapean las etiquetas textuales a valores numéricos:

* + Environmental → 0
  + Social → 1
  + Governance → 2

Balanceo de clases: dado que la distribución inicial de categorías presentaba diferencias, se ha igualado el número de ejemplos por clase (n = 1680 ejemplos por categoría) mediante técnicas de submuestreo aleatorio (resample), reduciendo así el riesgo de sesgo hacia la clase mayoritaria. Una vez el dataframe balanceado, este se ha divido en 3 conjuntos:   
entrenamiento (70%), validación (15%) y test (15%). La partición se realiza de forma estratificada para mantener la proporción de clases en cada subconjunto.

La arquitectura del segundo modelo es la misma que el primero, con la diferencia de la capa de clasificación, que en este caso cuenta con 3 neuronas, cada una de ellas correspondiente a una categoría ESG.

### ii.Procedimiento de entrenamiento

El proceso de entrenamiento se ha llevado a cabo utilizando la misma configuración que en el primer modelo. Durante el entrenamiento se ha monitorizado la pérdida sobre el conjunto de validación en cada época, aplicando criterios de early stopping implícito si se detectaban signos de sobreajuste.

Epoch 1/10 | Train Loss: 0.7331, Train Acc: 0.6465 | Val Loss: 0.3074, Val Acc: 0.8810

Epoch 2/10 | Train Loss: 0.2717, Train Acc: 0.9031 | Val Loss: 0.2848, Val Acc: 0.8876

Epoch 3/10 | Train Loss: 0.1673, Train Acc: 0.9419 | Val Loss: 0.2959, Val Acc: 0.8929

Epoch 4/10 | Train Loss: 0.1053, Train Acc: 0.9671 | Val Loss: 0.3049, Val Acc: 0.9021

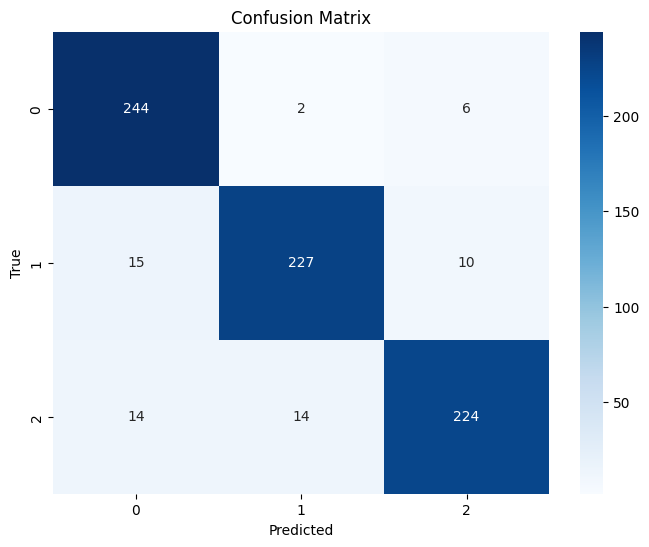
Epoch 5/10 | Train Loss: 0.0580, Train Acc: 0.9810 | Val Loss: 0.3674, Val Acc: 0.8981

Early stopping triggered!

### iii.Resultados y evaluación

Finalizado el entrenamiento, el modelo se evalúa sobre el conjunto de test independiente.  
 Los principales resultados han sido:

* Accuracy global: 91%
* Distribución de errores:  
  + Errores más frecuentes: confusión entre noticias de tipo *Social* y *Governance*, dada la cercanía semántica de ciertos temas.
  + Alta precisión para la clase *Environmental*, probablemente debido a la especificidad de los términos relacionados con medioambiente.



Estos resultados reflejan que el modelo es capaz de identificar de manera sólida la dimensión ESG predominante en las noticias, alcanzando niveles de precisión adecuados para alimentar los módulos posteriores de generación de scores y señales de trading.

Además de la métrica general de precisión global (91%), se han calculado otras métricas detalladas para cada una de las categorías ESG (Environmental, Social, Governance) sobre el conjunto de test. Esto es crucial para evaluar no solo el rendimiento global del modelo, sino también su capacidad para discriminar de forma efectiva entre las tres clases, especialmente considerando la naturaleza compleja y a menudo solapada de los temas ESG.

| **Categoría** | **Precisión (Precision)** | **Recall (Sensibilidad)** | **F1-Score** | **Support** |
| --- | --- | --- | --- | --- |
| Environmental | 0.89 | 0.97 | 0.93 | 252 |
| Social | 0.93 | 0.90 | 0.92 | 252 |
| Governance | 0.93 | 0.89 | 0.91 | 252 |
| **Accuracy Global** |  |  |  | **0.92** |

**b.Inferencia sobre el dataset completo**

Notebook: 06\_INFERENCE\_ROBERTA\_2.ipynb

Una vez entrenado el modelo RoBERTa especializado en la clasificación temática de noticias ESG, se procede a aplicar el modelo sobre el conjunto completo de noticias previamente identificadas como relevantes en materia ESG. El proceso se inicia con la carga del archivo ‘total\_news\_esg\_filtered.csv’, que contiene todas las noticias etiquetadas como relevantes (esg\_pred == 1). Para preparar las entradas al modelo, se ha utilizado como input la columna text, compuesta por el título de la noticia concatenado con los primeros 300 caracteres del contenido, asegurando un equilibrio entre la riqueza de información y la longitud máxima permitida por el modelo RoBERTa.

Cada noticia es etiquetada con un número (0 para Environmental, 1 para Social y 2 para Governance), que posteriormente se mapea a su correspondiente etiqueta textual para facilitar la interpretación de los resultados.

Tras el etiquetado completo,se puede observar la siguiente distribución de categorías. Environmental = 110702, Social = 151232 y Governance = 169286.

Para asegurar la homogeneización en la identificación de empresas, se han aplicado pequeñas correcciones a los nombres de compañías detectados en las noticias, tales como:

* "jp morgan" → "jpmorgan"
* "cvs health" → "cvs"
* "the walt disney company" → "disney"

El nuevo conjunto de datos, que incluye la clasificación temática ESG para cada noticia, se almacena en el archivo ‘news\_second\_classified.csv’. Este dataset constituye la base fundamental para las siguientes fases del proyecto, permitiendo construir scores diarios específicos por dimensión ESG y diseñar estrategias de inversión diferenciadas.

En conclusión, esta fase de inferencia completa la estructuración temática del corpus de noticias ESG, dotando al sistema de un flujo de información preciso, escalable y alineado con los objetivos de seguimiento dinámico de factores de sostenibilidad a nivel empresarial.

**6.TERCER MODELO NLP**

Despues de completar la clasificación temática de las noticias relevantes en materia ESG, se aborda el desarrollo de un tercer módulo de análisis orientado a evaluar el sentimiento que transmite cada noticia. El objetivo de esta fase ha sido enriquecer la base de datos con una variable adicional que capturara la percepción implícita que transmiten los medios financieros sobre las compañías, aspecto fundamental para incorporar una dimensión de análisis cualitativo a la modelización posterior de scores ESG.

El análisis de sentimiento se ha enfocado de manera específica a las noticias ESG, clasificando el tono general de cada noticia según tres posibles categorías:

* Bullish: noticia con impacto positivo esperado para la empresa. Ejemplos: anuncios de sostenibilidad exitosa, premios, mejora de políticas ESG
* Bearish: noticia con impacto negativo esperado. Ejemplos: sanciones regulatorias, problemas medioambientales, escándalos de gobernanza.
* Neutral: noticia sin impacto significativo previsible.

Para mantener la coherencia metodológica con los modelos anteriores, la generación de la muestra etiquetada se realiza siguiendo el mismo enfoque: LangChain + API de OpenAI (gpt-4o).

Notebook: 07\_LONGCHAIN\_3.ipynb

Prompt usado:

*‘’You are an expert financial analyst specialized in ESG-related news.*

*Analyze the following news article and classify its overall sentiment towards the company's performance as one of:*

*- bullish (positive impact expected)*

*- bearish (negative impact expected)*

*- neutral (no significant impact)*

*Only respond with one of these three words: bullish, bearish, or neutral.*

*Text:*

*{input}’’*

**a.Arquitectura y entrenamiento del modelo**

Notebook: 08\_SENTIMENT\_ANALYSIS.ipynb

## i.Entrenamiento y aplicación de un modelo propio de análisis de sentimiento

El conjunto de datos utilizado para esta tarea se encuentra en el archivo ‘sentiment\_sample\_e\_s\_g.csv’, obtenido en la fase previa mediante clasificación asistida por modelos de lenguaje. Cada noticia del dataset contiene la concatenación del título y el inicio del cuerpo de la noticia, así como la etiqueta de sentimiento predicha manualmente como bullish, bearish o neutral. La primera etapa del proceso consiste en preparar adecuadamente el conjunto de datos. Para ello:

* Se mapea el sentimiento textual a valores numéricos, asignando 0 a noticias bearish, 1 a noticias neutral y 2 a noticias bullish.
* Al observar un ligero desequilibrio entre clases, es necesario aplicar una estrategia de submuestreo aleatorio (downsampling) para igualar el número de ejemplos por clase, fijando un tamaño de muestra de 2.312 ejemplos para bearish y neutral, igual al número de ejemplos bullish disponibles.
* Se reorganiza el dataset en un formato estándar para clasificación, con dos columnas: body (texto de entrada) y label (sentimiento mapeado).

Posteriormente, se procede a la tokenización de los textos mediante el tokenizador oficial de RoBERTa (roberta-base), truncando y aplicando padding hasta una longitud máxima de 256 tokens. Se divide el conjunto de datos balanceado en tres subconjuntos: 70% para entrenamiento, 15% para validación y 15% para test, aplicando partición estratificada para conservar las proporciones de clases en cada subconjunto. Esta preparación permite construir TensorDataset y DataLoader para cada partición, facilitando el entrenamiento y evaluación eficientes en batches.

ii.Arquitectura del modelo

Para abordar la tarea de análisis de sentimiento multiclase sobre noticias financieras (clasificación en bearish, neutral, bullish), se ha desarrollado un modelo personalizado denominado NewsSentimentAttentionModel, diseñado sobre la base del preentrenado RoBERTa e integrado con capas adicionales de atención multi-cabeza y feed-forward que enriquecen la capacidad de captura de patrones complejos en los textos.

La arquitectura sigue un esquema modular compuesto por varias capas y bloques funcionales clave:

### Capa base: RoBERTa Encoder: El modelo inicia con la carga de ‘RobertaModel’ (configuración 'roberta-base'), que actúa como encoder principal y responsable de convertir el texto tokenizado (entrada) en una matriz de embeddings contextuales de dimensión (secuencia x 768).

### Atención multicabeza adicional: Para refinar y enriquecer aún más la representación generada por RoBERTa, se ha añadido una capa de nn.MultiheadAttention. Esta capa implementa un mecanismo de auto-atención, que permite al modelo ponderar diferentes partes de la secuencia de entrada en relación consigo misma, lo que mejora la capacidad para identificar patrones relacionales complejos dentro del texto (por ejemplo, relacionar sujetos y predicados distantes en la frase o identificar contexto emocional global). Justificación: Aunque RoBERTa ya contiene múltiples capas de atención, añadir una capa externa adicional puede actuar como una especie de reafinamiento, especialmente útil cuando se trabaja sobre un dominio específico (noticias financieras con matices emocionales).

* Feed\_Forward adicional: Después de la capa de atención, se implementa un bloque feed-forward que sigue la estructura típica de los bloques Transformer, compuesto por: Linear(768, 768) + ReLU + Dropout(0.1) + Linear(768, 768). Esto mejora la capacidad del modelo para capturar patrones no lineales y jerárquicos dentro de la representación textual.
* Segundo bloque de normalización + pooling: refuerza la robustez del modelo. La normalización garantiza que las activaciones permanezcan en rangos controlados, mientras que el pooling (por ejemplo, mean-pooling o max-pooling sobre la secuencia) condensa la información a un vector fijo de 768 dimensiones.
* Clasificación final: Tras obtener el vector final de 768 dimensiones, se aplica una secuencia de capas densas que terminan generando 3 clases de salida: bearish, neutral, bullish.

Input (input\_ids, attention\_mask)

→ RoBERTa (last\_hidden\_state)

→ Multi-Head Attention (4 heads)

→ Add & Norm

→ Feed-Forward + Dropout + ReLU

→ Add & Norm

→ Pooling (mean over sequence)

→ Dense Layer (768→128) + ReLU + Dropout

→ Dense Layer (128→3)

→ Output logits (para CrossEntropyLoss)

Estrategia de fine-tuning: Congelación parcial de capas: La congelación consiste en bloquear los parámetros (pesos) de ciertas capas del modelo para que no se actualicen durante el entrenamiento. Esto permite que partes del modelo con conocimiento ya adquirido (en este caso, del preentrenamiento general de RoBERTa) se mantengan estables, mientras se ajustan sólo las capas más relevantes para la tarea específica. En este contexto específico, se han congelado todas las capas excepto las últimas 4 del encoder: layer.8, layer.9, layer.10 y layer.11

El entrenamiento del modelo se ha llevado a cabo durante un máximo de 10 épocas, utilizando:

* Función de pérdida: CrossEntropyLoss, adecuada para problemas de clasificación multiclase.
* Optimizador: AdamW, especialmente efectivo en arquitecturas Transformer.
* Tasa de aprendizaje inicial: 2e-5.
* Tamaño de batch: 16 ejemplos.
* Aceleración mediante GPU cuando estuvo disponible.

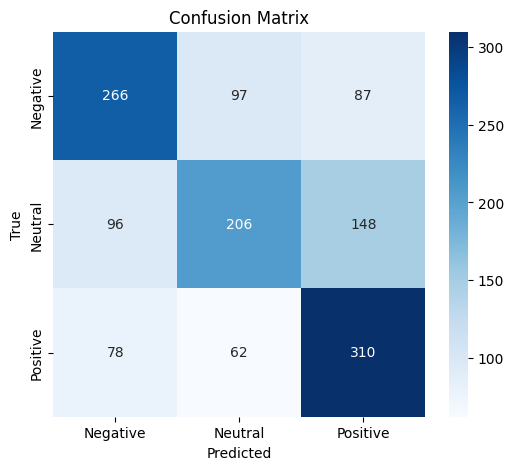
Para prevenir fenómenos de overfitting, se implementa un mecanismo de early stopping basado en la pérdida de validación, con una paciencia de 3 épocas. El modelo se guarda en disco cada vez que mejora la pérdida de validación, permitiendo recuperar la mejor versión entrenada una vez finalizado el proceso.

iii.Resultados

El modelo muestra un rendimiento sólido, especialmente en las clases Negative y Positive, donde se alcanzaron F1-scores de 0.83 y 0.84 respectivamente. Esto indica que resulta particularmente eficaz para identificar noticias con un sesgo negativo o positivo claro.

La clase Neutral, sin embargo, presenta un F1-score más bajo (0.68), lo que refleja una mayor dificultad para detectar con precisión este tipo de noticias. Este comportamiento es esperable en tareas de análisis de sentimiento, donde las noticias neutrales tienden a ser más ambiguas o menos definidas semánticamente, dificultando su clasificación incluso para modelos avanzados.

El accuracy global del modelo se sitúa en un 78%, lo que implica que aproximadamente 8 de cada 10 noticias son clasificadas correctamente en su categoría de sentimiento correspondiente



La evaluación del modelo ha evidenciado que el sistema es capaz de capturar adecuadamente el tono de las noticias, ofreciendo un rendimiento comparable al de modelos comerciales de análisis de sentimiento, pero con la ventaja de haber sido entrenado específicamente sobre un corpus de noticias ESG, alineado con los objetivos del proyecto.

Una vez validado el modelo, se ha procedido a aplicar la inferencia sobre el corpus completo de noticias relevantes (news\_second\_classified.csv). Para ello, se carga el modelo entrenado, se tokeniza todo el corpus con las mismas configuraciones utilizadas en entrenamiento, y se realiza la inferencia batch por batch, clasificando cada noticia en bearish, neutral o bullish.

bearish 149725

neutral 145726

bullish 135769

El resultado de esta fase ha sido la generación de una nueva columna de sentimiento (predicted\_sentiment\_final) asociada a cada noticia ESG del dataset global. Esta columna, junto con la clasificación temática (Environmental, Social, Governance), constituye la base de información necesaria para construir los scores ESG dinámicos que guiarán las estrategias de inversión en las fases posteriores del sistema.

En conclusión, el desarrollo de un modelo propio de análisis de sentimiento adaptado al dominio ESG permite cerrar el ciclo de procesamiento de noticias de forma completamente autónoma, robusta y escalable. Gracias a su entrenamiento cuidadoso, su diseño arquitectónico y su validación rigurosa, este modelo garantiza la calidad y fiabilidad necesarias para alimentar los módulos predictivos y de señalización que constituyen el núcleo del sistema de inversión basado en sostenibilidad.

**7.PREPARACIÓN DE LOS DATOS**

Notebook: 09\_DATA\_PREPARATION.ipynb

Con el objetivo de garantizar la calidad y consistencia del conjunto de datos de noticias ESG previo a la generación de puntuaciones dinámicas, se llevó a cabo un proceso de limpieza y estandarización de información.

El punto de partida fue el archivo ‘df\_finally\_labeled.csv’, el cual ya contiene para cada noticia los campos principales necesarios: identificador del ticker de la empresa, fecha de publicación, clasificación temática ESG (Environmental, Social o Governance) y sentimiento predicho (bullish, bearish o neutral).

En primer lugar, se realizó una limpieza estructural del dataset. Se eliminaron columnas innecesarias que no aportan valor al análisis posterior, tales como ‘description’, ‘sentiment\_body’ o ‘id’. Esta depuración permite reducir el tamaño del archivo y simplificar las operaciones posteriores. La siguiente fase se centra en garantizar la correcta asociación entre tickers y nombres de empresas. Se han detectado casos en los que, por problemas de extracción en las fases anteriores, algunas entradas tenían el ticker pero no el nombre de la empresa, o viceversa. Para resolver esta situación:

* Si una noticia tiene el ticker informado pero el company ausente, se imputa el nombre de empresa correspondiente utilizando un diccionario generado a partir de las coincidencias existentes en el dataset.
* De manera recíproca, si una noticia tiene el nombre de empresa pero el ticker ausente, se imputa el ticker correspondiente mediante el mapeo inverso.

Este proceso permite rellenar de manera precisa la gran mayoría de los valores ausentes en estos campos, asegurando así la trazabilidad correcta entre noticias y compañías.

No obstante, quedaban casos especiales donde el ticker aparecía como UNKNOWN. Para abordar esta problemática, se ha diseñado un proceso de corrección manual basado en el nombre de la empresa. Se aplicaron las siguientes sustituciones:

* "disney" → "DIS"
* "general motors" → "GM"
* "jpmorgan" → "JPM"
* "morgan stanley" → "MS"
* "wells fargo" → "WFC"
* "comcast" → "CMCSA"
* "american express" → "AXP"
* "the home depot" → "HD"

Esta imputación manual permite recuperar de manera precisa los tickers correctos en la mayoría de los casos de UNKNOWN, evitando la pérdida de información valiosa en empresas relevantes para el análisis.

Una vez resueltos los problemas de ticker y empresa, se ha abordado la imputación del sector de actividad (sector) para aquellas noticias cuyo sector estaba inicialmente ausente. Para ello, se ha construido un diccionario que asocia a cada ticker el sector más frecuente en el dataset. Este mapeo se aplica únicamente en los casos de ausencia de sector, evitando sobrescribir información válida. Gracias a esta estrategia, se ha logrado completar el campo sectorial en un porcentaje muy elevado de las noticias, permitiendo así posteriores análisis sectoriales de los scores ESG.

Finalmente, el conjunto de datos limpio, estructurado y enriquecido fue ordenado cronológicamente por fecha de publicación, asegurando una correcta alineación temporal para la construcción de ventanas de análisis dinámico. El dataset definitivo se ha almacenado en el archivo df\_cleaned.csv, sirviendo como base para la generación de los scores ESG diarios que alimentan las estrategias de inversión sostenibles.

En conclusión, esta fase de preparación de datos no solo permite consolidar toda la información extraída y procesada previamente, sino que establece las bases de integridad, consistencia y robustez necesarias para afrontar con garantías la modelización predictiva de la evolución ESG a partir de eventos de noticias. Cada corrección aplicada en esta fase representa un paso fundamental hacia la obtención de resultados fiables, replicables y científicamente válidos en las etapas posteriores del proyecto.

**8.REPUTACIÓN MEDIÁTICA**

Notebook: 10\_MEDIA\_REPUTATION.ipynb

Una vez completada la fase de limpieza del dataset de noticias, la siguiente consiste en el cálculo de la reputación de fuentes mediáticas. El objetivo de esta fase es ponderar cada noticia considerando la fiabilidad y la influencia de la fuente que la publica.

**a.Fundamento teórico**

La literatura académica ha demostrado que la exposición mediática repetida y la frecuencia de aparición en medios influyen de manera significativa en la percepción pública y en la construcción de la reputación de las empresas. Deephouse (2000) define la reputación mediática como un recurso estratégico que impacta directamente en el valor percibido de las compañías, y sostiene que la repetición mediática refuerza la legitimidad empresarial. Carroll (2010) también señala que los medios de comunicación actúan como amplificadores de la reputación, donde la frecuencia y la visibilidad son indicadores clave de influencia. Asimismo, Fombrun y Shanley (1990) argumentan que las empresas más cubiertas por medios consolidados disfrutan de una reputación más sólida y sostenida en el tiempo, lo cual justifica que la ponderación por frecuencia de aparición sea un proxy válido para medir la relevancia e impacto informativo de una fuente.

Partiendo de esta base teórica, se establece que las fuentes más prolíficas (en términos de número de noticias) deben tener un mayor peso específico en los análisis reputacionales y, por extensión, en el sistema de predicción ESG.

* Carroll, C. E. (2010). *Corporate reputation and the news media: Agenda-setting within business news coverage in developed, emerging, and frontier markets*. Routledge.
* Christophersen, M., & Jurish, B. (2021). *RapidFuzz: A fast string matching library*. Retrieved from https://maxbachmann.github.io/RapidFuzz/
* Deephouse, D. L. (2000). *Media reputation as a strategic resource: An integration of mass communication and resource-based theories*. Journal of Management, 26(6), 1091–1112.
* Fombrun, C. J., & Shanley, M. (1990). *What's in a name? Reputation building and corporate strategy*. Academy of Management Journal, 33(2), 233–258.

**b.Desarrollo**

Para abordar esta cuestión, se comienza por cargar el conjunto de datos ‘df\_cleaned.csv’, que contiene todas las noticias relevantes ya estructuradas y etiquetadas. Debido a la inconsistencia de la nomenclatura de las fuentes,que puede distorsionar futuros análisis, como "Reuters", "Reuters News Service" o "Reuters World News", se ha implementado un proceso de normalización basado en técnicas de "fuzzy matching".

Utilizando la librería ‘rapidfuzz’, se ha aplicado un algoritmo de comparación de cadenas que permite agrupar variantes similares bajo un único identificador. El proceso consiste en recorrer todas las fuentes y, para cada nueva fuente, buscar el match más similar entre las ya normalizadas, siempre que la similitud supere un umbral predefinido del 90%. De este modo, se construye un mapeo ‘source\_map’ que unifica todas las variantes bajo un nombre base representativo, normalmente el más corto o estándar.

Una vez normalizadas las fuentes, se calcula el volumen de noticias atribuido a cada fuente, desglosado además por sector. Para ello:

* Se agrupan las noticias por pares (sector, source).
* Se calcula el número de noticias correspondientes a cada combinación, obteniendo así una primera medida de la influencia relativa de cada fuente dentro de cada sector económico.

Sobre estos volúmenes de noticias, se ha diseñado una función de escalado para generar un reputation\_score para cada fuente. La asignación del score sigue una lógica de dos tramos:

* Las fuentes que contaban con 30 o más noticias en un sector determinado reciben un score máximo de 1, entendiendo que una alta frecuencia de aparición es indicativa de influencia y fiabilidad reconocida en ese ámbito.
* Para las fuentes con menos de 30 noticias, se aplica una función de escalado logarítmico que asignaba un score entre 0,5 y 1.

Esta función permite asignar scores de manera progresiva, penalizando ligeramente a las fuentes menos representativas sin excluirse totalmente del análisis. En resumen, la incorporación del ‘reputation\_score’ permite añadir una dimensión de calidad informativa al análisis de noticias ESG. Al ponderar la importancia de cada evento no solo en función de su contenido, sino también en función de la fiabilidad de la fuente que lo origina, se mejora la capacidad del sistema para priorizar señales relevantes y minimizar el impacto de eventos reportados por medios marginales o poco fiables. Esta estrategia contribuye a aumentar la robustez y la interpretabilidad del sistema global de predicción ESG basado en análisis de medios.

**9.EXPLORACIÓN DE DATOS**

Notebook: 11\_DATA\_EXPLORATION.ipynb

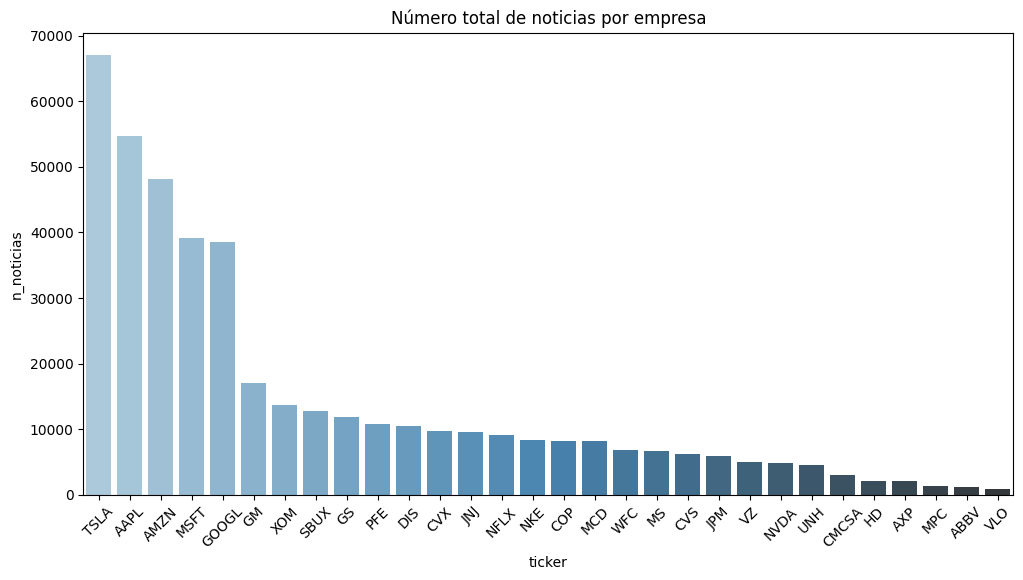
Antes de proceder a la construcción definitiva de los scores ESG dinámicos, se ha llevado a cabo un análisis exploratorio detallado del conjunto de noticias procesadas. El objetivo de esta fase es doble: por un lado, obtener una visión estructurada sobre la distribución temática, sectorial y sentimental de las noticias; por otro lado, identificar patrones que pudieran guiar la posterior definición de estrategias de generación de señales ESG.

**a.Preparación de datos**

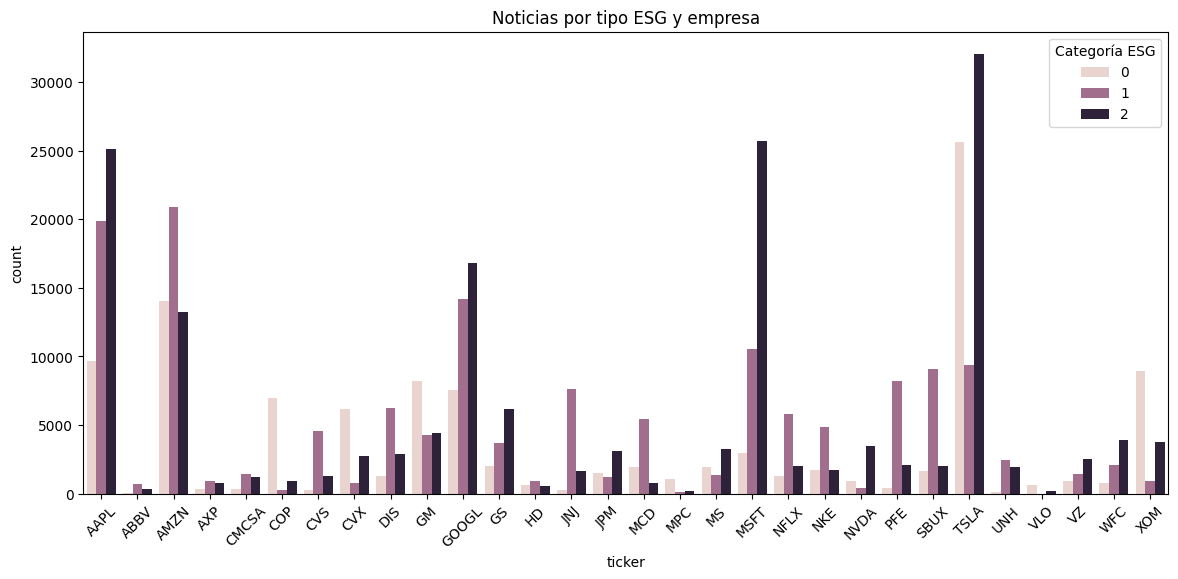
Se utilizan los dos conjuntos de datos principales, el que contiene las noticias y el de los scores ESG (NEWS\_+\_PRESS\_ESG.csv y ESG\_SCORES.csv), estandarizando las columnas en común.

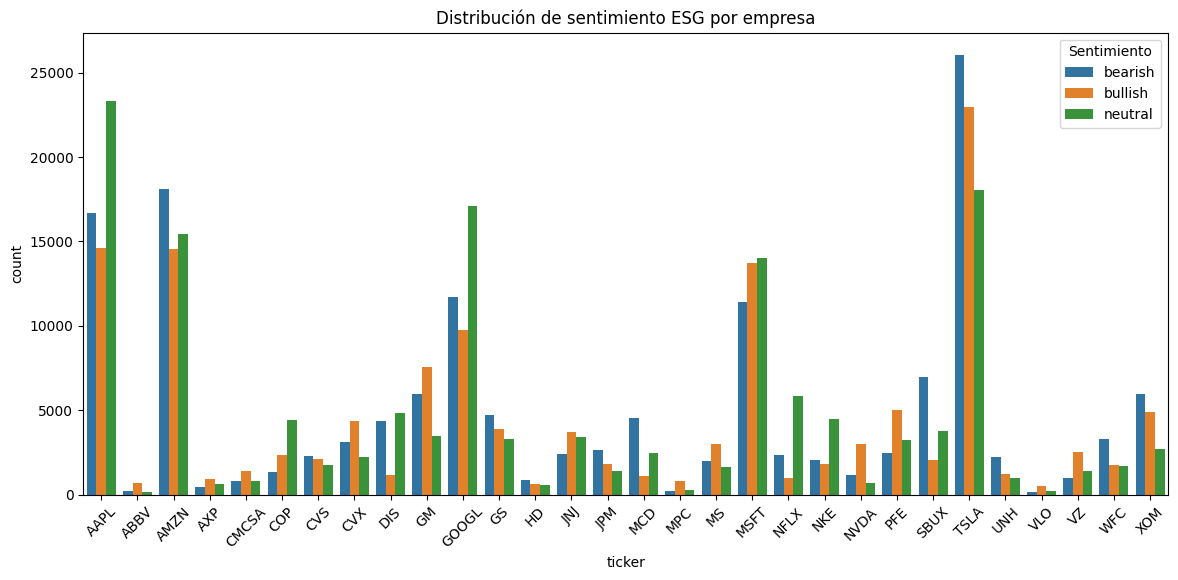
**b.Distribución de noticias por empresa y categoría ESG**

El primer paso consiste en cuantificar la cobertura mediática para cada una de las empresas, desglosando el volumen total de noticias según la categoría ESG correspondiente.



*Como podemos observar existen empresas como Tesla, Apple o Amazon que agrupan la mayor parte del total de noticias, mientras que otras como Valero o Abbvie ocupan un espacio más residual.*

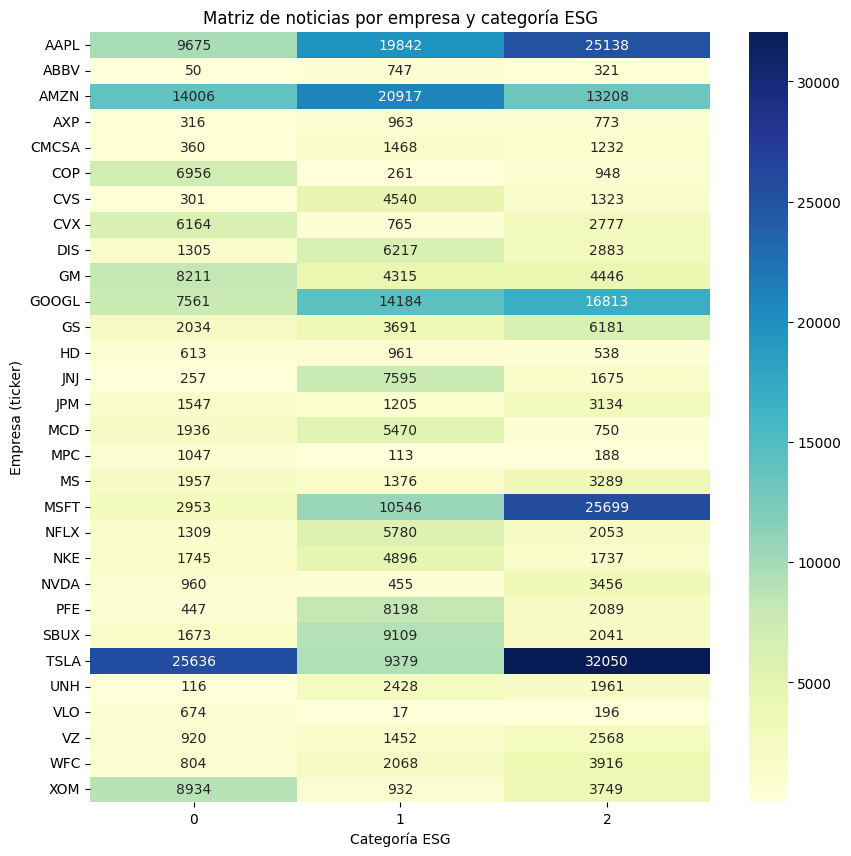




El análisis confirma la existencia de un notable desequilibrio en la cobertura ESG: mientras algunas empresas presentan una elevada exposición mediática distribuida de manera equilibrada entre las tres categorías ESG, otras se caracterizan por concentrar su visibilidad exclusivamente en aspectos ambientales o de gobernanza. Esta heterogeneidad sugiere que la percepción pública de las empresas puede estar fuertemente condicionada por los temas ESG más recurrentes en la prensa.

También se ha realizado un análisis de la evolución temporal del volumen de noticias ESG, donde podemos observar cómo desde mediados de 2023 a mediados de 2024 el volumen de estas crece considerablemente.





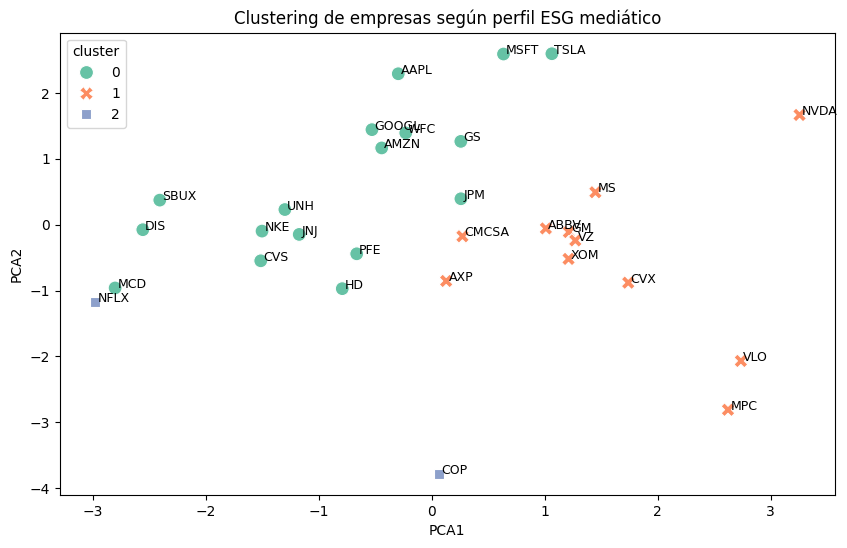
**c.Análisis de Clustering: Identificación de Perfiles ESG Mediáticos**

Con el objetivo de detectar patrones latentes de comportamiento ESG mediático, se ha llevado a cabo un análisis de clustering sobre las empresas. Este análisis permite agrupar a las compañías en función de características como el número total de noticias ESG, la distribución por categoría ESG y el sentimiento medio asociado a cada dimensión.

Para la segmentación se emplea el algoritmo K-Means, configurado para generar tres clusters. La elección de este algoritmo se justifica por su simplicidad, rapidez y eficacia en el manejo de datos numéricos estandarizados. La selección de tres clusters busca identificar perfiles diferenciados de empresas, potencialmente correspondientes a niveles alto, medio y bajo de exposición ESG mediática.

El análisis confirma la existencia de tres grupos bien definidos: un primer cluster compuesto por empresas con alta presencia mediática y distribución equilibrada en las tres categorías ESG; un segundo grupo de empresas con exposición moderada, generalmente dominada por una categoría concreta; y un tercer cluster que agrupa a compañías con baja visibilidad mediática o perfiles ESG poco definidos. Esta segmentación resulta especialmente útil para orientar estrategias empresariales futuras y para focalizar los análisis predictivos en función del perfil mediático de cada empresa.

Para facilitar la interpretación y visualización de los clusters obtenidos, se emplea la técnica de análisis de componentes principales (PCA), que permite reducir la dimensionalidad del espacio multivariado a dos componentes principales. Esta reducción mantiene la mayor parte de la varianza explicativa de los datos y posibilita representar gráficamente la estructura de los clusters en un plano bidimensional.



Finalmente, el dataset ‘df\_summary’, enriquecido con la asignación de clusters y las coordenadas PCA, ha sido guardado para ser utilizado en las siguientes fases de modelización.

En conclusión, esta fase de análisis exploratorio permite caracterizar de manera detallada el universo de empresas en función de su exposición mediática ESG, proporcionando una base sólida para entender la dinámica de generación de scores futuros. La combinación de análisis descriptivo, construcción de variables resumen y técnicas de clustering ha abierto la puerta a enfoques más sofisticados en la interpretación de señales ESG y en la personalización de estrategias de inversión sostenible.

**10.LSTM**

**a.Creación de embeddings**

Notebook: 12\_EMBEDING\_GENERATION.ipynb

El objetivo de esta sección es transformar el contenido de las noticias en representaciones numéricas densas (embeddings) que permitan capturar relaciones semánticas y contextuales entre las diferentes noticias. Esta fase de generación de embeddings constituye el primer paso para preparar la entrada del modelo LSTM. Esto consiste en transformar los textos diarios en representaciones vectoriales de dimensión fija que capturan tanto el contenido semántico como información adicional relevante para tareas de predicción.

i.Preparación de datos

Se trabaja con el conjunto de datos ‘NEWS\_+\_PRESS\_ESG.csv’, que contiene las noticias y comunicados ESG ya preprocesados y estructurados, y el resumen ‘DF\_SUMMARY.csv’, que aporta información complementaria como el cluster al que pertenece cada empresa (identificado en la fase de exploración previa).

Antes de proceder a la generación de embeddings, se realiza una validación y limpieza final de los datos, garantizando la ausencia de valores nulos y la correcta tipificación de las columnas clave. Esta fase resulta indispensable para asegurar la calidad de los embeddings y evitar la propagación de errores en las fases posteriores.

ii.Generación de embeddings

El proceso de generación se lleva a cabo utilizando la arquitectura SentenceTransformer (all-MiniLM-L6-v2). Cada noticia (formada por un título y/o cuerpo de texto) se transforma en un vector numérico de alta dimensión (768 dimensiones). La elección de Sentence Transformers se fundamenta en su capacidad para generar embeddings contextuales y significativos a nivel de frase o documento completo, superando las limitaciones de modelos tradicionales como Word2Vec o TF-IDF, que carecen de un entendimiento profundo del contexto.

iii.Agrupación y reducción dimensional

Posteriormente, se agrupan las noticias por ticker, fecha y categoría ESG, concatenando en un único texto todas las noticias relacionadas con una misma empresa y dimensión ESG en un mismo día. Esta estrategia se justifica por la necesidad de capturar de forma compacta toda la información disponible para cada combinación empresa-fecha-categoría, evitando la fragmentación de señales que podría producirse si se procesaran noticias individuales de manera aislada.

Una vez construido el corpus agregado, se ha procedido a la generación de embeddings utilizando un modelo de SentenceTransformer preentrenado (all-MiniLM-L6-v2). Esta elección se fundamenta en varias razones:

* Los modelos de tipo SentenceTransformer están específicamente diseñados para generar embeddings que preserven la semántica de frases y párrafos, no solo palabras sueltas.
* El modelo all-MiniLM-L6-v2 ofrece un equilibrio excelente entre calidad y eficiencia computacional, siendo capaz de generar embeddings de alta calidad con un coste razonable de recursos.
* Al tratarse de un modelo multilingüe, permite asegurar una cobertura adecuada incluso en el caso de noticias que contengan expresiones en inglés técnico o financiero.

Sin embargo, trabajar directamente con embeddings de tan alta dimensionalidad puede ser ineficiente e incluso perjudicial para el entrenamiento de redes neuronales posteriores, especialmente en datasets de tamaño limitado. Por ello, se aplica una reducción de dimensionalidad mediante Análisis de Componentes Principales (PCA), disminuyendo la representación de cada embedding a 100 dimensiones. La elección de PCA se justifica por su capacidad para:

* Preservar la mayor parte de la varianza informativa del conjunto original,
* Eliminar redundancias y ruido en los datos,
* Mejorar la eficiencia y estabilidad del entrenamiento del modelo LSTM posterior.

iv.Enriquecimiento de embeddings con información de clusters

Para dotar a los embeddings de un mayor contexto estructural, se combina cada embedding reducido con la información del cluster al que pertenece la empresa (obtenido previamente en el análisis de clustering). Esta combinación se implementa concatenando el vector reducido con el valor numérico del cluster, generando así un embedding enriquecido que integra tanto la representación semántica del texto como la segmentación estratégica identificada en fases anteriores.  
Finalmente, se ha almacenado el conjunto de embeddings generados en el archivo ‘EMBEDDINGS.pkl’. Cada fila del archivo contiene:

* El identificador de la empresa (ticker),
* La fecha del evento,
* La categoría ESG a la que pertenece el evento,
* El vector embedding combinado de 101 dimensiones (100 del contenido textual + 1 del cluster).

En conclusión, esta fase de generación de embeddings permite transformar de manera eficaz el flujo de noticias ESG diarias en una representación vectorial adecuada para su procesamiento secuencial mediante redes neuronales recurrentes. La combinación de información textual semántica y contexto empresarial estructurado sienta las bases para una modelización predictiva robusta, capaz de capturar las dinámicas temporales de percepción ESG en los mercados financieros.

**b.Creación de ventanas**

Notebook: 13\_WINDOW\_GENERATION.ipynb

Una vez creados los embeddings, estos se transforman en series temporales, ideales para su introducción en una red LSTM. Se parte del conjunto de datos EMBEDDINGS.pkl , que se ha construido en el apartado anterior.

i.Fundamento teórico

Cada ventana se diseña para contener una secuencia de 90 días de eventos ESG consecutivos. La elección de este tamaño de ventana responde a varias consideraciones:

* Permite capturar dinámicas de corto y medio plazo en la percepción ESG de las empresas.
* Cubre un rango temporal suficientemente amplio para incluir tanto eventos puntuales como tendencias sostenidas.
* Se alinea con la lógica de actualización trimestral de muchos scores ESG oficiales, facilitando comparaciones posteriores. En el caso de este trabajo, los scores ESG son los registrados en la SEC cada trimestre.

ii.Generación de ventanas con padding

Dado que en la práctica no todas las empresas publican noticias ESG diariamente, se implementa una estrategia de padding para ventanas incompletas. En concreto:

* Se permite que hasta un 20% de los días de la ventana (18 días) no tengan eventos reales.
* Cuando no existe evento para un día determinado, el embedding correspondiente se rellena con un vector de ceros.
* Si una empresa no dispone de al menos 72 días reales de eventos dentro de una ventana de 90 días, dicha ventana no se genera. Eliminar todas las ventanas con datos incompletos puede reducir drásticamente la cantidad de muestras disponibles, especialmente para empresas con baja cobertura mediática. La introducción de padding permite aprovechar la información disponible manteniendo la integridad estructural de la ventana, y al mismo tiempo ofrece al modelo la posibilidad de aprender a gestionar la ausencia de datos como una característica adicional.

### iii.Estadísticas y validaciones

Una vez generadas las ventanas, se realiza un análisis estadístico para validar la calidad del proceso. Entre las principales métricas evaluadas destacan:

* Número total de ventanas generadas: 63,401.
* Dimensión estándar de cada ventana: 90 días × 101 features.
* Número de ventanas que incluyen padding: 1,548.

Estas estadísticas confirman que la gran mayoría de las ventanas se generan a partir de datos completos, mientras que una fracción moderada incorpora padding, lo cual es coherente con la estrategia definida.

Ejemplo:

Ticker: TSLA

Category: Social

Target date: 2023-07-20

Primera fila del embedding de la ventana: [-0.03310936 -0.01596753 0.08599661 -0.02037817 0.0988157 -0.08739669

-0.19336331 -0.03057285 -0.00702647 -0.21213117]

### iv.Exportación de resultados

El dataset resultante se exporta en formato .pkl bajo el nombre WINDOWS.pkl. Este archivo contiene las siguientes estructuras:

* X: lista de matrices de 90×101 correspondientes a las ventanas.
* y: etiquetas o valores target asociados a cada ventana.
* tickers: lista de identificadores de empresas.
* categories: lista de categorías ESG (Environmental, Social, Governance).
* target\_dates: lista de fechas objetivo.

La exportación garantiza la persistencia del dataset en un formato eficiente y reutilizable para su uso inmediato en la fase de modelado predictivo. Este archivo servirá como input directo para la fase de construcción y entrenamiento del modelo LSTM, asegurando una organización eficiente y un acceso rápido a los datos durante los procesos de batch training.

En conclusión, la generación de ventanas temporales representa un paso crucial en el pipeline de modelización, permitiendo estructurar la información de eventos ESG de manera que el modelo LSTM pudiera aprender las dinámicas subyacentes a lo largo del tiempo. La combinación de una longitud de ventana adecuada, un mecanismo flexible de padding y una agrupación coherente por empresa y categoría sentó las bases para una predicción secuencial robusta y realista en el dominio ESG.

**c.Asociación de ventanas**

Notebook: 14\_MATCHING\_WINDOWS.ipynb

Una vez generadas las ventanas temporales de eventos ESG a partir de embeddings diarios, es necesario asociar cada ventana a un valor objetivo (target) que permita entrenar el modelo LSTM. Este valor objetivo correspondía al score ESG oficial asignado a la empresa en una fecha cercana a la fecha final de cada ventana.

i.Fundamento teórico

En los modelos de aprendizaje supervisado, la calidad de las etiquetas es crítica para asegurar la fiabilidad del entrenamiento y la capacidad de generalización del modelo. En este proyecto, las etiquetas son los scores ESG oficiales, que reflejan la valoración medioambiental, social y de gobernanza de cada empresa en fechas determinadas.

Dado que las ventanas se generan en función de las noticias y no siempre coinciden exactamente con las fechas de los scores ESG trimestrales, se aplica un proceso de matching tolerante. Este proceso busca la mejor coincidencia posible entre la fecha objetivo de la ventana y las fechas disponibles en la serie de scores, aceptando un margen máximo de 90 días. Este enfoque es esencial para maximizar la cantidad de datos utilizables y evitar descartar ventanas valiosas debido a pequeños desajustes temporales.

ii.Preparación de los datos

Se trabaja con los siguientes archivos:

* Los embeddings generados por día (EMBEDDINGS.pkl),
* Las ventanas de 90 días (WINDOWS.pkl),
* Los scores ESG oficiales (ESG\_SCORES.csv).

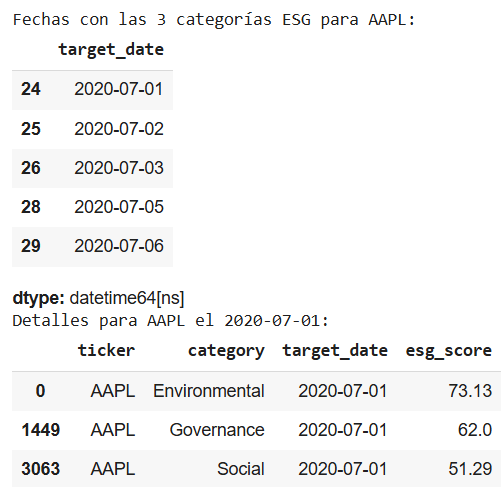
iii.Algoritmo de emparejamiento tolerante

El emparejamiento se implementa de la siguiente forma:

1. Para cada ventana, se identifican las variables clave:ticker, category y target\_date.
2. Se busca en el DataFrame de scores ESG una coincidencia exacta por ticker, category y date. Si esta no existe, se busca el score más próximo en el tiempo, siempre que la diferencia entre la fecha del score y la fecha de la ventana no supere un máximo de 90 días.
3. Si se encuentra un score válido (ya sea exacto o tolerante), se asigna como target (y) a la ventana correspondiente.
4. Si no se encuentra un score dentro del margen de tolerancia, la ventana se descarta para evitar etiquetas poco fiables.

Los scores ESG suelen actualizarse trimestralmente, por lo que puede existir un desfase natural entre las noticias diarias y las calificaciones oficiales. Permitir un margen de hasta 90 días asegura la coherencia temporal sin perder una cantidad significativa de datos, maximizando el uso del dataset disponible para entrenar modelos robustos.

Ejemplo:



### iv.Validación y exportación

Al finalizar el emparejamiento, se filtran todas las ventanas que no han conseguido asociarse a un score ESG válido, eliminando las entradas incompletas del dataset. Se realizan inspecciones aleatorias para validar que las correspondencias son correctas y coherentes (por ejemplo, revisando tickers, categorías y fechas emparejadas).

Finalmente, se exporta el dataset completo en formato .pkl bajo el nombre ESG\_TRAIN\_READY\_2.pkl. Este archivo incluye:

* X: las matrices correspondientes a las ventanas temporales (90 días × nº de features).
* y: los scores ESG emparejados (uno por ventana).
* tickers: identificadores de empresa.
* categories: categorias ESG (E, S o G).
* target\_dates: fechas objetivo de las ventanas.

Este dataset constituye la base definitiva para el entrenamiento del modelo de predicción, permitiendo realizar validaciones y backtests con datos reales y emparejados correctamente.

El emparejamiento preciso entre ventanas y scores ESG asegura que el modelo predictivo trabaje con datos consistentes y etiquetas verificadas. La estrategia de matching tolerante equilibra la necesidad de precisión temporal con la practicidad de maximizar el volumen de datos disponibles, un aspecto clave para robustecer el aprendizaje supervisado. Con este paso, se cierra la fase de preprocesamiento y se dispone de un dataset limpio, completo y estructurado para proceder al entrenamiento y evaluación del modelo.

**d.Entrenamiento**

Notebook: 15\_LSTM\_TRAINNING.ipynb

Esta fase tiene como propósito entrenar un modelo de red neuronal recurrente del tipo Long Short-Term Memory (LSTM), cuyo objetivo es predecir los scores ESG diarios a partir de las ventanas temporales generadas previamente. La arquitectura LSTM se selecciona por su capacidad para capturar patrones temporales complejos y gestionar dependencias a largo plazo en secuencias, algo crucial en la evolución mediática de los indicadores ESG.

### i.Carga y preparación de los datos

### El modelo se entrena sobre el dataset ESG\_TRAIN\_READY\_2.pkl, que contiene:

* X: ventanas temporales de 90 días × nº de features. Cada ventana contiene las características diarias (embeddings + cluster).
* y: scores ESG oficiales asociados a cada ventana.
* tickers, categories, target\_dates: metadatos para trazabilidad.

Antes de la alimentación al modelo, se aplica una normalización Min-Max a los datos, asegurando que todos los valores de entrada se encuentren en un rango [0, 1]. Esta técnica es clave para acelerar la convergencia del modelo y evitar problemas de escalado entre variables.

Se realiza una división train-test del 70%-30% para evaluar la capacidad de generalización del modelo, asegurando que no se produzcan fugas de información entre los conjuntos. La separación se ha realizado usando ‘train\_test\_split’, estratificando aleatoriamente para mantener la representatividad de la distribución de los scores ESG en cada subconjunto. Para facilitar la carga eficiente de datos, se ha definido la clase ESGDataset, que convierte las matrices de embeddings y los scores en tensores de PyTorch, con soporte para dispositivos CPU y GPU.

### ii.Diseño de la arquitectura LSTM

El modelo construido, denominado ESG\_LSTM, se diseñó para capturar eficazmente la dinámica temporal y semántica de las noticias ESG. La arquitectura se compone de varios bloques funcionales:

* Entrada: Cada input tiene forma (batch\_size, 90, 101), donde 90 representa los días consecutivos y 101 las características diarias (embeddings + cluster).
* LSTM: se implementó una LSTM con las siguientes caracteristicas:   
  + Dos capas (num\_layers=2), lo que incrementa la capacidad del modelo para capturar dependencias jerárquicas en la secuencia.
  + 128 unidades ocultas por capa (hidden\_dim=128),
  + Configuración bidireccional: esta configuración permite que la red procese la secuencia tanto en dirección temporal directa (pasado → futuro) como en dirección inversa (futuro → pasado), capturando así patrones que podrían manifestarse en ambos sentidos.
* Dropout:  
  + Se aplicó una regularización mediante Dropout(0.3) tras la salida de la LSTM, reduciendo así el riesgo de overfitting.
* Feedforward:  
  + Capa densa intermedia (Linear(256, 128)) seguida de una función de activación ReLU,
  + Capa final de salida (Linear(128, 1)) para producir un único valor de score ESG continuo.

Esta arquitectura se diseña para maximizar la capacidad del modelo de capturar la dinámica de eventos ESG a lo largo del tiempo, incorporando redundancia mediante bidireccionalidad y técnicas de regularización para mejorar su generalización.

### iii.Configuración de entrenamiento

El entrenamiento se ejecuta durante múltiples épocas (ajustable según los resultados obtenidos), recorriendo el dataset mediante DataLoaders que permiten optimizar la carga en batches. En cada iteración se registran la pérdida de entrenamiento y validación para monitorizar el proceso y evitar el sobreajuste.

* Función de pérdida: MSELoss (error cuadrático medio), que es la opción estándar para problemas de regresión continua. Esta función penaliza de manera cuadrática las desviaciones entre la predicción y el valor real, siendo especialmente eficaz para evitar grandes errores.
* Optimizador: Adam con tasa de aprendizaje 1e-3, que proporciona adaptación automática de la tasa de aprendizaje por parámetro.
* Número de épocas: 20.
* Batch size: 64.
* Dispositivo: GPU (cuda) cuando estuvo disponible.

Se fijan semillas aleatorias en NumPy y PyTorch (SEED=33) para garantizar la reproducibilidad de los resultados. Además, se configuró PyTorch en modo determinista para evitar variaciones entre ejecuciones.

Durante cada época:

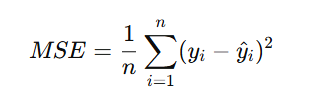
* El modelo se entrena sobre el conjunto de entrenamiento (train\_loader).
* Tras cada época, se evalúa sobre el conjunto de validación (val\_loader).
* Se calcula la pérdida media en entrenamiento y validación.
* Si la validación mejora respecto a la mejor anterior (best\_val\_loss), se guarda el modelo en disco (best\_model.pth).

Este procedimiento replica un esquema de early stopping manual, buscando preservar la generalización sin necesidad de recorrer un número excesivo de épocas.

### iv.Evaluación

Una vez finalizado el entrenamiento, se ha cargado el mejor modelo y se ha evaluado sobre el conjunto de test utilizando las métricas: Error cuadrático medio (MSE) y R 2 Score.

El MSE es una métrica estándar en problemas de regresión que mide la media de las diferencias elevadas al cuadrado entre las predicciones y los valores reales. Matemáticamente, se expresa como:



Un MSE de 0.85 indica que, en promedio, las predicciones del modelo tienen un error cuadrático muy bajo respecto a los scores ESG reales. Aunque el MSE penaliza más los errores grandes (por su naturaleza cuadrática), el resultado obtenido confirma que el modelo no comete grandes desviaciones y que su precisión es consistente incluso en casos extremos.  
  
 En un contexto de predicción ESG, un error bajo implica que las puntuaciones generadas son muy cercanas a las evaluaciones oficiales, lo que refuerza la fiabilidad del modelo como herramienta de apoyo para estrategias sostenibles e inversión responsable.

El R² Score, también conocido como coeficiente de determinación, mide la proporción de la varianza en la variable dependiente que es explicable por las variables independientes. Se calcula como:



donde:

* SSres​ es la suma de los residuos cuadrados,
* SStot​ es la suma total de las desviaciones respecto a la media.

Un R² de 0.986 implica que el 98.6% de la varianza observada en los scores ESG oficiales está siendo explicada por el modelo. Este valor está muy próximo al valor máximo posible (1.0), lo que confirma que el modelo captura con altísima precisión las relaciones y patrones temporales presentes en los datos.  
  
 El hecho de que casi la totalidad de la varianza esté explicada significa que el modelo no solo ajusta bien los datos conocidos, sino que tiene un alto poder predictivo sobre nuevos datos. En aplicaciones financieras, esto es especialmente importante porque sugiere que el modelo puede ser utilizado para monitorizar y anticipar dinámicas ESG de forma robusta.

Test MSE (real): 0.85

Test R² (real): 0.986

Sample of ESG predictions vs. actual scores on the test set:

Predicted: 66.52 | Actual: 66.14 | Absolute Error: 0.38

Predicted: 57.65 | Actual: 56.87 | Absolute Error: 0.78

Predicted: 47.30 | Actual: 45.81 | Absolute Error: 1.49

Predicted: 72.68 | Actual: 73.46 | Absolute Error: 0.78

Predicted: 66.49 | Actual: 66.45 | Absolute Error: 0.04

Predicted: 61.56 | Actual: 60.77 | Absolute Error: 0.79

Predicted: 50.59 | Actual: 50.74 | Absolute Error: 0.15

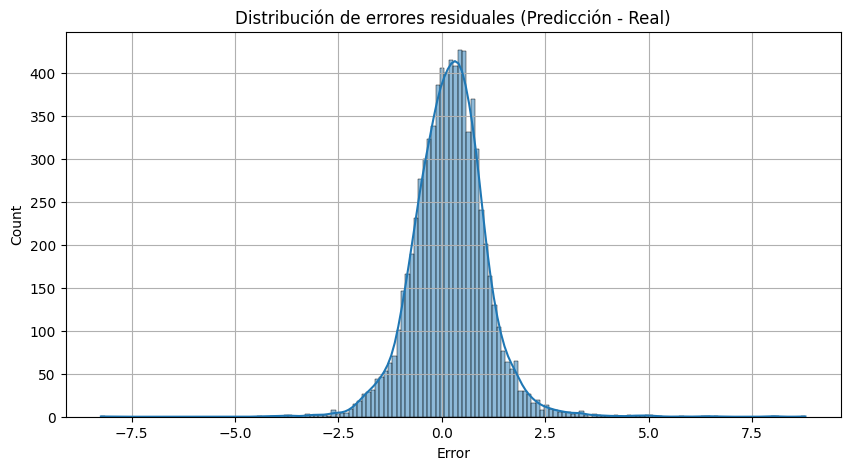
Predicted: 54.30 | Actual: 54.40 | Absolute Error: 0.10

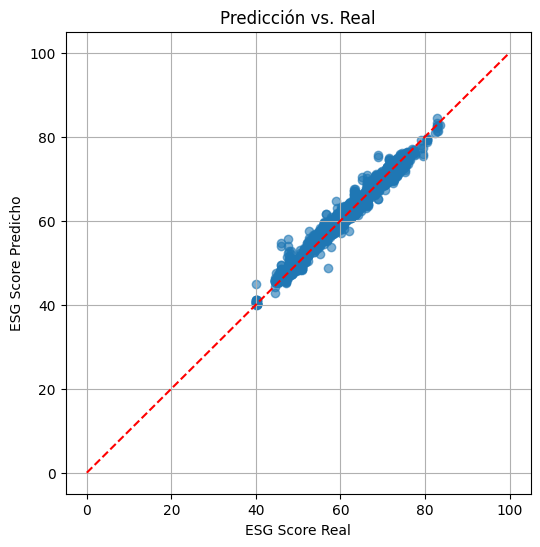
Predicted: 65.57 | Actual: 66.54 | Absolute Error: 0.97

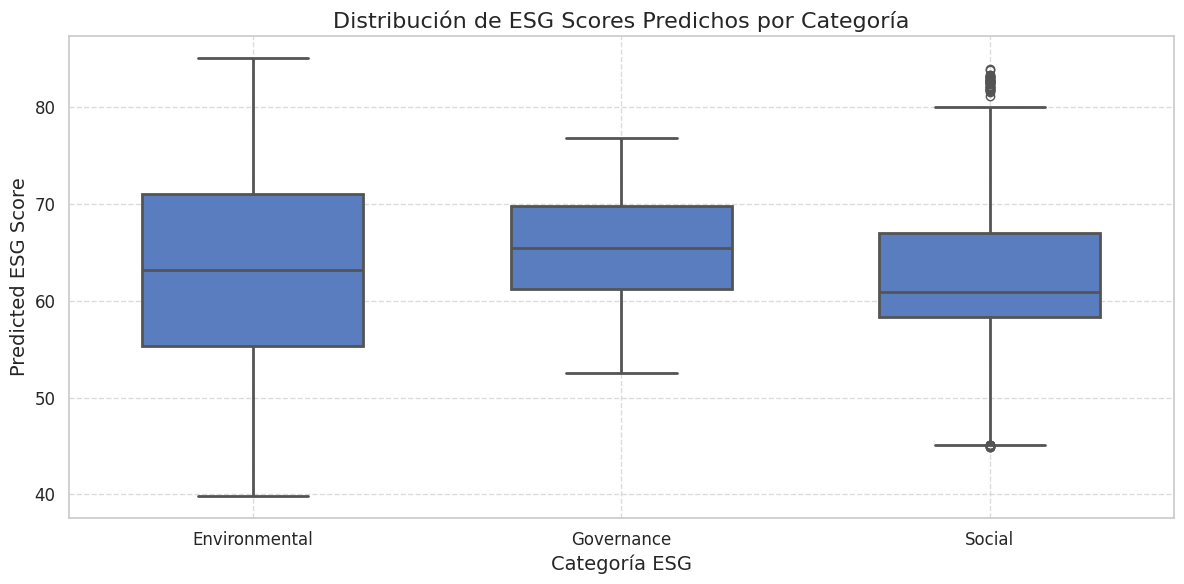
Predicted: 55.18 | Actual: 54.78 | Absolute Error: 0.40

Máx predicho: 84.41

Mín Predicho:39.89







Finalizada la evaluación, el modelo se aplicó a todas las ventanas disponibles, incluyendo aquellas ventanas para las cuales no existía score oficial, generando un df que recoge:

* ticker: Empresa asociada,
* category: Dimensión ESG (Environmental, Social, Governance),
* target\_date: Fecha correspondiente,
* predicted\_score: Score ESG predicho.

**11. CREACIÓN DE UN DATASET COMPLETO**

Notebook: 16\_COMPLETE\_DATASET.ipynb

La fase de inferencia da lugar a un conjunto de predicciones, que sumados a los scores ESG oficiales, forman un dataset casi completo, para el cual llevamos a cabo una serie de acciones que explico a continuación.

**a.Fundamento teórico**

En la práctica, los scores ESG oficiales suelen publicarse con periodicidad trimestral o anual, lo que deja lagunas temporales significativas en la cobertura de datos. Para suplir esta limitación, se ha diseñado un sistema donde las predicciones diarias generadas por el modelo LSTM rellenan estas lagunas, asegurando una serie temporal continua. Este enfoque es fundamental para ofrecer una visión actualizada y granular de la evolución ESG, algo cada vez más demandado en el ámbito financiero y de sostenibilidad.

El dataset final incluye una distinción explícita entre datos provenientes de fuentes oficiales y aquellos generados mediante predicción, garantizando la transparencia y trazabilidad de los datos.

### **b.Carga y preparación de los datos**

Se cargaron los siguientes conjuntos de datos:

* NEWS\_+\_PRESS\_ESG\_definitivo.csv: Dataset de noticias y comunicados procesados.
* DF\_SUMMARY.csv: Resumen estadístico de cada empresa.
* EMBEDDINGS.pkl: Embeddings diarios de eventos ESG por empresa y categoría.
* DAILY\_ESG\_PREDICTIONS.csv: Predicciones de scores ESG basadas en ventanas de 90 días.
* ESG\_SCORES.csv: Scores ESG oficiales históricos de las empresas analizadas.

### **c.Generación de nuevas ventanas de 7 días**

Para complementar las predicciones basadas en ventanas de 90 días, se generan también ventanas de 7 días de embeddings:

* Se utiliza la función ‘generate\_windows\_with\_padding’ para crear ventanas de 7 días con hasta un 20% de padding permitido (es decir, al menos 6 días reales de datos).
* Para cada ventana, se busca el score predicho más cercano dentro de un rango de ±3 días respecto al final de la ventana.
* Este proceso permite aumentar la granularidad temporal de las predicciones, mejorando la cobertura diaria de scores disponibles.

Se crea un conjunto de datos X\_lstm con las ventanas de 7 días, y sus correspondientes etiquetas y\_lstm obtenidas a partir de las predicciones anteriores.

### **d.Entrenamiento de un nuevo modelo LSTM**

Para estas nuevas ventanas de 7 días, se entrenó un modelo LSTM adicional que sigue la misma estructura del anterior. Este modelo dio los siguientes resultados sobre el conjunto de test:

Test MSE: 0.4569

Test R²: 0.9917

### Una vez disponibles las predicciones basadas en ventanas de 90 días y las generadas con ventanas de 7 días, se unifican ambas fuentes para maximizar la cobertura temporal.

### **e.Filtrado de empresas y categorías**

Para asegurar que el dataset final tuviera una calidad adecuada, se aplicaron criterios de filtrado:

* Se exigió que cada empresa tuviera al menos 100 días de predicciones disponibles por categoría ESG.
* Además, cada empresa debía disponer de al menos dos categorías ESG con cobertura mínima.

### **f.Interpolación de scores oficiales y predicciones**

Dado que no todas las fechas contaban con predicciones o scores oficiales, se diseñó un procedimiento específico para construir series temporales completas:

* Antes de la primera predicción disponible para una empresa y categoría:  
  + Se utilizaron los scores oficiales (ESG\_SCORES.csv),
  + Aplicando interpolación forward-fill (ffill) para días sin valor explícito.
* A partir de la primera predicción:  
  + Se utilizaron las predicciones del modelo LSTM.
  + Si en una fecha dada existía una predicción, se utilizaba la predicción.
  + Si no existía predicción para un día concreto tras el primer evento, se mantenía el último valor predicho.
* Para cada fecha y empresa-categoría, se registró si el valor provenía de una predicción (from\_official = False) o de un score oficial (from\_official = True).

Este procedimiento garantiza que cada empresa tenga un score ESG diario ininterrumpido a lo largo del periodo 2020–2025. La función fill\_esg\_scores\_filtered implemente este mecanismo, generando un dataframe final df\_esg\_final\_5y.

### **g.Dataset final generado**

El dataset final contiene:

* Ticker: Identificador de la empresa.
* Fecha: Día correspondiente.
* Categoría ESG: Environmental, Social o Governance.
* Score ESG diario (predicted\_score\_final): Predicción o valor interpolado.
* Indicador de origen (from\_official): True si el valor proviene de score oficial, False si proviene de predicción LSTM.

El conjunto de datos fue almacenado en ESG\_SCORES\_COMPLETOS\_2.csv, sirviendo como base definitiva para los análisis posteriores y para el diseño de estrategias de inversión basadas en la evolución de factores ESG.

Notebook: 17\_COMPLETE\_DATASET\_2.ipynb

Por último, se procede a asociar cada ventana temporal generada (con un tamaño de 7 días) con su correspondiente score ESG objetivo, garantizando así la correcta alineación entre las secuencias de entrada y los valores de salida para el entrenamiento supervisado. Para ello, se implementa un procedimiento de emparejamiento que, por cada ventana, busca el score ESG más cercano en el tiempo dentro de un margen de tolerancia de 3 días. Este margen ha sido definido para permitir una ligera flexibilidad temporal que mitigue posibles desajustes en las fechas de publicación o en la disponibilidad de predicciones, sin comprometer la precisión del emparejamiento. En caso de existir múltiples candidatos dentro de ese rango temporal, se selecciona el que presenta la menor distancia temporal respecto a la ventana analizada.

Este paso resulta fundamental para asegurar que cada secuencia de entrada esté emparejada de forma precisa con su respectivo target, lo que optimiza la calidad y la fiabilidad del aprendizaje supervisado en fases posteriores. Además, este procedimiento fortalece la coherencia interna del dataset al minimizar el riesgo de errores derivados de desfases temporales, un aspecto crítico cuando se trabaja con series temporales y datos financieros sensibles al contexto temporal.

# **12.DESARROLLO DE UN ALGORITMO DE INVERSIÓN**

Notebook: 18\_TRADING\_ALGO.ipynb

En esta sección se describe en profundidad el diseño, construcción, implementación y validación de un sistema de trading basado en la evolución de scores ESG dinámicos y eventos de noticias clasificadas. El objetivo principal fue desarrollar un enfoque de inversión sistemático que, basándose en señales extraídas del comportamiento ESG de las empresas, pudiera tomar decisiones de compra y venta con criterios objetivos, replicables y cuantificables.

## **a.Carga y preparación de los datos**

Para alimentar el sistema de trading, se utilizan tres conjuntos de datos principales:

* Noticias ESG (news): Incluía eventos clasificados como ambientales, sociales o de gobernanza, junto con un análisis de sentimiento.
* Precios de mercado (prices): Información diaria de precios de cierre, volumen y otros indicadores financieros de las empresas analizadas.
* Scores ESG diarios (esg\_scores): Dataset generado previamente que combinaba predicciones realizadas mediante redes LSTM con interpolaciones de scores oficiales históricos.

Todos los datasets han de ser correctamente alineados en el tiempo y transformados al formato datetime para garantizar operaciones consistentes y seguras durante la ejecución del algoritmo.

## **b.Diseño modular del sistema de trading**

El sistema se estructuró de manera modular, utilizando tres componentes principales:

### Generación de señales (SignalEngine)

La clase SignalEngine constituye el primer bloque funcional del sistema de trading y tiene como objetivo generar señales cuantitativas y cualitativas basadas en la evolución diaria de los scores ESG y en los eventos noticiosos clasificados para cada empresa. Este módulo es crucial porque transforma datos en indicadores operativos objetivos sobre los que se basan las decisiones de inversión posteriores.

El constructor de la clase (\_\_init\_\_) recibe como argumentos el dataset de scores ESG diarios, el dataset de noticias clasificadas y la categoría ESG sobre la cual operar (Environmental, Social o Governance). Adicionalmente, permite ajustar varios parámetros técnicos esenciales para el cálculo de señales, como la ventana de momentum (5 días por defecto), la ventana de Z-Score (20 días) y la ventana de percentiles para la clasificación de nivel ESG (90 días). La inclusión de estos parámetros asegura la flexibilidad del sistema, permitiendo adaptarlo fácilmente a distintos horizontes temporales o sectores.

El método central de la clase es get\_signals\_for\_day(date). Este método procesa la información correspondiente a una fecha concreta y construye, para cada empresa, un conjunto de señales cuantitativas y cualitativas. La primera parte del análisis se centra en los scores ESG. Se calculan cuatro indicadores clave: el momentum, que mide la variación del score respecto a una ventana móvil reciente y capta la tendencia de la empresa en términos ESG; el Z-Score, que evalúa cuán extremo es el score actual en relación con su media y desviación estándar histórica; y los percentiles móviles (P30 y P70), que clasifican cada score en niveles bajo, medio o alto en función de su posición relativa en la distribución histórica. Este último punto es fundamental para identificar situaciones estructurales como empresas persistentemente rezagadas o destacadas en sostenibilidad.

En paralelo, las noticias del día son analizadas y etiquetadas en función de su sentimiento predicho (bullish, bearish, neutral), transformándolas en categorías cualitativas interpretables: ‘muy\_positivo’, ‘muy\_negativo’ o ‘neutral’. Esta clasificación permite capturar el tono emocional o el impacto potencial que las noticias pueden tener sobre la percepción ESG de las empresas.

El resultado final es un diccionario estructurado en el que cada empresa (ticker) aparece asociada a un conjunto de señales clave: evento\_extremo, desacoplamiento, momentum, zscore y la categoría ESG correspondiente. Este formato permite la integración directa de las señales en la siguiente fase del sistema, facilitando la evaluación y ejecución de decisiones de trading.

En definitiva, SignalEngine cumple un papel estratégico dentro del sistema global, actuando como el cerebro que transforma datos desestructurados en indicadores operativos de alta calidad. Su diseño modular y flexible asegura que puede aplicarse simultáneamente a distintas categorías ESG o ampliarse para incorporar nuevas lógicas de señal en el futuro. Además, la combinación de técnicas estadísticas (como percentiles y Z-Score) con la interpretación semántica de las noticias ofrece un enfoque robusto y sofisticado que busca minimizar el sesgo y maximizar la relevancia de las señales generadas.

### Evaluación de señales (SignalEvaluator)

La clase SignalEvaluator representa la segunda pieza fundamental dentro del sistema modular de trading basado en señales ESG. Su función principal es traducir las señales individuales generadas por el módulo SignalEngine en una única señal agregada y, a partir de esta, determinar la decisión operativa concreta: comprar, vender o mantener. Este proceso es clave para la operatividad del sistema, ya que convierte múltiples métricas dispersas en una única directriz clara y cuantificable, apta para su ejecución automatizada.

El constructor de la clase (\_\_init\_\_) permite personalizar los pesos asignados a cada tipo de señal, distinguiendo entre las tres categorías ESG: Environmental, Social y Governance. Por defecto, la ponderación está definida de manera equilibrada para las tres categorías, dando mayor importancia al evento extremo (40%), seguido del desacoplamiento (30%), el momentum (20%) y finalmente el Z-Score (10%). Esta configuración inicial responde a la lógica de que las noticias muy positivas o muy negativas suelen tener un impacto más inmediato y determinante en la percepción ESG de una empresa, mientras que las señales cuantitativas como momentum o Z-Score aportan matices adicionales a la evaluación. Sin embargo, la arquitectura de la clase está diseñada para ser fácilmente ajustable, permitiendo adaptar estos pesos en función de backtests o preferencias específicas del usuario.

El método calcular\_signal\_final es el núcleo funcional de la clase. Para cada fila (empresa/día) del DataFrame recibido, este método extrae la categoría ESG asociada y aplica la combinación ponderada de señales según los pesos definidos. Cada tipo de señal (evento extremo, desacoplamiento, momentum y Z-Score) se multiplica por su peso correspondiente, sumándose de forma acumulativa para obtener una señal final única. Este proceso garantiza que la evaluación sea consistente y estructurada, integrando tanto la información cualitativa como la cuantitativa en un solo valor continuo.

Posteriormente, el método decision\_operativa traduce esta señal final en una decisión operativa discreta. La lógica empleada es sencilla pero efectiva: si la señal final supera un umbral positivo (por defecto, +0.25), se genera una señal de compra (1); si cae por debajo de un umbral negativo (–0.25), se emite una señal de venta (–1); y si se mantiene dentro del rango intermedio, la recomendación es mantener la posición actual (0). Esta política basada en umbrales permite filtrar señales débiles o ruidosas y centrarse solo en aquellas situaciones donde la confianza en la señal es suficientemente alta como para justificar una acción.

En resumen, SignalEvaluator actúa como el órgano decisor del sistema de trading, sintetizando las múltiples señales ESG en una directriz operativa clara y calibrada.

### Ejecución de órdenes (SignalExecutor)

La clase SignalExecutor constituye el tercer y último pilar del sistema modular de trading, siendo la encargada de materializar las decisiones de trading en operaciones concretas de compra y venta, simulando la ejecución realista de órdenes en el mercado financiero. Su diseño está orientado a reproducir con la máxima fidelidad las condiciones reales de mercado, incluyendo consideraciones como comisiones, deslizamiento (slippage) y la gestión activa de la cartera.

Desde un punto de vista funcional, el constructor (\_\_init\_\_) establece los parámetros clave de la simulación, como el capital inicial disponible (por defecto 1 millón de euros), la exposición máxima por activo (25% del capital), y los costes asociados a las operaciones: comisiones (0.1%), stop-loss (30%) y slippage (0.1%). Estos valores se han definido para reflejar escenarios realistas de inversión institucional, aunque su diseño permite personalización completa para adaptarse a diferentes perfiles de riesgo o contextos operativos.

El atributo portfolio mantiene el registro actualizado de las posiciones abiertas para cada activo, mientras que buy\_prices almacena los precios de compra, cruciales para calcular correctamente la activación de stop-loss. Además, se registra cada operación en la lista trades, y la evolución histórica de la cartera se almacena en history. La clase también mantiene la variable pending\_signals, que guarda las señales generadas hoy para ser ejecutadas en la siguiente sesión, simulando así el retraso natural entre la generación de señales y su ejecución real.

La función \_execute\_order implementa la lógica operativa para cada orden individual. Cuando se detecta una señal de compra (signal == 1), el algoritmo calcula el monto a invertir en función del capital disponible y la exposición máxima por activo, ajustando este monto mediante el "score" de confianza de la señal (esto introduce un matiz dinámico muy relevante). El número de acciones se calcula de forma entera (sin fracciones), y se verifica si existe suficiente capital disponible para cubrir la operación incluyendo comisiones. Si las condiciones son favorables, se realiza la compra y se actualizan las posiciones y el capital disponible. En caso de señal de venta (signal == -1), el sistema implementa una venta parcial aleatoria (entre 60% y 80% de la posición), reflejando la realidad operativa donde raramente se liquida la posición completa salvo fuerza mayor. También se contemplan las ventas forzadas por stop-loss, para proteger la cartera frente a caídas abruptas de precio.

El método update\_daily representa el núcleo operativo de la clase y simula el ciclo diario completo de trading. En primer lugar, ejecuta las señales pendientes que quedaron registradas el día anterior, aplicando el slippage correspondiente para aproximarse al precio real de ejecución. Posteriormente, almacena las nuevas señales generadas ese mismo día para ser ejecutadas en la siguiente sesión, simulando así un flujo realista de órdenes diferidas. A continuación, revisa diariamente todas las posiciones abiertas para verificar si alguna ha alcanzado el umbral de stop-loss, en cuyo caso ejecuta la venta automáticamente. Finalmente, actualiza el valor total de la cartera (sumando efectivo disponible más el valor de mercado de las posiciones abiertas) y lo registra en el historial.

Una característica destacada es la precisión con la que se calcula el valor de la cartera: para cada posición abierta, se consulta el precio de cierre del activo en el mercado ese día y se multiplica por la cantidad de acciones en cartera, permitiendo así monitorizar la evolución diaria del patrimonio.

Además, la clase incluye los métodos get\_history\_df y get\_trades\_df, que facilitan la exportación estructurada de los resultados operativos. get\_trades\_df añade información valiosa como el cálculo bruto y neto de cada operación (considerando comisiones), lo que permite realizar análisis financieros detallados como el cálculo del ratio de rotación de la cartera, el coste total de fricción o la rentabilidad ajustada por costes operativos.

En conjunto, SignalExecutor actúa como el motor transaccional del sistema, transformando señales en acciones concretas y gestionando de manera dinámica la cartera a lo largo del tiempo. Su diseño modular y su atención al detalle (slippage, stop-loss, exposición máxima, ejecución diferida) proporcionan una simulación robusta y realista que permite evaluar la viabilidad operativa y la rentabilidad de la estrategia ESG bajo condiciones de mercado plausibles. Además, al mantener un registro exhaustivo de todas las operaciones y su evolución histórica, esta clase sienta las bases para un análisis post-trading riguroso, imprescindible en cualquier proceso profesional de backtesting y optimización algorítmica.

1. Función para analizar resultados

La función analizar\_resultados representa el módulo de análisis post-operativo del sistema de trading, cuyo objetivo es evaluar de manera exhaustiva el desempeño de la estrategia implementada. Esta función sintetiza en un solo bloque toda la lógica necesaria para calcular tanto métricas financieras estándar (como rentabilidad y riesgo) como indicadores específicos sobre las operaciones realizadas (como win rate o factor beneficio), permitiendo una evaluación rigurosa y cuantificable de los resultados obtenidos.

Desde un punto de vista técnico, la función recibe varios datasets clave: el historial diario de la cartera (df\_history), el registro detallado de operaciones (df\_trades), y opcionalmente los precios diarios (df\_prices). La primera operación fundamental es asegurar la correcta alineación temporal: si se proporciona un dataset de precios, se filtran los registros históricos para conservar únicamente las fechas en las que hubo precios válidos, evitando así distorsiones debidas a días sin actividad de mercado.

Una vez limpiados y ordenados los datos, la función calcula la rentabilidad diaria de la cartera (columna returns) mediante el cambio porcentual del capital día a día. Esto constituye la base para derivar métricas clave como:

* Total Return: la rentabilidad total acumulada durante todo el periodo de backtest.
* Annualized Return: la tasa anualizada de retorno, que normaliza la rentabilidad al periodo de un año, permitiendo comparaciones directas con benchmarks.
* Volatilidad anualizada: un indicador de riesgo calculado como la desviación estándar de los retornos diarios escalada por la raíz de 252 (días hábiles anuales).
* Sharpe Ratio: una medida clásica de rendimiento ajustado al riesgo que compara el exceso de rentabilidad sobre la tasa libre de riesgo con la volatilidad. Este ratio es especialmente relevante para entender si la estrategia genera valor más allá de la simple exposición al mercado.

Además, la función incorpora un cálculo exhaustivo de drawdown: mide la mayor caída porcentual desde un pico histórico en la evolución del capital. La métrica max\_drawdown cuantifica la peor caída sufrida, mientras que calmar\_ratio relaciona la rentabilidad anualizada con la magnitud del drawdown, proporcionando un balance riesgo/rentabilidad complementario al Sharpe Ratio.

Otro aspecto clave es el análisis microestructural de las operaciones. Para ello, la función empareja cada operación de venta con la última operación de compra correspondiente (matching por ticker y categoría ESG) y calcula el beneficio neto (pnl) así como la duración de la operación (holding\_days). A partir de estos pares compra-venta, se derivan estadísticas operativas fundamentales:

* Número total de operaciones y distribución BUY vs SELL.
* Win Rate: proporción de operaciones cerradas con beneficio positivo, una métrica clave para evaluar la efectividad de la estrategia.
* Profit Factor: relación entre ganancias totales y pérdidas totales, considerada una medida robusta de la eficiencia global de la estrategia.
* Expectancy: beneficio medio esperado por operación, que permite entender el valor promedio de cada trade individual.
* Risk/Reward Ratio: mide la ganancia media por operación ganadora frente a la pérdida media por operación perdedora, proporcionando información sobre la asimetría de la estrategia.

También se calculan otros indicadores útiles, como la ganancia media por operación y la duración media de la tenencia (holding period), permitiendo así entender el ciclo de vida típico de una inversión en el sistema.

Este módulo de análisis se convierte así en un pilar metodológico imprescindible para asegurar que la estrategia de trading basada en ESG no solo funcione en términos teóricos, sino que sea financieramente viable y profesionalmente defendible frente a cualquier auditoría o revisión por parte de inversores o entidades reguladoras.

1. Comparación con el Benchmark

El módulo comparar\_con\_benchmark constituye un componente para evaluar la robustez y efectividad de la estrategia de inversión desarrollada, al comparar su rendimiento con un benchmark pasivo de referencia. Este tipo de análisis es indispensable en estudios financieros ya que permite contextualizar el desempeño de la estrategia frente a una inversión alternativa que replica la evolución natural del mercado (comúnmente conocida como estrategia Buy & Hold).

La función recibe como entradas los datos históricos de capital acumulado (df\_history), las operaciones ejecutadas (df\_trades), y los precios de mercado (df\_prices).

A partir de los precios iniciales, la función construye una cartera benchmark simulada, donde el benchmark se inicializa como un portafolio equitativamente ponderado (peso igual para cada activo). La cantidad de acciones compradas de cada activo se calcula dividiendo la porción correspondiente del capital inicial entre su precio inicial, asegurando que la inversión inicial total coincide con el capital teórico disponible.

A nivel visual, la función traza un gráfico comparativo que muestra la evolución conjunta de ambos portafolios. Esta visualización es crítica para captar de un vistazo si la estrategia logra superar al benchmark y en qué momentos clave se producen divergencias. Además, el gráfico se complementa con anotaciones que indican el valor final alcanzado por cada línea, reforzando la interpretación cuantitativa del rendimiento.

A nivel cuantitativo, la función reporta:

* Retorno final de la estrategia y del benchmark, expresado como porcentaje.
* Alpha y Beta: dos métricas fundamentales de análisis financiero. El Beta mide la sensibilidad de la estrategia respecto a los movimientos del benchmark (riesgo sistemático), mientras que el Alpha cuantifica el rendimiento adicional de la estrategia una vez descontado este riesgo sistemático. Valores de Alpha positivos indican que la estrategia ha generado valor más allá del mercado, mientras que un Beta superior a 1 refleja mayor volatilidad relativa.
* Correlación: mide el grado de relación lineal entre la evolución diaria de la estrategia y del benchmark. Una alta correlación (>0.8) sugiere que ambos portafolios se mueven de forma similar, mientras que valores bajos reflejan un comportamiento más independiente.

## **c.Estrategia de inversión por categorías ESG**

Una decisión clave del diseño es ejecutar estrategias separadas para cada categoría ESG (Environmental, Social, Governance). Esto se debe a que este enfoque permite evaluar el impacto diferencial que cada dimensión ESG podía tener sobre los resultados de inversión ademas de facilitar la posterior combinación o ponderación de las estrategias de acuerdo a perfiles de riesgo o intereses particulares de los inversores.

**d.Backtest operativo**

Una vez definido e implementado el sistema de trading basado en señales ESG, se procedió a realizar un backtest exhaustivo sobre el periodo comprendido entre marzo de 2020 y marzo de 2025. Este análisis tuvo como objetivo evaluar el rendimiento histórico de la estrategia simulando condiciones realistas de ejecución diaria, con la finalidad de validar su aplicabilidad práctica en contextos de inversión automatizada.

Cada jornada, el sistema genera señales de trading basadas en los eventos y la evolución ESG, las evaluaba según criterios predefinidos, y tomaba decisiones de compra, mantenimiento o venta para cada activo. Las operaciones se han ejecutado teniendo en cuenta comisiones de mercado (0,1%) y slippage (0,1%), y se han incluido mecanismos realistas como límites de exposición por activo (25%) y stop-loss (30%). Esta configuración logra simular con precisión el comportamiento de una estrategia implementada en un fondo de inversión real.

## **e.Análisis de resultados**



\Métricas para ENVIRONMENTAL:

Total Return: 1.2226899280990753

Annualized Return: 0.173210054627853

Volatility: 0.18957134588793575

Sharpe Ratio: 0.8609426380521341

Max Drawdown: -0.26975992365391854

Max DD Date: 2022-10-11 00:00:00

Drawdown Duration (days): 113

Calmar Ratio: 0.6420896487577166

Número total de operaciones: 11468

BUY vs SELL: {'BUY': 8044, 'SELL': 3424}

Días activos: 1254

Commission %: 0.001

Slippage %: 0.001

Win Rate: 0.5084696261682243

Profit Factor: 1.0656534721612096

Expectancy por trade: 28.44319966998007

Risk/Reward Ratio: 1.0301520928473948

Ganancia media por operación: 28.44319966998007

Holding medio (días): 7.674065420560748

Este resultado muestra un comportamiento sólido y rentable, destacando una estrategia que logra mantener un balance positivo entre rentabilidad y riesgo. El Sharpe superior a 0,85 y el Calmar Ratio cercano a 0,65 reflejan una eficiencia ajustada al riesgo razonable.

\Métricas para SOCIAL:

Total Return: 0.985545201045998

Annualized Return: 0.14703311348875125

Volatility: 0.15621725796798316

Sharpe Ratio: 0.8771957418228162

Max Drawdown: -0.2817811638699258

Max DD Date: 2022-10-11 00:00:00

Drawdown Duration (days): 252

Calmar Ratio: 0.5217989430855777

Número total de operaciones: 12443

BUY vs SELL: {'BUY': 6745, 'SELL': 5698}

Días activos: 1254

Commission %: 0.001

Slippage %: 0.001

Win Rate: 0.47595647595647594

Profit Factor: 0.8756718499926665

Expectancy por trade: -37.78969806072054

Risk/Reward Ratio: 0.9641431209727528

Ganancia media por operación: -37.78969806072054

Holding medio (días): 8.572130572130572

A pesar de lograr una buena rentabilidad acumulada y anualizada, los resultados muestran ciertas debilidades, como un profit factor por debajo de 1 y una expectativa negativa por operación. Esto sugiere que la estrategia basada en la dimensión social puede estar generando señales más ruidosas o menos consistentes, lo cual podría beneficiarse de un refinamiento posterior.

\Métricas para GOVERNANCE:

Total Return: 0.7279604824945818

Annualized Return: 0.11559551916537192

Volatility: 0.17177063341052365

Sharpe Ratio: 0.6147472188275842

Max Drawdown: -0.26981078201634134

Max DD Date: 2022-09-26 00:00:00

Drawdown Duration (days): 123

Calmar Ratio: 0.42843180061785213

Número total de operaciones: 12983

BUY vs SELL: {'BUY': 6648, 'SELL': 6335}

Días activos: 1254

Commission %: 0.001

Slippage %: 0.001

Win Rate: 0.4677190213101815

Profit Factor: 0.9063582662763944

Expectancy por trade: -31.691922859507855

Risk/Reward Ratio: 1.0314681315842047

Ganancia media por operación: -31.691922859507855

Holding medio (días): 8.477821625887925

Los resultados de la categoría Governance presentan un perfil de riesgo mayor, con menor rentabilidad y ratios de eficiencia más bajos, tanto en términos de Sharpe como de Calmar. Aunque el número de operaciones fue elevado, el sistema no logró generar ganancias medias positivas, lo que apunta a una posible sobreoptimización o falta de especificidad en las señales ESG de gobernanza.

El análisis histórico permite concluir que la dimensión Environmental ofrece el mejor desempeño ajustado al riesgo en el periodo analizado, seguido por Social y, en último lugar, Governance. Estas diferencias reflejan no solo el comportamiento de los factores ESG en el mercado, sino también la calidad de las señales derivadas de noticias y scores.

**f.Comparativa de la Estrategia ESG frente al Benchmark**

Para evaluar la eficacia de la estrategia de trading basada en señales ESG, se ha realizado una comparativa directa contra un benchmark representativo (estrategia buy & hold) desde marzo de 2020 hasta marzo de 2025. Este benchmark replica la evolución de un portafolio ponderado de los mismos activos utilizados en la estrategia, permitiendo contrastar de forma objetiva los resultados obtenidos.

Las métricas clave por categoría ESG son las siguientes:

* Environmental: La estrategia obtuvo un retorno acumulado del 123.36%, frente a un 207.08% del benchmark. Aunque la estrategia logra más de duplicar el capital inicial, no consigue superar la rentabilidad de un simple buy & hold. El alpha diario (-0.00006) es negativo, y la beta (0.79) indica que la estrategia tiene una sensibilidad moderada frente al mercado. La alta correlación (0.88) muestra que sigue de cerca la tendencia general.
* Social: Aquí, la estrategia presenta un retorno del 97.98% versus el 207.08% del benchmark. La beta es más baja (0.60), lo que sugiere una menor exposición al riesgo de mercado, pero el alpha es casi nulo (-0.00000), confirmando que no aporta valor añadido respecto a la evolución pasiva del mercado. La correlación (0.82) sigue siendo alta pero algo más débil.
* Governance: La categoría más rezagada, con un retorno del 74.67% comparado con el benchmark. La beta es de 0.71 y la correlación (0.88) mantiene un comportamiento muy similar al benchmark. El alpha (-0.00020) es el peor de las tres categorías, lo que indica que esta estrategia, pese a mantener cierta direccionalidad, no logra capitalizar en exceso sobre la base pasiva.

Es importante destacar que, a diferencia del benchmark buy & hold, que mantiene el capital completamente invertido en todo momento, la estrategia desarrollada no permanece siempre totalmente expuesta al mercado. Esto se debe a que las decisiones de inversión se basan en la generación de señales diarias específicas, y además existen limitaciones deliberadas sobre la exposición máxima por activo (25% del capital disponible). Asimismo, en ausencia de señales o cuando se activan stop-loss, la estrategia puede encontrarse parcial o totalmente en liquidez.

Este enfoque, aunque ofrece una mayor protección en escenarios adversos y contribuye a una mejor gestión del riesgo (limitando drawdowns), también penaliza la capacidad de capturar todas las subidas del mercado en los ciclos alcistas. Por ello, los resultados muestran cómo la estrategia, a pesar de lograr rentabilidades positivas, no iguala la performance acumulada del benchmark, que se beneficia de estar 100% invertido en todo momento.

Este matiz es crucial para entender la lógica detrás del comportamiento observado: la estrategia está diseñada para actuar únicamente cuando detecta oportunidades basadas en criterios ESG robustos, lo cual introduce períodos naturales de menor exposición.

**g.Interpretación y conclusiones.**

A nivel gráfico, se observa que todas las categorías mantienen un comportamiento positivo a lo largo del tiempo, pero ninguna de ellas logra batir al benchmark de manera consistente. En particular, la estrategia Governance muestra la mayor debilidad, evidenciando que este tipo de señales podría no ser tan eficaz en captar valor añadido en los mercados actuales.

Las betas moderadas reflejan que la estrategia introduce ciertos mecanismos de control de riesgo (como stop-loss y límites de exposición), lo que ayuda a suavizar la volatilidad, aunque también limita el potencial alcista en periodos de fuerte tendencia.

La comparación gráfica ilustra claramente estas diferencias, con un crecimiento acumulado siempre por debajo del benchmark a partir de 2021, algo especialmente notorio en la categoría Governance.

## 

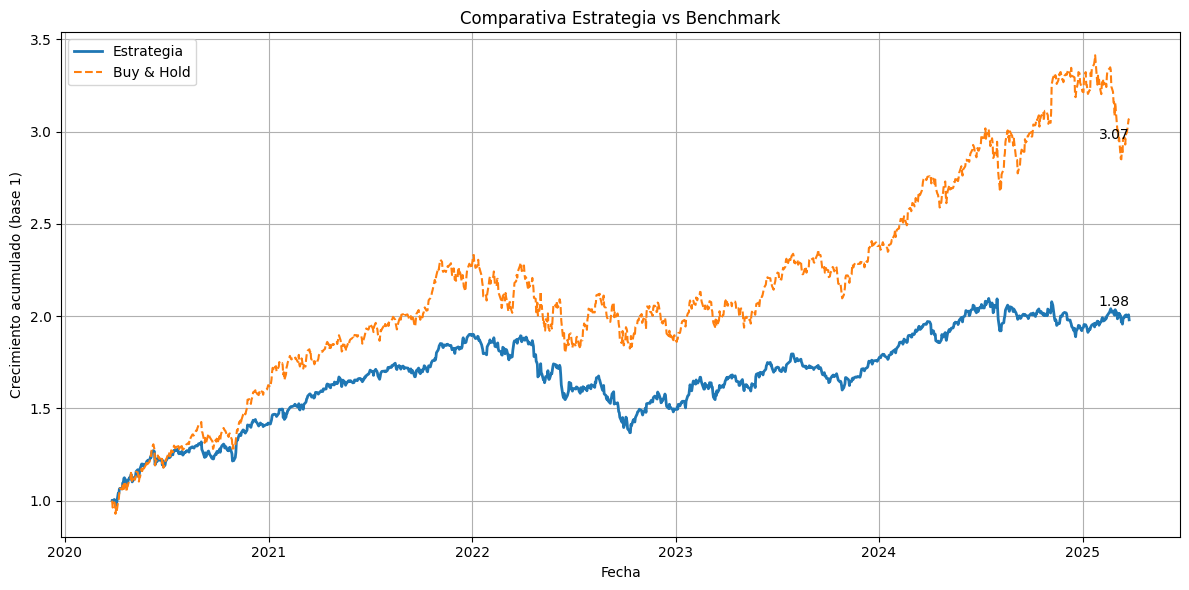
Retorno Estrategia: 123.36%

Retorno Benchmark: 207.08%

Alpha (diario): -0.00006

Beta: 0.79

Correlación: 0.88



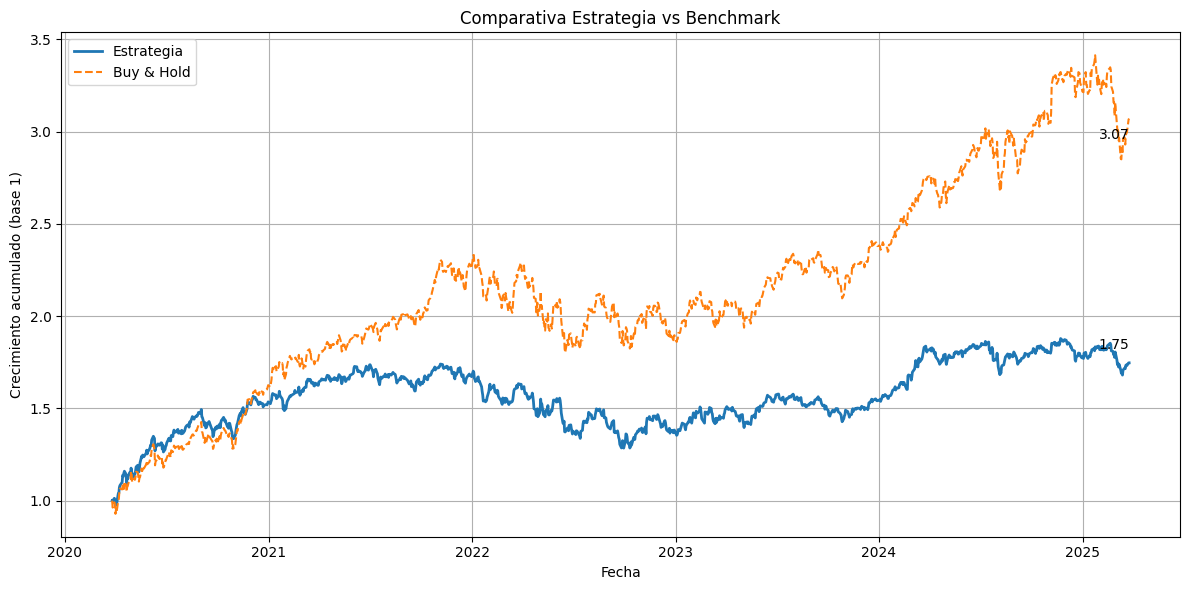
Retorno Estrategia: 97.98%

Retorno Benchmark: 207.08%

Alpha (diario): -0.00000

Beta: 0.60

Correlación: 0.82



Retorno Estrategia: 74.67%

Retorno Benchmark: 207.08%

Alpha (diario): -0.00020

Beta: 0.71

Correlación: 0.88

**h.Líneas de mejora y optimización de la exposición**

Una posible optimización futura de la estrategia consistiría en implementar una gestión dinámica de la liquidez para mejorar la eficiencia del capital no invertido. Por ejemplo, durante los períodos en los que no se detectan señales operativas claras, el algoritmo podría redistribuir temporalmente la liquidez hacia activos de bajo riesgo, como ETFs de renta fija o instrumentos del mercado monetario, con el objetivo de rentabilizar el capital o protegerlo frente a la inflación. Otra alternativa sería introducir una capa adicional de control sobre la exposición agregada, permitiendo ajustar dinámicamente el umbral máximo de inversión en función de condiciones de mercado (por ejemplo, mayor exposición en fases alcistas sostenidas y menor exposición en entornos más volátiles).

Además, se podría estudiar la incorporación de señales macroeconómicas o indicadores de mercado globales para decidir cuándo aumentar o reducir la exposición global, buscando así capturar más valor en mercados claramente tendenciales sin renunciar a la protección inherente del modelo basado en señales ESG.

**i.Resumen**

1. Datos : He trabajado con activos de renta variable (acciones de grandes empresas estadounidenses). El universo de inversión está compuesto por empresas del S&P 500, clasificadas en tres categorías ESG: Environmental, Social y Governance
2. Sesgo de supervivencia: este punto no ha podido ser tratado, al haber trabajado solo con 30 empresas que llevan en el índice todo el periodo seleccionado.
3. Homogeneización de los datos: conversión de fechas a datetime para alinear datasets,  
   uniformización de las categorías ESG (Environmental, Social, Governance) en minúsculas y sin espacios y conversión de sentimientos en variables numéricas coherentes.
4. Ajuste de la serie de datos: En este caso, como los datos provienen de precios de cierre ya ajustados, no se ha necesitado un tratamiento manual adicional.
5. Diseño de algoritmo de inversión:

* Criterio de selección de activos para entrar en mercado:
  + Basado en señales ESG (evento extremo, desacoplamiento, momentum y z-score).
  + Compra si la señal final supera +0.25.
* Criterio de asignación de capital disponible:
  + Máximo 25% del capital disponible por activo.
  + Ajuste proporcional al "score" de confianza de la señal.
* Criterio de selección para salir de mercado:
  + Venta si la señal final cae por debajo de -0.25.
  + Venta forzada si se activa un stop-loss (-30%).
* Criterio de desasignación de capital invertido:
  + Se vende entre el 60% y 80% de la posición, aleatoriamente, para simular salidas parciales.
  + No se requiere liquidar el 100% de la posición salvo stop-loss.

1. Gestión de la cartera:

* ¿Qué hacer con el capital disponible tras una venta?
  + Se mantiene en caja y se usa para nuevas oportunidades de compra futuras.
* ¿Qué hacer cuando no hay capital disponible?
  + No se ejecuta la compra si no hay suficiente efectivo (controlado por el SignalExecutor).
* Límites en la concentración de riesgo:
  + Máx. 25% del capital disponible por activo.
* Comisiones de compra-venta:
  + 0.1% por operación (simulado).
* Umbral de eficiencia:
  + Capital inicial: $1,000,000 (valor típico para simular carteras institucionales).
* Umbral de arrastre (impacto en el mercado):
  + No se simula impacto en profundidad, pero se incorpora slippage del 0.1% para reflejar cierta fricción de mercado.
* Tipo de órdenes:
  + Todas las órdenes se ejecutan en apertura del día siguiente al de la señal (T+1). La ejecución tiene en cuenta slippage para reflejar una ejecución realista.
* Gestión de tesorería:
  + Cálculo de flujos: se registra cada operación (BUY/SELL) y la evolución del capital diariamente.
  + Capital trabajado: nunca se asume dinero infinito; solo se invierte si hay liquidez disponible.
  + Reinversión de beneficios: los beneficios líquidos pasan a estar disponibles para reinvertir.
  + Reserva de efectivo: no se mantiene un % fijo para comisiones porque las comisiones se descuentan automáticamente al ejecutar cada operación.
* Ventana temporal de datos:
  + Rolling diario entre marzo de 2020 y marzo de 2025.
* Stop-loss:
  + 30% respecto al precio de compra. Si se alcanza, se vende toda la posición.
* Gestión de riesgo divisa:
  + No aplicable, ya que todos los activos están denominados en USD.

1. Backtesting:

* Análisis mínimo 5 años:
  + Se ha ejecutado el backtest sobre 5 años (2020-2025).
* Benchmark:
  + El benchmark utilizado es el índice S&P 500 (Buy & Hold) ajustado a las mismas empresas del universo de inversión.
  + Si no existe benchmark ESG puro, se justifica la elección por ser el índice más relevante en la renta variable estadounidense.
* Comparación con otros benchmarks:
  + No se ha comparado con otros fondos ESG por falta de datos equivalentes, pero se propone como mejora futura.
* Prueba de aleatoriedad:
  + El sistema es determinista en la generación de señales, pero introduce aleatoriedad controlada en la venta parcial (60-80%). La robustez ha sido verificada en múltiples ejecuciones.

1. Configuración del algoritmo: El sistema está diseñado como una estrategia dinámica:

* Se ajusta cada día en función de los nuevos datos de noticias y scores ESG.
* Combina señales dinámicas (momentum, z-score) y categóricas (evento extremo, desacoplamiento).
* La ejecución de órdenes también se adapta dinámicamente en función de la cartera y la disponibilidad de efectivo.

**13. STREAMLIT Y GOOGLE CLOUD**

**a.ESG Dashboard**

Carpeta: ESG APP

El objetivo de este apartado es desarrollar una aplicación en Streamlit que permita monitorizar la evolución ESG de distintas empresas, filtrar noticias relacionadas y proporcionar un entorno de análisis sencillo y eficaz. Todo ello apoyado en una infraestructura Cloud moderna que garantice escalabilidad, accesibilidad y reproducibilidad.

i.Objetivos

* Visualización interactiva: Crear un dashboard en Streamlit que permita la visualización dinámica de los scores ESG de varias empresas simultáneamente, categorizadas en Environmental, Social y Governance.
* Procesamiento de noticias: Incluir la capacidad de filtrar y analizar noticias clasificadas por sentimiento (bullish, bearish, neutral), así como por palabra clave y rango temporal.
* Comparativa sectorial: Permitir la comparación entre empresas dentro de un mismo sector para evaluar la competitividad ESG de manera sencilla.
* Arquitectura Cloud: Implementar y justificar una arquitectura Cloud escalable utilizando los servicios de Google Cloud Platform (GCP).
* Despliegue global: Publicar la aplicación de manera pública en Cloud Run, con acceso universal a través de una URL segura.





## ii.Arquitectura General del Proyecto

Para la implementación de este sistema se ha optado por una arquitectura modular basada en la nube. La arquitectura consta de los siguientes componentes:

* Frontend: La app Streamlit, desarrollada en Python, permite la interacción del usuario final con las visualizaciones ESG y las noticias filtradas.
* Backend: Un modelo NLP (basado en LSTM/RoBERTa) entrenado para clasificar noticias en categorías ESG y sentimiento.
* Infraestructura Cloud:
  + Cloud Run: Para desplegar la aplicación de manera serverless, permitiendo escalado automático.
  + Artifact Registry: Almacén seguro de imágenes Docker que facilita CI/CD.

Esta arquitectura permite la flexibilidad necesaria para ampliar el sistema en futuras versiones, incorporando nuevos modelos o capas de visualización.

## iii.Desarrollo de la Aplicación

El sistema se alimenta de dos datasets principales:

* df\_esg\_ready.pkl: contiene la evolución de los scores ESG por empresa, fecha y categoría (Environmental, Social, Governance).
* df\_news\_ready.pkl: dataset con las noticias clasificadas por empresa, fecha, título, contenido y sentimiento (bullish, bearish, neutral).

Ambos datasets han sido preprocesados y optimizados para facilitar su carga y visualización rápida en la aplicación.

El dashboard permite:

* Seleccionar múltiples empresas y categorías ESG.
* Filtrar noticias por sentimiento, palabra clave y rango de fechas.
* Visualizar los KPIs principales por empresa y categoría.
* Comparar fácilmente empresas dentro de un mismo sector.
* Ver noticias destacadas con enlaces externos y contenido resumido.

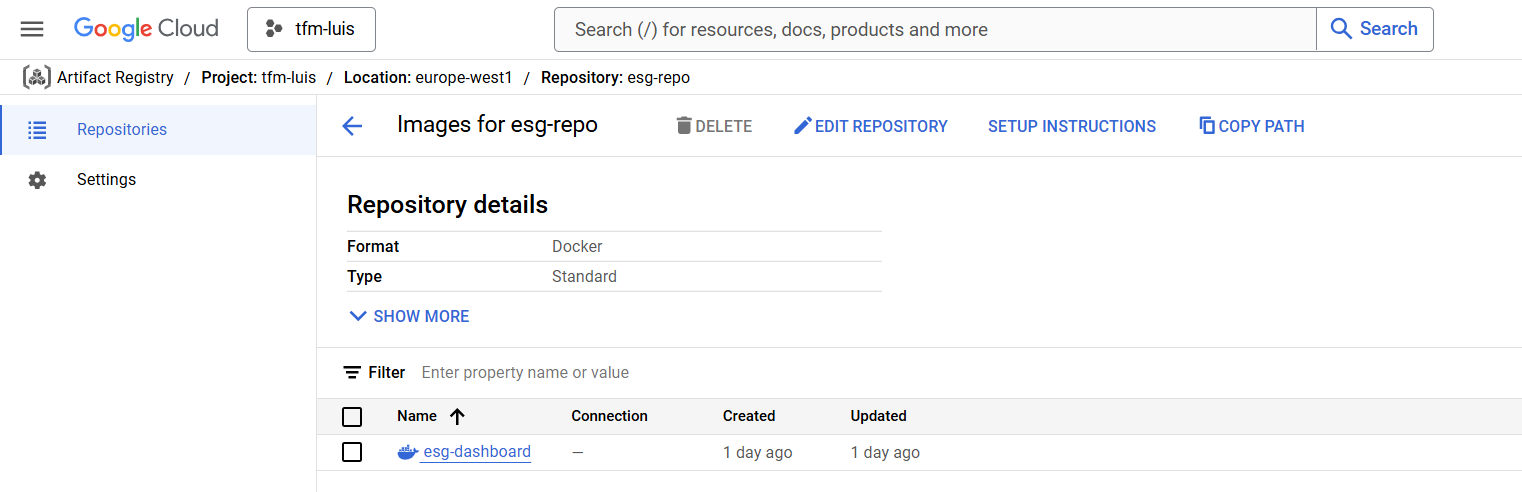
Todas estas funcionalidades están diseñadas para ser intuitivas y altamente configurables por parte del usuario.

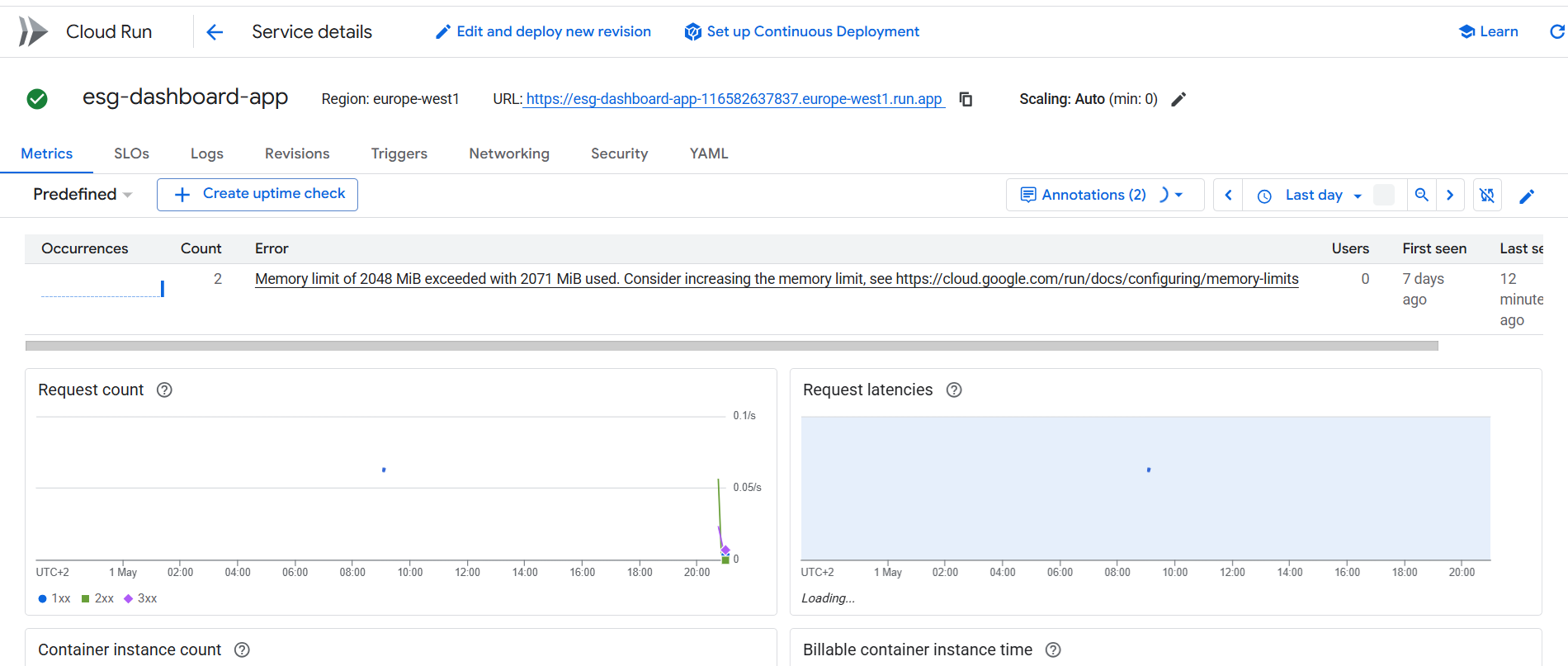
## iv.Justificación de la Arquitectura Cloud

Se ha optado por Google Cloud Platform debido a su amplio abanico de servicios gestionados y alta disponibilidad. Servicios como Cloud Run simplifican el despliegue y escalado sin necesidad de gestionar servidores. Alternativas como AWS o Azure fueron consideradas, pero GCP proporciona una curva de aprendizaje más sencilla y una integración natural con el ecosistema Python, muy utilizado en el proyecto.

### v.Despliegue

El contenedor Docker de la aplicación fue subido al Artifact Registry, lo que permite versionar y gestionar la imagen fácilmente. Posteriormente, la imagen se desplegó en Cloud Run con un aumento de memoria de 2 GiB a 4 GiB tras detectar errores de memoria insuficiente. Esto demuestra la flexibilidad del entorno cloud para adaptarse a necesidades reales.





Link: <https://esg-dashboard-app-116582637837.europe-west1.run.app>

**b.Simulador de estrategias ESG y generador de señales**

Carpeta: Trading app

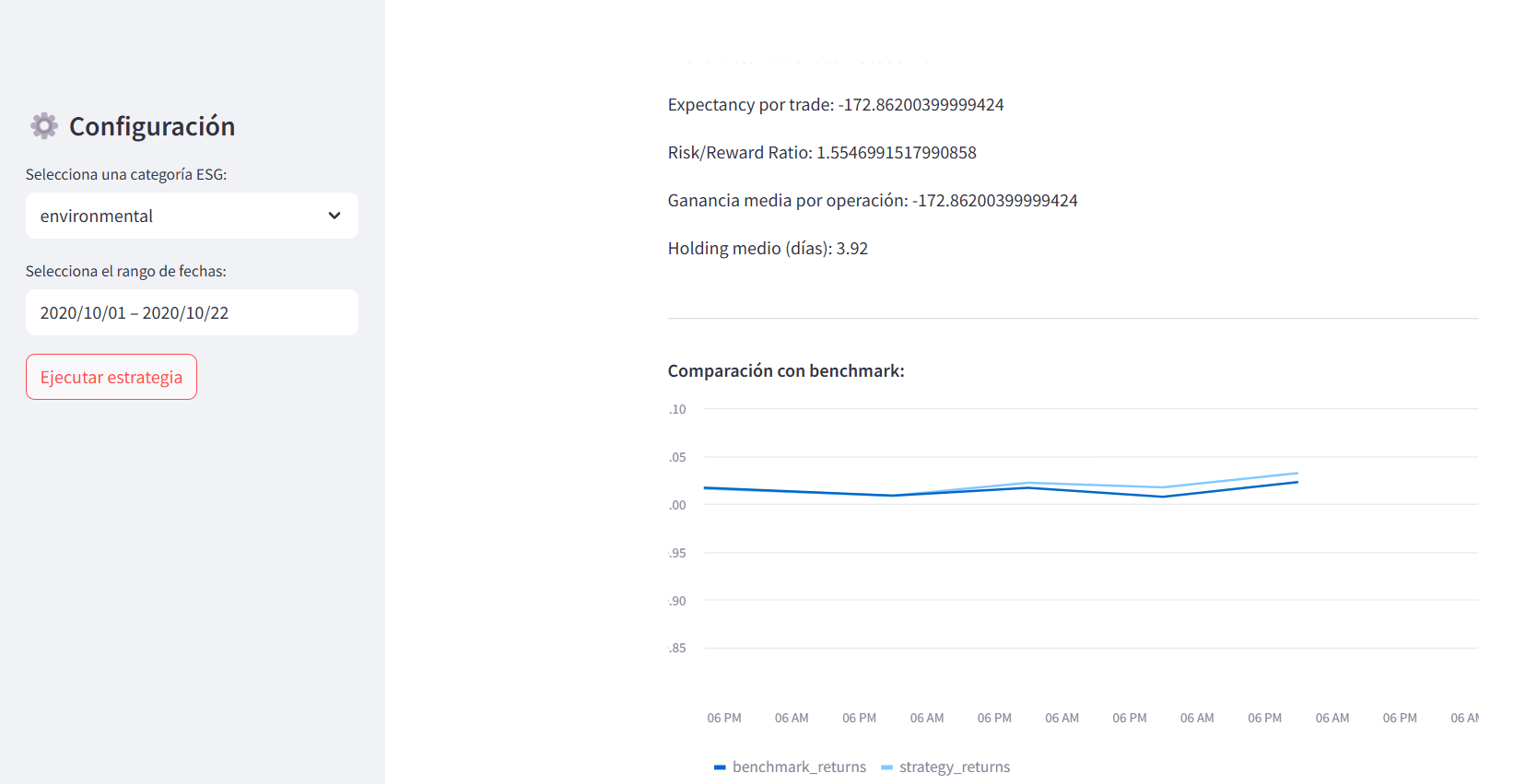
i.Objetivo

La aplicación desarrollada constituye una herramienta para la simulación y análisis de estrategias de inversión basadas en criterios ESG (Environmental, Social, Governance). Su propósito es proporcionar a los usuarios —en particular inversores institucionales y gestores de carteras— un entorno interactivo donde se evalúen las diferentes estrategias de trading basadas en señales ESG, se visualicen los resultados de estas y se acceda a un análisis detallado de señales diarias.

Las principales utilidades que la aplicación aporta al inversor son:

* Monitorización del desempeño ESG: La app permite hacer un seguimiento detallado de cómo los factores ESG afectan la rentabilidad de la cartera a lo largo del tiempo.
* Análisis en profundidad de señales: Cada señal generada se acompaña de una explicación desglosada (evento extremo, momentum, etc.), aportando transparencia en la toma de decisiones.
* Acceso directo a noticias relevantes: Para cada evento extremo detectado, se muestra la noticia original completa, lo cual facilita al inversor contextualizar la señal.
* Comparativa con benchmark: La aplicación compara automáticamente el rendimiento de la estrategia ESG frente a una estrategia Buy & Hold, lo que ayuda a evaluar la eficacia relativa de la inversión ESG.
* Sostenibilidad y fiabilidad: Al estar desplegada en Google Cloud Platform, la aplicación ofrece disponibilidad permanente y escalabilidad automática, lo que garantiza un acceso fiable y robusto.





ii.Desarrollo

Como en la anterior aplicación, el desarrollo y el despliegue se llevó a cabo en google cloud platform utilizando Artifact Registry y Cloud run.

Link: <https://trading-app-116582637837.europe-west1.run.app>

**c.Entrenamiento en Vertex AI**

El último punto del uso de servicios cloud, fue el entrenamiento del modelo de predicción de scores ESG, permitiendo su escalabilidad a otros usuarios y sistemas.

Se seleccionó Google Cloud Platform (GCP), en concreto la suite Vertex AI Workbench, por las siguientes razones:

* Integración Nativa con Big Data:  
   GCP ofrece una suite completa de servicios (Vertex AI, Cloud Storage, BigQuery) que permite gestionar datos de gran volumen y entrenar modelos de machine learning sin complicaciones adicionales.
* Facilidad de Uso y Similitud con Google Colab:  
   Vertex AI Workbench proporciona un entorno JupyterLab gestionado que resulta familiar y fácil de utilizar, lo que permite centrarse en la lógica del modelo sin preocuparse por la configuración de infraestructura.
* Flexibilidad y Escalabilidad:  
   La arquitectura permite escalar verticalmente (aumentando recursos como CPUs, GPUs y memoria RAM) y horizontalmente (añadiendo más instancias si fuera necesario).
* Coste y Región:  
   La posibilidad de utilizar centros de datos europeos (por ejemplo, europe-west1) garantiza la localización de datos y el cumplimiento con normativas como el GDPR.

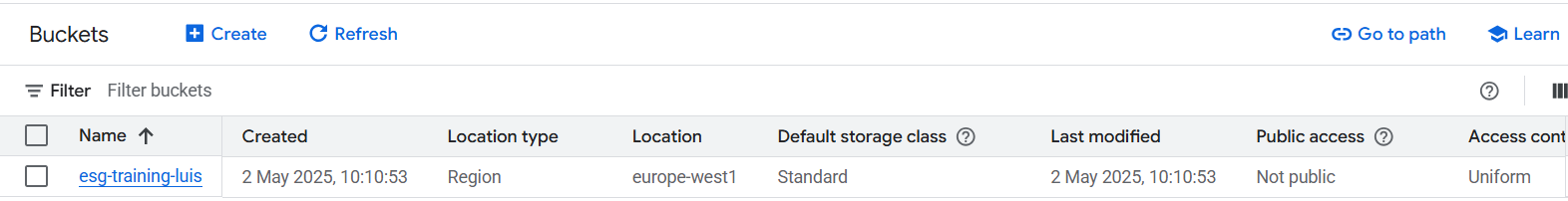
En resumen, la elección de GCP y Vertex AI responde tanto a criterios técnicos como operativos, maximizando la eficiencia y minimizando la complejidad del despliegue.

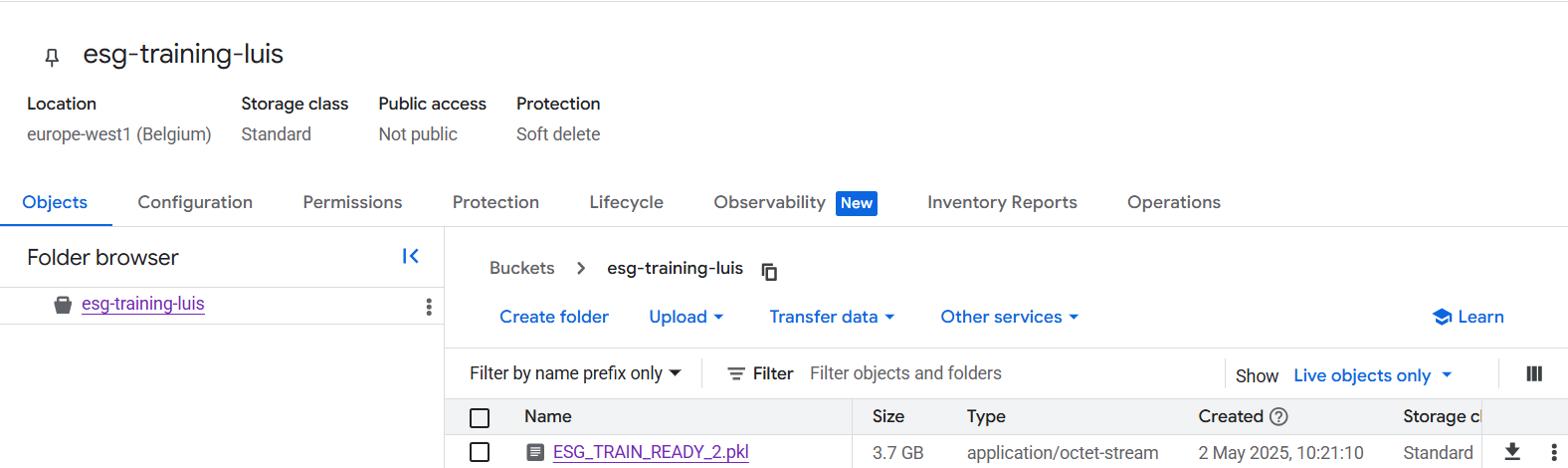
## i.Escalabilidad

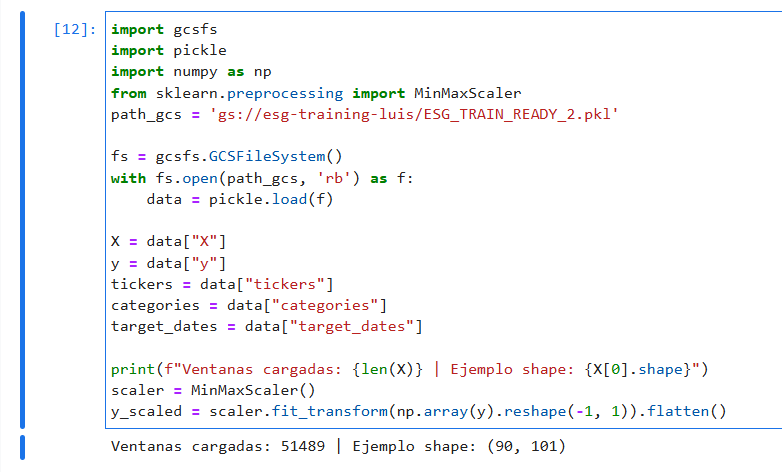
La solución ha sido diseñada pensando no solo en el entrenamiento actual, sino también en su capacidad de escalar para futuros escenarios donde se manejen datasets de más de 100 GB.

Componentes clave de la escalabilidad:

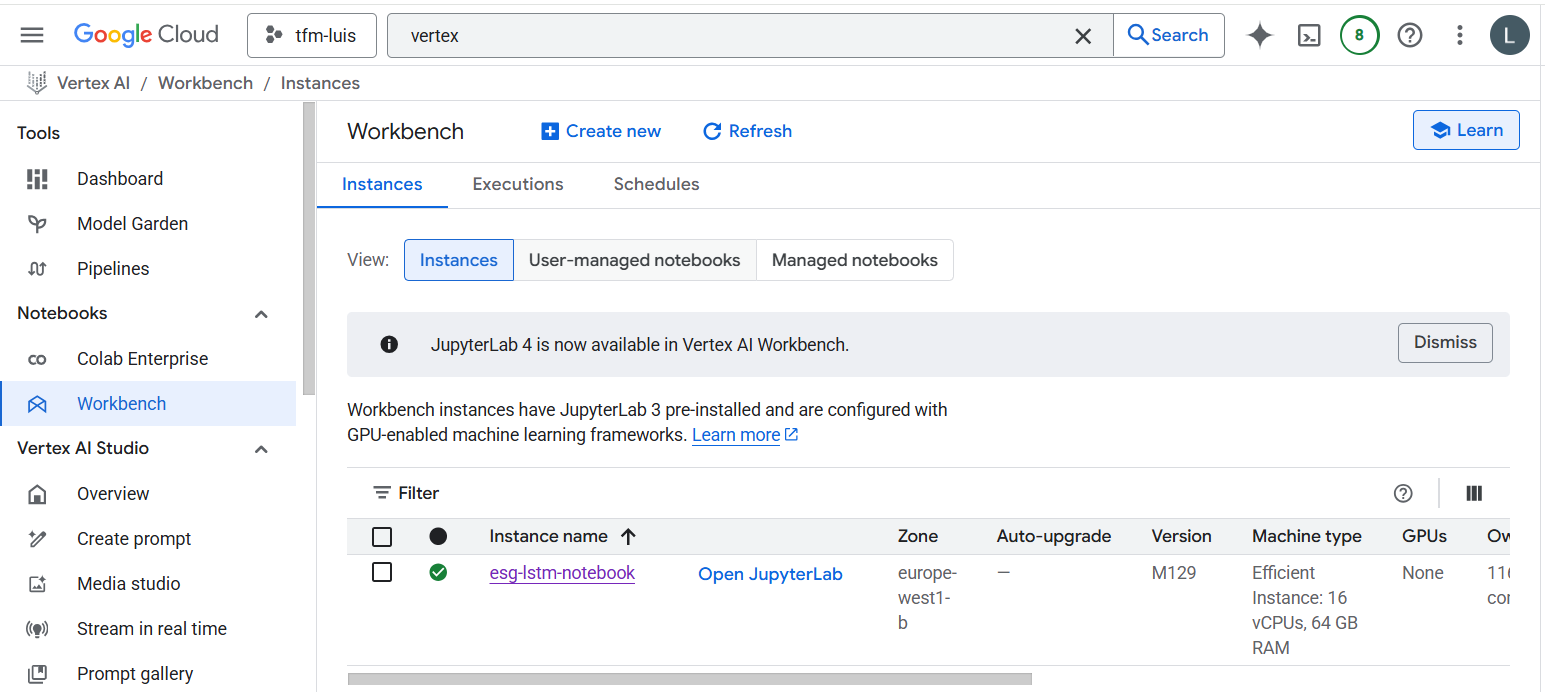
* Almacenamiento Escalable:  
   El dataset se encuentra almacenado en Google Cloud Storage (GCS), un sistema distribuido y redundante que soporta tamaños de archivo que pueden llegar a varios terabytes. La lectura de datos se realiza mediante la librería gcsfs, que permite acceder a los archivos de forma eficiente y segura.

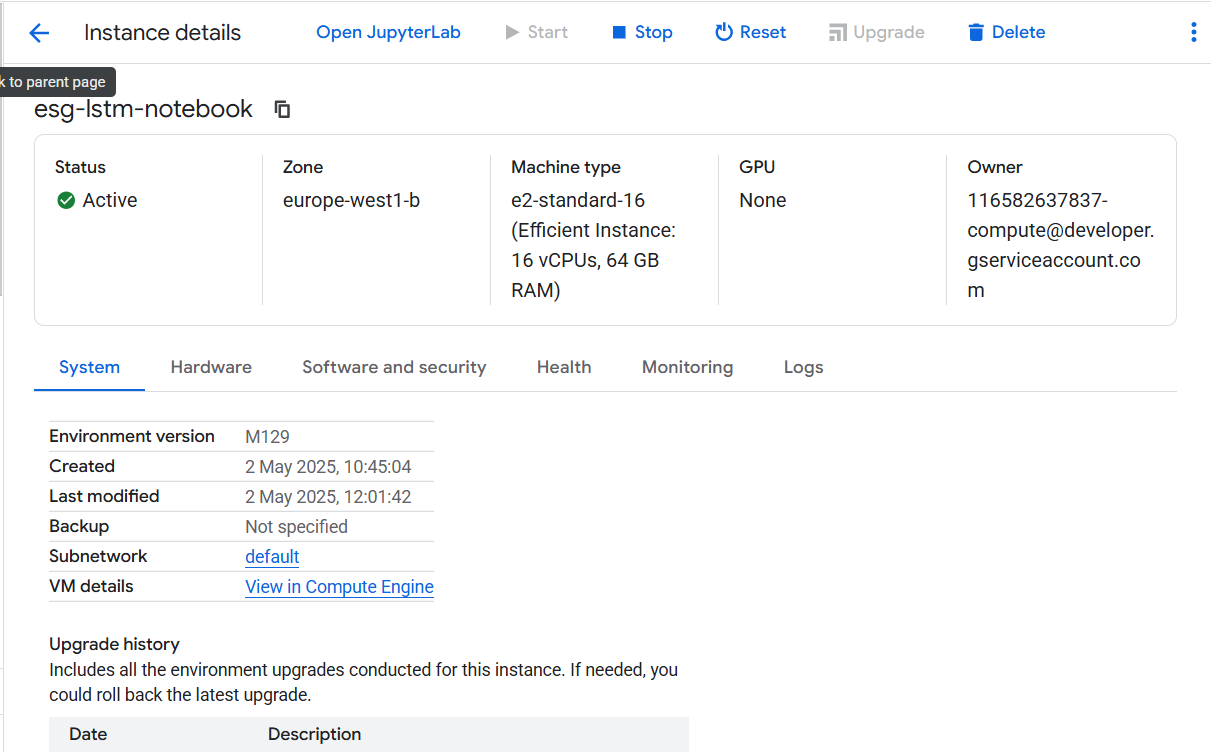






* Carga Bajo Demanda:  
   Para evitar la sobrecarga de memoria RAM, se ha optado por cargar los datos de manera batch-wise mediante la clase DataLoader de PyTorch. Esto permite procesar los datos en pequeños bloques (mini-batches), independientemente del tamaño global del dataset.
* Ajuste Dinámico de la Infraestructura:  
   Vertex AI permite modificar las características de la instancia (por ejemplo, cambiar de una máquina e2-standard-4 a una n1-standard-32 con varias GPUs) sin necesidad de modificar el código. Esto asegura que, ante un aumento de los requisitos de procesamiento, la solución sigue siendo válida.





* Posibilidad de Distribución:  
   Aunque no se ha implementado en esta fase, la arquitectura soporta la inclusión de servicios como Dataflow para procesamiento distribuido y BigQuery ML para entrenamientos paralelizados directamente sobre los datos.

## ii.Despliegue

Para facilitar el uso del modelo por otros usuarios o sistemas, se ha planificado un despliegue en dos fases:

* Persistencia del Modelo en Cloud Storage:  
   Cada vez que el modelo mejora su rendimiento (menor pérdida de validación), se guarda localmente en la instancia de Vertex AI y se sube automáticamente a Google Cloud Storage, concretamente al bucket:  
   gs://esg-training-luis/best\_model.pth
* Preparación para Inferencia y API:  
   El modelo almacenado en GCS está preparado para ser cargado en cualquier entorno que tenga acceso al bucket. Para su integración en producción, se propone desplegar un servicio API en Cloud Run o Vertex AI Endpoints, de forma que otros sistemas puedan enviar peticiones de predicción en tiempo real.

## iii.Conclusiones

El diseño y ejecución de este proyecto han demostrado la viabilidad y eficiencia de utilizar técnicas de Big Data en entornos Cloud para entrenar modelos de machine learning a gran escala. La arquitectura basada en Google Cloud (Vertex AI + Cloud Storage) no solo cumple con los requisitos funcionales, sino que también ofrece la flexibilidad y escalabilidad necesarias para adaptarse a futuros escenarios de mayor volumen de datos y necesidades de producción.

Se concluye que el modelo está listo para su integración en un entorno productivo mediante servicios Cloud, garantizando su accesibilidad y mantenibilidad a largo plazo.

**14.DESCARTES REALIZADOS**

**a.Uso de los ‘Sustainability Reports’ de las compañías**

Uno de los primeros enfoques considerados fue complementar el dataset de noticias y eventos ESG con la información interna proporcionada por las propias empresas, a través de sus reportes de sostenibilidad (*Sustainability Reports*). La hipótesis inicial era que estos informes, al detallar políticas medioambientales, sociales y de gobernanza, aportarían un valor añadido al modelo, permitiendo contrastar las percepciones externas (noticias y medios) con los compromisos formales y resultados declarados por las empresas.

Sin embargo, durante la fase de exploración surgieron varios desafíos que llevaron a descartar esta integración. El principal problema detectado fue la falta de estandarización entre las empresas y sectores: cada compañía publica estos reportes bajo formatos muy diversos, con distintas frecuencias, niveles de detalle y métricas no siempre comparables. Además, muchas veces estos reportes están redactados de forma cualitativa o en lenguaje corporativo que dificulta su uso directo en modelos cuantitativos sin un proceso complejo y costoso de estandarización y etiquetado previo.

Otro aspecto crítico fue la baja frecuencia temporal: al tratarse generalmente de informes anuales o semestrales, la granularidad temporal no era adecuada para alimentar un modelo como la red LSTM, diseñada para trabajar con series temporales de alta frecuencia (diaria en este caso). Esto hacía inviable su uso como input dinámico y actualizable para la predicción diaria de scores ESG.

Por todas estas razones (falta de estandarización, baja frecuencia y dificultad de procesamiento), se decidió finalmente no incorporar los *Sustainability Reports*, focalizando los esfuerzos en la explotación de fuentes externas más uniformes y frecuentes, como las noticias y eventos categorizados ESG.

**b.Información constante para complementar la red LSTM además del clúster**

Otra vía que se estudió fue enriquecer la entrada de la red LSTM con variables adicionales que aportaran contexto estático sobre cada empresa, complementando así la información dinámica obtenida a partir de las noticias y eventos diarios. Entre estas variables se consideraron factores como la capitalización bursátil, el sector, la puntuación ESG histórica promedio y otros indicadores fundamentales que pudieran representar una especie de firma ESG o financiera estable de cada compañía.

El objetivo era que la red LSTM no solo captara las variaciones temporales (ventanas roladas de 90 días), sino también se beneficiara de un perfil fijo que actuara como referencia base para contextualizar las fluctuaciones diarias. Esta técnica híbrida, combinando información estática y dinámica, ha sido utilizada con éxito en otros ámbitos de predicción financiera.

Sin embargo, tras varias pruebas preliminares, se identificaron ciertos problemas prácticos:

* La combinación de datos estáticos y dinámicos en una red LSTM aumentaba la complejidad del modelo y requería adaptar la arquitectura para gestionar entradas heterogéneas (secuencias temporales + vectores fijos).
* La calidad y disponibilidad de ciertas variables constantes no era homogénea para todas las empresas de la muestra, lo que introducía problemas de completitud y sesgo.
* En las primeras simulaciones, el valor añadido en términos de mejora de la predicción no fue suficientemente significativo como para justificar esta complejidad adicional en la fase actual del proyecto.

Por estos motivos, se decidió dejar esta línea de trabajo como potencial mejora futura. El modelo se centró finalmente en explotar al máximo la serie temporal dinámica de noticias ESG y en mantener una arquitectura más limpia y replicable.

**15.BIBLIOGRAFÍA**

* Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*
* Berg, F., Koelbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*
* Krueger, P., Sautner, Z., & Starks, L. T. (2020). The importance of climate risks for institutional investors. *Review of Financial Studies*
* Schmeling, M., & Wagner, C. (2019). Does ESG performance influence firm value? Evidence from the MSCI KLD 400 Social Index. *Finance Research Letters.*
* Vasilescu, L., & Chiş, T. (2022). ESG Scores and stock market performance: Evidence from the S&P 500. *Sustainability.*
* Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10‐Ks. *Journal of Finance.*
* Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT.*
* Brown, T., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems.*
* Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation.*