



VNIVERSIDAD
D SALAMANCA

TRABAJO DE FIN DE MÁSTER

**Estudio de la sostenibilidad en el
EuroStoxx50: un enfoque basado en
minería de texto y análisis
multivariante**

*Study of sustainability in the EuroStoxx50: an approach
based on text mining and multivariate analysis*

Autora:

Laura Sanz Martín

Tutor:

Pedro Ignacio Dorado Díaz

Tutor externo:

Javier Parra Domínguez

**Máster en Análisis Avanzado de Datos
Multivariantes y Big Data**

Salamanca, julio 2025



VNiVERSIDAD
D SALAMANCA

TRABAJO FIN DE MÁSTER

**Estudio de la sostenibilidad en el
EuroStoxx50: un enfoque basado en
minería de texto y análisis
multivariante**

*Study of sustainability in the EuroStoxx50: an approach
based on text mining and multivariate analysis*

Firma de la autora:

A handwritten signature in blue ink, reading "Laura Sanz", enclosed within a circular flourish.

Laura Sanz Martín

Tutor:

Pedro Ignacio Dorado Díaz

Tutor externo:

Javier Parra Domínguez

**Máster en Análisis Avanzado de Datos
Multivariantes y Big Data**

Salamanca, julio 2025

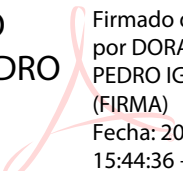
D *Pedro Ignacio Dorado Díaz*, profesor del Departamento de Estadística de la Universidad de Salamanca y D *Javier Parra Domínguez* profesor del Departamento de Administración y Economía de la Empresa,

HACEN CONSTAR:

Que el trabajo titulado “*Estudio de la sostenibilidad en el EuroStoxx 50: un enfoque basado en minería de texto y análisis multivariante*”, que se presenta, ha sido realizado por *Laura Sanz Martín* y constituye la memoria del trabajo realizado para la superación de la asignatura Trabajo de Fin de Máster en esta Universidad.

Salamanca, 10 de julio de 2025

DORADO
DIAZ, PEDRO
IGNACIO
(FIRMA)



Firmado digitalmente
por DORADO DIAZ,
PEDRO IGNACIO
(FIRMA)
Fecha: 2025.07.11
15:44:36 +02'00'

Fdo.: Pedro Ignacio Dorado Díaz

PARRA
DOMINGUEZ
JAVIER -
70935333K



Firmado digitalmente
por PARRA DOMINGUEZ
JAVIER - 70935333K
Fecha: 2025.07.11
17:07:50 +02'00'

Fdo.: Javier Parra Domínguez

Resumen

Este estudio examina la sostenibilidad de las empresas del EuroStoxx 50 desde dos perspectivas complementarias. En primer lugar, se utiliza un enfoque econométrico para analizar la relación entre los resultados ESG (medioambientales, sociales y de gobernanza) y la rentabilidad financiera de las empresas. Los resultados muestran que existe una asociación positiva y significativa entre las métricas ESG y los principales indicadores de rentabilidad, especialmente el ROE, destacando el impacto favorable de las dimensiones medioambiental y social, mientras que la gobernanza se asocia principalmente con la eficiencia operativa. En segundo lugar, se realiza minería de textos sobre los informes de sostenibilidad publicados por estas empresas, con el objetivo de identificar patrones y tendencias relevantes en sus estrategias y enfoques de sostenibilidad. Para ello, se aplican técnicas de análisis multivariante que permiten clasificar y comparar los enfoques adoptados por las distintas empresas y sectores. El análisis revela una evolución hacia estrategias más integradas y orientadas a resultados, adaptadas a las características de cada sector y con un énfasis creciente en la transparencia y la medición de impactos tangibles. Este estudio pone de relieve el valor de combinar métodos de análisis cuantitativos y textuales para comprender la sostenibilidad empresarial. Las conclusiones muestran que la integración de los factores ESG no sólo contribuye a la creación de valor financiero, sino también a la creación de valor financiero.

Palabras clave: Sostenibilidad; ESG; EuroStoxx 50; Análisis de panel; Minería de textos

Abstract

This study examines the sustainability of EuroStoxx 50 companies using an econometric approach to analyze the relationship between ESG (environmental, social and governance) performance and financial profitability. In addition, text mining is performed on the sustainability reports of these companies with the aim of extracting relevant patterns and analyzing them using multivariate analysis techniques. In a context where sustainability and corporate responsibility are increasingly relevant, the study evaluates how ESG factors contribute to corporate value creation beyond traditional financial indicators. The results show that there is a positive and statistically significant association between ESG metrics and the main profitability indicators, especially ROE. Environmental and social dimensions stand out for their favorable impact on financial performance, while governance mainly influences operational efficiency. In addition, the analysis of sustainability reports reveals an evolution towards more integrated and results-oriented strategies, adapted to sectoral characteristics and with an increasing emphasis on transparency and the measurement of tangible impacts. The paper demonstrates the potential of combining quantitative methods and text mining to analyze corporate sustainability, providing empirical evidence on the strategic importance of sustainability factors.

Keywords: Sustainability; ESG; EuroStoxx 50; Panel analysis; Text mining

Índice

1. Introducción	1
1.1. Objetivos	3
1.2. Justificación	4
1.3. Estructura del trabajo	4
2. Marco teórico y estado del arte	6
2.1. Las métricas ESG	6
2.2. Las métricas ESG y su impacto en el rendimiento financiero	12
2.3. El uso de la minería de textos en el análisis de los reportes sostenibles	15
2.4. Hipótesis del estudio econométrico tras la revisión del estado del arte	18
3. Material y métodos	20
3.1. Base de datos	20
3.2. Metodología	21
4. Resultados	35
4.1. Análisis exploratorio de la base de datos	35
4.2. Análisis de las métricas ESG y de los rendimientos económicos . . .	41
4.3. Minería de textos	45
5. Discusión	60
6. Conclusiones	63
7. Referencias	64
8. Anexos	70
8.1. Código	70
8.2. Publicación derivada del TFM	70

Índice de figuras

1.	Proporción de activos de inversión sostenible en relación con el total de activos gestionados, adaptación de Global Sustainable Investing Alliance (2022)	3
2.	Metodología LSEG para asignar puntuaciones ESG	20
3.	Correlaciones de las variables en estudio	37
4.	Evolución de las componentes ESG y de los rendimientos financieros	38
5.	Distribución de las puntuaciones ESG	39
6.	Relación entre las puntuaciones ESG categorizadas y el rendimiento financiero promedio	40
7.	Evolución de la relación entre las puntuaciones ESG categorizadas y el rendimiento financiero promedio	41
8.	Evolución	49
9.	Dendograma de la clusterización de temas	51
10.	Correlaciones de Pearson significativas entre los temas y las puntuaciones ESG de cada empresa por sector	54
11.	HJ-Biplot con PC1 y PC2	57
12.	HJ-Biplot con PC1 y PC3	58

Índice de tablas

1.	Clasificación de variables del dataset	21
2.	Definición de variables clave (Chen et al., 2023)	22
3.	Estadísticas descriptivas de variables financieras y ESG	36
4.	ESG y rendimiento financiero por sector económico	38
5.	Resumen de efectos aleatorios para la variable ESG	42
6.	Resumen de efectos aleatorios para la variable ambiental	42
7.	Resumen de efectos aleatorios para la variable social	43
8.	Resultados para la dimensión gobernanza con efectos aleatorios y fijos	44
9.	Resultados para las controversias con efectos aleatorios y fijos	45
10.	Temas predominantes en los reportes	46
11.	Evolución de la proporción temática por categoría por año	48
12.	Diferencias de medias en los cuartiles 1 y 4	55
13.	Cargas de los temas en los componentes	56

1. Introducción

La humanidad está en progreso continuo. Actualmente, con el rápido avance de la economía y la sociedad, el mundo enfrenta tanto oportunidades como desafíos significativos. Desde 2020, con la propagación global del COVID-19, los problemas medioambientales, sociales y de gobernanza han atraído una creciente atención a nivel mundial. El tema del desarrollo sostenible e integral se ha convertido en un tema candente de debate en todo el mundo. En respuesta a los graves problemas relacionados con el desarrollo sostenible en los ámbitos ambiental, social y financiero, se han implementado planes de acción como el desarrollo de los factores ESG (medioambientales, sociales y de gobernanza, por sus siglas en inglés) (Li et al., 2021). Estos factores representan un marco holístico para evaluar el rendimiento de una empresa más allá de las métricas financieras tradicionales. Esta visión más amplia reconoce que el éxito a largo plazo de una empresa depende no solo de su rentabilidad, sino también de su sostenibilidad ambiental, su responsabilidad social y sus prácticas de gobernanza (Friede et al., 2015). El sistema ESG, una extensión de la responsabilidad social empresarial (CSR), ha definido el marco para la transformación ecológica de las compañías. ESG no solo expande el concepto de inversión socialmente responsable, sino que también se ha establecido como un estándar a nivel mundial para medir el grado de desarrollo sostenible de las empresas. Esta perspectiva incentiva a las compañías a redimensionar su objetivo de incrementar el beneficio propio a incrementar el valor social. En este contexto, el Consejo de Normas de Contabilidad de Sostenibilidad de los Estados Unidos, el Programa de las Naciones Unidas para el Medio Ambiente, la *Global Reporting Initiative (GRI)*, muchas bolsas de valores de todo el mundo y muchas otras organizaciones gubernamentales y no gubernamentales han comenzado a promover activamente los principios ESG y sus estándares de divulgación (Jiang et al., 2024).

El creciente perfil de los criterios ESG ha sido evidente en las inversiones, a pesar de que la tasa de nuevas inversiones ha mostrado una caída reciente. Las entradas a fondos sostenibles aumentaron de 5 mil millones de dólares en 2018 a más de 50 mil millones de dólares en 2020, y luego a casi 70 mil millones en 2021. A mediados de 2022, los activos sostenibles globales sumaban aproximadamente 2.5 billones de dólares (Pérez et al., 2022). Los Principios para las Inversiones Responsables (PRI), documentaron 3,826 firmantes en su marco en 2021, con unos Activos Bajo Gestión (AUM) combinados de 121.3 billones de dólares. Las corporaciones también están otorgando cada vez más importancia a los temas de CSR y ESG. Grandes corporaciones están construyendo estrategias alineadas con los Objetivos de Desarrollo Sostenible de las Naciones Unidas, están invirtiendo cientos de millones de dólares en proyectos altruistas cada año e intentan aumentar la visibilidad de sus esfuerzos en sostenibilidad. Según el Informe del Décimo Aniversario emitido por la *Governance & Accountability Institute*, una consultora en sostenibilidad corporativa y ESG, el 92 % de las empresas del S&P 500 y el 70 % de las empresas del Russell 1000 publicaron sus informes de sostenibilidad en 2020. En resumen, la RSC y ESG parecen estar pasando de ser una preocupación periférica a una preocupación central tanto para los inversores como para las

empresas (Rau y Yu, 2024).

Las puntuaciones ESG se popularizaron rápidamente debido a su capacidad para mejorar la reputación de las empresas, reducir la presión legislativa, disminuir el riesgo financiero y atraer más capital. Para los inversores, las puntuaciones ESG ayudan a mitigar el riesgo asociado con escándalos ambientales o sociales, ya que permiten anticipar los impactos negativos sobre las inversiones. Sin embargo, son menos efectivos para evaluar los efectos positivos derivados de las acciones sostenibles de una empresa, como su contribución al bienestar social y ambiental. Aunque algunas pequeñas organizaciones también participan en actividades sostenibles, medir el impacto real de estas prácticas sigue siendo un desafío (Clément et al., 2025).

La Figura 1 ilustra la distribución de los activos de inversión sostenible en diferentes regiones desde 2014 hasta 2022, reflejando la evolución de la adopción y variación de estas regiones geográficas¹. Europa ha sido la región con mayor proporción de activos sostenibles, pero también se revela una tendencia decreciente a lo largo de estos años. Estados Unidos tiene un aumento en la proporción de activos sostenibles de 18 % en 2014 a 33 % en 2020, y una disminución dramática al 13 % en 2022. Esto puede ser el resultado de la volatilidad política y económica, especialmente relacionada con las políticas gubernamentales y los cambios en las prioridades en torno a la sostenibilidad bajo diferentes administraciones. Canadá tiene un crecimiento estable y continuo, aumentando su proporción de un 31 % en 2014 a un 47 % en 2022, lo que indica un fuerte compromiso con la sostenibilidad, impulsado por políticas favorables a la inversión sostenible. Australia y Nueva Zelanda muestran una tendencia fluctuante, alcanzando su porcentaje más alto en 2018 con más de un 60 % para luego disminuir casi un 20 % en 2022. Lo que podría reflejar una reacción a las incertidumbres económicas y políticas de la región, así como cambios en las normativas que afectan a las inversiones sostenibles. Por último, Japón presenta un patrón de adopción relativamente bajo en comparación con las otras regiones, comenzando en un 4 % en 2014, aumentando a un 30 % en 2020 y cayendo a un 2 % en 2022. Esto sugiere que, a pesar de algunos esfuerzos recientes, la región aún muestra una adopción moderada en términos de inversión sostenible. Esta limitada participación podría estar influenciada por factores culturales, la estructura de mercado y el enfoque relativamente conservador de Japón respecto a los cambios hacia modelos más sostenibles en su economía.

¹<https://www.gsi-alliance.org/>

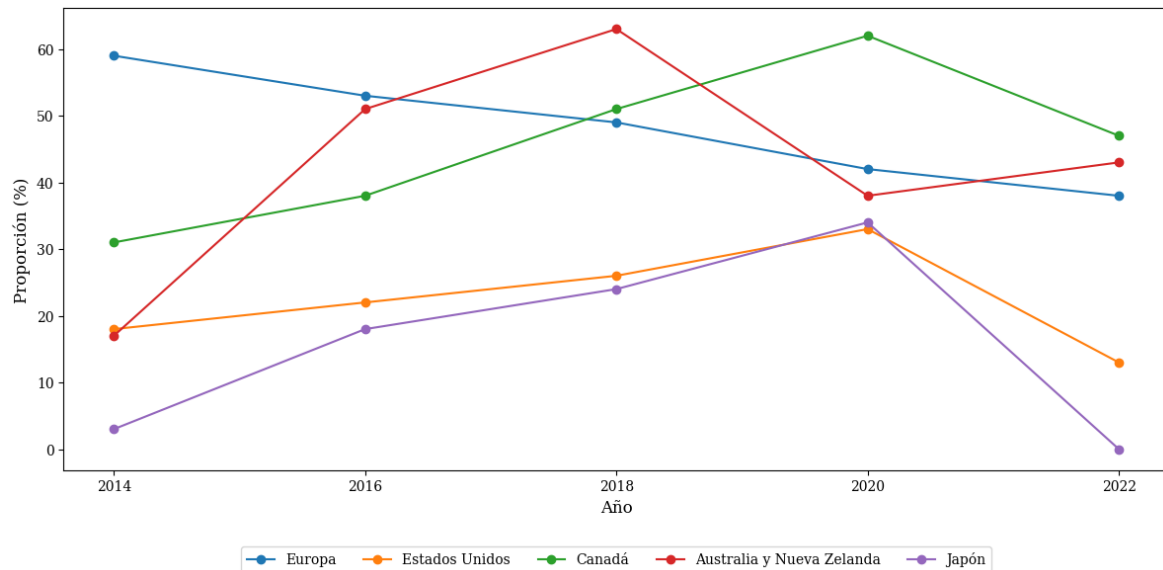


Figura 1. Proporción de activos de inversión sostenible en relación con el total de activos gestionados, adaptación de Global Sustainable Investing Alliance (2022)

1.1. Objetivos

Este trabajo tiene dos objetivos generales, que a su vez se pueden desglosar en objetivos específicos:

1. Analizar la relación entre las métricas de desempeño ESG y los rendimientos financieros de las empresas del EuroStoxx 50 mediante un análisis de datos de panel.
 - Examinar la relación entre las métricas de desempeño ESG y los indicadores financieros de las empresas del EuroStoxx 50 mediante modelos de datos de panel.
 - Verificar los supuestos clásicos de los modelos de datos de panel y seleccionar el modelo de panel más adecuado utilizando pruebas estadísticas.
2. Realizar un análisis de minería de textos de los reportes no financieros, o de sostenibilidad, publicados por las empresas del EuroStoxx 50 para identificar los temas clave y la información relevante sobre sus prácticas de sostenibilidad.
 - Identificar los principales temas y enfoques presentes en los informes de sostenibilidad, prestando especial atención a la centralidad de cuestiones como el impacto climático, la gobernanza, la gestión de riesgos, la diversidad y la alineación con marcos internacionales.
 - Analizar cómo varían las prioridades y los discursos de sostenibilidad a lo largo del tiempo, detectando tendencias emergentes y cambios en el énfasis temático, así como la integración de diferentes dimensiones ESG en la estrategia y comunicación corporativa.

- Examinar la estructura y las relaciones entre los distintos temas de sostenibilidad identificados en los informes, para comprender cómo se articulan y agrupan los diferentes aspectos de la sostenibilidad corporativa en el contexto europeo.

1.2. Justificación

Este Trabajo Fin de Máster se justifica por la creciente importancia de los criterios ESG en los mercados financieros europeos y la necesidad de comprender su impacto real en el rendimiento empresarial. El EuroStoxx 50, como índice de referencia de las principales empresas de la eurozona, ofrece un marco idóneo para analizar cómo las grandes corporaciones europeas están integrando la sostenibilidad en sus modelos de negocio y cómo esto afecta a su desempeño financiero.

Este trabajo busca contribuir al debate académico sobre la materialidad financiera de los factores ESG, aportando evidencia empírica basada en un análisis riguroso y multidimensional. La combinación de técnicas cuantitativas y cualitativas permitirá obtener una visión más completa y matizada de la relación entre sostenibilidad y rendimiento empresarial que los estudios previos centrados exclusivamente en uno de estos enfoques.

Los resultados de este estudio serán valiosos para inversores, al proporcionar información sobre qué factores tienen mayor impacto en el rendimiento financiero de las empresas europeas, directivos empresariales, al identificar las mejores prácticas en materia de sostenibilidad y reguladores, al evaluar la efectividad de los marcos de reporte no financiero existentes y su capacidad para transmitir información relevante a los mercados.

La originalidad de este trabajo radica en su enfoque integrado, que combina el análisis de datos ESG, la aplicación de técnicas avanzadas de análisis textual a los informes de sostenibilidad y el enfoque del EuroStoxx 50, que permite un análisis en profundidad de empresas que operan bajo un marco regulatorio común. Además, este trabajo resulta especialmente oportuno en el contexto actual, caracterizado por la implementación progresiva de la Directiva Europea de Información de Sostenibilidad Corporativa (CSRD) y los estándares ESRS, el creciente escrutinio sobre el *greenwashing* y la calidad de los datos ESG, y la evolución de los mercados financieros hacia la integración sistemática de factores de sostenibilidad en las decisiones de inversión. Esta investigación permitirá comprender mejor cómo las grandes empresas europeas están generando valor tanto para la sociedad como para los accionistas.

1.3. Estructura del trabajo

Este trabajo se estructura de la siguiente forma: primero, en la introducción, se presenta el contexto y la relevancia del tema, seguidamente de los objetivos del estudio y la justificación de su interés académico y práctico. El marco teórico y el estado del

arte desarrolla una revisión exhaustiva de las principales métricas ESG, los enfoques actuales hacia estos criterios, la relación entre la CSR y los criterios ESG, así como la evolución de ESG como estándar global. Además se analizan las regulaciones europeas en materia de ESG, su impacto en las prácticas empresariales y las bases de datos ESG más utilizadas, junto con una revisión de la literatura sobre el impacto de las métricas ESG en el rendimiento financiero y el uso de la minería de textos en el análisis de informes de sostenibilidad. Posteriormente, la sección de material y métodos describe la base de datos empleada y las metodologías aplicadas, que incluyen tanto el análisis de panel, como la minería de textos para el análisis cualitativo de los informes corporativos. En la parte de resultados se presentan el análisis exploratorio de la base de datos y los resultados obtenidos del análisis de panel y de la minería de textos. Finalmente, se discuten los resultados con sus limitaciones y recomendaciones para futuras investigaciones y se presentan las conclusiones del estudio.

2. Marco teórico y estado del arte

2.1. Las métricas ESG

2.1.1. Enfoques hacia los criterios ESG

En la década de 1950 nació la Responsabilidad Social Corporativa (CSR, por sus siglas en inglés) con el objetivo de que las empresas tuviesen responsabilidades más allá de la rentabilidad. En los 60 la conciencia social y ambiental creció y presionó a las empresas a abordar sus responsabilidades sociales. En los 70 se formalizó la CSR, enfocándose en la capacidad de respuesta proactiva a los problemas sociales y ambientales, en los 80, la CSR evolucionó hacia iniciativas estratégicas vinculadas al rendimiento corporativo, surgiendo la noción de "valor compartido" y la teoría de los *stakeholders*. En los 90 aumentó la rendición de cuentas, así como la globalización y se enfatizó la sostenibilidad. Posteriormente, en 2004, se desarrolló ESG en un informe titulado "*Who Cares Wins (WCW)*", iniciado por el Secretario General de la ONU y el Pacto Mundial de la ONU en colaboración con el gobierno suizo². En la década del 2010, los criterios ESG se generalizaron, integrándose en las decisiones de inversión tras la crisis financiera de 2008. En la actualidad, hay una transición hacia los ESG 2.0, donde la sostenibilidad se integra estratégicamente en las operaciones para impulsar la innovación y la ventaja competitiva y se espera un impacto medible y una profunda integración en la cultura competitiva (Passas, 2024).

Recientemente, ha surgido el término 'ESGT', que añade la dimensión tecnológica a los criterios ESG. Esta evolución responde a la necesidad de integrar herramientas avanzadas como inteligencia artificial, big data y blockchain para monitorizar, medir y optimizar el impacto de las empresas en tiempo real. La tecnología se convierte así en un habilitador clave de la sostenibilidad, permitiendo una gestión más rigurosa, transparente y alineada con los intereses de los accionistas e inversores. La ESGT representa un salto cualitativo en la gestión responsable, donde la tecnología no solo optimiza procesos y eficiencia operativa, sino que también facilita la toma de decisiones estratégicas basadas en datos y la medición del impacto real de las acciones de sostenibilidad. La competitividad empresarial a largo plazo se vincula cada vez más con la capacidad de ser sostenible, innovador y tecnológicamente avanzado³.

La CSR puede definirse como un compromiso continuado de las organizaciones y empresas de ir más allá de sus obligaciones legales, adoptando un comportamiento ético que contribuya al desarrollo económico y mejore de una forma sostenible la calidad de vida de la sociedad. Esta definición incorpora una dimensión multidimensional que abarca desde la implementación de programas voluntarios y asociaciones hasta el diseño de políticas que protejan el bienestar social y promuevan beneficios medioambientales. De esta forma, reconoce la

²<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/280911488968799581/who-cares-wins-connecting-financial-markets-to-a-changing-world>

³<https://www.pressreader.com/spain/actualidad-economica/20250413/281986088391588?srsltid=AfmBOoqgw4BCuNHZl6j6IYucUHmeOutZabJ-f-K7hJPnEUpnO4cXWfTX>

relevancia de extender estas prácticas al ámbito internacional y exige el establecimiento de criterios e indicadores claros para medir el grado de implicación de la empresa y evaluar su efecto sobre el bienestar social (Lindgreen y Swaen, 2010).

2.1.2. La relación entre CSR y ESG

La CSR puede considerarse el antecedente de las prácticas ESG, ya que ha sentado las bases para la adopción de sus principios en el entorno empresarial. La CSR es un enfoque estratégico que hace hincapié en la necesidad de responsabilizar a las empresas de sus actividades y garantizar que proporcionan resultados beneficiosos para amplia gama de partes interesadas, como el medio ambiente, los clientes, los trabajadores, las comunidades y el dominio público. Se hace hincapié en la importancia de la conducta ética, la sostenibilidad y la responsabilidad social. Al comparar ESG con CSR, podemos observar que ESG amplía las ideas de la CSR al ofrecer criterios cuantificables que pueden ser utilizados por los inversores y clientes para evaluar las políticas internas de una empresa. Este esfuerzo trasciende el ámbito del puro altruismo y traduce estas ideas en cifras y métricas tangibles. Esto permite a los *stakeholders* tener una comprensión más completa del rendimiento de una empresa con respecto a su impacto social y medioambiental. Estrategias como ESG y CSR están muy presentes en los entornos corporativos y comerciales contemporáneos, y desempeñan un papel crucial en la dirección y mejora de las iniciativas sociales y medioambientales emprendidas por las empresas (Nugroho et al., 2024).

Por tanto, CSR y ESG son ciertamente complementarias entre sí en el contexto de la aplicación de los objetivos de desarrollo sostenible (ODS). Mientras que la CSR pretende que una empresa sea responsable, los criterios ESG cuantifican los esfuerzos de dicha empresa. Sin embargo, la aplicación del concepto ESG parece más exigente que la realización de actividades en el ámbito de la CSR. Esto se debe a que requiere objetivos cuantificables, recopilación de datos y presentación de informes (Kaźmierczak, 2022).

2.1.3. Desarrollo de ESG como un estándar global

La publicación de informes de sostenibilidad se ha convertido en una práctica común en las empresas, especialmente después de que la *Global Reporting Initiative* (GRI)⁴ lanzara sus directrices para fomentar divulgaciones más comparables y confiables sobre temas sociales, ambientales y económicos. Desde su creación en 1997, GRI se ha consolidado como una de las organizaciones más influyentes en el campo de los informes de sostenibilidad. En 2016, GRI pasó de proporcionar directrices a establecer los primeros estándares globales para la presentación de informes de sostenibilidad. De hecho, según la Encuesta de Sostenibilidad 2020 de KMPG, dos tercios de las 5200 principales empresas en 52 países usan las directrices de GRI para sus informes, lo que puede deberse a que GRI fue la

⁴<https://www.globalreporting.org/>

primera en publicar guías de sostenibilidad y ha sido clave en el desarrollo de la Directiva Europea 2014/95/EU sobre informes integrados. Estas directrices están centradas en el concepto del *Triple Bottom Line* derivado del informe Brundtland de la ONU, que buscan reportar el desempeño económico, social y ambiental, así como en el principio de materialidad, que ayuda a las empresas a determinar que temas son lo suficientemente importantes como para ser reportados. Este principio es fundamental para garantizar que las empresas informen sobre sus impactos significativos en las áreas económica, social y ambiental. También previene la manipulación de informes, como el *cherry picking*, que implica seleccionar solo información positiva y omitir la negativa, o el *greenwashing*, que es una forma engañosa de promocionar una imagen más sostenible de una empresa sin que haya un compromiso real. Aunque GRI y otras organizaciones presionan a las empresas para que adopten este principio, existe una falta de estudios empíricos sobre su implementación práctica, especialmente en cuanto a los métodos, criterios y técnicas de involucramiento de los interesados (Machado et al., 2021, Yosifova y Petrova-Kirova, 2022).

La *Sustainability Accounting Standards Board (SASB)* es una organización sin ánimo de lucro fundada en 2011 que desarrolla estándares de contabilidad sostenible para guiar a las empresas en la divulgación de temas ambientales, sociales y de gobernanza (ESG) para los inversores en industrias específicas. A diferencia de iniciativas globales como GRI, SASB está enfocada en las empresas cotizadas de Estados Unidos y establece tres tipos de elementos en sus directrices: obligatorios, recomendados y opcionales. También especifica cómo deben alinearse los datos financieros y no financieros, cómo deben normalizarse las métricas y cómo se deben reportar incertidumbres y estimaciones (Hales, 2021).

2.1.4. Regulaciones europeas en ESG

En el contexto europeo, la regulación en materia ESG ha experimentado una evolución significativa en la última década, impulsada por la necesidad de mejorar la transparencia y la sostenibilidad en la actividad empresarial. La primera normativa fue la Directiva 2014/95/UE, que obligaba a ciertas grandes empresas a divulgar la información no financiera. Años después, la Directiva (UE) 2022/2464 estableció obligaciones más detalladas y estructuradas sobre la presentación de información en sostenibilidad. Y, finalmente, el Reglamento Delegado (UE) 2023/2772, desarrolló los estándares europeos de información sobre sostenibilidad, especificando los requisitos concretos de reporte y consolidando el marco normativo para el reporte ESG en Europa.

La Directiva 2014/95/UE sobre la divulgación de información no financiera y de diversidad, conocida comúnmente como la directiva sobre información no financiera (*Non-Financial Reporting Directive (NFRD)*), fue adoptada en 2014 y marcó un rumbo claro para una mayor transparencia empresarial y rendición de cuentas en cuestiones sociales y ambientales de la Unión Europea. Al combinar la rentabilidad a largo plazo con la justicia social y la protección del medio ambiente, se considera vital para gestionar el cambio hacia una economía global sostenible. Tiene como objetivo aumentar el nivel de transparencia de la información no

financiera divulgada por las grandes empresas. Contribuye a fomentar el crecimiento sostenible, financiar la transición verde y dirigir flujos financieros y capital hacia la inversión sostenible, un objetivo clave bajo el Pacto Verde Europeo. Esta Directiva exige que las empresas divulguen información no financiera. Concretamente, deben publicar informes sobre las políticas que implementan en relación con la responsabilidad social y tratamiento de los empleados, respeto de los derechos humanos, lucha contra la corrupción y el soborno y diversidad en los consejos de administración de las empresas. Además, requiere que las empresas divulguen información sobre sus modelos de negocio, políticas, resultados, riesgos y gestión de riesgos e indicadores clave de rendimiento. Esta directiva se aplicaba a grandes empresas cotizadas, bancos y compañías de seguros con más de 500 empleados, y estas empresas tuvieron que informar por primera vez en el año 2018, para el ejercicio financiero de 2017 ⁵.

La Directiva (UE) 2022/2464 del Parlamento Europeo y del Consejo de 14 de diciembre de 2022, llamada comúnmente Directiva sobre Información de Sostenibilidad Corporativa (*Corporate Sustainability Reporting Directive (CSRD)*), se considere una actualización y mejora de la Directiva 2014/95/UE. Esta Directiva busca aumentar la comparabilidad, fiabilidad y relevancia de la información sobre sostenibilidad y pretende asegurar que inversores, organizaciones de la sociedad civil otros tengan acceso a información adecuada para evaluar el impacto de las empresas y los riesgos a los que se enfrentan en cuestiones de sostenibilidad. Contribuye a la transición hacia una economía sostenible y a los objetivos del Pacto Verde Europeo. Esta Directiva modifica la anterior para aplicar los requisitos de información sobre sostenibilidad a grandes empresas y a pequeñas y medianas empresas (pymes) que sean de interés público, excluyendo microempresas. Las empresas deben incluir la información en su informe de gestión. Esta información debe ser suficiente para comprender el impacto de la empresa en las cuestiones de sostenibilidad y cómo estas cuestiones afectan a la empresa (el concepto de "doble significatividad"), considerando cada perspectiva de forma independiente. La información debe ser prospectiva y retrospectiva, cualitativa y cuantitativa; y debe cubrir la cadena de valor completa de la empresa. Los ámbitos de información incluyen modelo de negocio y estrategia, políticas de sostenibilidad, resultado de las políticas, riesgos relacionados con la sostenibilidad y su gestión, indicadores clave de resultados relevantes, función, conocimientos y capacidades de los órganos de administración, dirección y supervisión en cuestiones de sostenibilidad, y sistemas de incentivos vinculados a la sostenibilidad, información sobre los principales recursos inmateriales, información sobre prácticas de pago a proveedores y la información sobre la política de diversidad en los órganos de gobierno debe incluir el género. A diferencia de la NFRD, esta Directiva exige que las empresas informen de conformidad con normas comunes obligatorias de presentación de información sobre sostenibilidad de la UE. Estas normas serán adoptadas por la Comisión Europea mediante actos delegados y deben tener en cuenta las iniciativas de normalización globales ⁶.

⁵Directiva 2014/95/UE del Parlamento Europeo y del Consejo, de 22 de octubre de 2014, por la que se modifica la Directiva 2013/34/UE en lo que respecta a la divulgación de información no financiera y sobre diversidad por parte de determinadas grandes empresas y grupos

⁶Directiva (UE) 2022/2464 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022,

El Reglamento Delegado (UE) 2023/2772 (*Sustainable Finance Disclosure Regulation (SFDR)*) establece las Normas Europeas de Información sobre Sostenibilidad (NEIS), que cumplen los requisitos establecidos en la Directiva 2013/34/UE, modificada por la Directiva (UE) 2022/2464. Estas normas tienen como objetivo especificar la información sobre sostenibilidad que una empresa debe divulgar conforme a la NFRD, modificada por la CSRD. Las NEIS se estructuran en normas generales (NEIS 1 y NEIS 2) y normas técnicas (ambientales, sociales y de gobernanza)⁷. Sin embargo, este reglamento también presenta algunos retrocesos en cuanto a la reducción de los requisitos, como la disminución de hasta un 50 % en los criterios de divulgación. Esta flexibilización podría comprometer la efectividad de la CSRD, ya que permite que las empresas omitan información crucial o proporcionen datos menos detallados si no los consideran materialmente relevantes según su análisis de doble materialidad (Martínez-Torres, 2024). No obstante, la elaboración de informes de sostenibilidad no debe limitarse únicamente a la divulgación de datos, sino que debe estar estrechamente vinculada a las acciones concretas que las empresas llevan a cabo para promover el desarrollo sostenible. Este enfoque implica un cambio de paradigma en la manera en que las empresas gestionan sus operaciones, ya que deben integrar de forma coherente los aspectos ESG en su estrategia (Antonów, 2024).

Además, el 26 de febrero de 2025, la Comisión Europea adoptó un paquete de propuestas para simplificar las normas sobre sostenibilidad e inversiones, con el objetivo de reducir al menos un 25 % las cargas administrativas, y un 35 % en el caso de las pymes, y focalizar los requisitos, en las empresas más grandes, que son las que tienen mayor impacto ambiental y social. Las reformas propuestas tienen como objetivo crear un entorno más favorable para el crecimiento, la innovación y la creación de empleo de calidad, sin renunciar a los objetivos del Pacto Verde Europeo. La simplificación y el enfoque en las grandes empresas pretenden liberar recursos y facilitar la transición sostenible, especialmente para las pymes, al tiempo que se mantienen los compromisos con la descarbonización y la competitividad europea ⁸.

2.1.5. Impacto de las regulaciones europeas en las prácticas empresariales

Estas regulaciones europeas que hemos visto tienen un impacto en las prácticas empresariales de las corporaciones a las que afectan. Bradford, 2020 define el "Efecto Bruselas" como la capacidad que tiene la Unión Europea para establecer estándares globales en regulación y se basa en el hecho de que la UE es el bloque comercial más grande del mundo. En materia ESG esto también se cumple, puesto que las empresas que buscan operar o participar en el mercado europeo se ven obligadas a cumplir con sus estrictos estándares. Para evitar la complejidad y el coste de gestionar diferentes normativas en diversas jurisdicciones, las empresas

por la que se modifican el Reglamento (UE) n.º 537/2014, la Directiva 2004/109/CE, la Directiva 2006/43/CE y la Directiva 2013/34/UE, en lo que respecta a los informes de sostenibilidad de las empresas

⁷Reglamento Delegado (UE) 2023/2772 de la Comisión, de 31 de julio de 2023, por el que se completa la Directiva 2013/34/UE del Parlamento Europeo y del Consejo en lo que respecta a las normas de presentación de información sobre sostenibilidad

⁸https://ec.europa.eu/commission/presscorner/detail/es/ip_25.614

multinacionales a menudo optan por adoptar los estándares más rigurosos de la UE para todas sus operaciones a nivel global. De esta forma, la regulación europea se "exporta" a través de mecanismos de mercado, forzando o incentivando cambios en las prácticas empresariales a escala mundial, particularmente en áreas como la divulgación de información, la clasificación de inversiones, y la debida diligencia en las cadenas de suministro. Este impacto global se deriva de la amplitud del alcance de las propias normas, que a menudo involucran a países fuera de la UE, y de la participación de la UE en instituciones globales clave (Redondo Alamillos y De Mariz, 2022).

Este impacto se puede observar en estudios empíricos. Arvidsson y Dumay, 2022 centran su investigación en empresas suecas, que han sido históricamente líderes en reporte de sostenibilidad. En su estudio, examinan el efecto de la NFRD y muestran que, aunque la cantidad de reportes ESG ha aumentado significativamente con el tiempo, especialmente antes de que la Directiva fuera obligatoria, esta actividad se ha estabilizado en los últimos años. Además, la calidad de la información reportada ha mejorado de forma constante. Sin embargo, el desempeño ESG corporativo parece haberse estancado desde 2015. Los autores argumentan que hasta el momento, la Directiva de la UE ventrada en la divulgación regulada no ha mejorado significativamente el desempeño ESG. Las empresas tienden a reportar principalmente sobre políticas y riesgos, pero menos sobre los resultados de sus acciones ESG. Esto implica que la presión regulatoria ha cambiado las prácticas de reporte hacia una mayor cantidad y calidad en la divulgación, pero eso no implica que haya impulsado mejoras tangibles en el desempeño real.

2.1.6. Bases de datos ESG

Las bases de datos ESG se han convertido en herramientas fundamentales para inversores, empresas y reguladores que buscan evaluar el desempeño ambiental, social y de gobernanza de las compañías. Entre las diversas empresas que destacan asignando puntuaciones, nos podemos encontrar con LSEG, conocida también como Refinitiv, Bloomberg o Sustainalytics, entre otras.

LSEG cuenta con una de las operaciones de recopilación de contenido ESG más grandes, que procesan manualmente más de 870 medidas ESG para cada empresa. La base de datos se actualiza de forma continua, alineada con los patrones de informes corporativos y los datos se actualizan en los productos semanalmente incluyendo el recálculo de las puntuaciones ESG. Las noticias y controversias ESG se actualizan de forma continua a medida que ocurren y son recogidas por los medios ⁹.

Bloomberg recopila datos ESG disponibles públicamente y reportados por las empresas, sin incorporar estimaciones ni opiniones de analistas. Los datos son recopilados de diversas fuentes, como informes anuales, de sostenibilidad y otras divulgaciones corporativas. Tras este proceso, se someten a múltiples niveles de

⁹<https://www.lseg.com/en/data-analytics/sustainable-finance/esg-scores>

verificación de calidad y estandarización para asegurar la comparabilidad. Aunque la presentación de informes ESG suele ser anual, los datos se revisan, adquieren e incorporan de forma continua a medida que las empresas los publican. Las puntuaciones se calculan y publican de forma regular, varían entre 0 y 10 y también se proporcionan percentiles (de 0 a 100) ¹⁰.

Sustainalytics recopila información ESG de diversas fuentes públicas, incluyendo informes anuales y de sostenibilidad de las empresas, así como datos de organizaciones no gubernamentales, gobiernos y medios de comunicación. Esta información se somete a un proceso de análisis, en el que se evalúa la exposición de las empresas a riesgos ESG materiales y la calidad de su gestión de dichos riesgos. A partir de ese análisis, Sustainalytics emite sus *ESG Risk Ratings*, que representan una evaluación del riesgo ESG no gestionado de las empresas. Estas calificaciones se actualizan de forma regular para reflejar nueva información y cambios en el perfil de riesgo de las compañías, permitiendo a los inversores contar con una visión dinámica y fundamentada del riesgo ESG ¹¹.

2.2. Las métricas ESG y su impacto en el rendimiento financiero

Un gran conjunto de investigaciones ha explorado la correlación entre las puntuaciones ESG y los rendimientos financieros. Los primeros estudios sobre el tema mostraban resultados contradictorios, no concluyentes o ambiguos, algunos presentaban una relación positiva, otros una relación negativa y otros no hallaban ninguna correlación significativa. Esta incoherencia inicial puso de manifiesto la complejidad de la relación y las dificultades que conlleva medir con precisión tanto el desempeño ESG como los rendimientos financieros. Sin embargo, un metaanálisis realizado por Friede et al., 2015, avanzó de forma significativa en la comprensión de esta relación. Esta investigación sintetizó los datos de más de 2000 estudios y descubrieron que un 90 % revelaban una asociación no negativa entre ESG y los rendimientos financieros y la mayoría de estos estudios reportaban hallazgos positivos. Este impacto parece ser estable a lo largo del tiempo. También, encuentran una diferencia significativa entre los estudios que son de cartera y los que no, los estudios de cartera muestran menos hallazgos positivos que los no basados en cartera. A su vez los activos que no son de renta variable (bonos, inmuebles) y los mercados emergentes, muestran una mayor proporción de hallazgos positivos.

Las discrepancias observadas en estudios anteriores pueden deberse en gran medida a variaciones en fuentes de datos y en metodologías. Las investigaciones han empleado distintos enfoques metodológicos, como análisis de regresión (Ademi y Klungseth, 2022; Oprean-Stan et al., 2020) , que examinan la relación entre las puntuaciones ESG y los rendimientos financieros; las estrategias de cartera (Fiskerstrand et al., 2020), que evalúan el rendimiento financiero de las carteras construidas sobre la base de las puntuaciones ESG; y el modelado de ecuaciones estructurales (Cek y Eyupoglu, 2020), que explora las complejas

¹⁰<https://www.bloomberg.com/professional/solutions/sustainable-finance/>

¹¹<https://www.sustainalytics.com/esg-data>

relaciones entre múltiples variables. La elección de la agencia de calificaciones ESG también influye de forma significativa en los resultados, puesto que utilizan distintas metodologías, fuentes de datos y esquemas de ponderación (Hughes et al., 2021). Esta falta de estandarización dificulta la comparación directa de los resultados entre los distintos estudios. Además, también puede afectar la selección de las métricas de rendimiento financiero, así como el ámbito geográfico y el enfoque sectorial de los estudios (Zhao et al., 2018).

Aunque los estudios de correlación aportan información valiosa sobre la asociación entre los resultados ESG y los rendimientos financieros, para una comprensión profunda es importante establecer la causalidad. En este sentido hay un debate: ¿un desempeño superior en ESG conduce a mejores rendimientos financieros, o las empresas con éxito financiero tienen más recursos para invertir en iniciativas ESG? Hay estudios que sugieren una relación causal positiva, argumentando que un desempeño superior en ESG mejora la reputación de una empresa, fortalece las relaciones con los *stakeholders* y reduce los riesgos operativos, lo que conduce a un mejor rendimiento financiero (Ademi y Klungseth, 2022). Este punto de vista refuerza la idea de que incorporar consideraciones ESG a la estrategia empresarial crea valor a largo plazo. En cambio, algunas investigaciones plantean que el vínculo podría deberse principalmente al buen desempeño financiero, indicando que las compañías con solidez económica cuentan con mayores recursos para invertir en iniciativas ESG (Chams et al., 2021). Esta visión destaca la posibilidad del *greenwashing*, que consiste en la divulgación selectiva de información positiva sobre el desempeño de una empresa en materia ESG, sin divulgar completamente la información negativa, con el objetivo de crear una imagen corporativa excesivamente positiva (Lyon y Maxwell, 2011).

Varios mecanismos de mediación para explicar de qué forma los resultados ESG pueden tener un impacto en los resultados financieros. La mejora de la reputación y las mejoras en las relaciones con los *stakeholders* se mencionan con frecuencia como factores mediadores importantes. Un perfil ESG positivo puede atraer a los inversores que priorizan la sostenibilidad, puede mejorar la imagen de marca, aumentar la fidelidad del cliente, lo que puede crear una mejora de los sistemas financieros. Además, la reducción de los riesgos operativos asociados a las cuestiones sociales y medioambientales puede suponer un ahorro de costes, una mejora de la eficiencia y otras ventajas financieras (Khan et al., 2016). Por ejemplo, las empresas que son capaces de gestionar los riesgos medioambientales pueden evitar multas o demandas, mientras que aquellas que priorizan el bienestar de sus empleados pueden ver disminuir la rotación de personal y mejorar la productividad. La innovación en tecnologías sostenibles puede también ser considerada como factor mediador, permitiendo a las empresas la capacidad de desarrollar nuevos productos y servicios, un mejor acceso a nuevos mercados y conseguir una ventaja competitiva. Por último, unas prácticas de gobernanza sólidas pueden potenciar la eficiencia operativa, reducir costes de agencia y aumentar la transparencia (Zheng et al., 2022).

La relación entre las puntuaciones ESG y los rendimientos financieros no es uniforme en todos los contextos y hay varios factores que pueden configurar esta relación, influyendo en la fuerza y dirección de la asociación. Un factor crucial es el

tamaño de la empresa, empresas más grandes suelen tener más recursos para invertir en iniciativas ESG y pueden experimentar una relación positiva más fuerte entre el rendimiento financiero y las puntuaciones ESG (Kim y Li, 2021). El sector al que pertenece la empresa y sus características también desempeña un papel importante, ya que la relevancia de los factores ESG puede variar de forma considerable entre los distintos sectores (Koundouri et al., 2022). Otro factor es la calidad y disponibilidad de los datos, así como la elección de los proveedores de calificación y el nivel de divulgación ESG (Hughes et al., 2021).

La investigación existente sobre ESG y rendimiento financiero ha examinado el impacto de ESG en varios indicadores financieros. El rendimiento de activos (ROA), rendimiento de fondos propios (ROE), la Q de Tobin (ratio económico que compara el valor mercado de una empresa con su valor contable) y los ratios entre el mercado y los libros son medidas de rendimiento muy utilizadas (Ademi y Klungseth, 2022; Ionescu et al., 2019). No obstante, la magnitud de esta relación varía de forma considerable en distintos estudios y contextos. Otros estudios han explorado el impacto de los ESG en las calificaciones crediticias, el coste de capital y la rentabilidad de las acciones (Kim y Li, 2021; Hübel, 2022).

Esta relación entre los factores ESG y los rendimientos financieros no es uniforme en todos los sectores. Las características específicas de cada sector pueden influir en la relevancia y el impacto de los factores. En la literatura, podemos encontrar estudios enfocados en distintos sectores como el financiero (Kabir y Chowdhury, 2023), energía (Liu et al., 2022), salud (Kalia y Aggarwal, 2023) o en el sector público (Domanović, 2022). A parte de sectoriales, podemos encontrar estudios que se centran en distintos puntos geográficos, como en India (Dalal y Thaker, 2019), Estambul (Saygili et al., 2022), Noruega (Giannopoulos et al., 2022) o Europa (Koundouri et al., 2022).

La importancia y el impacto de los principios ESG varían significativamente entre las economías avanzadas y las economías en desarrollo, debido a diferencias en los marcos institucionales, las regulaciones, la sensibilización de los inversores y los niveles de crecimiento económico. La literatura científica sobre el rendimiento de la inversión ESG se concentra principalmente en los mercados financieros desarrollados, mientras que la investigación en mercados emergentes sigue siendo limitada, con la notable excepción de China. No obstante, países en desarrollo como China, India y Malasia están fortaleciendo sus marcos legales para los informes ESG, lo que evidencia una evolución en estas áreas (Citalán et al., 2025). Existen estudios específicos que han analizado la situación en países como China (Zhao et al., 2018; Zheng et al., 2022; Tang y Yang, 2023), Rusia (Koroleva et al., 2020; Basnet et al., 2022) y Egipto (Aboud y Diab, 2019), entre otros. Asimismo, los inversionistas ya consideran los criterios ESG en sus análisis de mercados emergentes para comprender mejor los aspectos culturales, políticos y sociales propios de cada contexto. Los resultados globales sobre el rendimiento de la inversión ESG son heterogéneos y dependen de la geografía; en particular, se ha encontrado que en algunos mercados emergentes, como Brasil, China, Rusia e India, las empresas con buenas prácticas ESG tienden a mostrar un rendimiento inferior al de las emisoras convencionales (Citalán et al., 2025).

A pesar de todos los avances que se han logrado en los últimos años, aún nos encontramos con ciertas limitaciones. La ausencia de métricas y bases de datos estandarizadas representa un desafío significativo, puesto que complica la comparación de resultados entre los distintos estudios (Hughes et al., 2021). Las diversas agencias de calificación ESG, utilizan metodologías, fuentes de información y esquemas de ponderación distintos, lo que hace más difícil sacar conclusiones de los estudios. La discusión de la causalidad continúa siendo un tema central de debate y la dirección de esta causalidad entre las métricas ESG y los rendimientos financieros sigue sin estar clara en muchos casos (Chams et al., 2021).

2.3. El uso de la minería de textos en el análisis de los reportes sostenibles

La minería de textos, o análisis de datos textuales, se utiliza cada vez más con el objetivo de analizar los informes de sostenibilidad o reportes de información no financiera de las empresas (Aureli, 2017; Contala et al., 2024). Puesto que se ha incrementado el número de informes de sostenibilidad, el interés de los *stakeholders* (Zhou et al., 2021) y el volumen de datos, el análisis sistemático de estos informes no puede realizarse de forma manual (Contala et al., 2024). En este sentido, la minería de textos ofrece un enfoque organizado para extraer información y tendencias de grandes volúmenes de datos textuales (Aureli, 2017), lo que permite explorar información no estructurada o semiestructurada (Zhou et al., 2021).

Esto permite a profesionales e investigadores identificar temas emergentes (Liew et al., 2014), evaluar el rendimiento en materia ESG (Caudron y Vrms, 2022) y comprender las intenciones estratégicas que están detrás de las decisiones de sostenibilidad (Ning et al., 2021). Este enfoque, ayuda a obtener información sobre el contenido y la estructura de los informes, así como a comprender el desarrollo de los problemas a lo largo del tiempo (Contala et al., 2024). Esto ayuda a superar las limitaciones del análisis de contenido manual tradicional, ya que procesa los informes de manera consistente, minimizando la subjetividad humana (Kiriú y Nozaki, 2020) y proporcionando un enfoque automatizado que supera los problemas asociados con la fiabilidad del análisis de contenido (Aureli, 2017). Además, estas herramientas pueden identificar temas que podrían pasarse por alto con una codificación manual, reduciendo el tiempo de análisis (Vaio et al., 2024) y ofreciendo una manera más eficiente y completa de evaluar los esfuerzos de sostenibilidad de las empresas (Contala et al., 2024).

Las técnicas de minería de textos se aplican de varias formas para analizar estos reportes, ofreciendo una comprensión más profunda de las prácticas de sostenibilidad corporativas (Kang y Kim, 2022). Esto incluye modelización de temas, análisis de sentimientos y la identificación de indicadores clave de rendimiento (Caudron y Vrms, 2022). El modelado temático es una aplicación de la minería de textos que ayuda a identificar los temas y asuntos principales que se tratan en estos reportes (Contala et al., 2024). Un algoritmo muy utilizado con este propósito es Asignación Latente de Dirichlet (LDA), que se utiliza para descubrir estructuras temáticas ocultas en una colección de documentos. Asume que cada

documento es una mezcla de unos cuantos temas latentes, y cada tema se caracteriza por una distribución de probabilidad sobre las palabras (Zhou et al., 2021). En este contexto, múltiples estudios han utilizado este algoritmo para sus investigaciones, Goloshchapova et al., 2019 utilizaron el análisis textual con LDA para detectar el número y la naturaleza de los temas incluidos en los informes de CSR de empresas de Reino Unido y Europa e identificar posibles sesgos sectoriales. Encontraron temas comunes como “seguridad de los empleados”, “apoyo a la formación de empleados”, “derechos humanos” y “atención médica y medicina”.

El análisis de sentimientos es el procesamiento computacional de los sentimientos y la subjetividad del texto. Esto implica determinar el tono o la emoción expresada en un texto. Una de sus aplicaciones principales es la detección de polaridad, que mide la tonalidad de un documento, la cual puede ser positiva, negativa o neutral. Esta técnica puede ser usada para evaluar cómo comunican las empresas sus esfuerzos de sostenibilidad y cómo los perciben los *stakeholders*. Mediante el análisis del sentimiento expresado en las redes sociales y otras plataformas, las empresas pueden obtener un entendimiento objetivo de su rendimiento en materia ESG y mejorar sus estrategias de comunicación (Kang y Kim, 2022; Lu y Jagoda, 2023). Anderson et al., 2024 evalúan y comparan la eficacia de una variedad de modelos de análisis de sentimientos para observar cómo la gente percibe y reacciona a las cuestiones ambientales y los esfuerzos de sostenibilidad, identificando diferentes sentimientos asociados con el discurso.

La minería de textos también puede ser utilizada para extraer indicadores clave de rendimiento asociados a la divulgación empresarial. Analizando la frecuencia y el contexto de los temas específicos, los investigadores pueden evaluar el rendimiento de una empresa en ámbitos como el impacto medioambiental, la responsabilidad social y la gobernanza. Esto permite una evaluación más objetiva y cuantitativa de los esfuerzos sostenibles que realizan las empresas, ayudando a los *stakeholders* a tomar decisiones informadas (Zhou et al., 2021). En este sentido, Ning et al., 2021, analiza cómo la divulgación voluntaria de información sobre sostenibilidad refleja la estrategia empresarial y su efecto en los resultados financieros. Para ello, utiliza LDA para identificar temas clave los informes, con un modelo de Markov latente comprende la relación con la intención estratégica y realiza un análisis empírico utilizando modelos estadísticos para evaluar cómo repercuten estas intenciones en los resultados financieros. El estudio concluye que la orientación al empleado tiene mayor impacto en el rendimiento que al cliente.

Para extraer la información de los reportes de sostenibilidad son muchas las herramientas y técnicas utilizadas. Estas incluyen enfoques basados en diccionarios, algoritmos de *Machine Learning (ML)* y software especializado. Los enfoques basados en diccionarios implican utilizar listas predefinidas de palabras y frases relacionadas con el tema en estudio para identificar información relevante en el texto. Baier et al., 2020 crea un diccionario específico de términos relacionados con ESG de 482 palabras, organizado en 40 subcategorías, lo que facilita un análisis detallado de los diferentes aspectos relacionados con ESG y lo emplea para evaluar de forma objetiva la actividad de reporte ESG en las empresas. Algoritmos de ML como LDA, *Support Vector Machines (SVM)* y redes neuronales, se utilizan para identificar de forma automática patrones y relaciones en los textos. Estos

algoritmos pueden ser entrenados para clasificar texto, extraer información o realizar predicciones. Raman et al. (2020) desarrolló un modelo de clasificación que categorizaba la relevancia de frases para ESG, utilizando un conjunto de datos de informes de sostenibilidad preentrenados y ajustados. Raghupathi et al. 2020 utilizaron una combinación de enfoques analíticos de texto, incluidos el agrupamiento y la clasificación para extraer información de las resoluciones de sostenibilidad de los accionistas.

En la actualidad, hay distintos softwares que se pueden utilizar para analizar contenido de minería de textos. Leximancer es un software especializado utilizado para el análisis automatizado de contenido. Está basado en la teoría bayesiana para identificar conceptos y mapear temas en datos textuales (Vaio et al., 2024). DICTION es un programa de análisis de contenido basado en diccionarios. Utiliza 33 diccionarios distintos y algoritmos con ponderación estadística para calcular puntuaciones (Kim y Kim, 2017). En cuanto a lenguajes de programación hay varias bibliotecas que se pueden utilizar. En R, el paquete *tm*, se utiliza específicamente para aplicar vocabularios a los reportes (Aureli, 2017), en Python la biblioteca *sentence transformer* permite trabajar con modelos pre-entrenados avanzados como BERT, DistilBERT y MiniLM para la creación de vectores de texto y análisis de sentimiento (Kang y Kim, 2022). Estos programas y lenguajes proporcionan las herramientas necesarias para implementar las técnicas de minería de textos y análisis de sentimientos, desde el preprocesamiento de los datos textuales hasta la aplicación de algoritmos de modelado temático, clasificación o análisis léxico-basado (Aureli, 2017).

La aplicación de la minería de textos en reportes de sostenibilidad ofrece grandes beneficios como el incremento de la eficiencia o de la objetividad. Además, automatiza el proceso de análisis de grandes volúmenes de datos de texto, lo que permite a los investigadores a identificar más rápidamente tendencias y temas clave, permitiendo una toma de decisiones más oportuna e informada (Zhou et al., 2021). También reduce el sesgo de subjetividad en el análisis de los informes, proporcionando una evaluación más objetiva de los resultados de sostenibilidad de las empresas (Kiriú y Nozaki, 2020). Utilizando diccionarios predefinidos y algoritmos de ML, el análisis se basa en datos cuantitativos y patrones estadísticos, en lugar de interpretaciones individuales (Kim y Kim, 2017).

La minería de textos puede descubrir patrones y relaciones ocultas que podrían no ser evidentes mediante un análisis manual (Vaio et al., 2024). De esta forma, se pueden obtener nuevos conocimientos sobre los factores que impulsan los resultados en materia de sostenibilidad, la eficacia de las distintas estrategias y la adecuación de la información divulgada por las empresas a las expectativas de los *stakeholders*. Ning et al., 2021 destacaron el uso de enfoques analíticos empresariales avanzados, concretamente, el modelado temático LDA, para codificar el conocimiento intangible incluido en los informes anuales de sostenibilidad con el objetivo de inferir la intención estratégica de una empresa detrás de la divulgación voluntaria.

La minería de textos efectiva requiere un entendimiento profundo del tema que se está analizando, incluyendo los conceptos clave, la terminología y los marcos

de reporte que se utilizan. Sin este conocimiento, es difícil desarrollar diccionarios apropiados, entrenar algoritmos de machine learning e interpretar los resultados del análisis. En ámbitos como la sostenibilidad ambiental, es importante dominar el tema (Aureli, 2017). La interpretación de los resultados, incluida la asignación de etiquetas a los temas identificados puede ser subjetiva si se basa únicamente en la opinión de investigador. Es importante que la interpretación tenga en cuenta el contexto en el que nos encontramos para no sacar conclusiones erróneas (Kang y Kim, 2022).

Herramientas de Inteligencia Artificial Generativa basadas en Grandes Modelos de Lenguaje (LLM) poseen el potencial de automatizar diversas tareas, lo que podría aplicarse a la generación de reportes. No obstante, la adopción de la IA Generativa suscita preocupaciones importantes. Estos sistemas, aunque capaces de generar contenido novedoso, pueden producir respuestas plausibles pero incorrectas, a menudo debido a fuentes de datos sesgadas, incompletas o erróneas con las que fueron entrenados. Esto introduce el riesgo de sesgo algorítmico, que puede amplificar patrones existentes o incluso generar contenido no ético. Esto motiva preocupación por el potencial del *greenwashing* y enfatiza la necesidad crítica de mecanismos de garantía sólidos. En consecuencia, una competencia clave para detectar anomalías es el juicio profesional, requiriendo habilidades analíticas avanzadas y pensamiento crítico (Anica-Popa et al., 2024).

Integrar la minería de textos con otras fuentes de datos, como puede ser datos financieros, medioambientales y de redes sociales, puede proporcionar una visión integral de los rendimientos sostenibles de las empresas. Esto puede ayudarnos a identificar vínculos entre los esfuerzos sostenibles y los resultados financieros, evaluando si el impacto de las actividades corporativas en el medioambiente y la sociedad, y entendiendo las percepciones de los *stakeholders* sobre los rendimientos sostenibles (Aureli, 2017).

2.4. Hipótesis del estudio econométrico tras la revisión del estado del arte

A partir de la revisión de la literatura académica sobre la relación entre el desempeño ESG y la rentabilidad financiera (apartado 2.2), se formulan las siguientes hipótesis que serán contrastadas empíricamente mediante modelos de datos de panel en la sección de resultados. Estas hipótesis se corresponden con el primer objetivo del estudio (ver sección 1.1), centrado en evaluar cuantitativamente dicha relación en el contexto de las empresas del EuroStoxx 50.

- H1: El desempeño ESG de las empresas tiene un efecto positivo sobre su rentabilidad financiera.
- H2: La dimensión ambiental del desempeño ESG tiene un efecto positivo sobre su rentabilidad financiera.
- H3: La dimensión social del desempeño ESG tiene un efecto positivo sobre su rentabilidad financiera.

- H4: La dimensión de gobernanza del desempeño ESG tiene un efecto positivo sobre su rentabilidad financiera.
- H5: Las controversias tienen un efecto negativo sobre la rentabilidad financiera de las empresas.

En todos los casos, la hipótesis nula (H_0) plantea la ausencia de efecto significativo entre las variables consideradas.

3. Material y métodos

3.1. Base de datos

Para hacer el análisis de las métricas ESG junto con los rendimientos financieros de las empresas del Eurostoxx 50, se han descargado los datos del espacio de trabajo de LSEG¹². La base de datos cuenta con 5 variables sobre las métricas ESG y con 12 de rendimientos financieros. Además, hay otras 4 variables que contienen el nombre de la empresa, el sector al que pertenece, el país y el año al que corresponden los datos de las variables. Los datos se corresponden a los últimos 7 años fiscales de los que hay datos disponibles, es decir, de 2017 a 2023. El año fiscal 2024 aún no se incluye, ya que los datos oficiales suelen publicarse entre mayo y junio del año siguiente. Por lo tanto, la base de datos está compuesta por 350 individuos y 24 variables.

La Figura 2 muestra un esquema de cómo LSEG evalúa el desempeño sostenible de las empresas a través de un sistema basado en tres pilares principales: ambiental, social y de gobernanza¹³. Cada pilar agrupa diferentes categorías, como uso de recursos y emisiones (ambiental), derechos humanos y comunidad (social), o gestión y responsabilidad social corporativa (gobernanza). Además, se incorpora una puntuación específica de controversias, que refleja incidentes negativos en todas las categorías ESG.

Para calcular la puntuación ESGC, se parte de más de 750 métricas, de las cuales se seleccionan 186 medidas comparables. Estas se agrupan en los tres pilares para obtener la puntuación ESG principal, y las controversias se suman en una puntuación adicional. Finalmente, la puntuación ESGC combina el desempeño ESG y la gestión de controversias, ofreciendo una visión integral y comparativa de la sostenibilidad y los riesgos reputacionales de cada empresa.

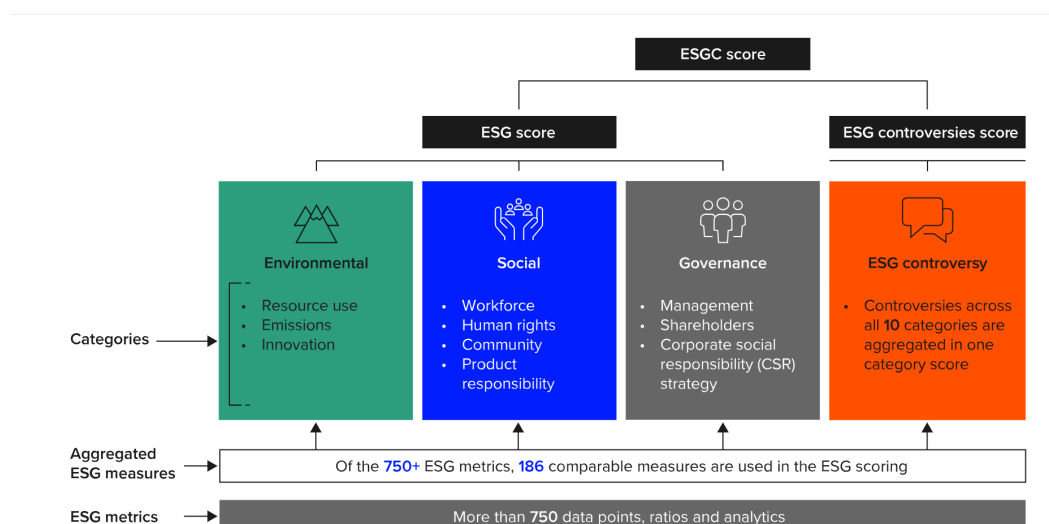


Figura 2. Metodología LSEG para asignar puntuaciones ESG

¹²<https://www.lseg.com/en/data-analytics>

¹³<https://www.lseg.com/en/data-analytics/sustainable-finance/esg-scores#methodology>

Categoría	Variable
Identificación	Identificador
	Nombre de la empresa
	Sector económico
	País de cotización
	Año
Variables ESG	Puntuación ESG
	Componente social
	Componente gobernanza
	Componente ambiental
	Controversias
Variables financieras	Retorno sobre activos (ROA)
	Retorno sobre patrimonio (ROE)
	Relación precio-valor contable
	Ingresos por actividades
	Flujo de caja
	Pasivos totales
	Activos totales
	Índice de cobertura de intereses
	Deuda total
	Costo ponderado de la deuda
	Otros activos no corrientes
	Activos corrientes
	Deuda sobre activos
	Ratio corriente

Tabla 1. Clasificación de variables del dataset

Para el apartado de minería de textos de los reportes de sostenibilidad en inglés de estas mismas empresas, se descargaron los archivos de la página web de cada empresa en el apartado correspondiente en el que publican los reportes. Se descargaron 316 reportes, los 34 faltantes se debe a que hay empresas que no los tienen publicados, otros no estaban en inglés o no fue posible acceder a su contenido.

3.2. Metodología

3.2.1. Análisis de panel

Con el objetivo de analizar la relación entre las métricas de desempeño ESG y los rendimientos financieros de las empresas del EuroStoxx 50, se realizó un análisis de panel. El análisis de datos de panel es una forma de análisis de datos longitudinales, que dota al análisis de regresión de una dimensión espacial y temporal. La dimensión espacial se refiere a un conjunto de unidades transversales de observación. Pueden ser países, estados, condados, empresas, materias primas, grupos de personas o individuos. La dimensión temporal se refiere a las observaciones periódicas de un conjunto de variables que caracterizan estas unidades transversales a lo largo de un periodo de tiempo determinado (Baltagi, 2008).

En este caso, el análisis se realiza sobre el panel de empresas del EuroStoxx50 observadas a lo largo de siete años, de 2017 a 2023. Para cada empresa y año, se dispone de una serie de variables financieras y ESG relevantes, tanto dependientes como independientes, algunas de las cuales se han transformado para capturar posibles relaciones no lineales.

Variables de control

Siguiendo el artículo de Chen et al., 2023 se han seleccionado las siguientes variables, que incluyen la relación precio-valor contable (PB), la rotación de activos (GR), la capacidad de flujo de caja (OCFA), el ratio de endeudamiento (LEV), la intensidad de activos fijos (TANG) y los tiempos de intereses ganados (TIE).

Tabla 2. Definición de variables clave (Chen et al., 2023)

Tipo de variable	Nombre de la variable	Símbolo	Descripción
Variables respuesta	Desempeño corporativo	ROA	Beneficio total dividido entre activos totales, métrica estándar de rentabilidad.
	Rentabilidad sobre el patrimonio	ROE	Beneficio neto dividido entre el patrimonio neto, indicador clave de rentabilidad para los accionistas.
Variable explicativa	Calificación ESG	ESG	Puntuación combinada de factores medioambientales, sociales y de gobernanza.
	Factor ambiental	ENV	Componente relacionado con el impacto medioambiental según los criterios ESG.
	Factor social	SOC	Componente relacionado con el impacto social según los criterios ESG.
	Factor de gobernanza	GOV	Componente relacionado con las prácticas de gobernanza según los criterios ESG.
	Controversias	CONT	Puntuación relacionada con las controversias o escándalos asociados con la empresa.
Variables de control	Relación precio/valor contable	PB	Ratio precio de mercado de la acción sobre valor contable del capital social.
	Rotación de activos	GR	Tasa de crecimiento interanual de los ingresos.
	Capacidad de flujo de caja	OCFA	Flujo de caja operativo neto sobre activos totales.
	Ratio de endeudamiento	LEV	Pasivos totales divididos entre activos totales.
	Intensidad de activos fijos	TANG	Activos fijos netos sobre activos totales.
	Índice de cobertura de intereses	TIE	Cociente EBIT sobre gastos financieros por intereses.

Algunas de estas variables no estaban disponibles en la base de datos de LSEG, por lo que se procedió a calcularlas a partir de otras que sí estaban disponibles.

Para este cálculo, se han utilizado ratios ampliamente aceptados en la literatura de análisis financiero (Damodaran, 2012; Brealey et al., 2014). Los ratios se han calculado de la siguiente manera:

- Rotación de activos:

$$GR_{it} = \frac{\text{Ingresos operativos}_{it}}{\text{Activos totales}_{it}} \quad (1)$$

- Capacidad de flujo de caja

$$OCFA_{it} = \frac{\text{Flujo de caja operativo}_{it}}{\text{Activos totales}_{it}} \quad (2)$$

- Ratio de endeudamiento

$$LEV_{it} = \frac{\text{Pasivos totales}_{it}}{\text{Activos totales}_{it}} \quad (3)$$

- Intensidad de activos fijos

$$TANG_{it} = \frac{\text{Activos fijos}_{it}}{\text{Activos totales}_{it}} \quad (4)$$

Estructura de los datos de panel

Una vez preparados los datos, se organiza la base en formato panel, identificando de manera única cada empresa (entidad) y cada año (tiempo). Esto se realiza mediante la creación de un índice compuesto por el nombre de la empresa y el año. Se verifica el número de observaciones, empresas y periodos disponibles, asegurando la viabilidad del análisis longitudinal. Una de las ventajas más importantes de este tipo de datos con respecto a otros es que permiten controlar diferencias inobservables constantes entre las unidades de análisis, lo cual mejora la calidad de las estimaciones econométricas al evitar sesgos por variables omitidas que no varían en el tiempo (Arellano, 2003).

Verificación de supuestos básicos

Antes de la estimación de modelos, se realiza una verificación de los supuestos clásicos de la regresión, como son la linealidad, la multicolinealidad y la normalidad, así como las estadísticas descriptivas. Para la linealidad se analizan las correlaciones entre la variable dependiente y las independientes. Para la multicolinealidad calculamos el Factor de Inflación de la Varianza (VIF) para cada variable explicativa, valores menores a 5 indican ausencia de multicolinealidad problemática. Para la normalidad se evalúa la normalidad de la variable dependiente mediante los test de Shapiro-Wilk y Jarque-Bera.

Para detectar si hay homocedasticidad, es decir, la varianza de los residuos es constante, o heterocedasticidad, que depende de los valores de las variables

explicativas, hemos aplicado el test de Breusch-Pagan (Breusch y Pagan, 1979). En los modelos clásicos de regresión se asume que los residuos (u_{it}) tienen varianza constante:

$$\text{Var}(u_{it}) = \sigma^2 \quad \forall i, t \quad (5)$$

Si esta condición no se cumple, los errores son heterocedásticos, lo que puede sesgar las inferencias estadísticas. Para la realización de este test, primero se estima el modelo de panel y se obtienen los residuos $\hat{u}_{it}$. A continuación, se realiza una regresión auxiliar, donde se modelan los residuos al cuadrado en función de las variables explicativas originales:

$$\hat{u}_{it}^2 = \alpha_0 + \sum_{j=1}^K \alpha_j x_{jit} + v_{it} \quad (6)$$

El estadístico de contraste para este test:

$$LM = n \cdot R_{aux}^2 \quad (7)$$

donde R_{aux}^2 es el coeficiente de determinación de la regresión auxiliar y n es el número de observaciones. Bajo la hipótesis nula de homocedasticidad, LM sigue una distribución χ^2 con K grados de libertad.

Para analizar la autocorrelación hemos utilizado el test de Durbin-Watson (Durbin y Watson, 1950), que evalúa si los residuos del modelo están correlacionados en el tiempo, lo que puede afectar la validez de los estimadores y las inferencias. En los modelos de panel, se da autocorrelación cuando:

$$E[u_{it}u_{is}] \neq 0 \quad \text{para } t \neq s \quad (8)$$

El estadístico se calcula como:

$$DW = \frac{\sum_{i=1}^N \sum_{t=2}^T (\hat{u}_{it} - \hat{u}_{i,t-1})^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2} \quad (9)$$

Donde:

- u_{it} son los residuos del modelo estimado para la observación i en el tiempo t
- N es el número total de unidades transversales en el panel, es decir, el número de empresas, individuos o entidades que se están analizando
- es el número total de períodos de tiempo para cada unidad i , el número de observaciones disponibles para cada entidad o individuo a lo largo del tiempo
- u_{it}^2 es el residuo al cuadrado para la observación i en el período t , que representa la varianza de los residuos en cada período

- $u_{i,t-1}$ es el residuo para la misma unidad i , pero en el período anterior $t - 1$

Es decir, el numerador es la suma de los cuadrados de todos los residuos, que captura la magnitud total de los errores en el modelo para todas las observaciones y períodos. Y el denominador es la suma de las diferencias entre los residuos consecutivos (u_{it} y $u_{i,t-1}$) para cada entidad y año. Este término mide la autocorrelación en los residuos, es decir, cómo los errores de un período están relacionados con los errores del período anterior. Si el valor se acerca a 2, no hay autocorrelación, si es menor que 2, hay autocorrelación positiva y si es mayor que 2, hay autocorrelación negativa.

La multicolinealidad se da cuando dos o más variables están altamente correlacionadas entre sí, lo que puede inflar la varianza de los coeficientes estimados y dificultar la interpretación de los resultados. En este sentido, se ha utilizado el Factor de Inflación de la Varianza (VIF, por sus siglas en inglés), que mide cuánto se incrementa la varianza de un estimador debido a la colinealidad con otros predictores (Wooldridge, 2016). Para cada variable x_j , el VIF se calcula como:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (10)$$

donde R_j^2 es el coeficiente de determinación de la regresión de x_j sobre el resto de variables explicativas. Si el VIF tiene un valor de aproximadamente 1, no hay colinealidad, si es mayor que 5, existe colinealidad moderada y en el caso de que sea mayor que 10, se recomendaría revisar la selección de variables o emplear técnicas de regularización, porque la colinealidad sería grave.

Estimación de modelos de datos de panel

Se aplican distintas especificaciones de modelos de datos de panel para analizar la relación entre las variables ESG y financieras siguiendo la metodología de Yaffee, 2005:

Modelo Pooled OLS (Mínimos Cuadrados Ordinarios Agrupados)

El modelo Pooled OLS es el método de mínimos cuadrados ordinarios agrupados. Este modelo ignora la estructura de panel, es decir, asume que no existen diferencias sistemáticas entre empresas ni a lo largo del tiempo. Se estima mediante mínimos cuadrados ordinarios (OLS) agrupando todos los datos, como si se tratara de una única muestra transversal grande.

$$y_{it} = \alpha + \beta' X_{it} + u_{it} \quad (11)$$

Donde:

- y_{it} es la variable dependiente para la empresa i en el año t
- α es el intercepto común para todas las empresas y períodos
- X_{it} es el vector de variables explicativas para la empresa i en el año t

- β' es el vector de coeficiente que muestra cómo las variables explicativas afectan y_{it}
- u_{it} es el término de error aleatorio que captura factores no observados que afectan a y_{it}

Las ventajas de este modelo es que es simple y de fácil interpretación, así como que es útil como punto de partida y para comparar. Sin embargo, si existen efectos no observados específicos de cada empresa, las estimaciones pueden estar sesgadas y ser ineficientes.

Modelo de Efectos Fijos (Fixed Effects)

El modelo de efectos fijos permite que cada empresa tenga su propio intercepto, controlando por factores invariables en el tiempo que podrían estar correlacionados con las variables explicativas. Así, se elimina el sesgo por omisión de variables inobservadas constantes.

$$y_{it} = \alpha_i + \beta'X_{it} + u_{it} \quad (12)$$

Donde:

- y_{it} es la variable dependiente para la empresa i en el año t
- α_i es el intercepto específico para cada empresa, capturando características inobservables que no varían en el tiempo
- X_{it} es el vector de variables explicativas para la empresa i en el año t
- β' es el vector de coeficiente que muestra cómo las variables explicativas afectan y_{it}
- u_{it} es el término de error aleatorio que captura factores no observados que afectan a y_{it}

Se puede estimar restando la media temporal de cada variable (“within transformation”) o incluyendo dummies para cada empresa. La estimación se realiza mediante mínimos cuadrados ordinarios sobre las variables transformadas.

Las ventajas de este modelo es que controla la heterogeneidad no observada entre empresas y es apropiado cuando se sospecha que las características inobservadas están correlacionadas con las variables explicativas. En cambio, no permite estimar el efecto de variables que no varían en el tiempo dentro de cada empresa y puede perder eficiencia si la variabilidad temporal es baja.

Modelo de Efectos Aleatorios (Random Effects)

Este modelo asume que las diferencias individuales entre empresas (μ_i) son aleatorias y no correlacionadas con las variables explicativas. Permite explotar tanto la variabilidad entre empresas como la variabilidad temporal.

$$y_{it} = \alpha + \beta'X_{it} + \mu_i + u_{it} \quad (13)$$

Donde:

- y_{it} es la variable dependiente para la empresa i en el año t
- α es el intercepto común
- X_{it} es el vector de variables explicativas para la empresa i en el año t
- β' es el vector de coeficiente que muestra cómo las variables explicativas afectan y_{it}
- μ_i es el efecto aleatorio individual, inobservable y no correlacionado con X_{it}
- u_{it} es el término de error aleatorio que captura factores no observados que afectan a y_{it}

Para la estimación se utiliza el método de “Generalized Least Squares” (GLS). Es más eficiente que el modelo de efectos fijos si el supuesto de no correlación se cumple. Presenta ventajas como que permite estimar el efecto de variables invariantes en el tiempo y es más eficiente si el supuesto de independencia es válido, pero si μ_i está correlacionado con X_{it} , las estimaciones serán inconsistentes.

La estimación se realiza utilizando la librería *linearmodels* de Python, seleccionando el tipo de covarianza adecuado para robustecer los errores estándar.

Selección del modelo adecuado

Para determinar si es más adecuado utilizar un modelo de efectos fijos o uno de efectos aleatorios, se utilizó el test de Hausman (Hausman, 1978), que compara los estimadores de ambos modelos. Para calcular este estadístico:

$$H = (\hat{\beta}_{RE} - \hat{\beta}_{FE})' [\text{Var}(\hat{\beta}_{FE}) - \text{Var}(\hat{\beta}_{RE})]^{-1} (\hat{\beta}_{RE} - \hat{\beta}_{FE}) \quad (14)$$

Donde:

- $\hat{\beta}_{RE}$ es el estimador de los efectos aleatorios.
- $\hat{\beta}_{FE}$ es el estimador de los efectos fijos.
- $\text{Var}(\hat{\beta}_{RE})$ es la matriz de varianza-covarianza del estimador de efectos aleatorios.
- $\text{Var}(\hat{\beta}_{FE})$ es la matriz de varianza-covarianza del estimador de efectos fijos.

Un valor significativo indica que son preferibles efectos fijos, puesto que hay correlación entre los efectos individuales y las variables explicativas.

Aunque el test de Hausman es el criterio estadístico para elegir entre efectos fijos y aleatorios, la estimación de los modelos Pooled OLS, efectos fijos y efectos aleatorios es fundamental para comparar resultados, entender la estructura de los datos y justificar la elección del modelo final. Además, permite detectar posibles sesgos y problemas de especificación, aportando rigor y transparencia al análisis.

Bondad de ajuste

En los modelos de panel, la bondad de ajuste se evalúa con el R^2 , que representa la proporción de la varianza total explicada por el modelo, considerando tanto la variación dentro como entre empresas (Baltagi, 2008). Se calcula:

$$R^2_{\text{overall}} = 1 - \frac{\sum_{i,t} (y_{it} - \hat{y}_{it})^2}{\sum_{i,t} (y_{it} - \bar{y})^2} \quad (15)$$

Donde $\hat{y}_{it}$ es el valor ajustado para cada observación. Da una visión global, aunque puede ser menos informativo en presencia de efectos fuertes. En este caso particular, el R^2 no es la mejor medida de fiabilidad del modelo al tratarse de un modelo no lineal, por eso también se calculan los criterios de información.

R^2_{within} mide la proporción de la varianza explicada por el modelo considerando únicamente la variación temporal dentro de cada empresa, es decir, cómo los cambios a lo largo del tiempo para cada entidad son explicados por el modelo. Se calcula como:

$$R^2_{\text{within}} = 1 - \frac{\sum_{i,t} (y_{it} - \hat{y}_{it})^2}{\sum_{i,t} (y_{it} - \bar{y}_i)^2} \quad (16)$$

donde $\bar{y}_i$ es la media de la variable dependiente para cada empresa i . Es especialmente relevante en modelos de efectos fijos, donde el interés está en los cambios internos de cada entidad.

El R^2_{between} refleja la proporción de la varianza explicada por el modelo considerando únicamente las diferencias promedio entre empresas, es decir, cómo las diferencias estructurales entre entidades son explicadas por el modelo. Se calcula como:

$$R^2_{\text{between}} = 1 - \frac{\sum_i (\bar{y}_i - \bar{\tilde{y}})^2}{\sum_i (\bar{y}_i - \bar{y})^2} \quad (17)$$

donde $\bar{y}_i$ es la media de la variable dependiente para la empresa i y $\bar{\tilde{y}}$ es la media de los valores ajustados para la empresa i .

3.2.2. Minería de textos

Para llevar a cabo el análisis de minería de textos se utilizó la misma muestra que para el análisis de panel, es decir, las empresas del EuroStoxx50 entre los años 2017 y 2023. Los reportes de sostenibilidad o no financieros de cada empresa se descargaron de sus páginas web oficiales en inglés. Algunas no estaban en inglés y otras tenían contraseñas, por lo que las frases no pudieron extraerse. En este caso, se ha seguido la metodología que utiliza Ferjančič et al., 2024 en su estudio, abordando una de las futuras líneas de investigación que propone en su estudio:

“en futuras investigaciones sería interesante repetir el análisis de los informes anuales en diferentes regiones”, y se ha ampliado el análisis implementando técnicas adicionales de reducción de la dimensionalidad y visualización. Concretamente, se ha realizado adicionalmente un Análisis de Componentes Principales (PCA, por sus siglas en inglés) y un posterior HJ-Biplot, para visualizar cómo se relacionan las empresas europeas con los distintos temas extraídos, identificando agrupaciones, valores atípicos y posibles estrategias de divulgación diferenciadas. Este enfoque visual y multivariante proporciona una perspectiva novedosa con respecto al estudio original, facilitando la interpretación de la posición relativa de cada empresa en el espacio temático ESG y permitiendo la detección de tendencias o alineamientos sectoriales y geográficos.

Para empezar con el análisis, todos los archivos pdf se convirtieron en archivos de texto (.txt) extrayendo el contenido de cada página pdf mediante la biblioteca PyPDF2. A continuación, se aplicó un proceso de limpieza y filtrado a las frases extraídas de los archivos de texto. Este proceso consistió en separar el texto en frases utilizando expresiones regulares para garantizar que cada frase cumpliera ciertos criterios de calidad, como una proporción mínima de letras respecto a caracteres, una proporción máxima de números respecto a palabras y una longitud mínima de frase. Además, se eliminaron URL y frases con caracteres no deseados, y se comprobó que la puntuación y el formato de las frases fueran correctos.

Modelización de temas

Para modelizar los temas de los informes no financieros de las empresas del EuroStoxx50 se ha utilizado la técnica BERTopic. Se ha seleccionado por sus ventajas frente a otros modelos lingüísticos, como su flexibilidad para trabajar con distintos modelos de incrustación, su capacidad para separar eficazmente las fases de agrupación y representación de temas, y su capacidad para generar representaciones de temas ricas y dinámicas mediante la técnica c-TF-IDF. Estas características permiten una mayor adaptabilidad y precisión en el análisis, especialmente cuando se trabaja con grandes volúmenes de datos no estructurados, como los informes de sostenibilidad. BERTopic es un modelo temático que combina modelos lingüísticos basados en transformadores, técnicas de agrupamiento y una variación de TF-IDF basada en clases para descubrir temas latentes en colecciones de documentos. Para ello, se generan representaciones de documentos (embeddings) mediante modelos lingüísticos preentrenados, como Sentence-BERT (SBERT), para crear representaciones vectoriales densas de cada documento. Se supone que los documentos que tratan del mismo tema son semánticamente similares y, por tanto, próximos en el espacio vectorial. A continuación, se reduce la dimensionalidad de las incrustaciones, por ejemplo mediante UMAP, para optimizar el proceso de agrupación. A continuación, los documentos se agrupan mediante una técnica de clustering como HDBSCAN, en la que cada grupo representa un tema diferente. HDBSCAN permite modelizar el ruido, como los valores atípicos. Para definir qué palabras son las más representativas de cada grupo, se aplica una versión modificada de TF-IDF denominada c-TF-IDF (TF-IDF basado en clases), este procedimiento modela la importancia de las palabras dentro de los grupos de documentos, en lugar de en

los documentos individuales (Grootendorst, 2022).

Evolución temporal de los temas

Para analizar cómo evolucionaron los temas de sostenibilidad a lo largo de los años, se calculó el porcentaje de cada tema en cada informe para cada año. Para ello, se extrajeron las proporciones de cada tema dentro de cada informe y se calcularon las varianzas de estas proporciones a lo largo de los años mediante la fórmula estándar para la varianza muestral:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (p_i - \bar{p})^2 \quad (18)$$

donde p_i es la proporción del tema en el informe del año i , $\bar{p}$ es la proporción media de todos los años, y n es el número de años analizados. Se identificaron los temas con mayor varianza, ya que indican una mayor fluctuación en su relevancia dentro de los informes de sostenibilidad. Por último, estos temas se representaron gráficamente para observar su evolución en el tiempo y detectar posibles cambios en las prioridades a lo largo del tiempo.

Clusterización de temas

Tras la extracción de tópicos, se calcularon los tópicos que más variaban, calculando la varianza de cada uno de ellos y, posteriormente, se realizó un clustering jerárquico de tópicos con el método de Ward (Ward Jr, 1963). El método de Ward busca, en cada paso de aglomeración, fusionar el par de clusters que resulte en la mínima pérdida de información. Esta pérdida se mide como el incremento mínimo de la suma total de cuadrados dentro de los conglomerados (o la varianza ponderada). Minimizar la varianza dentro de los conglomerados equivale a maximizar la varianza entre conglomerados. Antes de la agrupación, los vectores de palabras clave asociados a cada tema se normalizaron para garantizar que los temas con mayor número de palabras o mayor puntuación no dominaran el análisis. El número óptimo de clusters se determinó mediante una combinación de métricas de calidad de la agrupación, como el índice Silhouette y el coeficiente de correlación cofenética, que permitieron encontrar el número de clusters que mejor representaba las relaciones entre temas.

Las distancias entre temas se calcularon utilizando una matriz de similitud basada en la similitud coseno entre los vectores de palabras clave de cada tema. La similitud coseno mide la similitud entre dos vectores y se utiliza habitualmente en el análisis de textos para evaluar el grado de similitud entre dos documentos o, en este caso, dos temas.

Los vectores de cada tema se normalizaron para que todos tuvieran la misma longitud, independientemente del número de palabras clave de cada tema. Para ello, se añadieron ceros a los vectores más pequeños (relleno) o se recortaron los vectores más grandes.

Para calcular la similitud coseno entre los vectores de temas, se utiliza la

siguiente fórmula:

$$\text{cosine similarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (19)$$

donde A y B son los vectores sujetos, $A \cdot B$ son el producto punto de los dos vectores, y $\|A\|$ y $\|B\|$ son las longitudes (normas) de los vectores A y B , respectivamente. La similitud coseno varía entre -1 y 1, siendo el valor 1 que son completamente similares, -1 que son opuestos y 0 que no tienen similitud.

Una vez calculada la similitud coseno entre todos los pares de sujetos, la matriz de similitud resultante se convirtió en una matriz de distancias mediante la siguiente fórmula:

$$\text{distance}(A, B) = 1 - \text{cosine similarity}(A, B) \quad (20)$$

La matriz de distancias tiene valores comprendidos entre 0 y 2, donde 0 indica que los sujetos son idénticos, 1 indica que los sujetos son ortogonales y 2 indica que los sujetos son completamente opuestos. Para garantizar que la matriz de distancias es válida, se asegura que la diagonal de la matriz es exactamente cero, ya que la distancia de un tema a sí mismo debe ser cero. Se garantiza que la matriz de distancias sea simétrica, es decir, que la distancia entre el tema A y el tema B sea igual a la distancia entre el tema B y el tema A . Esta matriz de distancias se utiliza como entrada para el algoritmo de agrupación jerárquica aglomerativa, concretamente con el método de Ward, que busca minimizar la varianza dentro de los conglomerados en cada paso de fusión.

Correlaciones entre temas y puntuaciones ESG

Las correlaciones entre los temas generados por el modelado de temas y las puntuaciones ESG se calcularon mediante el coeficiente de correlación de Pearson, que mide la relación lineal entre dos variables (Pearson, 1896). En primer lugar, se calcularon las correlaciones entre las proporciones de temas generados por el modelo temático y las puntuaciones ESG de las empresas correspondientes. La correlación de Pearson mide la relación lineal entre dos variables y tiene valores en el rango de -1 a 1, donde 1 indica una correlación positiva perfecta, -1 una correlación negativa perfecta y 0 una ausencia de correlación lineal. Su ecuación es:

$$\rho(x, y) = \frac{E[xy]}{\sigma_x \sigma_y} \quad (21)$$

Para cada sector industrial, se calculó la correlación entre las puntuaciones ESG y las proporciones de temas generados por el modelo temático para cada empresa. Las correlaciones se calcularon independientemente para cada sector industrial. Para ello, se filtraron los datos según el sector y se calcularon las correlaciones de Pearson entre las puntuaciones ESG y las proporciones de temas para cada sector. Para garantizar la validez de las correlaciones, se requirieron al menos tres

observaciones por sector. Las correlaciones se analizaron y presentaron por sector, destacando las que eran estadísticamente significativas. Los resultados de las correlaciones se visualizaron mediante mapas de calor y tablas que mostraban las correlaciones por sector y tema, haciendo especial hincapié en las correlaciones más fuertes.

Diferencias en las proporciones de temas entre las empresas con las puntuaciones ESG más altas y más bajas

Para calcular las diferencias en las proporciones de temas entre las empresas con las puntuaciones ESG más altas y más bajas, ajustamos las puntuaciones ESG de las empresas por sector calculando la puntuación ESG media de cada sector. A continuación, calculamos la puntuación ESG corregida de cada informe calculando la diferencia entre la puntuación observada y la puntuación media del sector correspondiente y dividiendo la diferencia por la puntuación media del sector, es decir:

$$ESG_{corr} = \frac{ESG_{observed} - ESG_{sector\ mean}}{ESG_{sector\ mean}} \quad (22)$$

Para ajustar las proporciones de los temas por sector, primero normalizamos las proporciones de los temas ESG dentro de cada informe, de modo que para cada informe, la suma de las proporciones de los 30 temas interpretables sea igual a 100, y luego calculamos las proporciones normalizadas ajustadas por sector restando y dividiendo por la proporción normalizada específica del sector, es decir:

$$Adjusted\ proportion_{sector} = \frac{Proportion_{topic} - Proportion\ average\ sector}{Proportion\ average\ sector} \quad (23)$$

Este análisis se realizó específicamente para el año 2023, ordenamos los informes por puntuación ESG ajustada al sector y extrajimos el 25 % de los informes con las puntuaciones más altas (cuartil superior) y el 25 % de los informes con las puntuaciones más bajas (cuartil inferior) en ESG. A continuación, calculamos la proporción media ajustada por sector para estos dos grupos y medimos la diferencia entre estas proporciones. Al ajustar los datos por sector, se elimina la variabilidad inherente entre los distintos sectores, lo que permite una comparación más equitativa entre las empresas. Este ajuste garantiza que las diferencias observadas en las proporciones de temas ESG no estén sesgadas por las características particulares de cada sector, como el tamaño o la estructura de la empresa, lo que mejora la precisión y pertinencia del análisis. Al seleccionar las mejores y peores empresas (en términos de puntuaciones ESG corregidas) para el año 2023, eliminamos la tendencia general de aumento de las puntuaciones ESG. Además, al centrarnos en las diferencias entre las medias de los cuartiles superior e inferior de los datos (en términos de puntuaciones ESG corregidas) mejoramos el análisis de correlación, ya que mantenemos la magnitud del fenómeno subyacente y abordamos adecuadamente las relaciones lineales o no lineales (mientras que la

correlación de Pearson es simplemente una medida sin unidad de las relaciones lineales).

Análisis de componentes principales and HJ-Biplot

Se aplicó un análisis de componentes principales (ACP) a los 10 temas más frecuentes de los informes de sostenibilidad para reducir la dimensionalidad e identificar estructuras latentes en la información corporativa sobre ESG. El ACP transforma un conjunto de variables correlacionadas (temas) en un nuevo conjunto de componentes ortogonales que explican las principales fuentes de variación de los datos. Cada componente se asocia con un conjunto de cargas que indican la contribución de cada variable original al componente (Oprea, 2022).

Basándose en el ACP, se construyó un diagrama HJ-Biplot (Villardón, 1986) para proporcionar una representación conjunta de las empresas y los temas ESG en el mismo espacio de baja dimensión.

El HJ-Biplot, también conocido como RCMP-Biplot (*Row Column Metric Preserving*), se define como una representación gráfica multivariante de las líneas de una matriz $X_{n \times p}$ (donde n son los individuos/empresas y p las variables/temas) en un espacio de dimensión reducida. Para ello, utiliza marcadores $j_1, \dots, j_n$ para sus filas y $h_1, \dots, h_p$ para sus columnas, seleccionados de tal forma que ambos puedan superponerse en un mismo sistema de referencia con máxima calidad de representación. Este método facilita la interpretación de cómo se posicionan las empresas con respecto a los distintos temas de sostenibilidad. En concreto, la matriz de datos X , donde n es el número de observaciones (empresas) y p el número de variables (temas), se centró en columnas y se estandarizó posiblemente.

El objetivo del HJ-Biplot es aproximar la matriz X utilizando dos matrices de dimensión inferior A y B , de tal forma que el producto AB' se aproxime a X . El resultado es el siguiente

$$X = AB' + E \quad (24)$$

donde E es el error de aproximación. Las matrices A ($n \times q$) y B ($p \times q$) contienen las coordenadas de las empresas y variables, respectivamente, en el espacio reducido.

Estas coordenadas se derivan de la Descomposición en Valores Singulares (SVD) de X , como:

$$A = U(k)\Lambda(k) \quad \text{y} \quad B = V(k)\Lambda(k) \quad (25)$$

donde $U(k)$ y $V(k)$ son las matrices de vectores singulares, y $\Lambda(k)$ es una matriz diagonal con los valores singulares correspondientes a los primeros k componentes principales. La matriz A representa las posiciones de las empresas (observaciones) en el espacio reducido, y B contiene las cargas factoriales de las variables originales

(temas). Si los datos están normalizados, estas cargas pueden interpretarse como la correlación de cada variable con cada componente principal.

Esta representación biplot permite visualizar simultáneamente el posicionamiento de las empresas y las relaciones temáticas, facilitando la identificación de empresas con perfiles ESG similares y los temas específicos de sostenibilidad que conforman su posicionamiento estratégico en el contexto europeo.

Para la realización de todos estos análisis se han utilizado Python y Rstudio.

4. Resultados

Esta sección se divide en tres apartados, en el primero se realiza un análisis exploratorio de la base de datos y posteriormente se presentarán los resultados de todos los análisis realizados. En el segundo apartado se encuentran los datos correspondientes al análisis de las métricas ESG de las empresas escogidas y sus correspondientes rendimientos financieros, y en el último apartado, tenemos los resultados de los análisis realizados a partir de la minería de textos de los informes de las empresas.

4.1. Análisis exploratorio de la base de datos

Para comenzar con este apartado, hemos analizado qué variables tienen valores faltantes. La variable pasivos totales tenía 204 valores faltantes, lo que representaba un 58.29 % del total, un valor preocupante. Los activos corrientes tenían 69 valores perdidos, un 19.71 % y los otros activos no corrientes tenían un 8.57 % de sus valores faltantes. El resto de variables su porcentaje de valores perdidos era menor del 5 %, siendo el de la mayoría de ellas menor del 1 %. En este sentido, todos estos valores perdidos fueron imputados con Random Forest y las distribuciones de las variables antes y después de imputar se encuentran en los anexos. Una vez imputados los datos, podemos proceder al análisis exploratorio.

Es importante explicar la distribución de la base de datos. Si analizamos los datos por país, Francia es el país que más empresas tiene, con 17, seguido de Alemania con 14, Italia y Países Bajos con 5, España 4, Finlandia 3, y Bélgica y Suiza con 1. En cuanto al sector económico, consumo cíclico y el sector financiero cuentan con 10 empresas cada uno, el industrial y tecnológico 7, el no cíclico 5, sanidad y servicios públicos 3 cada uno, materiales básicos y energía 2 y el sector inmobiliario 1.

La tabla 3 presenta un resumen estadístico detallado de las variables financieras y de desempeño ESG incluidas en el análisis. En general, los puntajes promedio de ESG son altos (media de 78.32), indicando que la mayoría de las empresas evaluadas mantienen un desempeño favorable en aspectos ambientales, sociales y de gobernanza. Sin embargo, la desviación estándar de 12.86 y el valor mínimo cercano a 16 muestran la existencia de variabilidad significativa, con algunas compañías mostrando resultados considerablemente inferiores al promedio. Al desglosar las dimensiones del ESG, el componente social exhibe la media más elevada (81.60), seguido por el ambiental (80.93), mientras que la gobernanza presenta una media menor (72.37) y una mayor dispersión, lo que sugiere heterogeneidad en el cumplimiento de prácticas de buen gobierno corporativo. La variable controversias revela una media de 61.39 y una desviación estándar alta (35.50), reflejando diferentes niveles de exposición a eventos adversos o polémicos, con valores máximos que alcanzan la puntuación total, indicando casos con impactos significativos. Respecto a las variables financieras, se observa una amplia heterogeneidad en tamaño y desempeño entre las empresas. Los ingresos por actividades de negocio presentan una media de 4.93×10^{10} con una

desviación estándar similar, evidenciando la inclusión de empresas con volúmenes muy distintos. Esta variabilidad también se refleja en los pasivos y activos totales, lo que implica una muestra representativa tanto de grandes corporaciones como de entidades más pequeñas. El ratio precio-valor contable muestra una media de 3.80 pero una mediana inferior a 2, sugiriendo una distribución sesgada posiblemente por valores extremos. En cuanto a la rentabilidad financiera, el ROA y ROE promedian 5.59 % y 16.22 %, respectivamente, con rangos que incluyen valores negativos, lo cual indica que algunas empresas presentan pérdidas o bajos retornos. El flujo de efectivo incluye casos negativos, reflejando salidas netas en ciertos periodos. La cobertura de intereses varía ampliamente, con valores negativos en casos donde las ganancias antes de intereses y impuestos son insuficientes para cubrir gastos financieros, mientras que el costo promedio ponderado de la deuda (WACC) se mantiene bajo (2.69 %), mostrando condiciones financieras favorables en el periodo estudiado. Finalmente, las partidas relativas a activos no corrientes y corrientes muestran medias y desviaciones estándar elevadas, coherentes con la heterogeneidad observada en la muestra en términos de tamaño y estructura financiera.

Variable	Media	Desv. estándar	Mínimo	Mediana	Máximo
ESG	78.32	12.86	15.96	82.14	100.00
Controversias	61.39	35.50	1.04	70.00	100.00
Social	81.60	15.37	10.28	86.70	97.99
Gobernanza	72.37	17.79	12.20	76.11	85.59
Ambiental	80.93	16.66	0.00	85.00	94.12
ROA	0.0559	0.0461	-0.0064	0.0476	0.0792
ROE	0.1622	0.0996	-0.0174	0.1390	0.2000
Precio sobre valor contable	3.80	5.36	0.37	1.95	47.47
Ingresos por actividades de negocio	4.93e+10	5.06e+10	1.65e+09	3.32e+10	3.22e+11
Flujo de efectivo	7.59e+09	7.98e+09	-1.10e+10	5.36e+09	4.47e+10
Pasivos totales	2.32e+11	4.43e+11	1.50e+08	4.15e+10	2.57e+12
Activos totales	2.65e+11	4.58e+11	1.86e+09	9.41e+10	2.70e+12
Cobertura de intereses	23.90	43.72	-3.66	10.57	405.80
Deuda	4.60e+10	6.57e+10	1.60e+07	1.75e+10	3.35e+11
WACC - coste de la deuda	0.0269	0.0085	0.0070	0.0241	0.0575
Otros activos no corrientes	5.40e+09	1.92e+10	-1.00e+06	1.63e+09	2.09e+11
Activos corrientes	5.87e+10	6.78e+10	7.97e+08	2.40e+10	2.48e+11

Tabla 3. Estadísticas descriptivas de variables financieras y ESG

La Figura 10 muestra la matriz de correlaciones entre las variables ESG, la variable de controversias y diversas métricas financieras. Podemos observar que las variables ESG presentan correlaciones positivas fuertes entre sí, destacando la alta relación entre la puntuación ESG global y social (0.865), así como ESG global y ambiental (0.812); lo que sugiere que las empresas con un buen desempeño en un aspecto ESG, tienden a tener buenos resultados en los otros dos. En cambio, gobernanza no está tan relacionado con el componente social y el medioambiental, teniendo correlaciones positivas pero bajas. La variable de controversias se asocia de forma negativa con las métricas ESG, especialmente con la puntuación global y ambiental. En cuanto a las variables financieras, la rentabilidad sobre activos y sobre patrimonio muestran una correlación positiva entre sí, y también con el ratio

precio/valor contable. Las variables ESG y de rentabilidad financiera tienden a estar débil o moderadamente correlacionadas de forma negativa. La correlación ESG y ROA (-0.271) y ESG y ROE (-0.183) sugieren que un mejor desempeño ESG no se asocia necesariamente con una mayor rentabilidad en nuestra muestra. El flujo de caja tiene correlaciones positivas moderadas con las variables ESG y una correlación negativa notable con controversias, lo que indica que las empresas con menos controversias tienden a tener mejores flujos de caja.

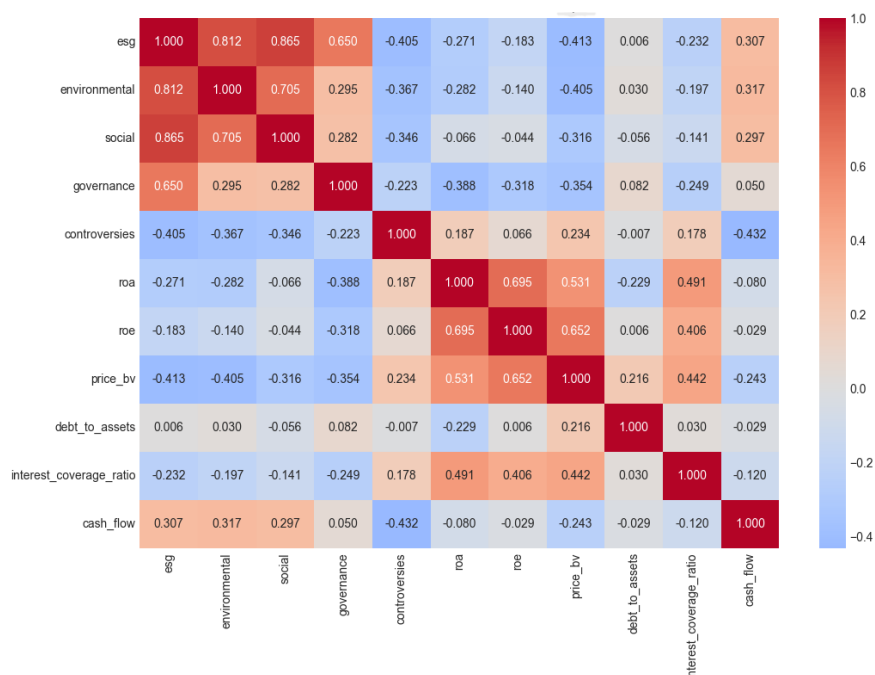


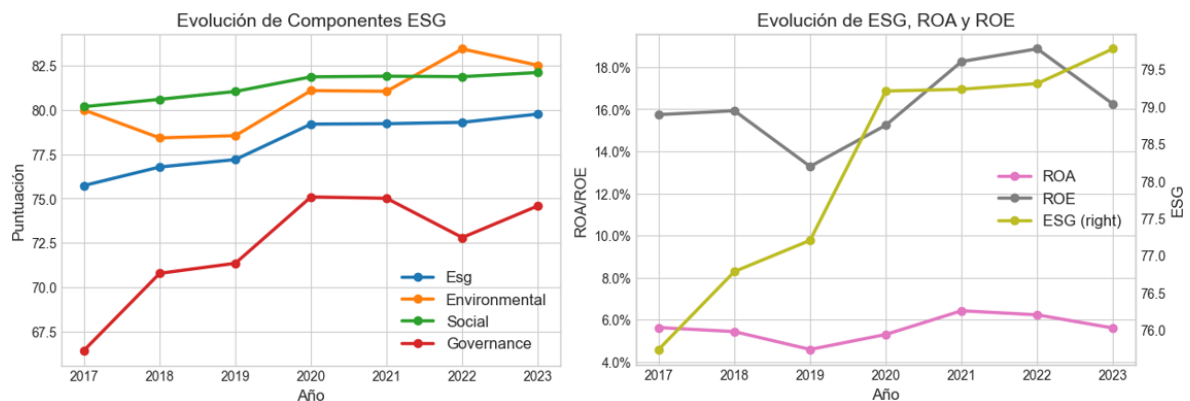
Figura 3. Correlaciones de las variables en estudio

La Tabla 4 revela las diferencias en las medias de los criterios ESG y rentabilidad financiera por sectores financieros. Los sectores energético y financiero destacan por sus altas puntuaciones ESG, con valores superiores a 83, posiblemente por la creciente regulatoria. Sin embargo, no son los sectores más rentables financieramente. En cambio, los sectores de bienes de consumo cíclicos y tecnológico, exhiben un ESG algo inferior, pero alcanzan las rentabilidades más elevadas. Estos resultados sugieren que la integración de criterios ESG debe abordarse con estrategias diferenciadas por sector, puesto que cada industria tiene particularidades propias.

Tabla 4. ESG y rendimiento financiero por sector económico

Sector económico	ESG	ROA	ROE
Energy	83.6521	0.0501	0.1298
Financials	83.4391	0.0079	0.1084
Healthcare	82.9691	0.0573	0.1221
Basic Materials	81.3902	0.0562	0.1277
Consumer Cyclicals	78.6605	0.0888	0.2111
Real Estate	77.4397	0.0296	0.0616
Technology	75.7652	0.0883	0.2068
Industrials	75.5884	0.0562	0.2096
Consumer Non-Cyclical	74.4568	0.0604	0.1397
Utilities	68.9511	0.0335	0.1198

La Figura 4 muestra dos gráficas, la primera con la evolución de componentes ESG a lo largo de los años en estudio, y la segunda, de la evolución de ESG, ROA y ROE. En la primera de ellas se observa que la puntuación ESG ha aumentado progresivamente a lo largo de los años en estudio. Tanto en el componente social como en el medioambiental, las empresas del EuroStoxx 50 tienen una puntuación alta. En cambio, en términos de gobernanza tienen peor puntuación, y aunque desde 2017 ha crecido casi 10 puntos, sigue estando muy por debajo del resto de componentes. Si observamos la evolución de ROA y ROE, su patrón es prácticamente igual. Sin embargo, si comparamos estos patrones con el que sigue ESG, vemos que no siguen el mismo, los rendimientos bajan en 2019, para luego aumentar en 2020 y volver a bajar a partir de 2021 en el caso del ROA, y a partir de 2022 en el caso del ROE.

**Figura 4.** Evolución de las componentes ESG y de los rendimientos financieros

La Figura 5 presenta 4 gráficos en los que podemos ver la distribución de las puntuaciones ESG y de sus componentes, la evolución de las puntuaciones ESG por año y las diferencias entre los sectores económicos.

El histograma muestra que la mayoría de las empresas tienen puntuaciones ESG entre 70 y 90, con un sesgo hacia valores altos. Hay pocos casos con puntuaciones altas, lo que sugiere que las empresas que componen el EuroStoxx 50 tienen buen desempeño ESG. El gráfico de densidad compara la distribución de los

tres componentes ESG. Se observa que los tres presentan una concentración en valores altos, aunque la gobernanza muestra una mayor dispersión y una cola más larga hacia puntuaciones más bajas, indicando variabilidad en ese aspecto, algo que podíamos intuir viendo la evolución de sus puntuaciones en la figura anterior. El primer boxplot muestra la evolución anual de las puntuaciones ESG. Las medianas se mantienen relativamente estables en valores altos, pero hay presencia de outliers en todos los años, especialmente en los valores bajos, lo que indica que aunque la mayoría de las empresas mantienen buenas prácticas en materia ESG, existen algunos casos rezagados. El boxplot inferior muestra la distribución de puntuaciones ESG por sector, en el que se pueden apreciar diferencias. Los sectores servicios y tecnológico presentan mayor dispersión y presencia de outliers con puntuaciones bajas; energía y salud muestran puntuaciones más homogéneas y elevadas; y el financiero y materiales básicos mantienen también puntuaciones altas, con un mayor rango intercuartílico.

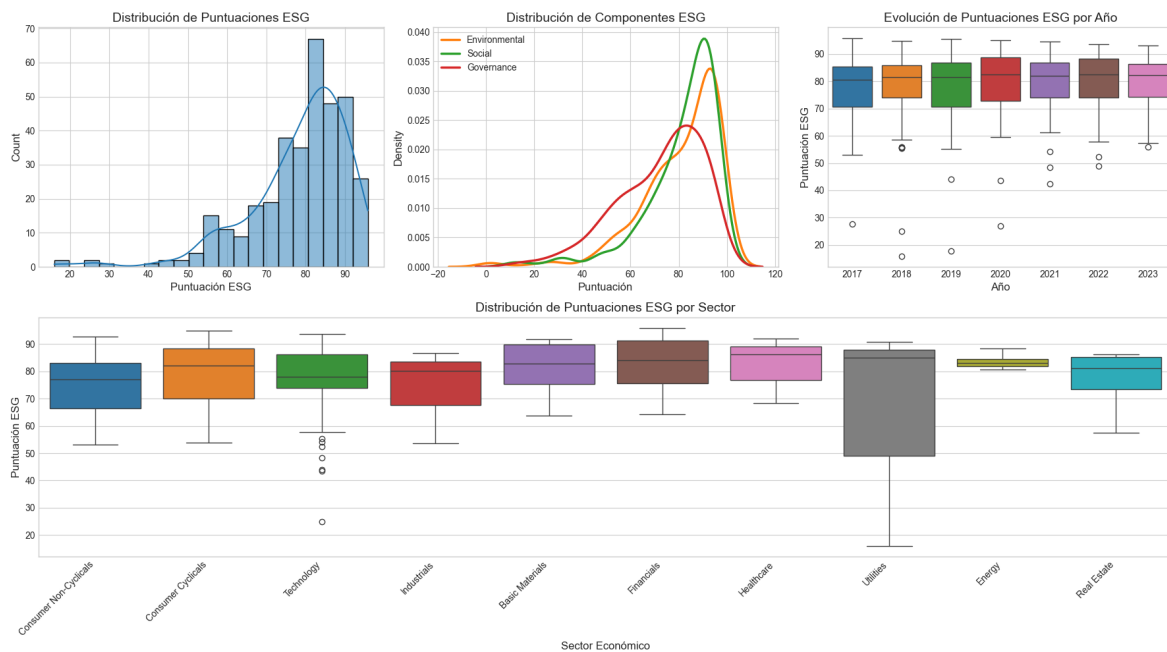


Figura 5. Distribución de las puntuaciones ESG

La Figura 6 presenta dos gráficos de barras que muestran la relación entre el rendimiento financiero, tanto ROA como ROE y la puntuación ESG de las empresas categorizado según su cuartil como bajo (Q4), medio-bajo (Q3), medio-alto (Q2) y alto (Q1). El primer gráfico muestra que el rendimiento sobre activos es mayor en las empresas con puntuaciones más bajas del EuroStoxx 50 y disminuye progresivamente a medida que aumenta la categoría ESG. Las empresas pertenecientes al primer cuartil de ESG presentan el ROA más bajo. El segundo gráfico, que representa el rendimiento sobre el capital, sigue una tendencia similar al ROA, las empresas con menor puntuación de nuestra muestra tienen un ROE más alto, y este indicador disminuye conforme aumenta la categoría ESG. Esto sugiere que las empresas con mejores prácticas ESG no necesariamente muestran una mayor eficiencia en la utilización de sus activos para generar beneficios. Este resultado puede deberse a varios factores, como los costes asociados a la

implementación de políticas ESG, la etapa de madurez de las inversiones sostenibles, o incluso la percepción del mercado sobre el valor de la sostenibilidad.

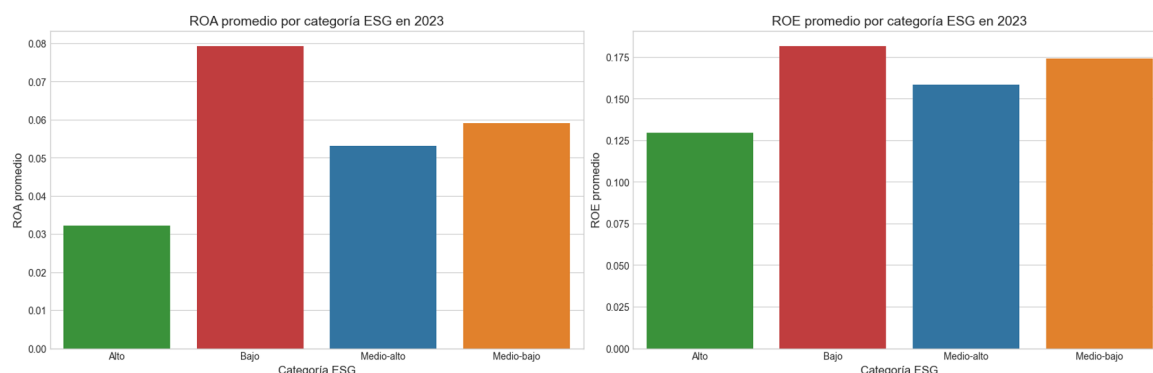


Figura 6. Relación entre las puntuaciones ESG categorizadas y el rendimiento financiero promedio

La Figura 7 presenta la evolución temporal de los rendimientos financieros ROA y ROE y observamos lo que se comentaba en el gráfico anterior, en ambos indicadores financieros las empresas con la puntuación ESG más baja de la muestra presentan sistemáticamente los valores promedio más altos durante todo el periodo analizado. Las empresas con puntuaciones ESG altas muestran los valores promedio más bajos tanto en ROA como en ROE, manteniéndose estables o incluso descendiendo ligeramente a lo largo de los años. Las categorías intermedias (medio-alto y medio-bajo) se sitúan entre ambos extremos, con una tendencia a acercarse más a las empresas de puntuación baja, llegando a superarlo en el caso del ROE en la categoría medio-bajo ESG en 2021 y 2022. También se observa una caída generalizada en los indicadores en 2019, seguida de una recuperación en los años siguientes, especialmente marcada en la categoría medio-bajo de ROE, que alcanza su máximo en 2022. El ROA muestra menos volatilidad que el ROE, pero la jerarquía entre categorías ESG y los patrones seguidos se mantienen en ambos casos. Esto sugiere que, en el periodo analizado, las empresas con peor desempeño ESG han obtenido una rentabilidad financiera superior a las empresas con mejores prácticas ESG. Esta tendencia puede deberse a varios factores: las inversiones en sostenibilidad y buen gobierno pueden tener un impacto negativo en la rentabilidad a corto plazo, factores sectoriales y coyunturales que favorecen a empresas menos comprometidas con criterios ESG.



Figura 7. Evolución de la relación entre las puntuaciones ESG categorizadas y el rendimiento financiero promedio

4.2. Análisis de las métricas ESG y de los rendimientos económicos

4.2.1. Análisis de panel

Se ha realizado un análisis de panel con el objetivo de evaluar la relación entre distintas dimensiones del desempeño ESG (ambiental, social, de gobernanza y controversias) y la rentabilidad financiera de las empresas, medida a través del ROA y el ROE. A lo largo del análisis de panel realizado para los cinco modelos: ESG, ambiental, social, gobernanza y controversias, se han seguido los principales supuestos econométricos que garantizan la validez y robustez de los resultados. El uso de datos de panel ha permitido controlar tanto la heterogeneidad no observada entre empresas como las variaciones temporales, incorporando efectos fijos a nivel de empresa y de tiempo, en línea con las mejores prácticas metodológicas. La elección de esta especificación se fundamentó en la prueba de Hausman, que confirmó en cada caso la idoneidad del modelo de efectos fijos o del de efectos aleatorios. Puesto que en ninguno de los casos fue posible verificar plenamente los supuestos de normalidad y homocedasticidad, se optó por la utilización de errores estándar robustos. En todos los casos el VIF fue menor que 5, por lo que no hubo problemas de multicolinealidad y el estadístico de Durbin-Watson fue cercano a 2, por lo que tampoco hubo problemas de autocorrelación.

En la Tabla 5 se observa que la puntuación ESG muestra una relación positiva y estadísticamente significativa con ambas métricas de rentabilidad financiera, siendo especialmente relevante en el caso del ROE. Este hallazgo respalda la hipótesis H1, sugiriendo que la integración de criterios ESG puede traducirse en ventajas económicas tangibles para las empresas, y no solo en respuesta a demandas sociales o regulatorias. Las variables de control tradicionales, como la relación precio/valor contable y la capacidad del flujo de caja (OCFA), mantienen su relevancia como determinantes clásicos de la rentabilidad, mientras que el ratio de endeudamiento (LEV) y la capacidad hipotecaria (TANG) no presentan significancia estadística en este modelo. El elevado R^2 overall indica que el modelo explica una proporción considerable de la variabilidad total de la rentabilidad,

destacando la importancia de las diferencias estructurales entre empresas para explicar la rentabilidad a largo plazo.

Tabla 5. Resumen de efectos aleatorios para la variable ESG

Variable	ROA (RE)	ROE (RE)
ESG	0.0002 (0.00006)**	0.0009 (0.0001)***
PB	0.0015 (0.0003)***	0.0077 (0.0027)***
GR	-0.0022 (0.0078)	0.0034 (0.0233)
OCFA	0.4831 (0.0648)***	0.7915 (0.1794)***
LEV	-0.0021 (0.0046)	0.0020 (0.0061)
TANG	-0.0265 (0.0523)	-0.0979 (0.1577)
TIE	0.0001 (0.00005)**	0.0002 (0.0001)
R² (Overall)	0.8920	0.8735
R² (Between)	0.9373	0.9260
R² (Within)	0.3114	0.3322

Notas: Errores estándar entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$.

La Tabla 6 muestra que la dimensión ambiental del desempeño ESG presenta un coeficiente positivo y altamente significativo tanto para ROA como para ROE, lo que respalda la hipótesis H2 y sugiere que las políticas ambientales avanzadas se asocian con mejores resultados financieros. Este efecto es especialmente relevante en el contexto europeo, caracterizado por una regulación ambiental estricta, lo que refuerza la idea de que la gestión ambiental puede ser una inversión estratégica y no solo un coste. El patrón de significancia de las variables de control es similar al modelo anterior, y los elevados valores de R^2 overall indican que la dimensión ambiental explica una parte relevante de la variabilidad en la rentabilidad empresarial.

Tabla 6. Resumen de efectos aleatorios para la variable ambiental

Variable	ROA	ROE
Ambiental	0.0002 (0.00004)***	0.0008 (0.0001)***
PB	0.0015 (0.0003)***	0.0078 (0.0028)***
GR	-0.0007 (0.0079)	0.0049 (0.0236)
OCFA	0.4756 (0.0641)***	0.7767 (0.1795)***
LEV	-0.0019 (0.0042)	0.0009 (0.0060)
TANG	-0.0274 (0.0542)	-0.0559 (0.1530)
TIE	0.0001 (0.00005)***	0.0002 (0.0001)
R² (Overall)	0.8897	0.8772
R² (Between)	0.9341	0.9302
R² (Within)	0.3199	0.3316

Notas: Errores estándar entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$.

La Tabla 7 muestra que la dimensión social del desempeño ESG tiene un impacto positivo y altamente significativo sobre ROA y ROE, lo que respalda la

hipótesis H3. Este resultado indica que la inversión en capital humano, bienestar laboral y relaciones con la comunidad contribuye a la rentabilidad empresarial, especialmente en sectores donde la gestión del talento y la reputación social son factores clave, y pone de manifiesto que la atención a los factores sociales aporta valor económico más allá de la perspectiva ética.

Tabla 7. Resumen de efectos aleatorios para la variable social

Variable	ROA	ROE
Social	0.0002 (0.00006)***	0.0008 (0.0002)***
PB	0.0015 (0.0003)***	0.0077 (0.0027)***
GR	-0.0025 (0.0078)	0.0035 (0.0224)
OCFA	0.4711 (0.0655)***	0.7796 (0.1825)***
LEV	-0.0026 (0.0045)	0.0029 (0.0074)
TANG	-0.0374 (0.0542)	-0.0575 (0.1595)
TIE	0.0001 (0.00005)***	0.0002 (0.0001)
R² (Overall)	0.8916	0.8728
R² (Between)	0.9362	0.9263
R² (Within)	0.3193	0.3212

Notas: Errores estándar entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$.

La Tabla 8 analiza los efectos de la dimensión gobernanza sobre la rentabilidad, en este caso con un modelo de efectos aleatorios sobre el ROA y de efectos fijos sobre el ROE. En el caso del ROA es positivo y altamente significativo, en cambio, en el ROE no es estadísticamente significativo. En el caso de la variable OCFA, podemos observar que su valor es mayor que en el resto de modelos realizados. Esto se puede deber a que las empresas con sólidos mecanismos de gobernanza corporativa suelen gestionar de forma más eficiente su liquidez. Estudios como Mensah et al., 2024 confirman que la capacidad de flujo de caja tiene mayor impacto en empresas con gobernanza robusta, donde la supervisión evita usos ineficientes del capital. Los $R^2_{overall}$ son altos en ambas especificaciones, aunque ligeramente menores en el modelo de efectos fijos, lo que es habitual al controlar exhaustivamente la heterogeneidad no observada entre empresas. En este sentido, para la hipótesis H4, los resultados son mixtos, el coeficiente de gobernanza es positivo y altamente significativo para ROA, pero solo marginalmente significativo para ROE, lo que sugiere un impacto más sólido sobre la rentabilidad operativa que sobre la rentabilidad del patrimonio.

Tabla 8. Resultados para la dimensión gobernanza con efectos aleatorios y fijos

Variable	ROA (RE)	ROE (FE)
Governance	0.0007 (0.0001)***	0.0006 (0.0003)*
PB	0.0074 (0.0027)***	0.0070 (0.0028)**
GR	0.0163 (0.0264)	0.0539 (0.0573)
OCFA	0.8402 (0.1792)***	0.8599 (0.2145)***
LEV	0.0070 (0.0048)	0.0057 (0.0053)
TANG	-0.0821 (0.1799)	-0.3704 (0.3401)
TIE	0.0002 (0.0001)	0.0002 (0.0002)
R² (Overall)	0.8651	0.8474
R² (Between)	0.9165	0.8964
R² (Within)	0.3350	0.3427

Notas: Errores estándar entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$.

Por último, la Tabla 9 analiza el impacto de las controversias sobre la rentabilidad, utilizando como en el caso anterior efectos aleatorios para ROA y efectos fijos para ROE. El coeficiente de la variable controversias es levemente negativo en el modelo de efectos fijos, aunque es marginalmente significativo. Esto sugiere que las controversias de las empresas del EuroStoxx50 no tienen un impacto claro o sistemático sobre la rentabilidad. Al igual que en el modelo en el que se examina la gobernanza, el OCFA también es bastante mayor en el caso del ROA, lo que puede significar que las empresas expuestas a controversias ESG dependen críticamente de su liquidez contra shocks reputacionales. El alto coeficiente en RE sugiere que aquellas con mayor cash flow resisten mejor a los eventos negativos sin comprometer su rentabilidad. El valor de $R^2_{overall}$ desciende en el modelo de efectos fijos como en el caso de la gobernanza. La relevancia de este análisis radica en que permite identificar si los eventos negativos asociados a ESG afectan a la rentabilidad y los resultados sugieren que el modelo puede absorber o descontar rápidamente estos efectos, o que las empresas tienen mecanismos para mitigar su impacto. En cuanto a la hipótesis H5, que plantea un efecto negativo de las controversias sobre la rentabilidad, los resultados muestran un coeficiente levemente negativo y marginalmente significativo únicamente para ROE, mientras que para ROA no se observa un efecto estadísticamente significativo.

Tabla 9. Resultados para las controversias con efectos aleatorios y fijos

Variable	ROA (RE)	ROE (FE)
Controversies	0.0000 (0.0001)	-0.0000 (0.0000)*
PB	0.0072 (0.0028)**	0.0010 (0.0003)***
GR	0.0398 (0.0294)	0.0165 (0.0200)
OCFA	0.8906 (0.1823)***	0.3635 (0.0630)***
LEV	0.0217 (0.0093)**	0.0013 (0.0018)
TANG	0.1661 (0.2396)	-0.3190 (0.1206)***
TIE	0.0002 (0.0002)	0.0001 (0.0000)**
R² (Overall)	0.8416	0.7081
R² (Between)	0.8923	0.7345
R² (Within)	0.3195	0.3696

Notas: Errores estándar entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$.

4.3. Minería de textos

La tabla 10 recoge los temas identificados en los reportes analizados, utilizando la metodología de modelado de tópicos BERTopic. El tema más frecuente es “Diversidad e Igualdad”, lo que evidencia una clara preocupación y prioridad en los reportes por aspectos como la igualdad de género, la inclusión de jóvenes y mujeres, y la promoción de la vida laboral equitativa. Las palabras clave asociadas (estudiantes, igualdad de género, bono verde, STEM, voluntariado) refuerzan la idea de un enfoque transversal en la diversidad, tanto en el ámbito social como profesional. A continuación, destacan temas relacionados con “Finanzas y Contabilidad” (4,22 %), “Movilidad Eléctrica”, y “Economía Circular”. Esto sugiere que, aunque la sostenibilidad social es prioritaria, los aspectos económicos y medioambientales tienen también un peso significativo. La presencia de términos como IFRS, valor razonable, reciclado, reutilización y movilidad eléctrica indica una integración de la sostenibilidad en la gestión financiera y operativa de las organizaciones. “Gobierno Corporativo” y “Ética y Cumplimiento” son también temas relevantes, reflejando la importancia de la transparencia, la rendición de cuentas y la integridad empresarial en los reportes. Las palabras clave apuntan a estructuras de supervisión, códigos éticos y mecanismos de denuncia de irregularidades, elementos clave para la gobernanza responsable. Temas como “Emisiones GEI”, “Energía Renovable”, “Riesgos Climáticos”, y “Biodiversidad y Bosques” muestran que la dimensión ambiental es tratada de forma amplia, abarcando desde la gestión de emisiones hasta la protección de ecosistemas y la transición energética. Temas menos frecuentes pero igualmente relevantes, como “Ciberseguridad”, “Salud y Seguridad Laboral”, “Derechos Humanos”, “Agricultura Sostenible”, “Compras Sostenibles” y “Patrimonio y Cultura”. Esto demuestra la diversidad temática de los reportes y la amplitud de los retos ESG que enfrentan las organizaciones. La distribución de los temas revela una clara orientación hacia la sostenibilidad social y la gobernanza, sin descuidar los aspectos ambientales y financieros.

Título	% Frases	Palabras clave
Diversidad e Igualdad	21.24 %	estudiantes, mujeres, jóvenes, préstamos, igualdad de género, vida laboral, bono verde, stem, voluntariado, informe global
Finanzas y Contabilidad	4.22 %	impuestos, ifrs, deterioro, pasivo, valor razonable, arrendamiento, financiero consolidado, deuda, estados financieros consolidados, fondo de comercio
Movilidad Eléctrica	3.05 %	carga, batería, coches, tráfico, baterías, flota, automoción, furgonetas, urbano, camiones
Economía Circular	2.79 %	reciclado, reutilización, plásticos, algodón, circularidad, residuos generados, residuos peligrosos, gestión de residuos, materiales reciclados, reutilizados
Gobierno Corporativo	2.64 %	consejo de supervisión, junta general, consejo ejecutivo, nombramiento, consejero independiente, junta de accionistas, no ejecutivo, votación, principal independiente, principal consejero independiente
Emisiones GEI	2.62 %	alcance de las emisiones, protocolo ghg, toneladas, alcance de las emisiones, emisiones de carbono, reducir las emisiones, emisiones indirectas, factores de emisión, alcance, emisiones
Energía Renovable	2.08 %	solar, consumo energético, descarbonización, fotovoltaica, fuentes renovables, electricidad renovable, iluminación, mwh, calefacción, calor
Ética y Cumplimiento	2.04 %	denuncia de irregularidades, cumplimiento de las normas éticas, acoso, soborno, código ético, denunciante de irregularidades, gestión del cumplimiento, cumplimiento del grupo, conducta empresarial, responsable del cumplimiento de las normas
Riesgos Climáticos	1.86 %	tsfd, tiempo, riesgo climático, riesgos relacionados con el clima, riesgos financieros relacionados con el clima, riesgos financieros relacionados, tiempo extremo, divulgaciones financieras relacionadas, divulgaciones financieras, riesgos climáticos
Biodiversidad y Bosques	1.58 %	especies, agricultores, bosque, árboles, hábitats, suelo, hectáreas, cultivo, bosques, restauración
Gestión de Riesgos	1.56 %	riesgo del grupo, apetito, apetito de riesgo, erm, cartografía del riesgo, control interno, riesgo empresarial, factores de riesgo, riesgos identificados, gestión del riesgo empresarial
Ciberseguridad	1.48 %	ciber, ciberseguridad, datos personales, privacidad de los datos, ciberseguridad, seguridad de la información, gdpr, ataques, amenazas, normativa de protección
Salud y Seguridad Laboral	1.43 %	salud laboral, salud ocupacional seguridad, accidentes, ehs, accidente, seguridad ocupacional, seguridad sanitaria, cultura de la seguridad, gestión de la seguridad, raíz
Derechos Humanos	1.40 %	pacto mundial, oit, derechos humanos en la empresa, derechos humanos en la empresa, cuestiones de derechos humanos, orientar la actividad empresarial, orientar la actividad empresarial hacia los derechos humanos, esclavitud, internacional del trabajo
Riesgos ESG	1.15 %	problemas esg, riesgos esg, csr, gobernanza esg, riesgo esg, social y gobernanza esg, aspectos esg, relacionado con esg, rendimiento esg, gestión de activos
Agricultura Sostenible	1.11 %	leche, stoxx, nutrición, lácteos, agricultores, regenerativa, agricultura regenerativa, animal, seguridad alimentaria, onda blanca
Remuneración	1.09 %	salario, bonificación, sueldo, remuneración variable, compensación variable, consejo de administración, política de remuneración, informe de remuneración, consejo de supervisión, igualdad de retribución
Artesanía y Materiales	1.06 %	cuero, savoir, faire, savoir faire, artículos de cuero, curtidurías, artesanía, artesanos, seda, textil
Gestión del Agua	1.03 %	consumo de agua, aguas residuales, estrés hídrico, refrigeración, uso del agua, gestión del agua, extracción, uso del agua, agua potable, agua utilizada
Compras Sostenibles	0.97 %	código de proveedores, código de conducta de proveedores, compras responsables, gestión de proveedores, contratación sostenible, suministro sostenible, cualificación, proceso de contratación, cadena de suministro sostenible, sostenibilidad de los proveedores
Patrimonio y Cultura	0.96 %	maisons, maison, lujo, belleza, museo, patrimonio, colecciones, exposición, perfumes, artístico
Auditoría Interna	0.89 %	auditoría interna, control interno, riesgo de auditoría, comité de auditoría, auditor, auditores externos, auditorías realizadas, auditorías internas, auditores legales, gestión del riesgo de auditoría
Transición Energética	0.87 %	transición energética, descarbonización, indicadores de rendimiento presupuestario, anexo, anexo presupuesto, indicadores presupuestarios, anexo indicadores presupuestarios, indicadores de rendimiento, indicadores, presupuesto
Infraestructura y Fuerza Laboral	0.81 %	autopistas, aeropuertos, mano de obra relacionada, informe consejo de administración, directores mano de obra relacionada, consejo de administración mano de obra, directores mano de obra, mano de obra relacionada social, concesiones, relacionada social medioambiental
Salud Visual	0.78 %	gafas, lentes, óptica, lentes, ojo, ojos, monturas, ban, grandvision, laboratorios
Cumplimiento	0.68 %	cumplimiento, partidos políticos, financiero consolidado, estados financieros consolidados, estado no financiero, estado financiero, multas, jurisdicciones, infracciones, trabajo basado
Dividendos y Capital	0.68 %	dividendo, capital social, dividendos, adquisición de derechos, consecución de objetivos, examen de sostenibilidad, información sobre estados financieros, examen de resultados, información sobre estados financieros
Reportes No Financieros	0.54 %	declaración no financiera, declaración, informe anual declaración, informe declaración no financiero, informe declaración, informe anual grupo, criminal, textil, fabricantes proveedores, productos químicos
Inversiones Sostenibles	0.54 %	gastos, actividades económicas, reglamentación taxonómica, objetivos medioambientales, gastos de capital, acto delegado, subvencionabilidad
Estrategia y Acción Climática	0.54 %	utilidad, sostenibilidad futura, acción por el clima, hace negocios, valores de propósito, dividendo, modelo energético, lucha contra el cambio climático, lucha contra el clima, sección capítulo

Tabla 10. Temas predominantes en los reportes

La Tabla 11 muestra la evolución temporal de la proporción relativa de distintas temáticas dentro de los reportes de sostenibilidad y no financieros, lo cual proporciona una perspectiva cuantitativa de las prioridades temáticas en la divulgación corporativa voluntaria o regulada sobre aspectos ESG. “Diversidad e igualdad” mantiene sistemáticamente la mayor proporción de mención entre todas las categorías. Sin embargo, presenta una ligera tendencia decreciente a partir de 2020, lo que podría reflejar una saturación informativa, un desplazamiento del foco hacia otros temas emergentes, o un cambio metodológico en la estructura de los reportes. “Finanzas y contabilidad” aunque se encuentra entre los temas más frecuentes en los primeros años, pierde protagonismo de forma continua, pasando del 11.5 % en 2017 al 5.9 % en 2023. Este comportamiento sugiere una paulatina desvinculación entre los informes no financieros y los aspectos contables tradicionales, reflejando el tránsito hacia métricas más cualitativas y ESG-centradas. “Emisiones GEI”, “Riesgos climáticos” y “Economía circular” muestran incrementos constantes en su presencia relativa, lo cual es coherente con el refuerzo normativo y el aumento del escrutinio público sobre cuestiones climáticas. Las emisiones de gases de efecto invernadero, en particular, crecen más del doble entre 2017 (2.48 %) y 2023 (5.43 %). “Inversiones sostenibles” parten de una presencia casi nula (0.03 % en 2017) y alcanzan el 1.92 % en 2023, lo cual indica su reciente consolidación como categoría emergente. Este ascenso puede atribuirse al auge de los fondos ESG y a la regulación europea (como SFDR o la taxonomía de la UE). “Derechos humanos” y “Ciberseguridad y protección de datos” también presentan una evolución positiva, en línea con los riesgos reputacionales y legales asociados a estos temas. “Ética y cumplimiento”, “Energía renovable”, “Gestión del agua” y “Auditoría interna” muestran variaciones pequeñas o cíclicas a lo largo del periodo. Esto podría interpretarse como una estabilización temática dentro de los reportes, donde estas cuestiones se tratan con cierta regularidad pero sin protagonismo creciente. “Salud visual”, “Cumplimiento”, “Reportes no financieros”, “Patrimonio y lujo” y “Transición energética” mantienen proporciones muy bajas y, en la mayoría de los casos, tienden a decrecer. En algunos casos puede tratarse de temas marginales o específicos de ciertos sectores (como lujo o salud visual), o bien reflejar una asimilación institucional que reduce su necesidad de ser destacados. El descenso de “Reportes no financieros” como categoría resulta paradójico, pero podría explicarse por un cambio de nomenclatura o enfoque (por ejemplo, desplazamiento hacia métricas ESG o sostenibilidad como términos dominantes). “Gobierno corporativo y gestión de riesgos” muestran una presencia significativa en los primeros años, aunque decrecen levemente. Este fenómeno podría reflejar su integración dentro de un enfoque más transversal del reporte ESG, especialmente ante el impulso de regulaciones como la CSRD (Corporate Sustainability Reporting Directive). “Estrategia y acción climática”, a pesar de su importancia estratégica, mantiene una participación relativamente baja y estable. Esto podría ser un indicio de la falta de planificación climática detallada en muchas empresas o de la dificultad para traducirla en métricas informables.

Tema	2017	2018	2019	2020	2021	2022	2023
Diversidad e igualdad	0.3418	0.3524	0.3330	0.3640	0.3443	0.3220	0.2961
Finanzas y contabilidad	0.1151	0.0950	0.0705	0.0502	0.0448	0.0499	0.0595
Movilidad eléctrica	0.0355	0.0387	0.0498	0.0478	0.0539	0.0542	0.0494
Economía circular	0.0294	0.0377	0.0400	0.0426	0.0491	0.0498	0.0499
Gobierno corporativo	0.0577	0.0483	0.0486	0.0402	0.0341	0.0329	0.0372
Emisiones GEI	0.0248	0.0276	0.0336	0.0366	0.0466	0.0510	0.0543
Energía renovable	0.0306	0.0322	0.0310	0.0313	0.0314	0.0364	0.0340
Ética y cumplimiento	0.0326	0.0327	0.0317	0.0316	0.0341	0.0317	0.0300
Riesgos climáticos	0.0188	0.0183	0.0255	0.0293	0.0356	0.0341	0.0351
Biodiversidad y bosques	0.0158	0.0206	0.0176	0.0234	0.0280	0.0311	0.0307
Gestión de riesgos	0.0259	0.0235	0.0223	0.0270	0.0241	0.0241	0.0249
Ciberseguridad y protección de datos	0.0171	0.0195	0.0214	0.0240	0.0263	0.0260	0.0249
Salud y seguridad laboral	0.0210	0.0204	0.0238	0.0251	0.0232	0.0220	0.0216
Derechos humanos	0.0174	0.0176	0.0187	0.0199	0.0242	0.0260	0.0255
Riesgos ESG	0.0117	0.0108	0.0128	0.0188	0.0215	0.0213	0.0238
Agricultura sostenible	0.0251	0.0242	0.0104	0.0154	0.0100	0.0177	0.0204
Remuneración	0.0130	0.0137	0.0190	0.0186	0.0171	0.0185	0.0184
Artesanía y materiales	0.0105	0.0158	0.0198	0.0216	0.0150	0.0157	0.0175
Gestión del agua	0.0128	0.0152	0.0161	0.0157	0.0165	0.0179	0.0170
Compras sostenibles	0.0175	0.0181	0.0157	0.0147	0.0145	0.0139	0.0139
Patrimonio y lujo	0.0249	0.0170	0.0170	0.0144	0.0132	0.0118	0.0118
Auditoría interna	0.0159	0.0168	0.0154	0.0128	0.0124	0.0121	0.0141
Transición energética	0.0166	0.0132	0.0131	0.0120	0.0131	0.0144	0.0139
Infraestructura y fuerza laboral	0.0088	0.0120	0.0120	0.0160	0.0135	0.0128	0.0127
Salud visual	0.0161	0.0136	0.0370	0.0076	0.0069	0.0055	0.0066
Cumplimiento	0.0132	0.0131	0.0115	0.0097	0.0088	0.0095	0.0106
Dividendos y capital	0.0121	0.0098	0.0114	0.0101	0.0104	0.0084	0.0128
Reportes no financieros	0.0102	0.0113	0.0095	0.0067	0.0091	0.0064	0.0078
Inversiones sostenibles	0.0003	0.0004	0.0006	0.0023	0.0115	0.0156	0.0192
Estrategia y acción climática	0.0079	0.0103	0.0113	0.0103	0.0069	0.0072	0.0065

Tabla 11. Evolución de la proporción temática por categoría por año

La Figura 8 ilustra la evolución temporal de los 13 temas con mayor variabilidad en los reportes de sostenibilidad del EuroStoxx50 entre 2017 y 2023. El análisis revela varias tendencias relevantes para comprender la dinámica de las prioridades en sostenibilidad de las grandes empresas europeas.

Diversidad e igualdad destaca como el tema dominante durante todo el periodo, manteniéndose hasta 2020, cuando comienza un descenso progresivo hasta situarse cerca del 30 % en 2023. Esta evolución sugiere que, aunque sigue siendo el eje central de los reportes, su peso relativo se ha reducido ligeramente en los últimos años, posiblemente en favor de otros temas emergentes. En contraste, finanzas y contabilidad muestra una tendencia descendente continua, lo que podría indicar un desplazamiento del foco desde los aspectos financieros tradicionales hacia cuestiones más sociales, ambientales y de gobernanza. Otros temas como emisiones GEI, economía circular, movilidad eléctrica y riesgos climáticos experimentan un crecimiento moderado o se mantienen estables, reflejando la consolidación de la agenda ambiental en los reportes. La economía circular y las emisiones GEI en particular muestran una ligera tendencia al alza, en línea con la creciente presión regulatoria y social sobre la descarbonización y la gestión de residuos. Por su parte, temas como gobierno corporativo, biodiversidad y bosques, inversiones sostenibles y riesgos ESG mantienen una presencia relativamente constante, lo que sugiere que, aunque no son los más destacados, forman parte del núcleo temático de la sostenibilidad empresarial.

Esta evolución evidencia un desplazamiento progresivo en las prioridades de sostenibilidad corporativa del EuroStoxx50. Mientras que en los primeros años

predominaban los marcos conceptuales y declarativos, como la diversidad e igualdad o las cuestiones financieras y de gobierno corporativo, en los últimos ejercicios se observa una mayor atención hacia temas operativos y de impacto ambiental concreto, como las emisiones GEI, la economía circular y los riesgos climáticos. Este cambio refleja cómo las empresas están adaptando sus estrategias de sostenibilidad para responder tanto a las crecientes expectativas de inversores y consumidores como a las exigencias regulatorias. Así, la sostenibilidad corporativa evoluciona desde el compromiso y la declaración de intenciones hacia la implementación de acciones tangibles en materia de descarbonización, transformación de materiales y gestión de riesgos ambientales, consolidando una agenda más orientada a resultados y transparencia.

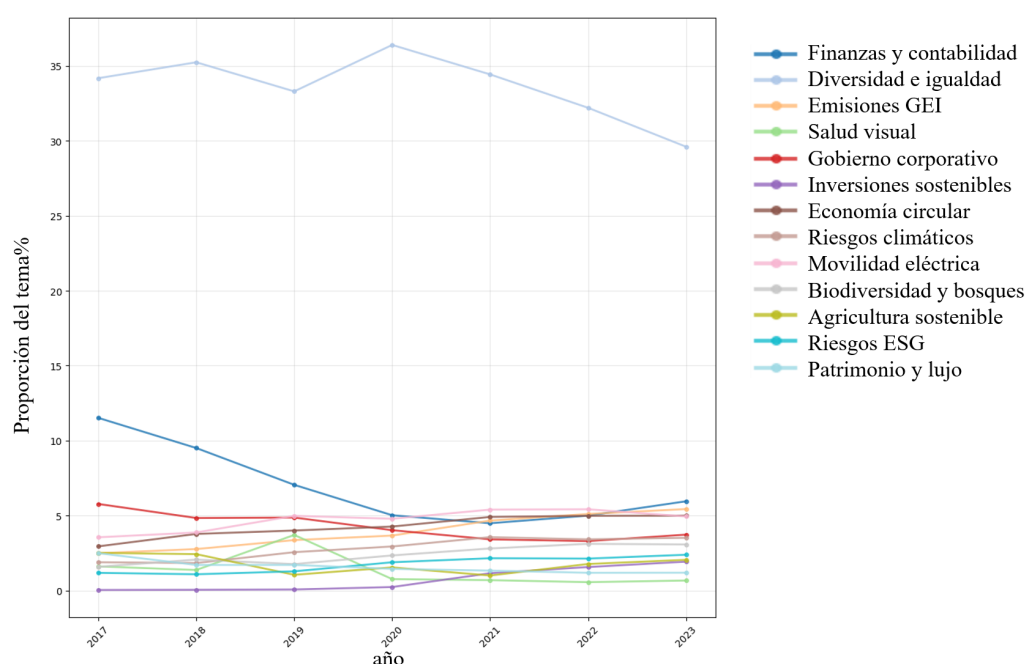


Figura 8. Evolución

La Figura 9 presenta el dendrograma resultante de la clusterización jerárquica de los temas identificados en los reportes de sostenibilidad del EuroStoxx50. Para determinar el número óptimo de clusters, se han empleado diversas métricas de evaluación, como el Silhouette Score, Davies-Bouldin Index, RMSSTD, R-squared y Semi-Partial R-squared, obteniéndose un coeficiente cofenético de 0,5892, que indica una representación adecuada de las distancias originales. Según estos indicadores, se ha seleccionado una partición en cuatro clusters, que ofrece el mejor equilibrio entre cohesión interna y separación entre grupos.

Este primer cluster agrupa temas centrados en la gestión interna de la organización, la innovación y la eficiencia operativa con perspectiva de sostenibilidad. Incluye temas como auditoría interna, ciberseguridad y protección de datos, cumplimiento, estrategia y acción climática, energía renovable, compras sostenibles, gestión del agua, infraestructura y fuerza laboral, biodiversidad y bosques, inversiones sostenibles, movilidad eléctrica y salud visual. La presencia de estos temas en un mismo grupo sugiere que las empresas tienden a abordar

conjuntamente la mejora de procesos internos, la adopción de tecnologías limpias y la gestión responsable de recursos. Además, destaca la integración de aspectos tecnológicos (como ciberseguridad) con cuestiones ambientales (energía renovable, gestión del agua), reflejando una visión holística de la sostenibilidad operativa.

El segundo cluster reúne temas de marcado carácter social y ético, así como la gestión de riesgos no estrictamente financieros. Aquí se agrupan diversidad e igualdad, finanzas y contabilidad, ética y cumplimiento, gestión de riesgos, salud y seguridad laboral, agricultura sostenible, riesgos climáticos, derechos humanos, transición energética, riesgos ESG y remuneración. Este agrupamiento refleja la interrelación entre la responsabilidad social corporativa, la igualdad de oportunidades, la protección de los derechos humanos y la gestión de riesgos emergentes. La inclusión de temas como riesgos climáticos y transición energética junto a aspectos sociales evidencia la creciente transversalidad de la sostenibilidad, donde los desafíos sociales y ambientales se gestionan de forma integrada.

El tercer cluster está formado exclusivamente por el tema de reportes no financieros. Su aislamiento respecto al resto de temas indica que este tipo de reportes presenta características y enfoques distintivos, probablemente relacionados con su función transversal de recopilación, divulgación y cumplimiento normativo en materia de sostenibilidad. Este resultado sugiere que los reportes no financieros actúan como un eje vertebrador, pero con suficiente especificidad para diferenciarse temáticamente del resto.

Por último, el cuarto cluster agrupa temas relacionados con la gobernanza corporativa y cuestiones ambientales de alto impacto, como patrimonio y lujo, gobierno corporativo, economía circular y emisiones de GEI. La proximidad entre economía circular y emisiones GEI refleja la creciente importancia de la transformación de materiales y la reducción de emisiones en la estrategia corporativa. La inclusión de patrimonio y lujo sugiere que la gestión de activos intangibles y culturales también se vincula a la gobernanza y la sostenibilidad ambiental en el contexto de las grandes empresas europeas.

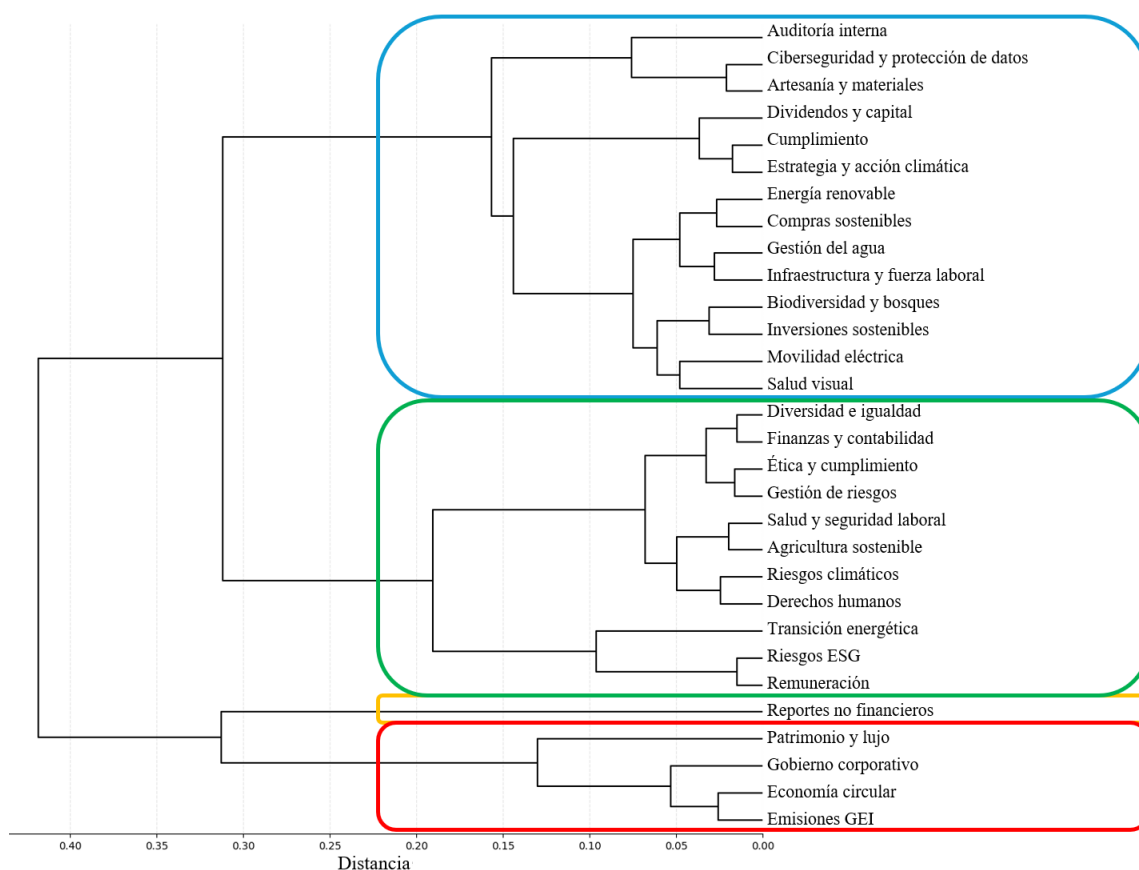


Figura 9. Dendrograma de la clusterización de temas

La Figura 10 muestra las correlaciones significativas entre las puntuaciones ESG y la proporción de temas por sector, presentando una visión de cómo las prioridades de sostenibilidad pueden variar según la naturaleza del sector industrial al que pertenezca cada empresa.

El análisis sectorial de las correlaciones significativas entre los temas ESG y las puntuaciones por industria revela patrones distintivos que reflejan tanto las prioridades como los desafíos de sostenibilidad propios de cada sector. A continuación, se desarrolla una visión profunda y argumentada para cada uno de los sectores principales, empleando la lista ordenada de temas y los resultados de correlación más relevantes.

En el sector de materiales básicos, la economía circular y los derechos humanos emergen como los temas más fuertemente asociados a un mejor desempeño ESG. La economía circular destaca por su altísima correlación positiva, lo que indica que las empresas que adoptan estrategias de reutilización y reciclaje de materiales tienden a obtener mejores puntuaciones ESG. Del mismo modo, los derechos humanos muestran una correlación positiva muy elevada, reflejando la presión social y regulatoria sobre la industria extractiva para garantizar condiciones laborales y comunitarias adecuadas. Sin embargo, la gobernanza corporativa y la diversidad e igualdad presentan correlaciones negativas notables, lo que sugiere que, aunque son temas abordados, no distinguen a las empresas líderes en ESG dentro de este sector, posiblemente por haberse convertido en requisitos mínimos

más que en factores diferenciadores.

En bienes cíclicos de consumo, los derechos humanos y la movilidad eléctrica aparecen como motores positivos del desempeño ESG. La correlación positiva con movilidad eléctrica y economía circular sugiere que las empresas que innovan en productos y procesos sostenibles logran mejores resultados en sostenibilidad. Sin embargo, la artesanía y materiales y la salud visual muestran correlaciones negativas, lo que puede interpretarse como una menor relevancia de estos aspectos en la percepción global de sostenibilidad del sector, que parece estar más enfocado en la cadena de suministro y la innovación ambiental.

En el sector de bienes no cíclicos de consumo, la agricultura sostenible es el único tema con correlación positiva significativa, lo que pone de manifiesto la importancia de la sostenibilidad en la producción primaria y el abastecimiento responsable. Por el contrario, la movilidad eléctrica, la diversidad e igualdad y la gestión del agua presentan correlaciones negativas, lo que puede indicar que la naturaleza tradicional y menos innovadora de este sector limita la integración de estos temas en su estrategia ESG, o bien que su impacto es menos visible para los evaluadores de sostenibilidad.

El sector energético, por su parte, muestra una fuerte correlación positiva con la reducción de emisiones GEI y la movilidad eléctrica, en línea con la transición energética global y la presión para descarbonizar la matriz energética. No obstante, la diversidad e igualdad y los derechos humanos presentan correlaciones negativas, lo que sugiere que, aunque importantes, estos temas no son los principales motores de la mejora ESG en energía, donde la innovación tecnológica y la gestión ambiental pesan más en la valoración global.

En el sector financiero, la gestión de riesgos ESG es el principal impulsor positivo, reflejando la creciente integración de criterios de sostenibilidad en la evaluación y financiación de proyectos. Sin embargo, la correlación negativa con reportes no financieros y finanzas y contabilidad sugiere que la transparencia y la gestión financiera responsable, aunque fundamentales, se han convertido en estándares básicos y no diferencian a las entidades con mejor desempeño ESG.

El sector salud destaca por una correlación positiva muy alta con diversidad e igualdad y biodiversidad y bosques, mostrando una orientación hacia la inclusión y la protección ambiental. Sin embargo, la salud visual presenta una correlación negativa extrema, lo que puede deberse a que este aspecto, aunque relevante, es percibido como menos crítico en la evaluación ESG general de las empresas del sector.

En la industria, la gestión del agua y la biodiversidad y bosques muestran correlaciones positivas, reflejando la importancia de la gestión ambiental en las operaciones industriales. Por el contrario, la salud y seguridad laboral y la diversidad e igualdad presentan correlaciones negativas, lo que puede indicar que estos temas, aunque relevantes, ya están ampliamente integrados y no diferencian a las empresas líderes en sostenibilidad industrial.

El sector inmobiliario ofrece un patrón singular: la biodiversidad y bosques tiene una correlación positiva casi perfecta con el desempeño ESG, lo que refleja la

sensibilidad del sector a la gestión del entorno natural y la integración de espacios verdes. Sin embargo, los derechos humanos y los reportes no financieros presentan correlaciones negativas extremas, lo que sugiere que estos temas han sido tan internalizados en la gestión sectorial que ya no actúan como elementos diferenciadores entre las empresas más avanzadas.

En tecnología, la diversidad e igualdad, la energía renovable y la ciberseguridad y protección de datos muestran correlaciones positivas significativas, evidenciando la prioridad de la inclusión, la innovación energética y la gestión de riesgos digitales en el sector. Por el contrario, la remuneración, la gobernanza corporativa y la auditoría interna presentan correlaciones negativas, lo que puede interpretarse como una madurez avanzada en estos temas, que han dejado de ser elementos diferenciadores en la valoración ESG de las empresas tecnológicas.

Finalmente, en el sector de servicios públicos, la diversidad e igualdad y la biodiversidad y bosques destacan como temas con correlaciones positivas elevadas, mientras que la gestión de riesgos climáticos y la gobernanza corporativa presentan correlaciones negativas. Esto sugiere que las empresas utilities que sobresalen en ESG tienden a priorizar la inclusión y la protección ambiental, mientras que los aspectos de gobernanza y gestión de riesgos climáticos ya están suficientemente integrados como para no marcar diferencias significativas en la puntuación ESG.

En conjunto, este análisis sectorial evidencia que la materialidad de los temas ESG es dinámica y específica para cada industria. Los temas con correlaciones positivas suelen reflejar áreas de innovación, presión regulatoria o demanda social, mientras que las correlaciones negativas pueden interpretarse como señales de madurez o estandarización, donde los requisitos mínimos ya no diferencian a las empresas líderes. Esta visión es clave para orientar la estrategia ESG de cada sector y maximizar el impacto de la sostenibilidad en la competitividad empresarial.

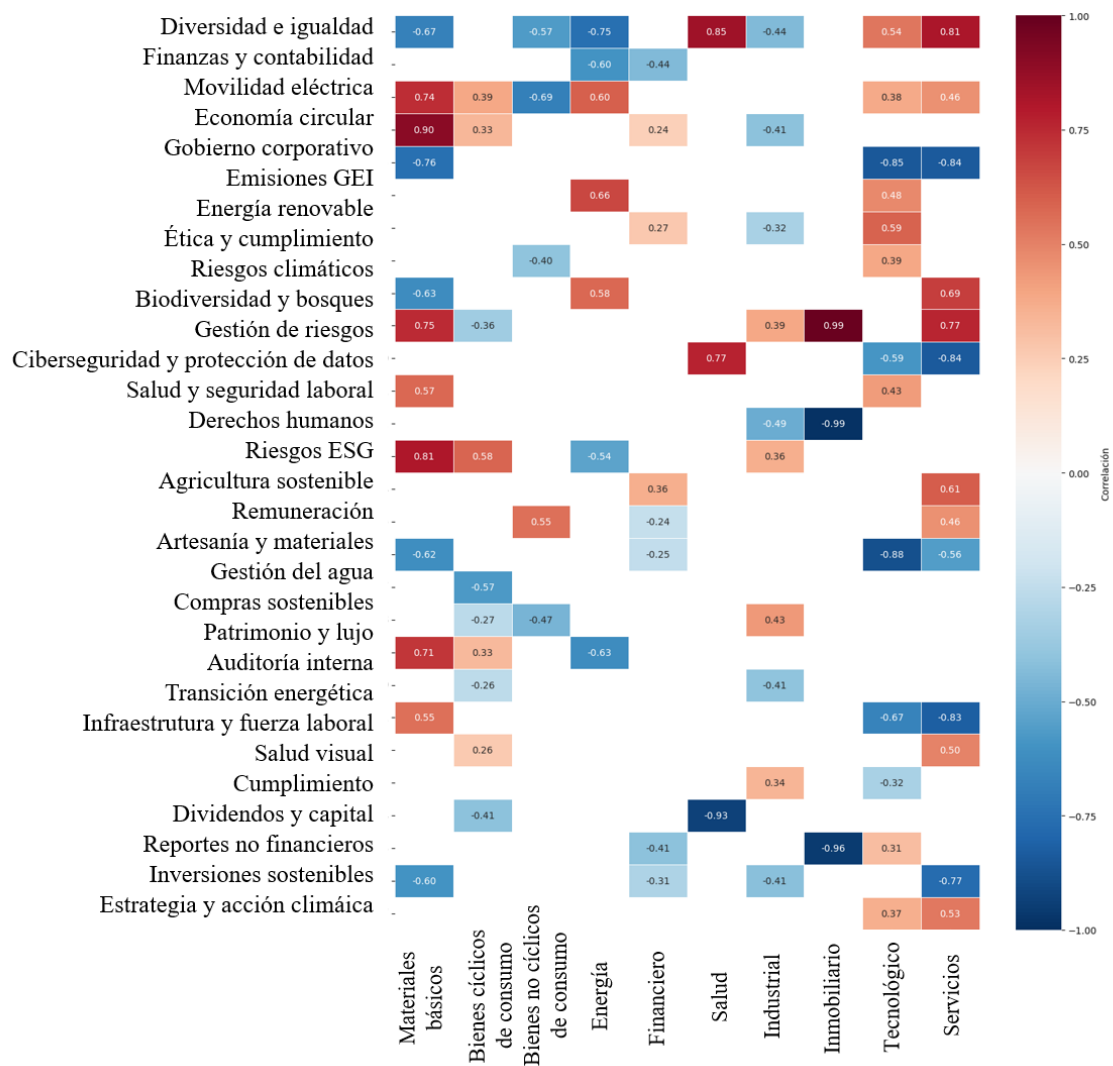


Figura 10. Correlaciones de Pearson significativas entre los temas y las puntuaciones ESG de cada empresa por sector

En la Tabla 12 podemos observar las diferencias temáticas ajustadas por industria entre empresas con altas y bajas puntuaciones ESG, comparando el cuartil superior con el cuartil inferior. Un valor positivo en la columna “Diferencia” indica que las empresas con mejores puntuaciones ESG (Q1) destacan más en ese tema respecto a las de menor puntuación (Q4), mientras que un valor negativo indica lo contrario. Los temas con mayores diferencias a favor del cuartil superior (Q1) son “Infraestructura y fuerza laboral” y “Finanzas y contabilidad”, seguidos de “Dividendos y capital”, “Patrimonio y lujo”, y “Movilidad eléctrica” y “Estrategia y acción climática”. Esto sugiere que las empresas líderes en ESG tienden a destacar significativamente en estos aspectos. Por el contrario, temas como “Gestión del agua”, “Riesgos climáticos” y “Biodiversidad y bosques” presentan diferencias negativas notables, lo que indica que las empresas con bajas puntuaciones ESG (Q4) tienen mejores medias en estos temas concretos. Hay temas donde la diferencia es muy baja o casi nula, como “Cumplimiento”, “Salud visual”, “Gobierno corporativo” y “Diversidad e igualdad”, lo que sugiere que en estos aspectos no hay una brecha significativa entre los extremos de puntuación ESG.

Las diferencias más marcadas en temas financieros y de infraestructura pueden reflejar que las empresas con mejores prácticas ESG también suelen tener una gestión más sólida en estos ámbitos, o que estos factores son especialmente valorados en la evaluación ESG. Las diferencias negativas en temas ambientales como agua, clima y biodiversidad podrían indicar que algunas empresas con bajas puntuaciones ESG aún mantienen buenas prácticas en áreas ambientales específicas, o bien que estos temas no son el principal foco de las empresas mejor posicionadas en ESG. La escasa diferencia en temas de cumplimiento y gobierno corporativo puede deberse a regulaciones mínimas obligatorias que igualan a la mayoría de las empresas independientemente de su puntuación ESG.

Tema	Media Q1	Media Q4	Diferencia
Infraestructura y fuerza laboral	2.7023	-0.9415	3.6438
Dividendos y capital	0.8967	-0.5114	1.4081
Movilidad eléctrica	0.3882	-0.5656	0.9538
Gestión del agua	-0.2078	0.6736	-0.8814
Riesgos climáticos	-0.1859	0.5735	-0.7594
Biodiversidad y bosques	-0.1629	0.5376	-0.7005
Riesgos ESG	0.3185	-0.1814	0.4999
Auditoría interna	0.2231	-0.2718	0.4949
Gestión de riesgos	-0.0818	0.3581	-0.4399
Energía renovable	-0.1316	0.3057	-0.4373
Ciberseguridad y protección de datos	-0.1531	0.2022	-0.3553
Compras sostenibles	0.1297	0.4267	-0.2970
Salud y seguridad laboral	0.0612	-0.2151	0.2762
Economía circular	0.2866	0.0710	0.2157
Inversiones sostenibles	-0.9364	-0.7720	-0.1644
Ética y cumplimiento	-0.8070	-0.9699	0.1629
Emisiones GEI	-0.0428	0.1040	-0.1468
Agricultura sostenible	0.0200	0.1626	-0.1426
Reportes no financieros	-0.8837	-0.9680	0.0843
Derechos humanos	-0.9472	-0.8679	-0.0793
Remuneración	-0.0459	0.0331	-0.0791
Transición energética	0.0682	0.0143	0.0540
Artesanía y materiales	0.0138	0.0551	-0.0413
Gobierno corporativo	-0.9607	-0.9224	-0.0383
Diversidad e igualdad	-0.1537	-0.1662	0.0125
Cumplimiento	-0.8687	-0.8645	-0.0042
Salud visual	-0.4193	-0.4184	-0.0009
Finanzas y contabilidad	2.7023	-0.9415	3.6438
Patrimonio y lujo	0.8967	-0.5114	1.4081
Estrategia y acción climática	0.3882	-0.5656	0.9538

Tabla 12. Diferencias de medias en los cuartiles 1 y 4

Se ha realizado un PCA con los temas que más aparecen en los informes de sostenibilidad, para ver cómo se asocian las empresas con estos temas en el último

año en estudio. PC1 representa el 27,76 % de la varianza. Este primer componente diferencia claramente a las empresas que priorizan la diversidad y la igualdad, las emisiones GEI, la energía renovable, los riesgos climáticos y la biodiversidad y los bosques; de las que se enfocan en el gobierno corporativo y finanzas y contabilidad. Esto indica que las empresas con mayor puntuación en este componente están más enfocadas en aspectos sociales y ambientales, integrando la sostenibilidad ambiental y la equidad social en sus estrategias. Las que tienen puntuaciones bajas en PC1 se centran más en el gobierno corporativo y la gobernanza ética, lo que refleja una orientación hacia estructuras de control, cumplimiento normativo y gestión ética, posiblemente con menor énfasis en aspectos ambientales y sociales operativos.

PC2, representa un 17,5 % de la varianza. El segundo componente está fuertemente asociado a la integración de la sostenibilidad ambiental en la estrategia empresarial. Las empresas con puntuaciones altas en PC2 destacan por la presencia de temas como economía circular, gestión de riesgos climáticos y biodiversidad, lo que indica una internalización avanzada de criterios ambientales en la toma de decisiones y en la gestión de riesgos. Una puntuación alta en este componente sugiere que la empresa no solo reporta, sino que también implementa acciones concretas para mitigar impactos ambientales, gestionar emisiones y promover la conservación de recursos naturales. Por el contrario, una puntuación baja en este componente indica una fuerte asociación con riesgos climáticos y con diversidad e igualdad.

Por último, PC3 representa un 14,59 % de la varianza total y presenta una asociación negativa con la movilidad eléctrica y la energía renovable. Las puntuaciones altas en este componente indican alineación con emisiones GEI y con biodiversidad y bosques. PC3 diferencia a las empresas que priorizan la gestión de emisiones y la biodiversidad, pero que muestran un menor compromiso con la movilidad eléctrica y las energías renovables, posiblemente por pertenecer a sectores menos innovadores o tradicionales en términos energéticos.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Diversidad e igualdad	0.316	-0.408	-0.100	0.359	-0.003	0.226	0.618	-0.122	-0.377	0.064
Finanzas y contabilidad	-0.498	-0.120	0.274	-0.081	0.171	-0.196	0.135	0.069	-0.237	0.713
Movilidad eléctrica	-0.091	0.103	-0.438	-0.496	-0.620	-0.107	0.134	-0.117	-0.335	0.060
Economía circular	0.086	0.504	-0.204	0.474	0.070	-0.560	0.035	0.218	-0.322	-0.028
Gobierno corporativo	-0.518	-0.070	0.220	-0.088	0.182	-0.030	0.121	-0.082	-0.408	-0.670
Emisiones GEI	0.370	-0.110	0.354	-0.399	-0.031	-0.189	0.204	0.686	-0.067	-0.112
Energía renovable	0.354	0.178	0.089	-0.407	0.443	-0.326	0.202	-0.570	0.014	0.023
Ética y cumplimiento	0.059	-0.184	-0.562	-0.217	0.552	0.183	-0.321	0.252	-0.303	0.055
Riesgos climáticos	0.236	-0.491	0.213	0.111	-0.210	-0.360	-0.574	-0.221	-0.306	-0.007
Biodiversidad y bosques	0.215	0.481	0.373	-0.030	-0.048	0.529	-0.226	-0.063	-0.478	0.131

Tabla 13. Cargas de los temas en los componentes

En las Figuras 11 y 12 se muestran los HJ-Biplots en los que se representan con el color verde las empresas que se encuentran en el primer cuartil de mejores puntuaciones ESG del EuroStoxx50 en 2023, con rojo las que están en el último cuartil, y de gris, el resto. Este enfoque permite identificar asociaciones clave entre las empresas y las principales dimensiones temáticas del ámbito ESG.

PC2 está vinculado a la internalización de la sostenibilidad ambiental en la estrategia empresarial. Las empresas con valores positivos en PC2 se asocian con economía circular, biodiversidad y gestión de riesgos climáticos, lo que indica una integración avanzada de criterios ambientales. En esta zona se destacan empresas como LOREAL, BASF, ADIDAS o ENI, reflejando una orientación fuerte hacia la sostenibilidad ambiental y un compromiso más tangible con la transformación ecológica. En cambio, puntuaciones negativas en PC2 implican una menor internalización estratégica de estos factores ambientales. En este extremo se encuentran compañías del sector financiero como ALLIANZ, MUNICHRE, SANTANDER, ING y BBVA, lo cual sugiere que, a pesar de posibles avances en gobernanza o reportes, estas empresas presentan rezagos en la implementación operativa de prácticas ambientales sólidas.



57

de emisiones GEI y biodiversidad. Así, las empresas con valores negativos en PC3 son aquellas que muestran una mayor alineación con la transición energética y una integración más fuerte de tecnologías limpias como la movilidad eléctrica. En este grupo se encuentran compañías como MERCEDES, VOLKSWAGEN, STELLANTIS y SAP, todas con puntuaciones ESG bajas, lo que sugiere que, si bien pueden estar alineadas con ciertas temáticas transformadoras, su desempeño ESG global sigue siendo insuficiente o limitado por otros factores (como gobernanza o aspectos sociales). Por el contrario, las empresas con valores positivos en PC3, como TOTALENERGIES, ENI, BBVA, BNPP y DEUTSCHEBOERSE, están más asociadas a temas como la gestión de emisiones GEI y biodiversidad, pero presentan menor compromiso relativo con movilidad eléctrica y energías renovables. Esto podría estar vinculado a su perfil sectorial (energía tradicional o financiero), o a una estrategia menos centrada en tecnologías limpias, aunque algunas de ellas —como DEUTSCHEBOERSE— destacan por su desempeño ESG global gracias a su fortaleza en gobernanza.

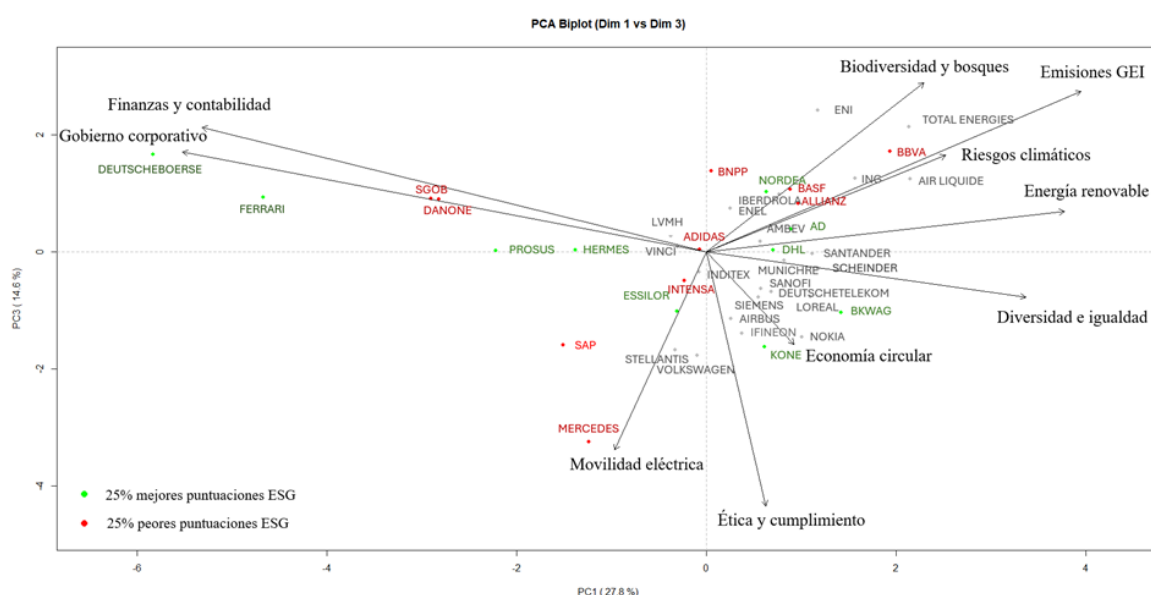


Figura 12. HJ-Biplot con PC1 y PC3

Desde una perspectiva sectorial, las empresas del sector financiero, como BBVA, ING, SANTANDER, ALLIANZ, MUNICHRE, BNP, SGOB e INTESA, tienden a agruparse en zonas de PC1 negativas y PC2 positivas, lo que indica un enfoque predominante en la ética y cumplimiento, riesgos climáticos y diversidad e igualdad. DEUTSCHEBOERSE, en cambio, se posiciona en la zona negativa del PC1, asociándose con gobierno corporativo y finanzas y contabilidad.

En el sector automoción e industrial, se observan diferentes niveles de madurez en la transición sostenible. Empresas como MERCEDES, VOLKSWAGEN y STELLANTIS presentan valores negativos en PC3, lo que sugiere una alineación con la movilidad eléctrica y ética y cumplimiento. Sin embargo, sus puntuaciones generales ESG no siempre reflejan esta orientación. MERCEDES permanece con puntuaciones ESG deficientes y por otro lado, FERRARI, si bien destaca en

gobernanza, se sitúa en zonas menos favorables respecto a factores ambientales, lo que matiza su posicionamiento ESG.

El análisis del sector consumo y lujo (con empresas como LVMH, ADIDAS, DANONE, L'ORÉAL, HERMÈS, ESSILOR, INDITEX y ESSILOR) muestran puntuaciones positivas en PC2. En el caso de L'OREAL y ADIDAS, destacan por integrar sostenibilidad ambiental en su estrategia. En contraste, DANONE presenta una puntuación ESG negativa pese a su reputación sostenible, lo que podría reflejar debilidades en la implementación o una desconexión entre narrativa e impacto real. ESSILOR y HERMÈS, por su parte, se sitúan en posiciones más consistentes tanto en alineación temática como en desempeño ESG.

Desde la perspectiva geográfica, las empresas francesas (como TOTAL ENERGIES, L'ORÉAL, SANOFI, AIR LIQUIDE, BNPP y LVMH) se agrupan con valores positivos en PC1, lo que indica una orientación clara hacia los temas ambientales y sociales, especialmente en energía renovable, biodiversidad y gestión climática. No obstante, SAINT GOBAIN o DANONE, se encuentran en posiciones negativas de PC1 y con bajas puntuaciones ESG.

En el caso de las empresas alemanas —como DEUTSCHEBOERSE, VOLKSWAGEN, SIEMENS, MUNICHRE, SAP e INFINEON— predomina un enfoque más conservador, con mayor concentración en PC1 negativo, lo que refleja un énfasis en gobernanza y cumplimiento normativo. Su posicionamiento en PC2 y PC3 es más heterogéneo, y algunas como SAP y MUNICHRE exhiben una débil integración de factores ambientales, lo cual se traduce en puntuaciones ESG desfavorables.

5. Discusión

El presente estudio destaca tanto el potencial del análisis econométrico como la combinación de la minería de textos con el análisis multivariante. Los resultados del análisis de panel muestran una relación positiva y estadísticamente significativa entre el desempeño ESG y la rentabilidad financiera (ROA y ROE, con mayor énfasis en ROE), lo que permite rechazar la hipótesis nula de ausencia de efecto y, por tanto, encontrar evidencia a favor de la hipótesis de investigación H1, que plantea un efecto positivo. De igual modo, las políticas ambientales avanzadas muestran una contribución positiva y significativa a la rentabilidad, permitiendo rechazar la hipótesis nula y aportar evidencia a favor de la hipótesis de investigación correspondiente H2. De manera similar, la dimensión social impacta positivamente en ROA y ROE, lo que también lleva al rechazo de la hipótesis nula y respalda la hipótesis de investigación H3, destacando la importancia de los factores sociales. En cuanto a la gobernanza, su efecto es positivo y significativo para ROA, pero solo marginalmente significativo para ROE; por lo tanto, la hipótesis nula se rechaza para ROA, lo que apoya la hipótesis de investigación H4 en este indicador, mientras que la evidencia es menos concluyente para ROE, indicando una mayor influencia en la rentabilidad operativa. Finalmente, el impacto de las controversias ESG sobre la rentabilidad es débil o poco sistemático, con un coeficiente ligeramente negativo y marginalmente significativo solo para ROE, por lo que no se puede rechazar de forma robusta la hipótesis nula de ausencia de efecto negativo, es decir, no se encuentra suficiente evidencia a favor de la hipótesis de investigación H5 en este caso.

Estos resultados son consistentes con la literatura previa, que ha documentado una asociación positiva entre el desempeño ESG y la rentabilidad financiera en empresas europeas, especialmente en lo relativo a las dimensiones ambiental y social. No obstante, el impacto de la gobernanza y, sobre todo, de las controversias, sigue mostrando una mayor heterogeneidad y dependencia del contexto. Cabe señalar que, como ocurre en buena parte de la investigación empírica sobre ESG, la relación observada entre la puntuación ESG y la rentabilidad podría estar influida por factores no capturados por el modelo, lo que plantea el reto de identificar la dirección de la causalidad. Aunque los modelos de datos de panel permiten controlar la heterogeneidad no observada y las variaciones temporales, establecer con certeza si una mayor puntuación ESG impulsa la rentabilidad, o si las empresas más rentables son las que pueden invertir más en iniciativas ESG, sigue siendo una cuestión abierta. Este aspecto representa una interesante línea de investigación futura para clarificar la naturaleza y el sentido de la relación entre sostenibilidad y desempeño financiero.

Además, la calidad y homogeneidad de los datos ESG sigue siendo un reto. Las diferencias en la forma de reportar, la falta de estandarización y la posible presencia de información poco veraz puede afectar la precisión y fiabilidad de los resultados. Y aunque los valores elevados de $R^2_{between}$ indican que las diferencias estructurales entre empresas explican buena parte de la variabilidad en la rentabilidad, la capacidad del modelo para captar cambios interanuales dentro de cada empresa es más limitada, lo que refuerza la necesidad de prudencia a la hora

de interpretar la relación entre ESG y el desempeño financiero.

Por otro lado, este trabajo también destaca la capacidad de la minería de textos para descubrir tendencias y patrones en los informes de sostenibilidad de las empresas europeas. Sin embargo, a pesar de sus ventajas, el uso de la minería de textos en los informes de sostenibilidad se enfrenta a varios retos y limitaciones. Entre ellos figuran los problemas de calidad de los datos, la necesidad de conocimientos específicos y la interpretación errónea de los resultados. La calidad de los informes puede variar considerablemente, lo que puede afectar a la exactitud y fiabilidad de los resultados. Las incoherencias en los formatos, el uso de un lenguaje ambiguo, la inclusión de información irrelevante o la falta de información veraz pueden afectar (Contala et al., 2024; Liew et al., 2014; Goloshchapova et al., 2019).

Además, el estudio se ve limitado por la subjetividad inherente a la asignación de etiquetas semánticas a los temas identificados por BERTopic. Aunque las palabras clave proporcionan orientación, la interpretación y la denominación de los temas dependen del juicio del investigador, lo que puede introducir sesgos. Para validar la estabilidad temática de los resultados obtenidos con BERTopic, se replicó el análisis utilizando LDA, una técnica clásica basada en distribución de probabilidad. Aunque los temas identificados no coincidieron de forma exacta, sí se observaron agrupaciones temáticas similares, lo que refuerza la consistencia de los hallazgos. No obstante, debido a su mayor capacidad de captura semántica contextual, se optó por presentar los resultados finales basados en BERTopic. Otra limitación es el posible sesgo lingüístico, ya que sólo se analizaron informes en inglés. Esto puede subestimar los matices del discurso de sostenibilidad de las empresas que informan principalmente en otras lenguas europeas. Las investigaciones futuras deberían considerar la incorporación de informes en varios idiomas y la validación de las etiquetas temáticas mediante la revisión de expertos para mejorar la fiabilidad y la aplicabilidad transcultural.

Es fundamental reconocer que, aunque este análisis ofrece una valiosa perspectiva de la evolución de la sostenibilidad en las grandes empresas europeas, tiene ciertas limitaciones inherentes a su alcance y metodología. Una limitación notable es su dependencia exclusiva de las empresas del EuroStoxx50, lo que puede limitar la generalizabilidad de los resultados. La muestra se compone de grandes empresas consolidadas que están sujetas a un mayor escrutinio normativo y de las partes interesadas, por lo que puede no ser representativa del entorno empresarial europeo en general, en particular de las empresas más pequeñas o menos intensivas en capital.

Las investigaciones futuras deberían tratar de ampliar este análisis a un conjunto más diverso de empresas, como las incluidas en el EuroStoxx600, y durante un periodo de tiempo más largo. De este modo se obtendría una visión más completa de las tendencias de sostenibilidad y se podría comprender mejor la resistencia y la evolución a largo plazo de las estrategias de sostenibilidad de las empresas.

Por último, aunque la minería de textos ofrece potentes capacidades analíticas, su aplicación debe adaptarse cuidadosamente al panorama cambiante y sectorial

de la sostenibilidad empresarial. Las conclusiones de este estudio acentúan la importancia de abordar tanto las oportunidades como los retos inherentes al uso de técnicas automatizadas para comprender las prioridades y prácticas cambiantes de las grandes empresas europeas.

6. Conclusiones

Basándonos en los objetivos presentados en el apartado 1.1 y en los resultados del apartado 4, podemos presentar las siguientes conclusiones:

1. El análisis de datos de panel realizado sobre las empresas del EuroStoxx50 revela una relación positiva y estadísticamente significativa entre las métricas de desempeño ESG y los indicadores financieros de rentabilidad, especialmente en el caso del ROE. Tanto la puntuación combinada ESG como sus dimensiones ambiental y social muestran impactos positivos y robustos sobre la rentabilidad. En cambio, la gobernanza destaca por su influencia en la eficiencia operativa. Las controversias ESG presentaron un efecto limitado o marginalmente negativo sobre la rentabilidad, lo que sugiere que las grandes empresas pueden mitigar el impacto de efectos adversos.
2. Los modelos tienen valores elevados de $R^2_{overall}$ y $R^2_{between}$, que indica que explican una proporción muy alta de la variabilidad total y de las diferencias estructurales entre empresas. En cambio, el R^2_{within} es considerablemente menor, lo que sugiere que los cambios interanuales dentro de cada empresa tienen un peso más limitado en la explicación de la rentabilidad. Las variables de control como el precio sobre el valor contable y el flujo de caja sobre activos resultaron altamente significativas y positivas, confirmando que son determinantes clásicos de la rentabilidad.
3. El análisis de minería de textos combinado con el análisis multivariante revela que los informes de sostenibilidad de las empresas del EuroStoxx 50 han evolucionado hacia un enfoque más operativo y orientado a resultados, alejándose del cumplimiento declarativo. En cuanto a la evolución temporal, se observa una clara tendencia a la integración de la sostenibilidad en la estrategia empresarial, pasando de un enfoque fragmentado a uno holístico que abarca gestión interna, innovación, eficiencia operativa y aspectos sociales y éticos.
4. La estructura temática de los informes evidencia una articulación cada vez más compleja y sectorializada. Las empresas líderes en ESG destacan por anticipar y abordar cuestiones emergentes como infraestructuras sostenibles, fuerza laboral del futuro, agricultura sostenible e inversiones responsables. La materialidad de los temas ESG varía significativamente por sector, lo que subraya la importancia de estrategias adaptadas a las características de cada industria. Prácticas que antes eran diferenciadoras, como la gobernanza y la gestión de riesgos, se han convertido en estándares mínimos esperados generalizándose. Por último, existe una tendencia clara hacia una mayor transparencia y un enfoque en resultados tangibles, con informes que reflejan cada vez más la implementación de acciones concretas y la medición de su impacto. El compromiso con la acción y la rendición de cuentas se ha convertido en un elemento clave para la legitimidad y la competitividad de las grandes empresas.

7. Referencias

- Li, T.-T., Wang, K., Sueyoshi, T., & Wang, D. D. (2021). ESG: Research progress and future prospects. *Sustainability*, 13(21), 11663. <https://doi.org/https://doi.org/10.3390/su132111663>
- Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of sustainable finance & investment*, 5(4), 210-233. <https://doi.org/https://doi.org/10.1080/20430795.2015.1118917>
- Jiang, P.-C., Feng, G.-F., Wang, H.-J., & Chang, C.-P. (2024). CSR from different perspectives: The global ESG indexes updated. *Corporate Social Responsibility and Environmental Management*, 31(5), 4694-4714. <https://doi.org/https://doi.org/10.1002/csr.2814>
- Pérez, L., Hunt, V., Samandari, H., Nuttall, R., & Biniek, K. (2022). Does ESG really matter—and why. *McKinsey Quarterly*, 60(1).
- Rau, P. R., & Yu, T. (2024). A survey on ESG: investors, institutions and firms. *China Finance Review International*, 14(1), 3-33. <https://doi.org/https://doi.org/10.1108/CFRI-12-2022-0260>
- Clément, A., Robinot, É., & Trespeuch, L. (2025). The use of ESG scores in academic literature: a systematic literature review. *Journal of Enterprising Communities: People and Places in the Global Economy*, 19(1), 92-110. <https://doi.org/https://doi.org/10.1108/JEC-10-2022-0147>
- Passas, I. (2024). The Evolution of ESG: From CSR to ESG 2.0. *Encyclopedia*, 4(4), 1711-1720. <https://doi.org/https://doi.org/10.3390/encyclopedia4040112>
- Lindgreen, A., & Swaen, V. (2010). Corporate Social Responsibility. *International Journal of Management Reviews*, 12, 1-7. <https://doi.org/10.1111/j.1468-2370.2009.00277.x>
- Nugroho, D. P., Hsu, Y., Hartauer, C., & Hartauer, A. (2024). Investigating the interconnection between environmental, social, and governance (ESG), and corporate social responsibility (CSR) strategies: An examination of the influence on consumer behavior. *Sustainability*, 16(2), 614. <https://doi.org/https://doi.org/10.3390/su16020614>
- Kaźmierczak, M. (2022). A literature review on the difference between CSR and ESG. *Zeszyty Naukowe. Organizacja i Zarządzanie/Politechnika Ślaska*, (162).
- Machado, B. A. A., Dias, L. C. P., & Fonseca, A. (2021). Transparency of materiality analysis in GRI-based sustainability reports. *Corporate Social Responsibility and Environmental Management*, 28, 570-580. <https://doi.org/10.1002/csr.2066>
- Yosifova, D., & Petrova-Kirova, M. (2022). The New EU Corporate Sustainability Reporting Framework in the context of GRI Standards. *Economic and Social Development: Book of Proceedings*, 85-95.
- Hales, J. Sustainability accounting standards board (SASB). En: *En World Scientific Encyclopedia of Climate Change: Case Studies of Climate Risk, Action, and Opportunity Volume 3*. World Scientific, 2021, pp. 37-41. https://doi.org/https://doi.org/10.1142/9789811213960_0007
- Martínez-Torres, V. (2024). Regulatory Developments on Sustainability Issues in Light of Delegated Regulation (EU) 2023/2772. *Rev. Eur. & Comp. L.*, 58, 223. <https://doi.org/https://doi.org/10.31743/recl.17459>

- Antonów, D. (2024). The Concept of Sustainability from the Perspective of Tax Law. *Studia Iuridica Lublinensia*, 33(4), 87-101. <https://doi.org/10.17951/sil.2024.33.4.87-101>
- Bradford, A. (2020). *The Brussels effect: How the European Union rules the world*. Oxford University Press.
- Redondo Alamillos, R., & De Mariz, F. (2022). How can European regulation on ESG impact business globally? *Journal of risk and Financial Management*, 15(7), 291. <https://doi.org/https://doi.org/10.3390/jrfm15070291>
- Arvidsson, S., & Dumay, J. (2022). Corporate ESG reporting quantity, quality and performance: Where to now for environmental policy and practice? *Business strategy and the environment*, 31(3), 1091-1110. <https://doi.org/https://doi.org/10.1002/bse.2937>
- Ademi, B., & Klungseth, N. J. (2022). Does it pay to deliver superior ESG performance? Evidence from US S&P 500 companies. *Journal of global responsibility*, 13(4), 421-449. <https://doi.org/https://doi.org/10.1108/JGR-01-2022-0006>
- Oprean-Stan, C., Oncioiu, I., Iuga, I. C., & Stan, S. (2020). Impact of sustainability reporting and inadequate management of ESG factors on corporate performance and sustainable growth. *Sustainability*, 12(20), 8536. <https://doi.org/https://doi.org/10.3390/su12208536>
- Fiskerstrand, S. R., Fjeldavli, S., Leirvik, T., Antoniuk, Y., & Nenadić, O. (2020). Sustainable investments in the Norwegian stock market. *Journal of Sustainable Finance & Investment*, 10(3), 294-310. <https://doi.org/https://doi.org/10.1080/20430795.2019.1677441>
- Cek, K., & Eyupoglu, S. (2020). Does environmental, social and governance performance influence economic performance? *Journal of Business Economics and Management*, 21(4), 1165-1184. <https://doi.org/https://doi.org/10.3846/jbem.2020.12725>
- Hughes, A., Urban, M. A., & Wójcik, D. (2021). Alternative ESG ratings: How technological innovation is reshaping sustainable investment. *Sustainability*, 13(6), 3551. <https://doi.org/https://doi.org/10.3390/su13063551>
- Zhao, C., Guo, Y., Yuan, J., Wu, M., Li, D., Zhou, Y., & Kang, J. (2018). ESG and corporate financial performance: Empirical evidence from China's listed power generation companies. *Sustainability*, 10(8), 2607. <https://doi.org/https://doi.org/10.3390/su10082607>
- Chams, N., García-Blandón, J., & Hassan, K. (2021). Role reversal! financial performance as an antecedent of ESG: The moderating effect of total quality management. *Sustainability*, 13(13), 7026. <https://doi.org/https://doi.org/10.3390/su13137026>
- Lyon, T. P., & Maxwell, J. W. (2011). Greenwash: Corporate environmental disclosure under threat of audit. *Journal of economics & management strategy*, 20(1), 3-41. <https://doi.org/https://doi.org/10.1111/j.1530-9134.2010.00282.x>
- Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *The accounting review*, 91(6), 1697-1724. <https://doi.org/https://doi.org/10.2308/accr-51383>
- Zheng, J., Khurram, M. U., & Chen, L. (2022). Can green innovation affect ESG ratings and financial performance? Evidence from Chinese GEM listed

- companies. *Sustainability*, 14(14), 8677.
<https://doi.org/https://doi.org/10.3390/su14148677>
- Kim & Li, Z. (2021). Understanding the impact of ESG practices in corporate finance. *Sustainability*, 13(7), 3746. <https://doi.org/https://doi.org/10.3390/su13073746>
- Koundouri, P., Pittis, N., & Plataniotis, A. (2022). The impact of ESG performance on the financial performance of European area companies: An empirical examination. *Environmental Sciences Proceedings*, 15(1), 13. <https://doi.org/https://doi.org/10.3390/environsciproc2022015013>
- Ionescu, G. H., Firoiu, D., Pirvu, R., & Vilag, R. D. (2019). The impact of ESG factors on market value of companies from travel and tourism industry. *Technological and Economic Development of Economy*, 25(5), 820-849.
- Hübel, B. (2022). Do markets value ESG risks in sovereign credit curves? *The quarterly review of economics and finance*, 85, 134-148. <https://doi.org/https://doi.org/10.1016/j.qref.2020.11.003>
- Kabir, M. A., & Chowdhury, S. S. (2023). Empirical analysis of the corporate social responsibility and financial performance causal nexus: Evidence from the banking sector of Bangladesh. *Asia Pacific Management Review*, 28(1), 1-12. <https://doi.org/https://doi.org/10.1016/j.apmrv.2022.01.003>
- Liu, P., Zhu, B., Yang, M., & Chu, X. (2022). ESG and financial performance: A qualitative comparative analysis in China's new energy companies. *Journal of Cleaner Production*, 379, 134721. <https://doi.org/https://doi.org/10.1016/j.jclepro.2022.134721>
- Kalia, D., & Aggarwal, D. (2023). Examining impact of ESG score on financial performance of healthcare companies. *Journal of Global Responsibility*, 14(1), 155-176. <https://doi.org/https://doi.org/10.1108/JGR-05-2022-0045>
- Domanović, V. (2022). The relationship between ESG and financial performance indicators in the public sector: empirical evidence from the Republic of Serbia. *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies*, 27(1), 69-80.
- Dalal, K. K., & Thaker, N. (2019). ESG and corporate financial performance: A panel study of Indian companies. *IUP Journal of Corporate Governance*, 18(1), 44-59.
- Saygili, E., Arslan, S., & Birkan, A. O. (2022). ESG practices and corporate financial performance: Evidence from Borsa Istanbul. *Borsa Istanbul Review*, 22(3), 525-533. <https://doi.org/https://doi.org/10.1016/j.bir.2021.07.001>
- Giannopoulos, G., Kihle Fagernes, R. V., Elmarzouky, M., & Afzal Hossain, K. A. B. M. (2022). The ESG disclosure and the financial performance of Norwegian listed firms. *Journal of Risk and Financial Management*, 15(6), 237. <https://doi.org/https://doi.org/10.3390/jrfm15060237>
- Citalán, R. E. M., Tinoco, J. D., & Castro, A. M. (2025). Revisión de la Literatura sobre el Rendimiento de la Inversión ESG: un Estudio Bibliométrico. *Revista Mexicana de Economía y Finanzas (REMEF): nueva época*, 20(1), 10. <https://doi.org/10.21919/remef.v20i1.1210>
- Tang, T., & Yang, L. (2023). Shaping corporate ESG performance: role of social trust in China's capital market. *China Finance Review International*, 14(1), 34-75. <https://doi.org/https://doi.org/10.1108/CFRI-07-2023-0187>

- Koroleva, E., Baggieri, M., & Nalwanga, S. (2020). Company performance: Are environmental, social, and governance factors important. *International Journal of Technology*, 11(8), 1468-1477. <https://doi.org/https://doi.org/10.14716/ijtech.v11i8.4527>
- Basnet, A., Blomkvist, M., & Galariotis, E. (2022). The role of ESG in the decision to stay or leave the market of an invading country: The case of Russia. *Economics Letters*, 216, 110636. <https://doi.org/https://doi.org/10.1016/j.econlet.2022.110636>
- Aboud, A., & Diab, A. (2019). The financial and market consequences of environmental, social and governance ratings: The implications of recent political volatility in Egypt. *Sustainability Accounting, Management and Policy Journal*, 10(3), 498-520. <https://doi.org/https://doi.org/10.1108/SAMPJ-06-2018-0167>
- Aureli, S. (2017). A comparison of content analysis usage and text mining in CSR corporate disclosure. *International Journal of Digital Accounting Research*, 17, 1-32. https://doi.org/10.4192/1577-8517-v17_1
- Contala, T., Gerke, A. M., Hoettler, J., & Buettner, R. (2024). Topic Trends in Sustainability Disclosure of German DAX 40 Companies-A Text Mining-Based Analysis. *IEEE Access*, 12, 77300-77335. <https://doi.org/10.1109/ACCESS.2024.3404368>
- Zhou, Y., Wang, X., & Yuen, K. F. (2021). Sustainability disclosure for container shipping: A text-mining approach. *Transport Policy*, 110, 465-477. <https://doi.org/10.1016/j.tranpol.2021.06.020>
- Liew, W. T., Adhitya, A., & Srinivasan, R. (2014). Sustainability trends in the process industries: A text mining-based analysis. *Computers in Industry*, 65, 393-400. <https://doi.org/10.1016/j.compind.2014.01.004>
- Caudron, E., & Vrin, F. (2022). Measuring ESG Performance: A Text Mining Approach. *Louvain School of Management, Université catholique de Louvain. CFA Institute*, [Online] available: <https://www.cfainstitute.org/-/media/documents/protected/esg-candidate/pdf/2021-Chapter3.pdf> [2022 August 20].
- Ning, X., Yim, D., & Khuntia, J. (2021). Online Sustainability Reporting and Firm Performance: Lessons Learned from Text Mining. *Sustainability*, 13(3). <https://doi.org/10.3390/su13031069>
- Kiriu, T., & Nozaki, M. (2020). A Text Mining Model to Evaluate Firms' ESG Activities: An Application for Japanese Firms. *Asia-Pacific Financial Markets*, 27, 621-632. <https://doi.org/10.1007/s10690-020-09309-1>
- Vaio, A. D., Hassan, R., D'Amore, G., & Strologo, A. D. (2024). Digital Technologies for Sustainable Waste Management On-Board Ships: An Analysis of Best Practices From the Cruise Industry. *IEEE Transactions on Engineering Management*, 71, 12715-12728. <https://doi.org/10.1109/TEM.2022.3197241>
- Kang, H., & Kim, J. (2022). Analyzing and Visualizing Text Information in Corporate Sustainability Reports Using Natural Language Processing Methods. *Applied Sciences (Switzerland)*, 12. <https://doi.org/10.3390/app12115614>
- Goloshchapova, I., Poon, S.-H., Pritchard, M., & Reed, P. (2019). Corporate social responsibility reports: topic analysis and big data approach. *The European*

- Journal of Finance*, 25(17), 1637-1654.
<https://doi.org/https://doi.org/10.1080/1351847X.2019.1572637>
- Lu, J., & Jagoda, K. (2023). Sentiment analysis in sustainability accounting reporting: does the tone reveal future environmental performance. *International Journal of Accounting and Finance*, 11, 202-219. <https://doi.org/10.1504/ijaf.2023.10060039>
- Anderson, T., Sarkar, S., & Kelley, R. (2024). Analyzing public sentiment on sustainability: A comprehensive review and application of sentiment analysis techniques. *Natural Language Processing Journal*, 100097. <https://doi.org/https://doi.org/10.1016/j.nlp.2024.100097>
- Baier, P., Berninger, M., & Kiesel, F. (2020). Environmental, social and governance reporting in annual reports: A textual analysis. *Financial Markets, Institutions & Instruments*, 29(3), 93-118. <https://doi.org/https://doi.org/10.1111/fmii.12132>
- Kim & Kim, S. (2017). Sustainable supply chain based on news articles and sustainability reports: Text mining with Leximancer and DICTION. *Sustainability (Switzerland)*, 9. <https://doi.org/10.3390/su9061008>
- Anica-Popa, I.-F., Vrîncianu, M., Anica-Popa, L.-E., Cișmașu, I.-D., & Tudor, C.-G. (2024). Framework for integrating generative AI in developing competencies for accounting and audit professionals. *Electronics*, 13(13), 2621. <https://doi.org/https://doi.org/10.3390/electronics13132621>
- Baltagi, B. H. (2008). *Econometric Analysis of Panel Data*. John Wiley & Sons.
- Chen, S., Song, Y., & Gao, P. (2023). Environmental, social, and governance (ESG) performance and financial outcomes: Analyzing the impact of ESG on financial performance. *Journal of environmental management*, 345, 118829. <https://doi.org/https://doi.org/10.1016/j.jenvman.2023.118829>
- Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset*. John Wiley & Sons.
- Brealey, R. A., Myers, S. C., & Allen, F. (2014). *Principles of corporate finance*. McGraw-hill.
- Arellano, M. (2003). *Panel data econometrics*. OUP Oxford.
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, 1287-1294. <https://doi.org/https://doi.org/10.2307/1911963>
- Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4), 409-428. <https://doi.org/https://doi.org/10.2307/2332391>
- Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach 6rd ed*. Cengage learning.
- Yaffee, R. A. (2005). *A Primer for Panel Data Analysis* [Originally published September 2003].
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251-1271. <https://doi.org/https://doi.org/10.2307/1913827>
- Ferjančič, U., Ichev, R., Lončarski, I., Montariol, S., Pelicon, A., Pollak, S., Šuštar, K. S., Toman, A., Valentinčič, A., & Žnidaršič, M. (2024). Textual analysis of corporate sustainability reporting and corporate ESG scores.

- International Review of Financial Analysis*, 96, 103669.
<https://doi.org/https://doi.org/10.1016/j.irfa.2024.103669>
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*. <https://doi.org/https://doi.org/10.48550/arXiv.2203.05794>
- Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301), 236-244. <https://doi.org/https://doi.org/10.1080/01621459.1963.10500845>
- Pearson, K. (1896). VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, (187), 253-318.
<https://doi.org/https://doi.org/10.1098/rsta.1896.0007>
- Oprea, A. (2022). The Use of Principal Component Analysis (PCA) in Building Yield Curve Scenarios and Identifying Relative-Value Trading Opportunities on the Romanian Government Bond Market. *Journal of Risk and Financial Management*, 15(6). <https://doi.org/10.3390/jrfm15060247>
- Villardón, M. P. G. (1986). Una alternativa de representació simultànea: HJ-Biplot. *Qüestió: quaderns d'estadística i investigació operativa*, 13-23.
- Mensah, L., Arhinful, R., & Owusu-Sarfo, J. S. (2024). Enhancing cash flow management in Ghanaian financial institutions through effective corporate governance practices. *Corporate Governance: The International Journal of Business in Society*.
<https://doi.org/https://doi.org/10.1108/CG-01-2024-0016>

8. Anexos

8.1. Código

El código fuente desarrollado para este Trabajo de Fin de Máster está disponible públicamente en el repositorio de GitHub accesible a través del siguiente enlace: <https://github.com/laurasanzmartin/TFM/>

8.2. Publicación derivada del TFM

A partir del presente Trabajo de Fin de Máster, y en particular del análisis de informes no financieros y de sostenibilidad mediante técnicas de minería de textos, se ha elaborado un artículo científico titulado “Evolving Sustainability priorities in EuroStoxx 50 Companies: A Text Mining and Multivariate Analysis”, el cual ha sido enviado para su evaluación al Journal of Innovation and Knowledge, en concreto al número especial “Transforming Finance and Governance: Knowledge, Innovation, and Sustainability”.

A continuación, se presenta el manuscrito enviado:

Highlights

Evolving Sustainability Priorities in EuroStoxx 50 Companies: A Text Mining and Multivariate Analysis

- European companies are moving from declarative sustainability commitments to operational, results-oriented ESG strategies, reflecting increased regulatory and societal expectations.
- Environmental issues such as GHG emissions reduction and the circular economy are gaining prominence over traditional social priorities, signalling a shift in corporate sustainability agendas.
- The materiality of ESG topics varies by industry, highlighting the need for tailored sustainability strategies that address sector-specific challenges and opportunities.
- Companies with the highest ESG scores stand out for their proactive approach to emerging issues and their adoption of innovative solutions, setting new standards for corporate sustainability.

Evolving Sustainability Priorities in EuroStoxx 50 Companies: A Text Mining and Multivariate Analysis

Abstract

In recent years, sustainability and ESG criteria have become central concerns for companies, investors, and regulators worldwide, reflecting a shift towards more responsible and transparent business practices. As sustainability reporting becomes more widespread and complex, advanced analytical tools such as text mining are increasingly used to systematically analyze large volumes of non-financial corporate disclosures. This study examines the evolution of sustainability priorities among EuroStoxx50 companies from 2017 to 2023 by applying topic modeling and dimensionality reduction techniques to their sustainability and non-financial reports. The analysis reveals a clear transition from predominantly declarative reporting to more operational and results-oriented sustainability strategies. While social themes such as diversity and equality have historically been prominent, recent years have seen a growing emphasis on environmental topics, including GHG emissions reduction, the circular economy, and climate risk management. Thematic clustering and multivariate analysis highlight the integration of internal management, innovation, social responsibility, and governance, as well as the sector-specific nature of ESG priorities. The results indicate that large European companies are adopting a more holistic and tangible approach to sustainability, with increasing focus on measurable actions and impacts.

Keywords: ESG, Text Mining, Sustainability, EuroStoxx50, BERTopic

1. Introduction

Humanity is in continuous progress. Today, with the rapid advancement of economy and society, the world faces both significant opportunities and challenges. Since 2020, with the global spread of COVID-19, environmental, social and governance issues have attracted increasing attention worldwide. The issue of sustainable and integral development has become a hot topic of

debate around the world. In response to the serious problems related to sustainable development in the environmental, social and financial fields, action plans have been implemented such as the adoption of ESG (environmental, social and governance) factors [1]. These factors represent a holistic framework for evaluating a company's performance beyond traditional financial metrics. This broader view recognises that a company's long-term success depends not only on its profitability, but also on its environmental sustainability, social responsibility and governance practices [2].

In 2004, the ESG was developed into a report entitled "Who Cares Wins (WCW)", initiated by the UN Secretary General and the UN Global Compact in collaboration with the Swiss government. In it, ESG is seen as an extension of corporate social responsibility (CSR), which has defined the framework for the ecological transformation of business ¹. ESG not only extends the concept of socially responsible investment, but has also established itself as a global standard for measuring the degree of sustainable development of companies. This perspective encourages companies to shift their focus from increasing their own profit to increasing social value. In this context, the U.S. Sustainability Accounting Standards Board, the United Nations Environment Programme, the Global Reporting Initiative (GRI), many stock exchanges around the world and many other governmental and non-governmental organizations have begun to actively promote ESG principles and their disclosure standards [3].

The rising profile of ESG criteria has been evident in investments, even though the rate of new investments has shown a recent drop. Inflows to sustainable funds increased from USD 5 billion in 2018 to over USD 50 billion in 2020, and then to almost USD 70 billion in 2021. By mid-2022, global sustainable assets totalled approximately \$2.5 trillion [4]. The Principles for Responsible Investment (PRI) documented 3,826 signatories to its framework in 2021, with combined Assets Under Management (AUM) of \$121.3 trillion. Corporations are also placing increasing importance on CSR and ESG issues. Large corporations are building strategies aligned with the UN Sustainable Development Goals, are investing hundreds of millions of dollars in altruistic projects each year and are trying to increase the visibility of their

¹<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/280911488968799581/who-cares-wins-connecting-financial-markets-to-a-changing-world>

sustainability efforts. According to the Tenth Anniversary Report issued by the Governance & Accountability Institute ², a corporate sustainability and ESG consultancy, 92% of S&P 500 companies and 70% of Russell 1000 companies published their sustainability reports by 2020. In short, CSR and ESG appear to be moving from being a peripheral concern to a central concern for both investors and companies [5].

ESG scores quickly became popular because of their ability to enhance corporate reputation, reduce legislative pressure, reduce financial risk and attract more capital. For investors, ESG scores help mitigate the risk associated with environmental or social scandals by anticipating negative impacts on investments. However, they are less effective in assessing the positive effects derived from a company's sustainable actions, such as its contribution to social and environmental well-being. Although some small organizations also engage in sustainable activities, measuring the real impact of these practices remains a challenge [6].

Although recent research, has advanced the identification and analysis of ESG topics in annual reports using advanced topic modelling techniques, the literature lacks studies that transfer this approach to the set of continental European companies, especially in the framework of the new regulations (CSRD³). Moreover, no studies have been found that explore the multivariate relationship between companies and ESG topics using dimensionality reduction techniques such as Principal Component Analysis (PCA) or that use biplots to visualize and analyze how companies position themselves with respect to the different ESG topics in the most recent period. Therefore, there is a relevant gap in the literature regarding the application of multivariate tools for the comparative and visual study of ESG disclosure in Europe, which limits the understanding of sustainability communication dynamics and strategies at the company and sectoral level.

In our research, based on the corpus of EURO STOXX 50 companies' annual reports for the period 2017 to 2023, our aim is to identify and interpret

²<https://www.ga-institute.com/research/research/sustainability-reporting-trends/2021-sustainability-reporting-in-focus/>

³Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU as regards corporate sustainability reporting (Corporate Sustainability Reporting Directive, CSRD), Official Journal of the European Union, L 322, 16 December 2022, pp. 15–80.

the key ESG topics that emerge from the automated text analysis; analyse the hierarchical structure of the topics and identify the most important ESG concepts; observe the development of ESG topic proportions over time and explain any structural breaks identified; observe and interpret the correlation between the ESG score and the ESG topic proportions by industry; and analyse the difference in ESG topic proportions between the companies with the highest and lowest ESG scores. In addition to replicating and adapting the methodology of [7] to the European context, we have extended the analysis by implementing additional dimensionality reduction and visualization techniques. Specifically, we have applied a Principal Component Analysis (PCA) on the distribution matrix of topics identified by BERTopic for the last year of the sample. This technique has allowed us to synthesize the high dimensionality information and explore latent patterns in the disclosure of ESG topics.

Additionally, by using biplots, we have visualized how European companies relate to the different ESG topics, identifying clustering, outliers and possible differentiated disclosure strategies. This visual and multivariate approach provides a novel perspective with respect to the original study, facilitating the interpretation of the relative position of each company in the ESG thematic space and allowing the detection of sectorial and geographical trends or alignments. This represents the first comprehensive application of PCA and biplot visualization to ESG topic analysis in the European context, providing unique insights into company positioning and strategic differentiation within the sustainability landscape.

This paper is structured as follows: first, we present the state of the art, followed by the methodology. Next, we describe the results, and finally, we discuss the findings and present our conclusions.

2. State of the art

Text mining, or textual data analysis, is increasingly used for the purpose of analyzing corporate sustainability reports or non-financial information reports [8, 9]. Since the number of sustainability reports, the interest of stakeholders [10] and the volume of data have increased, the systematic analysis of these reports cannot be performed manually [9]. In this sense, text mining offers an organized approach to extract information and trends from large volumes of textual data [8], which allows exploring unstructured or semi-structured information [10].

This enables practitioners and researchers to identify emerging issues [11], assess ESG performance [12] and understand the strategic intentions behind sustainability decisions [13]. This approach helps to gain insight into the content and structure of reports, as well as to understand the development of issues over time [9]. This helps overcome the limitations of traditional manual content analysis by processing reports consistently, minimizing human subjectivity [14] and providing an automated approach that overcomes the problems associated with the reliability of content analysis [8]. In addition, these tools can identify issues that might be missed with manual coding, reducing analysis time [15] and providing a more efficient and comprehensive way to evaluate companies' sustainability efforts [9].

Text mining techniques are applied in various ways to analyze these reports, providing a deeper understanding of corporate sustainability practices [16]. This includes topic modeling, sentiment analysis, and the identification of key performance indicators [12]. Topic modeling is an application of text mining that helps identify the main themes and topics covered in these reports [9]. A widely used algorithm for this purpose is Latent Dirichlet Allocation (LDA), which is used to discover hidden topic structures in a collection of documents. It assumes that each document is a mixture of a few latent topics, and each topic is characterized by a probability distribution over words [10]. In this context, multiple studies have used this algorithm for their research, [17] used textual analysis with LDA to detect the number and nature of topics included in CSR reports of UK and European companies and identify possible sectoral biases. They found common themes such as "employee safety", "support for employee training", "human rights" and "health care and medicine".

Text mining can also be used to extract key performance indicators associated with corporate disclosure. By analyzing the frequency and context of specific topics, researchers can assess a company's performance in areas such as environmental impact, social responsibility and governance. This allows for a more objective and quantitative assessment of the sustainable efforts being made by companies, helping stakeholders to make informed decisions [10]. In this regard, [13], analyzes how voluntary sustainability disclosure reflects corporate strategy and its effect on financial performance. It uses LDA to identify key issues in the reports, a latent Markov model to understand the relationship with strategic intent, and an empirical analysis using statistical models to assess the impact of these intentions on financial performance. The study concludes that employee orientation has a greater impact

on performance than customer orientation.

Many tools and techniques are used to extract information from sustainability reports. These include dictionary-based approaches, Machine Learning (ML) algorithms and specialized software. Dictionary-based approaches involve using predefined lists of words and phrases related to the topic under study to identify relevant information in the text. [18] creates a specific dictionary of ESG-related terms of 482 words, organized into 40 subcategories, which facilitates a detailed analysis of different ESG-related aspects and uses it to objectively assess ESG reporting activity in companies. ML algorithms such as LDA, Support Vector Machines (SVM) and neural networks are used to automatically identify patterns and relationships in texts. These algorithms can be trained to classify text, extract information or make predictions.

The application of text mining in sustainability reporting offers great benefits such as increased efficiency or objectivity. In addition, it automates the process of analyzing large volumes of text data, enabling researchers to more quickly identify key trends and themes, allowing for more timely and informed decision making. It also reduces subjectivity bias in report analysis, providing a more objective assessment of companies' sustainability performance [14]. Using predefined dictionaries and ML algorithms, the analysis is based on quantitative data and statistical patterns, rather than individual interpretations [19].

Text mining can uncover hidden patterns and relationships that might not be evident through manual analysis [15]. In this way, new insights can be gained into the drivers of sustainability performance, the effectiveness of different strategies, and the alignment of corporate disclosures with stakeholder expectations. [13] highlighted the use of advanced business analytical approaches, namely LDA thematic modeling, to code intangible knowledge included in annual sustainability reports with the objective of inferring a company's strategic intent behind voluntary disclosure.

Effective text mining requires a thorough understanding of the subject matter being analyzed, including the key concepts, terminology and reporting frameworks being used. Without this knowledge, it is difficult to develop appropriate dictionaries, train machine learning algorithms and interpret the results of the analysis. In areas such as environmental sustainability, it is important to master the subject matter [8]. The interpretation of the results, including the assignment of labels to the identified themes can be subjective if based solely on the researcher's opinion. It is important that the interpre-

tation takes into account the context in which we find ourselves so as not to draw erroneous conclusions [16].

As organizations increasingly turn to advanced technologies to manage and analyze large volumes of sustainability-related information, generative AI tools based on Large Language Models (LLM) have the potential to automate various tasks, which could be applied to report generation. However, the adoption of Generative AI raises significant concerns. These systems, while capable of generating novel content, may produce plausible but incorrect answers, often due to biased, incomplete or erroneous data sources with which they were trained. This introduces the risk of algorithmic bias, which may amplify existing patterns or even generate unethical content. This raises concerns about the potential for “greenwashing” and emphasizes the critical need for robust assurance mechanisms. Consequently, a key competency for detecting anomalies is professional judgment, requiring advanced analytical skills and critical thinking [20].

Integrating text mining with other data sources, such as financial, environmental and social media data, can provide a holistic view of sustainable business performance. This can help us identify links between sustainable efforts and financial results, assessing whether the impact of corporate activities on the environment and society, and understanding stakeholders’ perceptions of sustainable performance [8].

3. Methodology

Our sample is composed with the companies of the EuroStoxx50 and the study years chosen were from 2017 to 2023. The sustainability or non-financial reports of each company were downloaded from their official pages in English language. Some were not available and others were not in English, others had password and their sentences could not be extracted. Both Python and Rstudio were used to conduct these analyses.

To begin with, all PDF files were converted to text files (.txt) by extracting content from each PDF page using the PyPDF2 library. Next, a cleaning and filtering process was applied to the sentences extracted from the text files. This process consisted of separating the text into sentences using regular expressions to ensure that each sentence met certain quality criteria, such as a minimum ratio of letters to characters, a maximum ratio of numbers to words, and a minimum sentence length. In addition, URLs, phrases with

unwanted characters were removed, and sentences were checked for correct punctuation and formatting.

Finally, the valid sentences were stored in output files and detailed statistics were generated on the number of files and sentences processed, as well as accepted and rejected sentences, for each company and by year, which can be seen in Table 1. It can be seen that, over the years, the number of sentences processed has increased considerably, with a notable growth from 2017 to 2023. In 2017, 126,170 sentences were processed, while in 2023 this figure exceeded 212,000 sentences, indicating that the volume of content extracted from the reports has more than doubled in this period. This increase could reflect a greater amount of information generated in sustainability reports over the years, possibly due to greater inclusion of detail in the reports and expansion of the topics covered. Despite this increase, the percentage of accepted sentences has remained relatively stable at around 45-50%, indicating that, although more sentences have been processed, the proportion of those that meet established quality criteria (such as letter-to-character ratio and number-to-word ratio) has remained constant. This suggests that the quality of the sentences has not been compromised by the increase in data volume.

Year	Processed files	Processed sentences	Accepted sentences	Accepted %
2017	42	126170	59318	47.01
2018	45	135921	62054	45.65
2019	44	126658	61991	48.94
2020	45	140749	69322	49.25
2021	44	162201	75411	46.49
2022	44	180817	83809	46.35
2023	41	212016	94367	44.51

Table 1: Summary of processed, accepted, and filtered sentences by year.

ESG scores were extracted from LSEG, a widely recognized and authoritative provider of ESG data. This ensures that the quantitative evaluation of companies' ESG performance is based on reliable, standardized, and comparable metrics ⁴.

3.1. Topic modelling

The BERTopic technique was used to model the topics of the non-financial reports of the EuroStoxx50 companies. This has been selected due to its ad-

⁴<https://www.lseg.com/en>

vantages over other language models, such as its flexibility to work with different embedding models, its ability to efficiently separate the clustering and topic representation phases, and its ability to generate rich and dynamic topic representations using the c-TF-IDF technique. These features allow for greater adaptability and accuracy in analysis, especially when working with large volumes of unstructured data, such as sustainability reports. BERTopic is a topic model that combines transformer-based language models, clustering techniques and a class-based variation of TF-IDF to discover latent topics in document collections. To do so, document representations (embeddings) are generated using pre-trained language models, such as Sentence-BERT (SBERT), to create dense vector representations for each document. Documents dealing with the same topic are assumed to be semantically similar and thus close in vector space. Subsequently, the dimensionality of the embeddings is reduced, e.g., using UMAP, to optimize the clustering process. Then, the documents are grouped using a clustering technique such as HDBSCAN, where each group represents a different topic. HDBSCAN allows modeling noise, such as outliers. To define which words are most representative of each group, a modified version of TF-IDF called c-TF-IDF (class-based TF-IDF) is applied, this procedure models the importance of words within document clusters, rather than in individual documents. The selection of BERTopic over more traditional models such as LDA was motivated by several practical and methodological advantages. Unlike LDA, BERTopic leverages pre-trained transformer-based embeddings (e.g., SBERT), which capture contextual semantics more effectively and allow better topic separation even in short or heterogeneous documents. Additionally, BERTopic separates the embedding, dimensionality reduction, and clustering steps, offering more flexibility and interpretability. This is especially valuable in the context of sustainability reporting, where topics are nuanced, multidimensional, and often exhibit semantic overlap. Comparative studies have shown that BERTopic outperforms LDA in terms of topic coherence and stability in complex corpora [21], making it particularly suitable for analyzing non-financial corporate disclosures.

3.2. Time series of topics

In order to analyze how sustainability topics evolved over the years, the percentage of each topic in each report was calculated for each year. To do this, the proportions of each topic within each report were extracted and

the variances of these proportions over the years were calculated using the standard formula for sample variance:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (p_i - \bar{p})^2 \quad (1)$$

where p_i is the proportion of the topic in the report for year i , $\bar{p}$ is the mean proportion across all years, and n is the number of years analyzed. The topics with the greatest variance were identified, as they indicate a greater fluctuation in their relevance within the sustainability reports. Finally, these topics were graphically represented to observe their evolution over time and to detect possible changes in priorities over time.

3.3. Clustering of topics

After the extraction of topics, we calculated the topics that varied the most, calculating the variance of each one of them and, subsequently, a hierarchical clustering of topics was done with Ward's [22] method. Ward's method seeks, at each agglomeration step, to merge the pair of clusters that results in the minimum loss of information. This loss is measured as the minimum increase in the total sum of squares within the clusters (or the weighted variance). Minimizing the variance within clusters is equivalent to maximizing the variance between clusters. Prior to clustering, the keyword vectors associated with each theme were normalized to ensure that themes with higher word counts or higher scores did not dominate the analysis. The optimal number of clusters was determined using a combination of clustering quality metrics such as the Silhouette index and the co-phenetic correlation coefficient, which allowed finding the number of clusters that best represented the relationships between themes.

Distances between topics were calculated using a similarity matrix based on the cosine similarity between the keyword vectors of each topic. Cosine similarity measures the similarity between two vectors and is commonly used in text analysis to assess how similar two documents or, in this case, two topics are.

The vectors for each topic were normalized so that they were all the same length, regardless of how many keywords each topic had. This was done by adding zeros to the smaller vectors (padding) or trimming the larger vectors.

To calculate the cosine similarity between topic vectors, the following formula is used:

$$\text{cosine similarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (2)$$

where A and B are the subject vectors, $A \cdot B$ is the dot product of the two vectors, and $\|A\|$ and $\|B\|$ are the lengths (norms) of the vectors A and B , respectively. The cosine similarity varies between -1 and 1, with the value 1 meaning that they are completely similar, -1 meaning that they are opposites, and 0 meaning that they have no similarity.

Once the cosine similarity between all pairs of subjects was calculated, the resulting similarity matrix was converted to a distance matrix using the following formula:

$$\text{distance}(A, B) = 1 - \text{cosine similarity}(A, B) \quad (3)$$

The distance matrix has values in the range 0 to 2, where 0 indicates that the subjects are identical, 1 indicates that the subjects are orthogonal, and 2 indicates that the subjects are completely opposite. To ensure that the distance matrix is valid, it is ensured that the diagonal of the matrix is exactly zero, since the distance from a topic to itself must be zero. It is ensured that the distance matrix is symmetric, i.e., the distance between topic A and topic B is equal to the distance between topic B and topic A . This distance matrix is used as input to the agglomerative hierarchical clustering algorithm, specifically with Ward's method, which seeks to minimize the variance within clusters at each merge step.

3.4. Correlation of ESG

The correlations between the topics generated by topic modeling and ESG scores were calculated using Pearson's correlation coefficient, which measures the linear relationship between two variables [23]. First, correlations were calculated between the proportions of topics generated by the topic modeling model and the ESG scores of the corresponding companies. Pearson's correlation measures the linear relationship between two variables and has values in the range of -1 to 1, where 1 indicates a perfect positive correlation, -1 a perfect negative correlation, and 0 an absence of linear correlation. Its equation is:

$$\rho(x, y) = \frac{E[xy]}{\sigma_x \sigma_y} \quad (4)$$

For each industry sector, the correlation between ESG scores and proportions of issues per company was calculated. Correlations were calculated independently for each industry. To do this, the data were filtered according to industry and Pearson correlations between ESG scores and proportions of themes were calculated for each industry. To ensure the validity of the correlations, at least three observations per sector were required. The correlations were analyzed and presented by industry, highlighting those that were statistically significant. The results of the correlations were visualized using heat maps and tables showing correlations by sector and topic, with special emphasis on the strongest correlations.

3.5. Differences in topic proportions between companies with the highest and lowest ESG scores

To calculate the differences in topic proportions between companies with the highest and lowest ESG scores, we adjusted the ESG scores of companies by industry by calculating the average ESG score for each industry. Then, we calculated the corrected ESG score for each report by calculating the difference between the observed score and the corresponding industry mean score and dividing the difference by the industry mean score, i.e,

$$ESG_{\text{corr}} = \frac{ESG_{\text{observed}} - ESG_{\text{sector mean}}}{ESG_{\text{sector mean}}} \quad (5)$$

To adjust the proportions of the topics by industry, we first normalize the proportions of the ESG topics within each report, so that for each report, the sum of the proportions of the 30 interpretable topics equals 100, and then calculate the adjusted normalized proportions by industry by subtracting and then dividing by the industry-specific normalized proportion, i.e,

$$\text{Adjusted proportion}_{\text{sector}} = \frac{\text{Proportion}_{\text{topic}} - \text{Proportion average sector}}{\text{Proportion average sector}} \quad (6)$$

This analysis was done specifically for the year 2023, we sorted the reports by industry-adjusted ESG score and extracted the top 25% of reports with the highest scores (top quartile) and the bottom 25% of reports with the lowest scores (bottom quartile) on ESG. We then calculated the industry-adjusted mean proportion for these two groups and measured the difference between these proportions. By adjusting the data by industry, the inherent variability between different industries is removed, allowing for a more

equitable comparison across companies. This adjustment ensures that the observed differences in the proportions of ESG topics are not biased by the particular characteristics of each industry, such as company size or structure, which improves the accuracy and relevance of the analysis. By selecting the best and worst companies (in terms of corrected ESG scores) for the year 2023, we eliminate the general trend of increasing ESG scores. Furthermore, by focusing on the differences between the averages of the top and bottom quartiles of the data (in terms of adjusted ESG scores) we improve the correlation analysis, as we maintain the magnitude of the underlying phenomenon and adequately address linear or nonlinear relationships (whereas Pearson correlation is simply a unitless measure of linear relationships).

3.6. Principal Component Analysis (PCA) and HJ-Biplot

A Principal Component Analysis (PCA) was applied to the 10 most frequent topics in the sustainability reports in order to reduce dimensionality and identify latent structures in corporate ESG disclosures. PCA transforms a set of correlated variables (topics) into a new set of orthogonal components that explain the main sources of variation in the data. Each component is associated with a set of loadings indicating the contribution of each original variable to the component [24].

Based on the PCA, an HJ-Biplot [25] was constructed to provide a joint representation of companies and ESG topics in the same low-dimensional space. This method facilitates the interpretation of how firms are positioned with respect to different sustainability themes. Specifically, the data matrix X , where n is the number of observations (companies) and p the number of variables (topics), was column-centered and possibly standardized.

The goal of the HJ-Biplot is to approximate the matrix X using two lower-dimensional matrices A and B , such that the product AB' closely approximates X :

$$X = AB' + E \quad (7)$$

where E is the approximation error. The matrices A ($n \times q$) y B ($p \times q$) contain the coordinates of the companies and variables, respectively, in the reduced space.

These coordinates are derived from the singular value decomposition (SVD) of X , as:

$$A = U(k)\Lambda(k) \quad \text{y} \quad B = V(k)\Lambda(k) \quad (8)$$

where $U(k)$ and $V(k)$ are the matrices of singular vectors, and $\Lambda(k)$ is a diagonal matrix with the singular values corresponding to the first k principal components. The matrix A represents the positions of the companies (observations) in the reduced space, and B contains the factor loadings of the original variables (topics). If the data are standardized, these loadings can be interpreted as the correlation of each variable with each principal component.

This biplot representation enables the simultaneous visualization of company positioning and topic relationships, facilitating the identification of companies with similar ESG profiles and the specific sustainability themes that shape their strategic positioning within the European context.

4. Results

The table 2 gathers the topics identified in the analyzed reports, using the BERTopic topic modeling methodology. The most frequent topic is “Diversity and Equality”, which evidences a clear concern and priority in the reports for aspects such as gender equality, inclusion of young people and women, and promotion of an equitable working life. The associated keywords (students, gender equality, green bonus, STEM, volunteering) reinforce the idea of a cross-cutting approach to diversity, both socially and professionally. Next, topics related to “Finance and Accounting”, “Electric Mobility”, and “Circular Economy” stand out. This suggests that, although social sustainability is a priority, economic and environmental aspects also have a significant weight. The presence of terms such as IFRS, fair value, recycling, reuse and electric mobility indicates an integration of sustainability into the financial and operational management of organizations. “Corporate Governance” and “Ethics and Compliance” are also relevant topics, reflecting the importance of transparency, accountability and corporate integrity in reporting. Keywords point to oversight structures, codes of ethics and whistleblower mechanisms, key elements for responsible governance. Topics such as “GHG Emissions”, “Renewable Energy”, “Climate Risks”, and “Biodiversity and Forests” show that the environmental dimension is treated broadly, ranging from emissions management to ecosystem protection and

energy transition. Less frequent but equally relevant topics, such as “Cybersecurity”, “Health and Safety at Work”, “Human Rights”, “Sustainable Agriculture”, “Sustainable Procurement” and “Heritage and Culture”. This demonstrates the thematic diversity of the reports and the breadth of ESG challenges facing organizations. The distribution of topics reveals a clear orientation towards social sustainability and governance, without neglecting environmental and financial aspects.

Topic	Keywords
Diversity and equality	students, female, young people, loans, gender equality, work life, green bond, stem, volunteering, global reporting
Finance and accounting	taxes, ifrs, impairment, liabilities, fair value, lease, consolidated financial, debt, consolidated financial statements, goodwill
Electric mobility	charging, battery, cars, traffic, batteries, fleet, automotive, vans, urban, trucks
Circular economy	recycled, reuse, plastics, cotton, circularity, waste generated, hazardous waste, waste management, recycled materials, reused
Corporate governance	supervisory board, general meeting, executive board, nomination, independent director, shareholders meeting, non executive, voting, lead independent, lead independent director
GHG emissions	scope emissions, ghg protocol, tons, emissions scope, carbon emissions, reduce emissions, indirect emissions, emission factors, scope scope, emis
Renewable energy	solar, energy consumption, decarbonization, photovoltaic, renewable sources, renewable electricity, lighting, mwh, heating, heat
Ethics and compliance	whistleblowing, ethics compliance, harassment, bribery, code ethics, whistleblower, compliance management, group compliance, business conduct, compliance officer
Climate risks	tcfd, weather, climate risk, climate related risks, climate related financial, related financial, extreme weather, related financial disclosures, financial disclosures, climate risks
Biodiversity and forests	species, farmers, forest, trees, habitats, soil, hectares, crop, forests, restoration
Risk management	group risk, appetite, risk appetite, erm, risk mapping, internal control, enterprise risk, risk factors, risks identified, enterprise risk management
Cybersecurity and data protection	cyber, cybersecurity, personal data, data privacy, cyber security, information security, gdpr, attacks, threats, protection regulation
Occupational health and safety	occupational health, occupational health safety, accidents, ehs, accident, occupational safety, safety health, safety culture, safety management, root
Human rights	global compact, ilo, business human rights, business human, rights issues, human rights issues, guiding business, guiding business human, slavery, international labour
ESG risks	esg issues, esg risks, esr, governance esg, esg risk, social governance esg, esg aspects, esg related, esg performance, asset management
Sustainable agriculture	milk, stoxx, nutrition, dairy, farmers, regenerative, regenerative agriculture, animal, food safety, whitewave
Remuneration	salary, bonus, wage, variable remuneration, variable compensation, executive board, remuneration policy, remuneration report, supervisory board, equal pay
Crafts and materials	leather, savoir, faire, savoir faire, leather goods, tanneries, craftsmanship, craftspeople, silk, textile
Water management	water consumption, wastewater, water stress, cooling, water use, water management, withdrawal, use water, drinking water, water used
Sustainable procurement	supplier code, supplier code conduct, responsible purchasing, supplier management, sustainable procurement, sustainable supply, qualification, procurement process, sustainable supply chain, supplier sustainability
Heritage and luxury	maisons, maison, luxury, beauty, museum, heritage, collections, exhibition, perfumes, artistic
Internal audit	internal audit, internal control, audit risk, audit committee, auditor, external auditors, audits conducted, internal audits, statutory auditors, audit risk management
Energy transition	energy transition, decarbonization, balance sheet performance indicators, attachment, attachment balance sheet, balance sheet performance indicators, attachment balance sheet performance indicators, performance indicators, indicators, balance sheet
Infrastructure and workforce	autoroutes, airports, workforce related, report board directors, directors workforce related, board directors workforce, directors workforce, workforce related social, concessions, related social environmental
Visual health	eyewear, lenses, optical, lens, eye, eyes, frames, ban, grandvision, laboratories
Compliance	group does, political parties, consolidated financial, consolidated financial statements, non financial statement, financial statement, fines, jurisdictions, breaches, based work dividend, naspers, share capital, dividends, vesting, target achievement, sustainability review, financial statements information, performance review, statements information
Dividends and equity	statement non financial, statement non, annual report statement, report statement non, report statement, group annual report, criminal, textile, manufacturers suppliers, chemical products
Non-financial reports	capex, opex, expenditure, economic activities, taxonomy regulation, environmental objectives, capital expenditure, delegated act, dnsh, eligibility
Sustainable investments	utility, future sustainability, climate action, does business, purpose values, dividend, energy model, fight climate change, fight climate, section chapter
Climate strategy and action	

Table 2: Predominant topics in the reports

Figure 1 illustrates the time evolution of the 13 topics with the highest variability in EuroStoxx50 sustainability reporting between 2017 and 2023. The analysis reveals several trends relevant to understanding the dynamics of sustainability priorities of large European companies.

Diversity and equality stands out as the dominant theme throughout the period, remaining so until 2020, when it begins a progressive decline to stand at around 30% in 2023. This evolution suggests that, although it remains the central focus of the reports, its relative weight has declined slightly in recent years, possibly in favor of other emerging themes. In contrast, finance and accounting shows a continuing downward trend, which could indicate a shift in focus from traditional financial aspects to more social, environmental and governance issues. Other topics such as GHG emissions, circular economy, electric mobility and climate risks experience moderate growth or remain stable, reflecting the consolidation of the environmental agenda in the reports. The circular economy and GHG emissions in particular show a slight upward trend, in line with the growing regulatory and social pressure on decarbonization and waste management.

For their part, topics such as corporate governance, biodiversity and forests, sustainable investments and ESG risks maintain a relatively constant presence, suggesting that, although they are not the most prominent, they are part of the thematic core of corporate sustainability.

This evolution is evidence of a progressive shift in the corporate sustainability priorities of the EuroStoxx50. While the early years were dominated by conceptual and declarative frameworks, such as diversity and equality or financial and corporate governance issues, recent years have seen a greater focus on operational issues and specific environmental impacts, such as GHG emissions, the circular economy and climate risks. This shift reflects how companies are adapting their sustainability strategies to respond to both growing investor and consumer expectations and regulatory requirements. Thus, corporate sustainability is evolving from commitment and declaration of intent to the implementation of tangible actions in decarbonization, materials transformation and environmental risk management, consolidating a more results-oriented and transparent agenda.

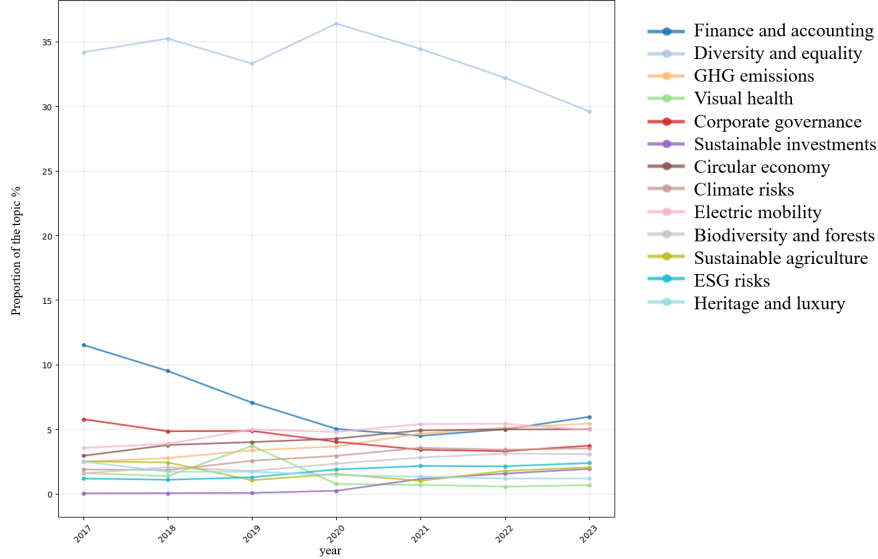


Figure 1: Evolution of the most varying topics

Figure 2 presents the dendrogram resulting from the hierarchical clustering of the topics identified in the EuroStoxx50 sustainability reports. To determine the optimal number of clusters, several evaluation metrics have been used (Figure 2), such as the Silhouette Score, Davies-Bouldin Index, RMSSTD, R-squared and Semi-Partial R-squared, obtaining a cophenetic coefficient of 0.5892, which indicates an adequate representation of the original distances. According to these indicators, a partition into four clusters has been selected, which offers the best balance between internal cohesion and separation between groups.

This first cluster groups together topics focused on internal management of the organization, innovation and operational efficiency with a sustainability perspective. It includes topics such as internal audit, cybersecurity and data protection, compliance, climate strategy and action, renewable energy, sustainable procurement, water management, infrastructure and workforce, biodiversity and forestry, sustainable investments, electric mobility and visual health. The presence of these topics in the same group suggests that companies tend to address internal process improvement, adoption of clean technologies and responsible resource management together. It also highlights the integration of technological aspects (such as cybersecurity) with environmental issues (renewable energy, water management), reflecting a holistic

vision of operational sustainability.

The second cluster brings together issues of a markedly social and ethical nature, as well as the management of risks that are not strictly financial. Diversity and equality, finance and accounting, ethics and compliance, risk management, occupational health and safety, sustainable agriculture, climate risks, human rights, energy transition, ESG risks and remuneration are grouped here. This grouping reflects the interrelationship between corporate social responsibility, equal opportunity, human rights protection and emerging risk management. The inclusion of topics such as climate risks and energy transition alongside social aspects is evidence of the increasing cross-cutting nature of sustainability, where social and environmental challenges are managed in an integrated manner.

The third cluster is formed exclusively by the topic of non-financial reports. Its isolation from the rest of the topics indicates that this type of reporting presents distinctive characteristics and approaches, probably related to its cross-cutting function of collection, disclosure and regulatory compliance in sustainability. This result suggests that non-financial reports act as a backbone, but with sufficient specificity to differentiate themselves thematically from the rest.

Finally, the fourth cluster groups together topics related to corporate governance and high-impact environmental issues, such as heritage and luxury, corporate governance, circular economy and GHG emissions. The proximity between circular economy and GHG emissions reflects the growing importance of materials transformation and emissions reduction in corporate strategy. The inclusion of heritage and luxury suggests that intangible and cultural asset management is also linked to governance and environmental sustainability in the context of large European companies.

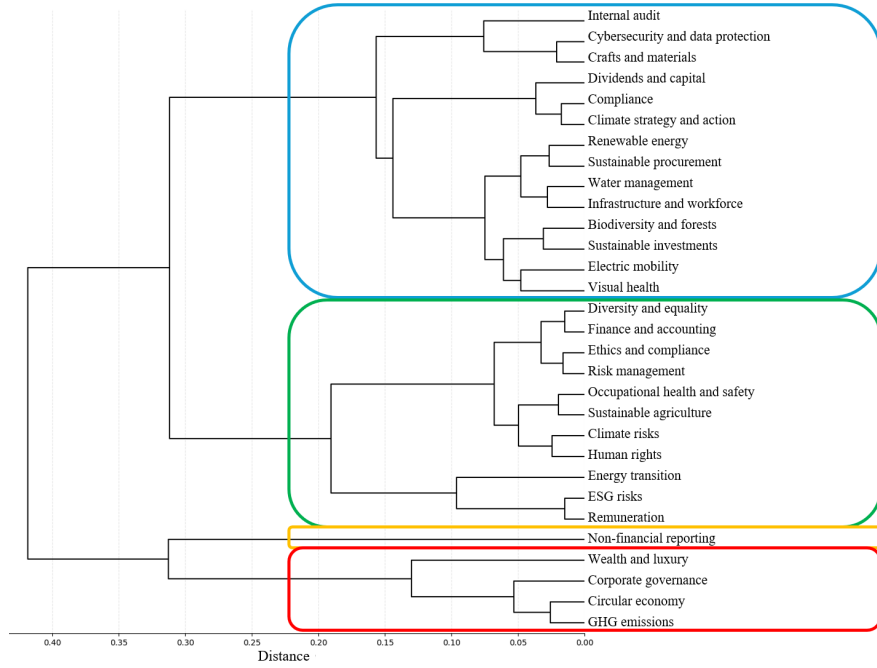


Figure 2: Hierarchical clustering of topics

Figure 3 shows significant correlations between ESG scores and the proportion of issues by industry, highlighting how sustainability priorities vary across sectors. Sector analysis reveals that in basic materials, circular economy and human rights are strongly associated with higher ESG scores, while corporate governance and diversity and equality do not show a positive association with ESG scores. In cyclical consumer goods, human rights and electric mobility drive ESG performance, whereas crafts and materials and visual health are not linked to higher ESG scores.

Non-cyclical consumer goods are most influenced by sustainable agriculture, with electric mobility, diversity and equality, and water management not showing a positive relationship with ESG scores. The energy sector's ESG scores are positively correlated with GHG emissions reduction and electric mobility, while diversity and equality and human rights are not associated with higher ESG scores. In the financial sector, ESG risk management is the main positive driver, while non-financial reporting and finance and accounting are not linked to improved ESG performance. In the health sector, diversity and equality and biodiversity and forests are important, but visual

health is not associated with higher ESG scores.

In industry, water management and biodiversity and forests are positively correlated with ESG scores, while occupational health and safety and diversity and equality are not linked to higher ESG scores. Real estate is unique, with biodiversity and forests closely linked to ESG performance, but human rights and non-financial reporting are not associated with higher ESG scores. In technology, diversity and equality, renewable energy, and cybersecurity and data protection are positively correlated, while remuneration, corporate governance, and internal audit are not linked to higher ESG scores. For utilities, diversity and equality and biodiversity and forests are important, while climate risk management and corporate governance are not associated with higher ESG scores.

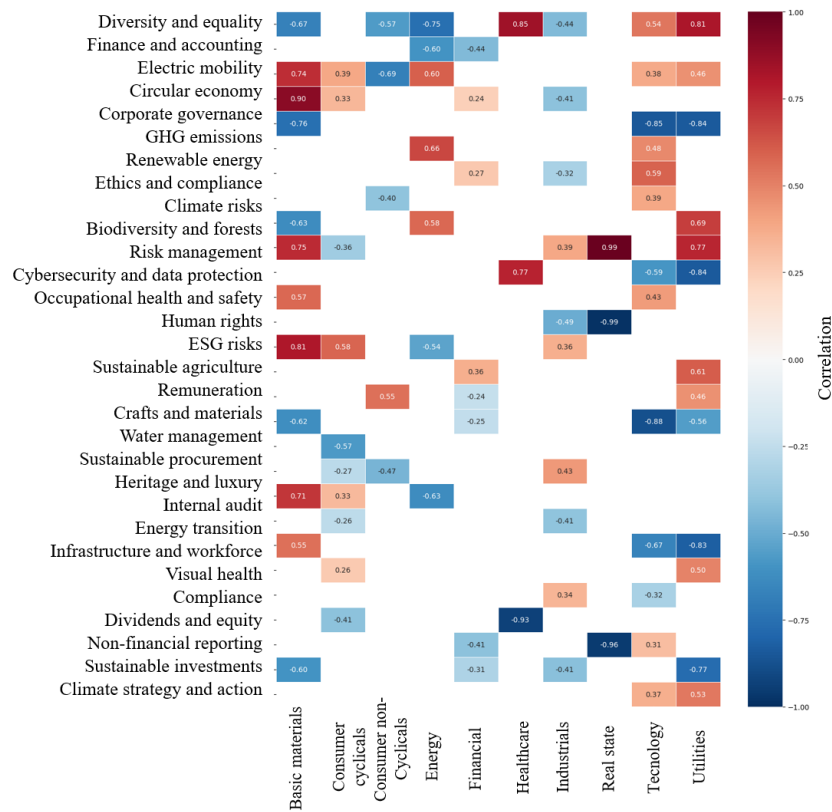


Figure 3: Correlations between ESG scores and the proportion of topics by sector

In Table 3, which shows the industry-adjusted topic differences between companies with high and low ESG scores in 2023, we can see how topic focus and priorities vary between the top quartile (Q1) and the bottom quartile (Q4) in terms of ESG.

A positive value indicates that companies with better ESG scores (in the top quartile) are more likely to address certain topics compared to lower-scoring companies (bottom quartile). For example, the topic “Infrastructure and workforce” has a difference of 1.6363, suggesting that companies with higher ESG scores in 2023 address infrastructure and workforce issues to a greater extent.

Topics such as “Sustainable agriculture” (1.3075) and “Sustainable investments” (0.7475) also have positive differences, reflecting a greater focus of companies with better ESG scores on sustainability issues and responsible agricultural practices in 2023. Thus, we can infer that companies with a stronger ESG focus tend to have a greater concern for energy transition, sustainable agriculture and responsible investment, which is consistent with global trends in sustainability.

Conversely, negative values indicate that companies with lower ESG scores (in the bottom quartile) are more involved in these topics. For example, “Dividends and equity” (-1.3796) has a significant negative difference, suggesting that companies with lower ESG scores pay more attention to dividend and equity issues in 2023. Similarly, “GHG emissions” (-0.3186) and “Risk management” (-0.028) show that companies with low ESG performance in 2023 have a higher proportion of discussion on GHG emissions and risk management.

Topic	Top Quartile Avg	Bottom Quartile Avg	Difference
Infrastructure and workforce	1.1406	-0.4958	1.6363
Dividends and equity	-0.0821	1.2975	-1.3796
Non-financial reports	0.7798	-0.5981	1.3779
Sustainable agriculture	1.2292	-0.0783	1.3075
Energy transition	0.7569	-0.4140	1.1709
Sustainable investments	0.4984	-0.2491	0.7475
Visual health	0.3856	-0.3152	0.7008
Climate strategy and action	0.3465	-0.2891	0.6356
Internal audit	0.4510	-0.0643	0.5152
Remuneration	0.1147	0.6202	-0.5056
Heritage and luxury	0.1626	-0.3327	0.4953
Biodiversity and forests	0.1324	-0.3206	0.4530
Water management	-0.1256	-0.5230	0.3974
ESG risks	-0.0912	0.2283	-0.3195
GHG emissions	-0.2425	0.0761	-0.3186
Ethics and compliance	-0.2371	0.0618	-0.2989
Circular economy	0.0665	-0.2169	0.2834
Corporate governance	0.3158	0.5980	-0.2821
Compliance	0.2054	0.4604	-0.2551
Climate risks	-0.2865	-0.0657	-0.2209
Cybersecurity and data protection	-0.2205	-0.0103	-0.2102
Crafts and materials	0.1869	0.3764	-0.1895
Electric mobility	0.0173	-0.1216	0.1389
Sustainable procurement	-0.1147	0.0125	-0.1272
Finance and accounting	0.3242	0.4428	-0.1186
Occupational health and safety	-0.0271	-0.1330	0.1059
Human rights	0.0322	-0.0672	0.0994
Diversity and equality	-0.0834	-0.1440	0.0607
Risk management	-0.0256	0.0023	-0.0280
Renewable energy	-0.1714	-0.1597	-0.0117

Table 3: Differences in the proportions of topics between companies with high and low ESG scores (2023)

A PCA has been carried out with the topics that appear most frequently in the sustainability reports, to see how the companies are associated with these topics in the last year under study. PC1 accounts for 27.76% of the variance. This first component clearly differentiates companies that prioritize diversity and equality, GHG emissions, renewable energy, climate risks, and biodiversity and forests from those that focus on corporate governance and ethical governance. This indicates that companies with higher scores in this component are more focused on social and environmental aspects, integrating environmental sustainability and social equity into their strategies. Those with low scores on PC1 focus more on corporate governance and ethical governance, reflecting an orientation towards control structures,

regulatory compliance and ethical management, possibly with less emphasis on environmental and social operational aspects.

PC2 accounts for 17.5% of the variance. The second component is strongly associated with the integration of environmental sustainability into business strategy. Companies with high scores in PC2 stand out for the presence of topics such as circular economy, climate risk management and biodiversity, which indicates an advanced internalization of environmental criteria in decision making and risk management. A high score in this component suggests that the company not only reports, but also implements concrete actions to mitigate environmental impacts, manage emissions and promote the conservation of natural resources. Conversely, a low score in this component indicates a strong association with climate risks and with diversity and equality.

Finally, PC3 accounts for 14.59% of the total variance and has a negative association with electric mobility and ethics and compliance. High scores in this component indicate alignment with GHG emissions and with biodiversity and forests. PC3 differentiates companies that prioritize emissions management and biodiversity, but show less commitment to electric mobility and ethics and compliance, possibly because they belong to less innovative or traditional sectors in terms of energy.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Diversity and equality	0.316	-0.408	-0.100	0.359	-0.003	0.226	0.618	-0.122	-0.377	0.064
Finance and accounting	-0.498	-0.120	0.274	-0.081	0.171	-0.196	0.135	0.069	-0.237	0.713
Electric mobility	-0.091	0.103	-0.438	-0.496	-0.620	-0.107	0.134	-0.117	-0.335	0.060
Circular economy	0.086	0.504	-0.204	0.474	0.070	-0.560	0.035	0.218	-0.322	-0.028
Corporate governance	-0.518	-0.070	0.220	-0.088	0.182	-0.030	0.121	-0.082	-0.408	-0.670
GHG emissions	0.370	-0.110	0.354	-0.399	-0.031	-0.189	0.204	0.686	-0.067	-0.112
Renewable energy	0.354	0.178	0.089	-0.407	0.443	-0.326	0.202	-0.570	0.014	0.023
Ethics and compliance	0.059	-0.184	-0.562	-0.217	0.552	0.183	-0.321	0.252	-0.303	0.055
Climate risks	0.236	-0.491	0.213	0.111	-0.210	-0.360	-0.574	-0.221	-0.306	-0.007
Biodiversity and forests	0.215	0.481	0.373	-0.030	-0.048	0.529	-0.226	-0.063	-0.478	0.131

Table 4: Loadings of topics in the components

Figures 4 and 5 present HJ-Biplots of EuroStoxx50 companies, with the top ESG performers highlighted in green, the bottom in red, and the rest in gray. This approach allows for a clear identification of the main associations between companies and the principal thematic dimensions of the ESG domain.

Figure 4, which displays the first and second principal components, reveals two major dimensions. The first principal component (PC1) sepa-

rates companies that prioritize diversity and equality, GHG emissions, renewable energy, climate risks, and biodiversity—such as **TOTALENERGIES** and **AIR LIQUIDE**—from those that focus more on corporate governance and finance and accounting, like **DEUTSCHEBOERSE**, **FERRARI**, **SGOB**, and **DANONE**. This highlights the different strategic emphases within the index, with some firms integrating social and environmental aspects more deeply than others.

PC2 is linked to the internalization of environmental sustainability in business strategy. Companies with positive values on this axis, such as **L'ORÉAL**, **BASF**, **ADIDAS**, and **ENI**, demonstrate a strong alignment with circular economy, biodiversity, and climate risk management. In contrast, financial sector companies like **ALLIANZ**, **MUNICHRE**, **SANTANDER**, **ING**, and **BBVA** are positioned at the negative end of PC2, indicating a lesser strategic focus on environmental factors but greater focus on climate risks and diversity and equality.

Figure 5, which combines the first and third principal components, adds a third dimension to the analysis. PC3 is negatively associated with electric mobility and ethics and compliance, and positively associated with GHG emissions management and biodiversity. This analysis enables the identification of companies such as **MERCEDES**, **VOLKSWAGEN**, **STELLANTIS**, and **SAP**, which are strongly aligned with the energy transition and the integration of clean technologies. These firms demonstrate a clear strategic focus on advancing electric mobility and driving innovation in sustainability. In contrast, companies such as **TOTALENERGIES**, **ENI**, **BBVA**, **BNP**, and **DEUTSCHEBOERSE** are positioned at the positive end of PC3, indicating a greater emphasis on emissions management and biodiversity. While these companies are actively engaged in addressing climate impacts and preserving natural ecosystems, their strategic orientation places comparatively less weight on electric mobility and renewable energy initiatives.

From a sectoral and geographic perspective, financial sector companies tend to cluster in areas associated with ethics, compliance, and climate risk management. Automotive and industrial companies display varying degrees of alignment with the energy transition and environmental themes. In the consumer and luxury sector, certain companies, such as **L'ORÉAL** and **ADIDAS**, stand out for their integration of environmental sustainability into their strategies. French companies generally show a clear orientation toward environmental and social priorities, while German companies tend to emphasize governance and compliance, with differing levels of environmental integra-

tion.

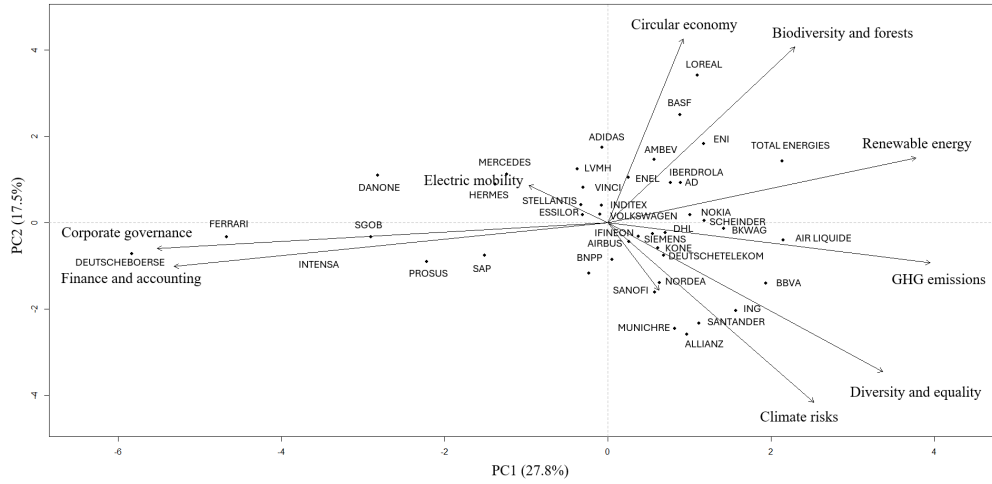


Figure 4: HJ-Biplot

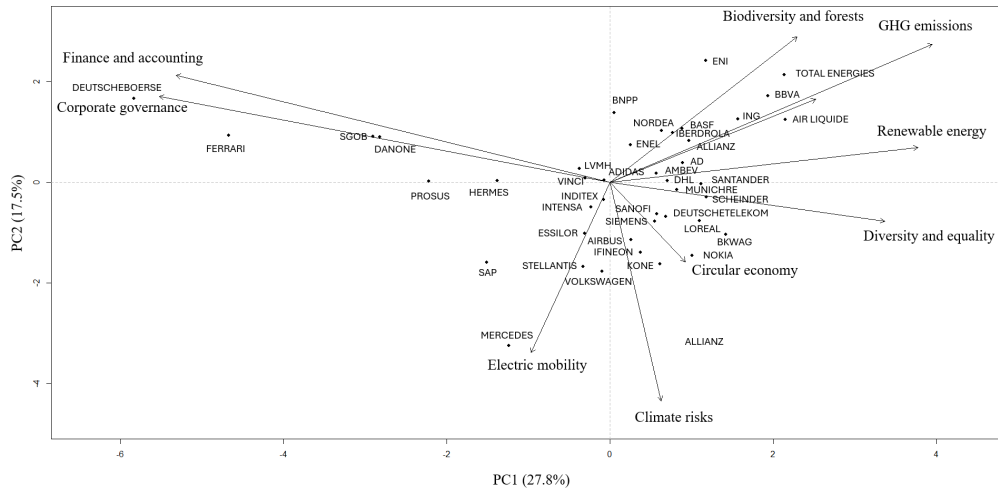


Figure 5: HJ-Biplot

5. Discussion

This study highlights the potential of text mining to uncover trends and patterns in sustainability reports from european companies. However, de-

spite its benefits, the use of text mining in sustainability reporting faces several challenges and limitations. These include data quality issues, the need for specific knowledge, and misinterpretation of results. The quality of the reports can vary significantly, which can affect the accuracy and reliability of the results. Inconsistencies in formats, use of ambiguous language, inclusion of irrelevant information, or lack of truthful information can affect [9, 11, 17].

Additionally, the study is limited by the inherent subjectivity involved in assigning semantic labels to the topics identified by BERTopic. Although keywords provide guidance, the interpretation and naming of topics rely on the researcher’s judgment, which can introduce bias. Another limitation concerns potential language bias, as only English-language reports were analyzed. This may underrepresent nuances in sustainability discourse from companies reporting primarily in other European languages. Future research should consider incorporating reports in multiple languages and validating topic labels through expert review to enhance reliability and cross-cultural applicability.

It is crucial to recognize that, while this analysis provides valuable insight into the evolution of sustainability in large European companies, it has certain limitations inherent in its scope and methodology. A notable limitation is its exclusive reliance on EuroStoxx50 companies, which may limit the generalizability of the findings. The sample consists of large, well-established firms that are subject to greater regulatory and stakeholder scrutiny, and thus may not be representative of the broader European business environment, particularly smaller or less capital-intensive companies.

Future research should aim to extend this analysis to a more diverse set of companies, such as those included in the EuroStoxx600, and over a longer time period. This would provide a more comprehensive view of sustainability trends and enable a deeper understanding of the resilience and long-term evolution of corporate sustainability strategies.

Finally, while text mining offers powerful analytical capabilities, its application must be carefully adapted to the evolving and sector specific landscape of corporate sustainability. The insights from this study accentuate the importance of addressing both the opportunities and the challenges inherent in using automated techniques to understand the changing priorities and practices of large European companies.

6. Conclusions

Sustainability in large European companies is undergoing a significant transformation from a predominantly declarative to a more operational and results-based approach. This evolution reflects a more mature understanding of ESG challenges, as well as higher regulatory and societal expectations.

There is a clear shift in sustainability priorities. Previously dominant issues such as diversity and equality remain relevant, but have ceded some of their relative weight to high-impact environmental issues. GHG emissions reduction, the circular economy, and climate risk management have gained prominence, signaling a growing urgency and commitment to concrete environmental action.

Leading companies now address sustainability in a holistic manner, integrating a wide range of issues including internal management, innovation, operational efficiency, and social and ethical aspects, as well as governance. This comprehensive perspective is crucial for the effective management of ESG challenges. However, the materiality of ESG issues is not uniform; it varies significantly by sector. For example, the circular economy is central in the basic materials industry, while electric mobility and decarbonization are priorities in the energy sector. This sector specificity underscores the need for sustainability strategies tailored to each industry.

Companies with the best ESG scores are distinguished by their ability to anticipate and address emerging issues such as sustainable infrastructure, the workforce of the future, sustainable agriculture, and responsible investments. This proactive approach and adoption of innovative solutions are hallmarks of leading sustainability companies. Practices that were once differentiators, such as governance and risk management, have now become expected minimum standards, indicating a generalization of these core practices across the corporate landscape.

There is a clear strategic differentiation in the European business environment: some companies prioritize the integration of social and environmental aspects, such as renewable energy and biodiversity, while others emphasize governance and ethical compliance. Regardless of the approach, the general trend is toward greater transparency and a focus on tangible results. Sustainability reports increasingly reflect a commitment to implementing concrete actions in areas such as decarbonization, materials transformation, and environmental risk management.

References

- [1] T.-T. Li, K. Wang, T. Sueyoshi, D. D. Wang, Esg: Research progress and future prospects, *Sustainability* 13 (21) (2021) 11663.
- [2] G. Friede, T. Busch, A. Bassen, Esg and financial performance: aggregated evidence from more than 2000 empirical studies, *Journal of sustainable finance & investment* 5 (4) (2015) 210–233.
- [3] P.-C. Jiang, G.-F. Feng, H.-J. Wang, C.-P. Chang, Csr from different perspectives: The global esg indexes updated, *Corporate Social Responsibility and Environmental Management* 31 (5) (2024) 4694–4714.
- [4] L. Pérez, V. Hunt, H. Samandari, R. Nuttall, K. Biniek, Does esg really matter—and why, *McKinsey Quarterly* 60 (1) (2022).
- [5] P. R. Rau, T. Yu, A survey on esg: investors, institutions and firms, *China Finance Review International* 14 (1) (2024) 3–33.
- [6] A. Clément, É. Robinot, L. Trespeuch, The use of esg scores in academic literature: a systematic literature review, *Journal of Enterprising Communities: People and Places in the Global Economy* 19 (1) (2025) 92–110.
- [7] U. Ferjančič, R. Ichev, I. Lončarski, S. Montariol, A. Pelicon, S. Pollak, K. S. Šuštar, A. Toman, A. Valentinčič, M. Žnidaršič, Textual analysis of corporate sustainability reporting and corporate esg scores, *International Review of Financial Analysis* 96 (2024) 103669.
- [8] S. Aureli, A comparison of content analysis usage and text mining in csr corporate disclosure, *International Journal of Digital Accounting Research* 17 (2017) 1–32. doi:10.4192/1577-8517-v17_1.
- [9] T. Contala, A. M. Gerk, J. Hoettler, R. Buettner, Topic trends in sustainability disclosure of german dax 40 companies-a text mining-based analysis, *IEEE Access* 12 (2024) 77300–77335. doi:10.1109/ACCESS.2024.3404368.
- [10] Y. Zhou, X. Wang, K. F. Yuen, Sustainability disclosure for container shipping: A text-mining approach, *Transport Policy* 110 (2021) 465–477. doi:10.1016/j.tranpol.2021.06.020.

- [11] W. T. Liew, A. Adhitya, R. Srinivasan, Sustainability trends in the process industries: A text mining-based analysis, *Computers in Industry* 65 (2014) 393–400. doi:10.1016/j.compind.2014.01.004.
- [12] E. Caudron, F. Vrans, Measuring esg performance: A text mining approach, Louvain School of Management, Université catholique de Louvain. CFA Institute,[Online] available: <https://www.cfainstitute.org/-/media/documents/protected/esg-candidate/pdf/2021-Chapter3.pdf> [2022 August 20] (2022).
- [13] X. Ning, D. Yim, J. Khuntia, Online sustainability reporting and firm performance: Lessons learned from text mining, *Sustainability* 13 (3) (2021). doi:10.3390/su13031069.
URL <https://www.mdpi.com/2071-1050/13/3/1069>
- [14] T. Kiriü, M. Nozaki, A text mining model to evaluate firms' esg activities: An application for japanese firms, *Asia-Pacific Financial Markets* 27 (2020) 621–632. doi:10.1007/s10690-020-09309-1.
- [15] A. D. Vaio, R. Hassan, G. D'amore, A. D. Strologo, Digital technologies for sustainable waste management on-board ships: An analysis of best practices from the cruise industry, *IEEE Transactions on Engineering Management* 71 (2024) 12715–12728. doi:10.1109/TEM.2022.3197241.
- [16] H. Kang, J. Kim, Analyzing and visualizing text information in corporate sustainability reports using natural language processing methods, *Applied Sciences (Switzerland)* 12 (6 2022). doi:10.3390/app12115614.
- [17] I. Goloshchapova, S.-H. Poon, M. Pritchard, P. Reed, Corporate social responsibility reports: Topic analysis and big data approach (2019).
URL <https://www.unpri.org/about/the-six-principles>.
- [18] P. Baier, M. Berninger, F. Kiesel, Environmental, social and governance reporting in annual reports: A textual analysis, *Financial Markets, Institutions & Instruments* 29 (3) (2020) 93–118.
- [19] D. Kim, S. Kim, Sustainable supply chain based on news articles and sustainability reports: Text mining with leximancer and diction, *Sustainability (Switzerland)* 9 (6 2017). doi:10.3390/su9061008.

- [20] I.-F. Anica-Popa, M. Vrîncianu, L.-E. Anica-Popa, I.-D. Cişmaşu, C.-G. Tudor, Framework for integrating generative ai in developing competencies for accounting and audit professionals, *Electronics* 13 (13) (2024) 2621.
- [21] M. Grootendorst, Bertopic: Neural topic modeling with a class-based tf-idf procedure, arXiv preprint arXiv:2203.05794 (2022).
- [22] J. H. Ward Jr, Hierarchical grouping to optimize an objective function, *Journal of the American statistical association* 58 (301) (1963) 236–244.
- [23] K. Pearson, Vii. mathematical contributions to the theory of evolution.—iii. regression, heredity, and panmixia, *Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character* (187) (1896) 253–318.
- [24] A. Oprea, The use of principal component analysis (pca) in building yield curve scenarios and identifying relative-value trading opportunities on the romanian government bond market, *Journal of Risk and Financial Management* 15 (6) (2022). doi:10.3390/jrfm15060247.
URL <https://www.mdpi.com/1911-8074/15/6/247>
- [25] M. P. G. Villardón, Una alternativa de representació simultànea: H_j-biplot, *Qüestiió: quaderns d'estadística i investigació operativa* (1986) 13–23.