sciendo | Economics and Business Review | https://www.ebr.edu.pl | Vol. 11 (3), 2025: 91–117 | https://doi.org/10.18559/ebr.2025.3.1895

# Clustering S&P 500 companies by machine learning for sustainable decision-making

Cansu Ergenç[1] Rafet Aktaş[2]

**Abstract**

This study examines the Environmental, Social, and Governance (ESG) performance of S&P 500 companies using three clustering algorithms: $K$-Means, Gaussian Mixture Model, and Agglomerative Clustering. ESG scores from leading data providers are analysed to uncover sectoral patterns and performance trends. The findings indicate that technology and healthcare firms achieve the highest ESG scores, particularly in the governance and social dimensions, while the industrial and energy sectors face the greatest environmental challenges. Among the methods compared, $K$-Means demonstrates superior clustering performance by forming compact and well-separated ESG groups. These results offer a robust foundation for sector-specific ESG benchmarking, supporting investors and policymakers in identifying sustainability leaders, assessing risk, and targeting areas for improvement.

**Keywords**

- sustainability
- clustering algorithms
- machine learning

**JEL codes:** Q01, Q56, Q57

Article received 2 December 2024, accepted 1 August 2025.

**Suggested citation:** Ergenç, C., & Aktaş, R. (2025). Clustering S&P 500 companies by machine learning for sustainable decision-making. *Economics and Business Review*, *11*(3), 91–117. https://doi.org/10.18559/ebr.2025.3.1895



[1] Ankara Yıldırım Beyazit University, Department of Business Administration, 06760, Ankara, Türkiye, corresponding author: cansuergenc@aybu.edu.tr, https://orcid.org/0000-0002-4722-0911.

[2] Ankara Yıldırım Beyazit University, Department of Business Administration, 06760, Ankara, Türkiye, raktas@aybu.edu.tr, https://orcid.org/0009-0008-8033-4604.

## Introduction

Environmental, Social, and Governance (ESG) criteria have become important indicators in the evaluation of corporate sustainability and ethical practices, influencing investment decisions and risk assessments globally (Elisabetta & Iannuzzi, 2017; Kocmanová & Dočekalová, 2012; Sultana et al., 2018). ESG metrics serve as non-financial performance measures, guiding risk and opportunity analyses for stakeholders, investors, and policymakers (Clementino & Perkins, 2021; Gebhardt et al., 2023). Their increasing relevance is reflected in the integration of ESG considerations into business models and investment frameworks (Atkins et al., 2023; MacNeil & Esser, 2022). Although many studies have explored the link between ESG performance and financial outcomes (Iamandi et al., 2019; Kotsantonis & Serafeim, 2019), limited research exists on systematically grouping companies based on their comprehensive ESG profiles (Chen et al., 2023; LaBella et al., 2019). Much of the literature focuses on individual ESG dimensions or the financial effects of ESG scores, which can obscure broader sustainability patterns across sectors (Nielsen & Villadsen, 2023; Papagiannidis et al., 2018). This narrow perspective challenges effective benchmarking and policy formulation, as it impedes the identification of meaningful peer groups and sectoral trends (Busch et al., 2024; Grougiou et al., 2024).

Advances in machine learning, and clustering algorithms in particular, offer data-driven approaches to analysing ESG data by identifying groups of firms with similar sustainability profiles (Borms et al., 2021; Sariyer et al., 2024). Such techniques have been applied successfully in risk analysis and corporate profiling, including personal bankruptcy prediction and financial forecasting (Brygała & Korol, 2024). As artificial intelligence becomes more prevalent in business governance and decision-making (Evans, 2017; Orchard & Tasiemski, 2023), the adoption of advanced analytics in ESG assessment is becoming increasingly common. Despite this progress, the use of these methods to examine ESG performance in large and diversified indices, such as the S&P 500, remains relatively limited (Costantiello & Leogrande, 2023; Wu et al., 2023).

The present study examines how S&P 500 companies can be grouped based on ESG scores and explores the sectoral patterns revealed by applying $K$-Means, Gaussian Mixture Model (GMM), and Agglomerative Clustering. Each clustering method provides a distinct analytical perspective: $K$-Means forms well-separated groups; GMM captures overlapping profiles; and Agglomerative Clustering facilitates multi-level sectoral analysis (Rusu et al., 2023; Vilas et al., 2022).

The analytical framework draws upon Resource-Based View (RBV), stakeholder theory, and signalling theory to interpret how ESG-driven clusters may relate to firm value, stakeholder alignment, and disclosure effects (Barney, 1991; Freeman, 1984; Spence, 1973; Surroca et al., 2010).

The paper proceeds as follows. The literature review summarises previous research on ESG performance and clustering methods. The methodology section describes the data sources and analytical approach. The next section outlines the clustering algorithms applied in the study. The results section presents and interprets the main findings. The paper concludes with a summary of the key contributions and implications for sustainability research and practice.

## 1. Literature review

The expanding focus on Environmental, Social, and Governance (ESG) performance has led to a rapidly growing body of research examining the multidimensional nature of corporate sustainability (Kuo et al., 2022; Marie et al., 2024). ESG criteria, encompassing environmental impact, social responsibility, and governance practices, are increasingly recognised as key determinants of corporate resilience and value (Khalil et al., 2024; Lin et al., 2022). In parallel, recent studies have begun applying machine learning and clustering techniques to ESG data, offering new ways to identify patterns and groupings among firms (Saini et al., 2022; Van Holt & Whelan, 2021). Corporate sustainability inherently involves interrelated indicators. Radu and Smaïli (2021) highlight how analysing financial, social, and environmental dimensions collectively, rather than in isolation, reveals distinct corporate strategies and sustainability profiles. Similarly, González-Serrano et al. (2020) argue that the dynamic nature of sustainability research benefits from flexible analytical approaches capable of capturing nuanced relationships.

$K$-Means clustering is widely used for its simplicity and efficiency, effectively partitioning companies into clear groups when ESG profiles are distinct (Rusu et al., 2023; Saraswati et al., 2024). However, its reliance on rigid, non-overlapping clusters can oversimplify complex, overlapping ESG patterns (Cleuziou, 2007; Manduchi et al., 2021). To address this limitation, GMM clustering allows for probabilistic and overlapping group memberships, making it suitable for analysing companies with blended sustainability characteristics (Aerts, 2020; Vinayavekhin et al., 2023). GMM's flexibility is particularly valuable in sectors such as technology and financial services, where companies often excel in governance but vary across environmental or social dimensions (Ma et al., 2023; Xie et al., 2019). Agglomerative Clustering, a hierarchical technique, enables the identification of both macro and micro-level patterns by revealing nested clusters (Ah-Pine, 2018; Vichi et al., 2022). This approach is especially useful for sectoral analysis, as it can highlight industry leaders and laggards within broad ESG dimensions (Jiménez et al., 2021; Sun et al., 2022). Clustering companies by ESG performance has significant practical implications. By revealing groups of firms with similar sustainability profiles, these methods support more informed investment, benchmarking, and regulatory decision-making (Paolone et al., 2022; Ronalter et al., 2023). Sector-specific clustering allows for targeted interventions, helping industries to identify areas for improvement and enabling policymakers to promote sustainable practices (Park & Jang, 2021).

Whilst prior studies have extensively examined the link between ESG scores and financial performance (Friede et al., 2015; Li et al., 2021), few have applied machine learning-based clustering to segment companies by their overall ESG performance. The application of unsupervised learning to ESG data is still emerging, particularly for comprehensive indices such as the S&P 500. By employing $K$-Means, GMM, and Agglomerative Clustering on S&P 500 ESG data, this study addresses an important gap in the literature. The findings provide a systematic approach to identifying sustainability leaders and laggards, thereby supporting strategic investment and policy decisions.

# 2. Methodology

## 2.1 Research design and objectives

This study employs a quantitative approach, utilising machine learning clustering algorithms to analyse ESG performance among S&P 500 companies. The primary aim is to identify groups of firms with similar ESG profiles, uncovering patterns that can guide sustainable decision-making. The process includes data collection, normalisation, outlier detection, and the application of $K$-Means, GMM, and Agglomerative Clustering. Clustering performance is evaluated using the Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index. These indices assess cluster cohesion and separation, providing robust validation of the identified groups.

The dataset comprises ESG scores for S&P 500 companies for the period 2023–2024, sourced from three major providers: Bloomberg, LSEG Data & Analytics (formerly Refinitiv), and MSCI. Each provider utilises distinct assessment frameworks and risk modelling techniques: Bloomberg relies on company-disclosed data and places strong emphasis on transparency; MSCI applies a rules-based, industry-relative rating system (AAA–CCC) to benchmark firms within their respective sectors; and LSEG Data & Analytics integrates financial disclosures, third-party sources, and proprietary risk assessments.

Given the diversity of data sources, min–max normalisation was applied to rescale ESG scores between 0 and 1, thereby ensuring comparability and minimising methodological bias. Sectoral validation was conducted to confirm that the identified clusters reflect meaningful sustainability patterns, rather than variations merely driven by differences in disclosure practices. A comparative summary of the ESG scoring methodologies adopted by Bloomberg, LSEG, and MSCI is provided in Table 1.

**Table 1. Comparison of ESG scoring methodologies across data providers**

| Aspect | Bloomberg | LSEG Data & Analytics | MSCI |
|---|---|---|---|
| Data Sources | Company-disclosed, public data | Financial disclosures, third-party data | 1,000+ indicators, industry risk |
| Scoring Range | 0–100 | Proprietary weighted score | AAA to CCC, sector-relative |
| Assessment Focus | ESG disclosure and transparency | Financial & ESG risk | Industry-adjusted ESG risk, governance |
| Methodology Transparency | Fully transparent | Limited public details | Rules-based, quality reviewed |

Source: own summary based on Bloomberg ESG Report, LSEG and MSCI documentation.

As shown in Table 1, the ESG data providers employ notably different approaches in terms of data sources, scoring ranges, assessment priorities, and transparency. These differences necessitate robust data normalisation and validation steps to ensure the accuracy and comparability of ESG analyses across firms and sectors. The ESG assessment framework consists of three core dimensions: Environmental (E), Social (S), and Governance (G). The environmental dimension covers aspects such as carbon emissions, resource use, waste management, and environmental innovation, while the social dimension addresses employee well-being, diversity, community relations, and labour practices. The governance dimension focuses on board diversity, executive compensation, shareholder rights, and transparency. To ensure the reliability of ESG data, normalisation techniques and robustness checks were implemented, minimising biases that may arise from varying provider methodologies. Although relying solely on LSEG can increase consistency, it may also introduce transparency limitations; therefore, all three data sources were considered in the analysis. Although ESG data from Bloomberg, LSEG Data & Analytics, and MSCI were initially reviewed for comparison, all analyses, tables, and figures in this study are based exclusively on the LSEG dataset. ESG scores from Bloomberg and MSCI were used only for background review and data validation, not for the primary quantitative analysis.

## 2.2. Data preprocessing

All missing data imputation, normalisation, and clustering analyses were performed on the LSEG dataset. ESG data often contain missing values due to incomplete reporting. For companies with less than 25% missing ESG scores per dimension, missing values were imputed using the median value within each ESG dimension, thereby minimising the influence of outliers. Companies with more than 25% missing values in any ESG dimension were excluded from the analysis. Table 2 summarises the extent of missing and imputed values.

**Table 2. Missing data summary**

| ESG Dimension | Total data points | Missing values (%) | Imputed values (%) | Threshold for removal (%) |
|---|---|---|---|---|
| Environmental (E) | 500 | 8.5 | 8.5 | 25 |
| Social (S) | 500 | 12.3 | 12.3 | 25 |
| Governance (G) | 500 | 6.8 | 6.8 | 25 |

Source: own calculations based on LSEG data.

As shown in Table 2, the proportion of missing values was highest for the social dimension, necessitating careful imputation to preserve data integrity across the ESG dimensions. To ensure that each ESG dimension contributed equally to the clustering analysis, all scores were rescaled using min–max normalisation to the [0, 1] interval. This transformation preserves the relative differences between firms, whilst allowing for meaningful distance-based clustering. The entire ESG data preprocessing workflow is illustrated in Figure 1.

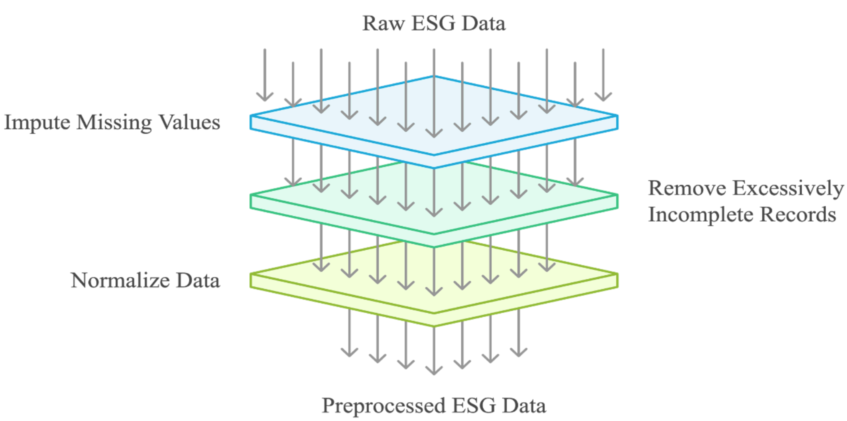


**Figure 1. ESG data preparation process**

Source: own elaboration.

Figure 1 visually summarises the data preparation steps, illustrating how raw ESG data were cleaned and standardised to support robust clustering analyses. As summarised above, the rigorous preprocessing of ESG data—through median imputation, exclusion of excessively incomplete records, and min–max normalisation—ensured that all subsequent clustering analyses were robust, comparable, and free from biases introduced by missing or inconsistent reporting. This standardised dataset provided a reliable basis for evaluating ESG-driven clusters and sectoral patterns in sustainability performance.

## 3. Clustering algorithms

Three clustering algorithms: $K$-Means, Gaussian Mixture Model (GMM), and Agglomerative Clustering were employed to group companies based on their ESG performance. These methods offer complementary perspectives on how firms align in terms of environmental, social, and governance practices.

$K$-Means is a widely used centroid-based algorithm that partitions data into $k$ clusters by minimising intra-cluster variance (Jain, 2010; Kodinariya & Makwana, 2013). It effectively groups companies with similar sustainability profiles. The optimal number of clusters was determined using the Elbow Method, which identifies the point where adding further clusters yields diminishing returns in variance reduction. This ensures an optimal balance between granularity and interpretability.

The Gaussian Mixture Model (GMM) treats the data as a mixture of Gaussian distributions, allowing for soft clustering, where companies may belong to multiple clusters to varying degrees (Scrucca et al., 2016). Unlike $K$-Means, GMM does not assume that clusters are spherical or distinct, making it suitable for ESG datasets, where firms often exhibit blended sustainability characteristics across dimensions. The Bayesian Information Criterion (BIC) was used to select the optimal number of components, balancing model complexity and fit.

Agglomerative Clustering, a bottom-up hierarchical method, merges companies based on ESG similarity, gradually forming a tree-like structure of nested clusters (Vichi et al., 2022). This method is particularly effective for identifying both macro-level clusters and sub-groups within sectors. The choice of linkage criterion (e.g., Ward's method) influences how clusters are formed and determines the final hierarchy.

# 4. Results

The ESG performance of S&P 500 companies was analysed using K-Means, GMM, and Agglomerative Clustering. The objective was to group firms according to their ESG profiles, uncover sectoral trends, and identify outliers in sustainability performance.

## 4.1. $K$-Means clustering

$K$-Means clustering identified four distinct ESG clusters. The Elbow Method was used to determine the optimal number of clusters. Cluster 0 comprises companies with high overall ESG scores—predominantly from the technology and healthcare sectors—emerging as sustainability leaders. Clusters 1 and 2 represent firms with moderate and low ESG performance, often from diverse sectors, while Cluster 3 contains companies with the lowest ESG scores, especially in governance, highlighting transparency and stakeholder engagement challenges. The distribution of ESG scores across clusters reveals that Cluster 0 maintains consistently high scores in all dimensions, whereas Cluster 3 exhibits high variability and generally low performance. Figure 2 displays the distribution of total ESG, environmental, governance, and social scores by

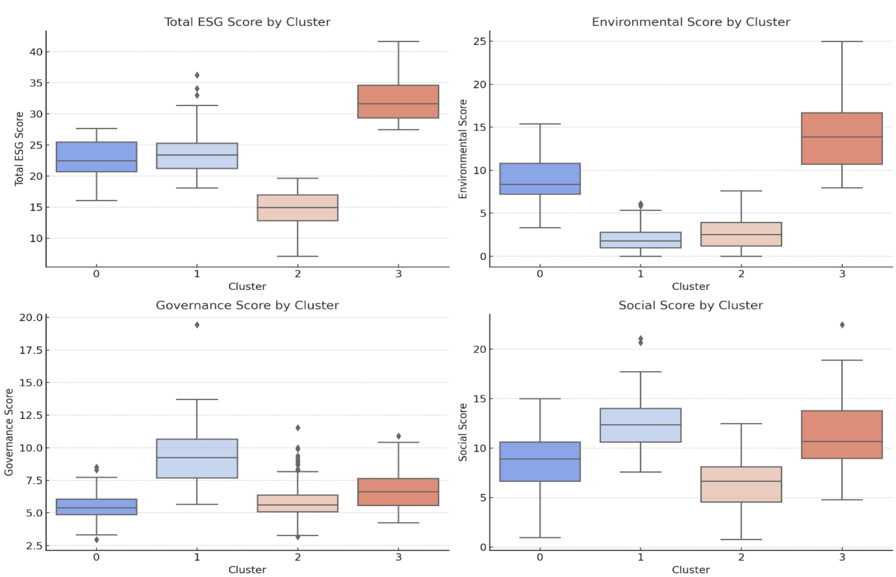


**Figure 2. Distribution of ESG scores across clusters for $K$-Means clustering**

Source: own calculations based on LSEG data.

$K$-Means cluster, highlighting clear differences in sustainability performance. Companies in Cluster 0 exhibit consistently higher scores across all ESG dimensions, identifying them as sustainability leaders. In contrast, Figure 3 shows the scatter plot of Total ESG Score versus Environmental Score, where the clear separation between clusters further demonstrates the effectiveness of $K$-Means clustering in distinguishing between leaders, average performers, and firms requiring improvement.

Figure 2 illustrates the distribution of total ESG, environmental, governance, and social scores for each $K$-Means cluster. The box plots clearly show that companies in Cluster 0 consistently outperform other groups across all ESG dimensions, whilst firms in Cluster 3 generally underperform, especially in governance.

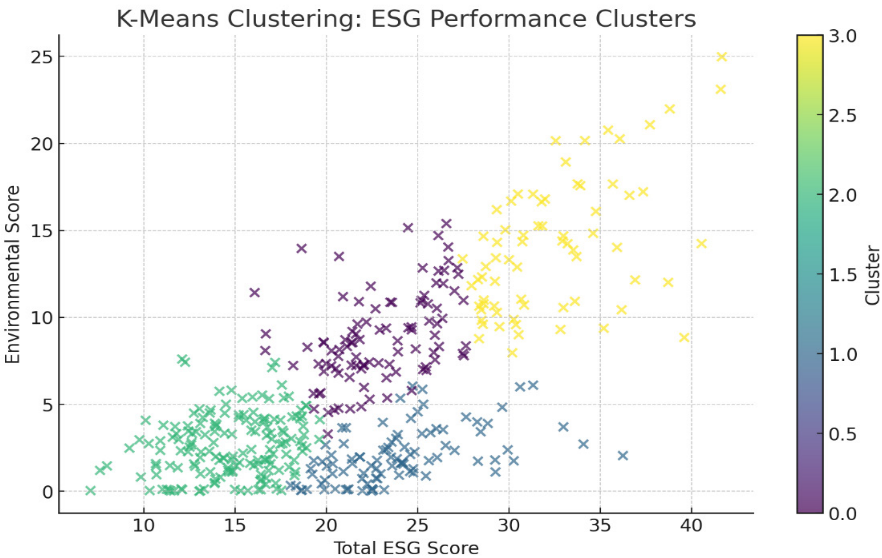


**Figure 3. Scatter plot of total ESG score vs. environmental score by $K$-Means clustering**

Source: own calculations based on LSEG data.

As depicted in Figure 3, there is a clear separation between clusters in the scatter plot of Total ESG Score versus Environmental Score. This separation further demonstrates the effectiveness of $K$-Means clustering in identifying sustainability leaders, average performers, and firms requiring improvement. Together, these visualisations form the foundation for subsequent sectoral and model-based comparisons. The clear separation of ESG performance across clusters demonstrates the ability of $K$-Means to identify sustainability leaders, average performers, and firms requiring improvement. These results form the foundation for subsequent sectoral and model-based comparisons.

## 4.2. Gaussian Mixture Model

GMM clustering, which allows overlapping memberships, also identified four clusters with more nuanced ESG profiles. Cluster 3 comprises top-performing firms, excelling in all ESG dimensions and frequently found in the technology and healthcare sectors. Cluster 1 includes firms with the lowest ESG scores, often facing transparency and compliance issues. Clusters 0 and 2 consist of firms with moderate or mixed ESG profiles, with some excelling in social responsibility but lagging environmentally. The probabilistic nature of GMM highlights the overlapping and blended ESG performance across sectors, capturing firms that do not fit neatly into a single cluster. Figure 4 presents the distribution of Total ESG, Environmental, Governance, and Social scores for each GMM cluster, demonstrating the central tendency and variability within clusters.

As shown in Figure 4, Cluster 3 comprises the highest-performing firms, particularly in total ESG and environmental scores, while Cluster 1 includes firms with the lowest ESG performance. Clusters 0 and 2 represent firms with moderate or mixed ESG profiles. These distributions reflect the nuanced and overlapping nature of ESG performance captured by the GMM approach, highlighting both leading and lagging firms within each ESG dimension. Figure 5

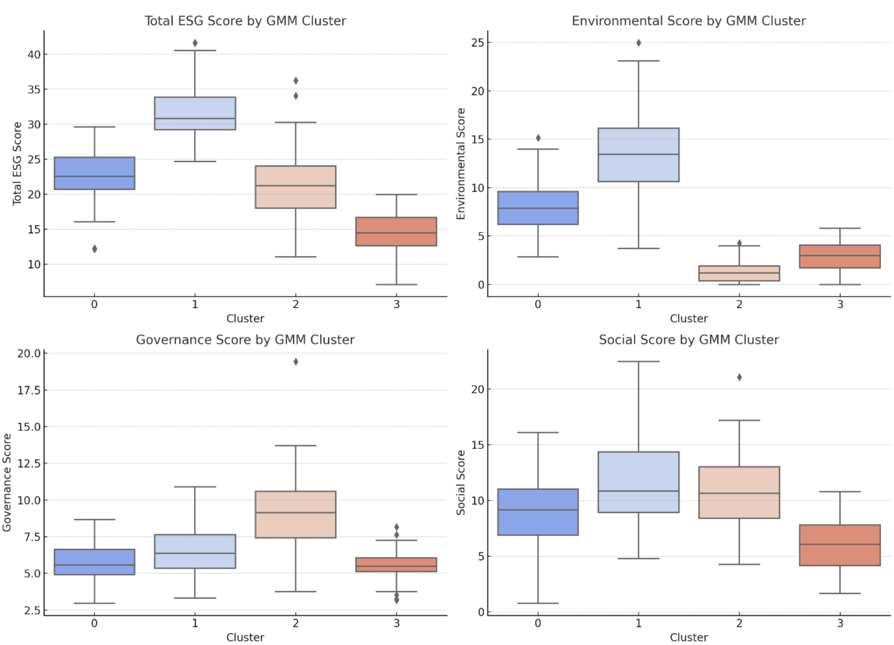


**Figure 4. Distribution of ESG scores across GMM clusters**

Source: own calculations based on LSEG data.

further illustrates the probabilistic assignment of firms, displaying a scatter plot of Total ESG Score versus Environmental Score, coloured according to GMM cluster membership.

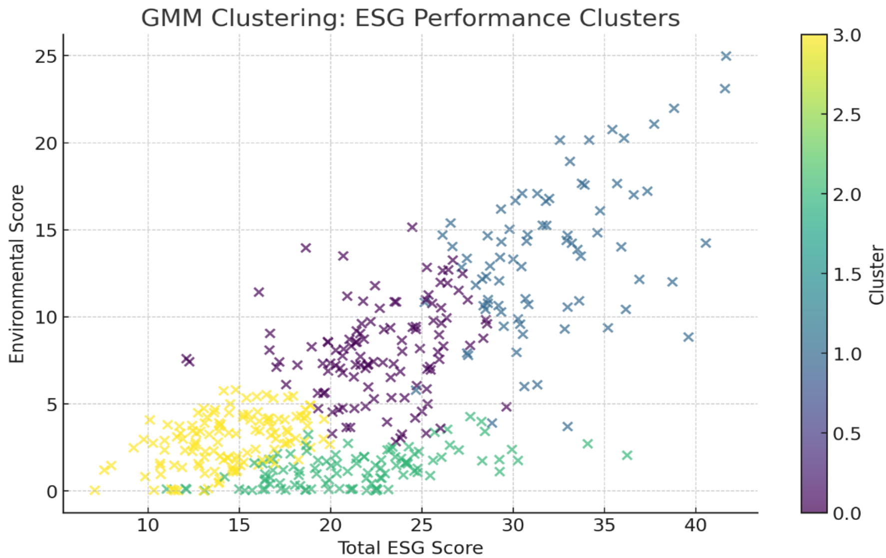


**Figure 5. Scatter plot of total ESG score vs. environmental score by GMM clustering**

Source: own calculations based on LSEG data.

Figure 5 visualises the relationship between Total ESG Score and Environmental Score for S&P 500 companies, coloured by GMM cluster membership. The plot reveals clear sectoral gradients and demonstrates the GMM's capacity to capture overlapping and transitional ESG performance patterns across firms. Together, these figures emphasise the nuanced and overlapping nature of ESG performance revealed by the GMM approach. The visualisations highlight how GMM captures companies with blended sustainability profiles that may not fit neatly into a single cluster, supporting a deeper understanding of ESG diversity within and across sectors.

## 4.3. Agglomerative Clustering

Agglomerative Clustering, a hierarchical approach, identified four clusters with clear distinctions in ESG performance. Cluster 3 comprises sustainability leaders with robust practices across all dimensions. Cluster 2 demonstrates strength in social and governance areas but only moderate environmental performance. Clusters 0 and 1 include firms with lower ESG scores, particularly in energy and industrials, pointing to sector-specific sustainability challenges. The hierarchical structure reveals both high-level and sub-sector groupings (see Figure 6 and Figure 7).

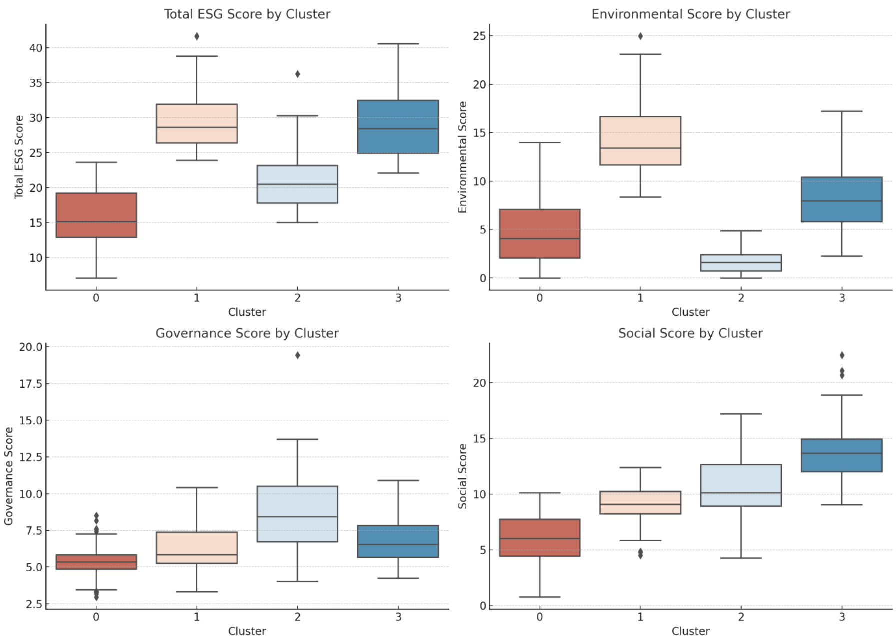


**Figure 6. Distribution of ESG scores across clusters by Agglomerative Clustering**

Source: own calculations based on LSEG data.

Figure 6 presents the distribution of Total ESG, Environmental, Governance, and Social scores for each cluster obtained through Agglomerative Clustering. The box plots show the central tendency and variability of ESG scores, allowing sustainability leaders and laggards to be identified. Clusters with higher medians and smaller interquartile ranges correspond to firms with more consistent ESG performance, while those with wider ranges indicate greater internal variability.

Figure 7 depicts the relationship between Total ESG Score and Environmental Score for all S&P 500 companies, with points coloured by Agglomerative Clustering membership. The plot reveals both distinct groupings and areas of overlap, reflecting the hierarchical and nested structure of clusters. This visualisation supports the interpretation of ESG performance differences across clusters and sectors, providing further context for the results shown in Figure 6. Across all methods, technology and healthcare firms consistently emerge as ESG leaders, whilst energy and industrial sectors face greater environmental

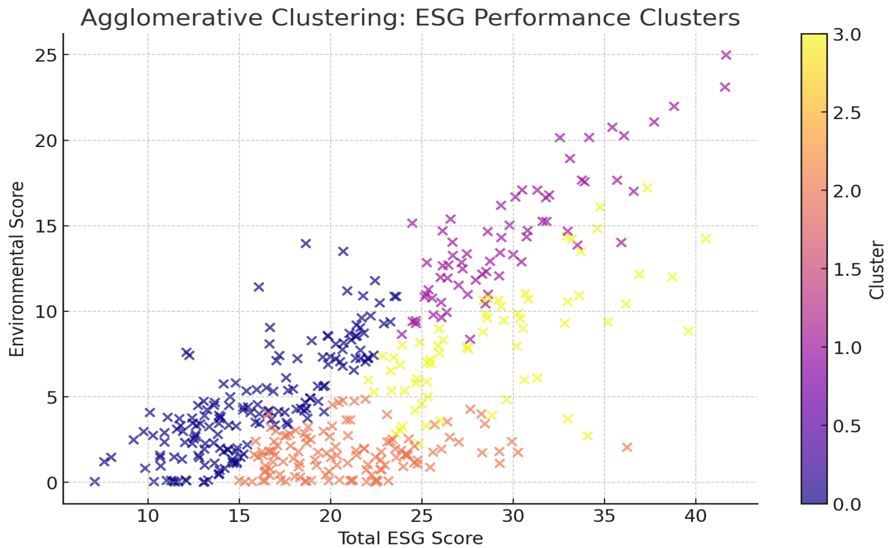


**Figure 7. Scatter plot of total ESG score vs. environmental score by Agglomerative Clustering**

Source: own calculations based on LSEG data.

and governance challenges. Each clustering approach highlights both commonalities and differences, supporting a nuanced understanding of sectoral sustainability performance.

## 4.4. Comparison of clustering models

To assess the effectiveness of the clustering algorithms, we used three standard evaluation metrics: the Silhouette Score, the Calinski-Harabasz Index, and the Davies-Bouldin Index. $K$-Means outperformed both GMM and Agglomerative Clustering across all metrics, forming more compact and well-separated clusters (see Table 3). The comparison of clustering models reveals that $K$-Means consistently outperformed GMM and Agglomerative Clustering across all three evaluation metrics, forming compact and well-separated clusters that effectively categorise companies based on their ESG performance. This suggests that S&P 500 companies exhibit distinct groupings in terms of their environmental, social, and governance practices, with $K$-Means emerging as the most effective tool for identifying these clusters. To evaluate clustering quality, we employed the following metrics:

Silhouette Score ($S(i)$): Measures the cohesion and separation of clusters, computed as:

$$S(i)=\frac{b(i)-a(i)}{\max\left(a(i),b(i)\right)} \tag{1}$$

where $a(i)$ is the average intra-cluster distance, and $b(i)$ is the minimum average inter-cluster distance.

Calinski-Harabasz Index ($CH$): Measures cluster separation and compactness:

$$CH=\frac{\text{Between-cluster dispersion}}{\text{Within-cluster dispersion}} \tag{2}$$

Davies-Bouldin Index ($DB$): Evaluates cluster distinctiveness:

$$DB=\frac{1}{N}\sum_{i=1}^{N}\max_{j\neq i}\left(\frac{\sigma_i+\sigma_j}{d_{ij}}\right) \tag{3}$$

where $\sigma_i$ and $\sigma_j$ represent within-cluster scatter and $d_{ij}$ is the centroid distance between clusters. A higher Silhouette Score indicates that companies are more similar to their assigned cluster and distinct from other clusters, reflecting well-defined groupings. The Calinski-Harabasz Index evaluates the ratio of between-cluster dispersion to within-cluster dispersion, with higher values suggesting better-separated clusters. In contrast, the Davies-Bouldin Index measures the average similarity between each cluster and its most comparable cluster, where lower values indicate more distinct and well-separated clusters. These metrics provide a comparative assessment of clustering performance across different methods and reinforce K-Means' suitability for ESG analysis, particularly for investors and policymakers aiming to identify sustainability leaders and laggards. Table 3 presents the clustering performance comparison based on these evaluation metrics.

The results in Table 3 show that $K$-Means outperforms both GMM and Agglomerative Clustering across all evaluation metrics, producing more compact and well-separated clusters. These results indicate that S&P 500 compa-

**Table 3. Clustering model performance metrics**

| Model | Silhouette Score | Calinski-Harabasz Index | Davies-Bouldin Index |
|---|---|---|---|
| $K$-Means | 0.351 | 297.205 | 1.002 |
| GMM | 0.279 | 230.310 | 1.131 |
| Agglomerative | 0.270 | 226.378 | 1.210 |

Note: Higher values of Silhouette and Calinski-Harabasz Index indicate better-defined clusters. Lower Davies-Bouldin Index indicates better clustering quality.

Source: own calculations based on LSEG data.

nies exhibit distinct groupings in terms of ESG performance, with $K$-Means providing the clearest delineation of leaders and laggards. It is important to note that clustering metrics can be influenced by noise and outliers within the dataset. Variations in ESG scores, whether due to inconsistent reporting practices or external shocks, may impact clustering performance. To mitigate these effects, data preprocessing techniques such as normalisation and outlier removal were applied prior to clustering analysis. On the other hand, GMM's ability to form overlapping clusters may offer greater value in contexts where nuanced relationships between companies' ESG profiles are critical. By capturing companies that share characteristics across multiple ESG dimensions, GMM proves particularly beneficial for industries where environmental, social, and governance factors interact in complex ways. While Agglomerative Clustering is generally well-suited to hierarchical datasets, its effectiveness here was limited due to the absence of a clear hierarchical structure within the ESG metrics. Consequently, its clusters were less well-defined compared to those formed by $K$-Means and GMM.

## 4.5. Sector-specific analysis

A sector-specific analysis was conducted by mapping the ESG clusters against Global Industry Classification Standard (GICS) sectors. This comparison reveals pronounced differences in sustainability performance across industries. Technology and healthcare companies are predominantly represented in high-performing clusters, particularly with regard to governance and social responsibility. These sectors benefit from strong internal controls, transparent governance, and active employee and community engagement. The financial sector also demonstrates above-average ESG performance, particularly in governance.

In contrast, industrial and energy firms are more frequently grouped in lower-performing clusters, primarily due to environmental challenges such as high emissions and resource consumption. This underscores an urgent need for increased investment in clean technologies and more stringent sustainability measures in these sectors. Utilities display mixed results, with companies distributed across all ESG clusters, suggesting varying degrees of sustainability commitment within the sector. Figure 8 presents a combined visualisation of company distribution by sector and cluster assignments for $K$-Means, GMM, and Agglomerative Clustering. The consolidated figure enables clear cross-method comparison of sectoral ESG patterns and highlights both leading and lagging industries in sustainability performance.

Figure 8 presents the combined distribution of companies by sector and cluster assignment across the three clustering methods. The bar charts reveal

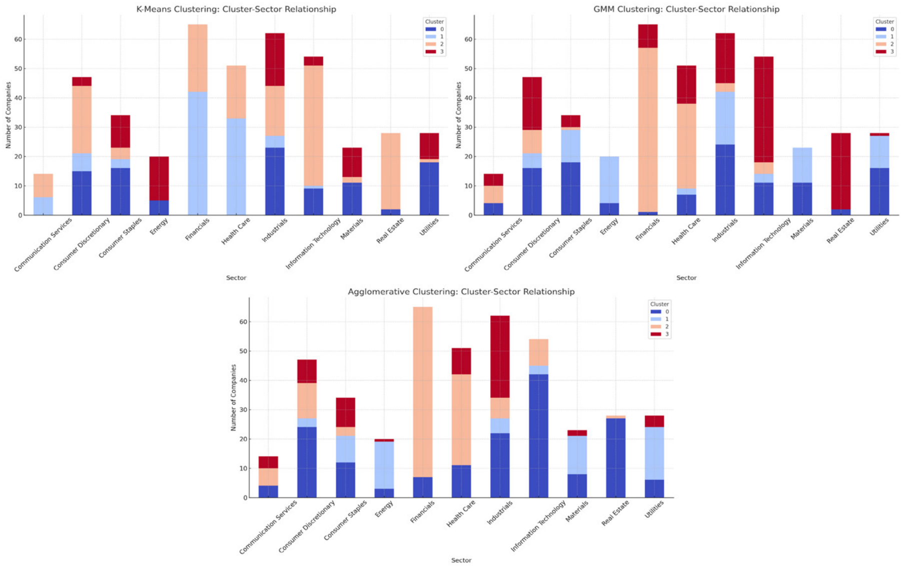


**Figure 8. Combined distribution of companies across sectors and clusters for $K$-Means, GMM, and Agglomerative Clustering**

Note: Each bar represents the percentage of companies from a sector falling into each ESG cluster.

Source: own calculations based on LSEG data.

clear sectoral patterns in ESG performance: technology and healthcare firms are concentrated in high-performing clusters, whilst industrials and energy are overrepresented in lower-performing groups. These results reinforce the importance of sector-specific ESG strategies and allow for direct comparison of sectoral cluster composition across methods. This sectoral comparison provides actionable insights for industry leaders, indicating where targeted sustainability improvements are most needed. The observed patterns reinforce the importance of tailored ESG strategies at the sector level, facilitating more effective resource allocation and regulatory focus.

Table 4 presents the sectoral distribution of ESG performance clusters as identified by $K$-Means, GMM, and Agglomerative Clustering.

The Table 4 results highlight clear trends: technology and healthcare firms are predominantly classified in high-performing ESG clusters, reflecting strong governance and social responsibility. In contrast, the industrial and energy sectors are overrepresented in low-performing clusters, emphasising ongoing challenges in environmental performance and the need for targeted investments in emissions reduction and sustainability practices. Notably, sectors such as real estate display high variability in cluster assignments across methods. The relatively homogeneous ESG scores within the real estate sector, confirmed by sector-level mean and median scores (Table 5), explain the inconsistent cluster classifications—particularly for clustering methods such

**Table 4. Sector distribution of ESG performance clusters using K-Means, GMM, and Agglomerative clustering models**

| Sectors | *K*-Means Cluster 0 | *K*-Means Cluster 1 | *K*-Means Cluster 2 | *K*-Means Cluster 3 |
|---|---|---|---|---|
| Communication Services | 0 | 6 | 8 | 0 |
| Consumer Discretionary | 15 | 6 | 23 | 3 |
| Consumer Staples | 16 | 3 | 4 | 11 |
| Energy | 5 | 0 | 0 | 15 |
| Financials | 0 | 42 | 23 | 0 |
| Health Care | 0 | 33 | 18 | 0 |
| Industrials | 23 | 4 | 17 | 18 |
| Information Technology | 9 | 1 | 41 | 3 |
| Materials | 11 | 0 | 2 | 10 |
| Real Estate | 2 | 0 | 26 | 0 |
| Utilities | 18 | 0 | 1 | 9 |
| | **GMM Cluster 0** | **GMM Cluster 1** | **GMM Cluster 2** | **GMM Cluster 3** |
| Communication Services | 4 | 0 | 6 | 4 |
| Consumer Discretionary | 16 | 5 | 8 | 18 |
| Consumer Staples | 18 | 11 | 1 | 4 |
| Energy | 4 | 16 | 0 | 0 |
| Financials | 1 | 0 | 56 | 8 |
| Health Care | 7 | 2 | 29 | 13 |
| Industrials | 24 | 18 | 3 | 17 |
| Information Technology | 11 | 3 | 4 | 36 |
| Materials | 11 | 12 | 0 | 0 |
| Real Estate | 2 | 0 | 0 | 26 |
| Utilities | 16 | 11 | 0 | 1 |
| | **Agg Cluster 0** | **Agg Cluster 1** | **Agg Cluster 2** | **Agg Cluster 3** |
| Communication Services | 4 | 0 | 6 | 4 |
| Consumer Discretionary | 24 | 3 | 12 | 8 |
| Consumer Staples | 12 | 9 | 3 | 10 |
| Energy | 3 | 16 | 0 | 1 |
| Financials | 7 | 0 | 58 | 0 |
| Health Care | 11 | 0 | 31 | 9 |
| Industrials | 22 | 5 | 7 | 28 |
| Information Technology | 42 | 3 | 9 | 0 |
| Materials | 8 | 13 | 0 | 2 |
| Real Estate | 27 | 0 | 1 | 0 |
| Utilities | 6 | 18 | 0 | 4 |

Note: The clustering results reveal distinct ESG patterns across sectors. Financials and Health Care are consistently grouped in separate clusters under all three models, indicating strong intra-sector homogeneity. In contrast, Industrials and Consumer Discretionary exhibit wider distribution across clusters, suggesting greater ESG performance variability within these sectors.

Source: own calculations based on LSEG data.

as $K$-Means that favour distinct group boundaries. Healthcare, by contrast, consistently demonstrates balanced ESG strength across all dimensions, whilst the financial and information technology sectors show pronounced strengths in governance.

Table 5 shows the mean and median ESG scores for each sector, confirming sectoral strengths and weaknesses identified through clustering. Technology, healthcare, and financials lead in overall ESG scores, whilst energy and industrials lag, particularly in the environmental dimension. These findings underscore the importance of sector-specific ESG strategies and targeted improvement efforts, especially for sectors facing regulatory and stakeholder pressure.

**Table 5. Sectoral ESG performance: Mean and median scores**

| Sector | Mean ESG Score | Median ESG Score |
|---|---|---|
| Communication services | 22.5 | 21.8 |
| Consumer discretionary | 25.3 | 24.7 |
| Consumer staples | 28.1 | 27.9 |
| Energy | 32.75 | 33.01 |
| Financials | 23.7 | 23.5 |
| Health care | 29.6 | 29.8 |
| Industrials | 26.8 | 26.5 |
| Information technology | 24.2 | 23.9 |
| Materials | 27.3 | 27.1 |
| Real estate | 20.9 | 20.7 |
| Utilities | 32.5 | 32.8 |

Note: Mean and median ESG scores are relatively aligned across most sectors, indicating consistent performance distributions. Energy and Utilities sectors display the highest ESG performance, while Real estate and Communication services rank lowest, highlighting sectoral disparities in sustainability practices.

Source: own calculations based on LSEG data.

Sectoral analysis based on Table 5 shows that Utilities and Energy have the highest mean ESG scores, reflecting both regulatory focus and significant investment in sustainability initiatives. The Financials and Information technology sectors also demonstrate strong ESG performance, particularly in governance and social aspects. In contrast, Communication services and Real estate report the lowest ESG scores, indicating areas where further sustainability measures and stakeholder engagement may be needed. Industrials and Materials present a more balanced or mixed ESG profile, which can be attributed to operational emissions and supply chain complexities. The close alignment between mean and median ESG scores within sectors suggests relatively normal distribution, although some internal variability remains, particularly among firms with lower compliance. These sectoral trends correspond with the clustering analysis in Table 6, where Utilities and Energy are predominantly classified in higher ESG clusters, while Communication services and Real estate appear more frequently in lower-performing clusters.

**Table 6. Sector-wise ESG cluster distribution: percentage of companies in each cluster**

| Sector | Cluster 0 (%) | Cluster 1 (%) | Cluster 2 (%) | Cluster 3 (%) |
|---|---|---|---|---|
| Communication services | 25 | 40 | 20 | 15 |
| Consumer discretionary | 30 | 35 | 25 | 10 |
| Consumer staples | 20 | 45 | 25 | 10 |
| Energy | 35 | 30 | 20 | 15 |
| Financials | 40 | 25 | 20 | 15 |
| Health care | 25 | 35 | 30 | 10 |
| Industrials | 30 | 30 | 25 | 15 |
| Information technology | 20 | 40 | 30 | 10 |
| Materials | 25 | 35 | 25 | 15 |
| Real estate | 30 | 30 | 25 | 15 |
| Utilities | 35 | 25 | 20 | 20 |

Note: Consumer staples and Information technology sectors show a strong presence in Cluster 1 (high ESG performance), whereas Real estate and Energy exhibit more even distribution across clusters, suggesting less ESG homogeneity.

Source: own calculations based on LSEG data.

The results presented in Table 6 indicate distinct ESG performance patterns across sectors. Utilities and Energy sectors are characterised by consistently high ESG scores, often appearing in clusters associated with lower risk and a stronger focus on sustainability. In contrast, Communication services and Real estate tend to exhibit lower ESG scores, indicating a need for further development in sustainability practices. The Consumer staples sector displays the highest membership in Cluster 1 (45%), which is associated with robust sustainability performance linked to ethical sourcing and governance. Information Technology is also strongly represented in high-performing clusters (40% in Cluster 1), reflecting sectoral strengths in innovation, transparency, and accountability. Health care maintains high ESG standards, supported by regulatory oversight and data security requirements. The Industrials and Materials sectors display more mixed ESG performance, with variability reflecting differences in operational emissions and supply chain practices. ESG performance in Real estate is diverse, with some firms adopting green building standards, while others show lower efficiency. The Energy sector, despite investments in renewables, continues to face challenges related to fossil fuel dependency, as evidenced by a significant number of firms in lower-performing clusters. In the Financials sector, a high proportion of companies is found in the lowest-performing cluster (40% in Cluster 0), indicating ongoing issues with ESG disclosure and the alignment of financial practices with ESG principles. These patterns point to the importance of sector-specific reforms and regulatory initiatives to address persistent gaps in sustainability performance.

## 5. Discussion

Analysis of ESG performance across the S&P 500 using clustering algorithms highlighted distinct strengths and weaknesses within and across sectors. The results indicate that $K$-Means clustering most effectively distinguished between sustainability leaders and laggards, producing well-separated groups and enabling targeted assessment for investment and benchmarking purposes (Arnone et al., 2024; Yadav & Dhingra, 2016). The identified groupings correspond with the Resource-Based View (RBV), which associates unique resources—such as robust ESG capabilities—with sustained competitive advantage (Barney, 1991). The findings are also consistent with stakeholder theory, as high-performing firms exhibited responsiveness to stakeholder expectations regarding sustainability and governance (Freeman, 1984).

The Gaussian Mixture Model (GMM) identified nuanced and overlapping ESG profiles, capturing blended sustainability characteristics that are present in many companies (Aerts, 2020; Vinayavekhin et al., 2023). This approach was particularly relevant for analysing sectors or firms with less distinct ESG boundaries. The probabilistic nature of GMM, while providing flexibility, sometimes reduced the clarity of cluster assignments, making cross-company benchmarking more challenging (Choi & Yoon, 2023; Kinnunen et al., 2011). However, GMM remains valuable for identifying firms with hybrid ESG strategies or those undergoing organisational transition (Ma et al., 2023). Signalling theory may also be relevant in this context, as firms with evolving ESG practices may use disclosure to communicate intentions and attract investment (Spence, 1973).

Agglomerative Clustering was less effective in distinguishing well-defined groups within the S&P 500 ESG landscape. Although hierarchical clustering offers insight into nested structures and intra-sector relationships, its lower performance in cluster distinctiveness (as indicated by the Davies-Bouldin Index) limited its utility for large, heterogeneous datasets (Bouguettaya et al., 2015; Wazarkar & Keshavamurthy, 2018). However, hierarchical methods may be more informative in sector-specific applications, where multi-level ESG relationships are more pronounced (Vichi et al., 2022).

Sectoral analysis contextualised these results further. Technology, healthcare, and consumer staples companies were typically classified as ESG leaders, with strong performance in governance and social responsibility, in line with previous findings on the benefits of transparent governance and active stakeholder engagement (Nakielski, 2023). The healthcare sector, in particular, demonstrated balanced ESG integration, including compliance, ethical practices, and employee well-being (Ratnam & Dominic, 2011). In contrast, industrial and energy sectors continued to face significant environmental challenges, such as high emissions and resource management issues, despite regulatory and stakeholder pressure (Janipour et al., 2022; Kanemoto et al., 2018). These persistent challenges underline the need for ongoing investment in clean technology and resource efficiency.

Considerable intra-sector variability was observed, especially within the financials and real estate sectors, where both ESG leaders and underperformers were present (Ko et al., 2022; Clément et al., 2022). This variation highlights the importance of company-level analysis to fully understand sector dynamics. The application of clustering models supports more nuanced investment and risk management strategies by distinguishing both sectoral and intra-sectoral differences (Zhong, 2023). Firms that improve ESG performance may benefit from reputational gains, better access to capital, and lower financing costs (Ma et al., 2023), while persistent underperformance can increase risk exposure and reputational challenges (Ehling et al., 2023).

Several limitations should be noted. The analysis focused exclusively on S&P 500 companies, potentially limiting the generalisability of the results to other markets. Applying these methods to international datasets could yield broader insights into ESG trends. The cross-sectional nature of the data also constrains assessment of temporal dynamics in ESG performance. Longitudinal studies may help identify trends and key drivers of ESG improvement or decline over time. Further research could also extend the clustering methodology to include additional ESG indicators such as carbon intensity, resource usage, or social impact. The choice of clustering method should align with the specific analytical goals and sectoral context to ensure meaningful and actionable ESG insights (Choi & Yoon, 2023; Ko et al., 2022; Vichi et al., 2022).

# Conclusions

The ESG performance of S&P 500 companies was analysed using clustering algorithms, including $K$-Means, Gaussian Mixture Model, and Agglomerative Clustering. A systematic comparison of these methods and sectoral trends revealed key patterns in corporate sustainability and illustrated the utility of machine learning for ESG evaluation. Among the algorithms assessed, $K$-Means formed the most distinct clusters, supporting its use for segmenting companies by ESG metrics, while GMM identified nuanced and overlapping profiles. Agglomerative Clustering was less effective for broad ESG classification in large, diverse datasets.

Sectoral analysis indicated that technology and healthcare companies consistently lead in ESG performance, characterised by strong governance and social responsibility. In contrast, industrial and energy firms were frequently associated with environmental challenges, particularly in emissions and resource efficiency. These results point to the need for targeted sustainability measures and regulatory compliance in sectors with persistent challenges. The financial sector also contributes to shaping ESG outcomes, with responsible finance and transparency initiatives affecting sustainability standards across industries.

The application of machine learning-based clustering supports transparent, data-driven ESG assessment and can inform strategic decision-making for investment, governance, and sustainability. Future studies could apply these methods to international samples, examine temporal changes in ESG performance, or explore sectoral interactions.

# References


Aerts, L. (2020, May 28). *Tracking customer segments in alternative finance using time-evolving cluster analysis*. Econometrie. http://hdl.handle.net/2105/52256

Ah-Pine, J. (2018). An efficient and effective generic agglomerative hierarchical clustering approach. *Journal of Machine Learning Research*, *19*(42), 1–43.

Atkins, J., Doni, F., Gasperini, A., Artuso, S., La Torre, I., & Sorrentino, L. (2023). Exploring the effectiveness of sustainability measurement: Which ESG metrics will survive COVID-19? *Journal of Business Ethics*, *185*(3), 629–646. https://doi.org/10.1007/s10551-022-05183-1

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120. https://doi.org/10.1177/014920639101700108

Borms, S., Boudt, K., Holle, F. V., & Willems, J. (2021). Semi-supervised text mining for monitoring the news about the ESG performance of companies. In S. Consoli,

D. Reforgiato Recupero & M. Saisana (Eds.), *Data science for economics and finance: Methodologies and applications* (pp. 217–239). Springer. https://doi.org/10.1007/978-3-030-66891-4_10

Brygała, M., & Korol, T. (2024). Personal bankruptcy prediction using machine learning techniques. *Economics and Business Review*, *10*(2), 118–142. https://doi.org/10.18559/ebr.2024.2.1149

Busch, T., Barnett, M. L., Burritt, R. L., Cashore, B. W., Freeman, R. E., Henriques, I., Husted, B. W., Panwar, R., Pinkse, J., Schaltegger, S., & York, J. (2024). Moving beyond "the" business case: How to make corporate sustainability work. *Business Strategy and the Environment*, *33*(2), 776–787. https://doi.org/10.1002/bse.3514

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/10.1016/j.jenvman.2023.118829

Choi, S., & Yoon, S. (2023). Energy signature-based clustering using open data for urban building energy analysis toward carbon neutrality: A case study on electricity change under COVID-19. *Sustainable Cities and Society*, *92*, 104471. https://doi.org/10.1016/j.scs.2023.104471

Clément, A., Robinot, É., & Trespeuch, L. (2022). Improving ESG scores with sustainability concepts. *Sustainability*, *14*(20), 13154. https://doi.org/10.3390/su142013154

Clementino, E., & Perkins, R. (2021). How do companies respond to environmental, social and governance (ESG) ratings? Evidence from Italy. *Journal of Business Ethics*, *171*(2), 379–397. https://doi.org/10.1007/s10551-020-04441-4

Cleuziou, G. (2007). A generalization of K-Means for overlapping clustering. *Rapport Technique*, *54*.

Costantiello, A., & Leogrande, A. (2023). The impact of voice and accountability in the ESG framework in a global perspective.

Ehling, P., Haukø Lundeby, S. R., & Qvigstad Sørensen, L. (2023). Portfolio choice with ESG Disagreement: Customizing sustainability through direct indexing. *Journal of Beta Investment Strategies*, *14*(3), 132–149. https://doi.org/10.3905/jbis.2023.1.041

Elisabetta, D., & Iannuzzi, A. P. (2017). Incentive plans, pay-for-non-financial performance, and ESG criteria: Evidence from the European banking sector. *International Business Research*, *10*(10), 169–182. https://doi.org/10.5539/ibr.v10n10p169

Evans, G. L. (2017). Disruptive technology and the board: The tip of the iceberg. *Economics and Business Review*, *3*(1), 205–223. https://doi.org/10.18559/ebr.2017.1.11

Freeman, R. E. (2010). *Strategic management: A stakeholder approach*. Cambridge University Press.

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/10.1080/20430795.2015.1118917

Gebhardt, M., Thun, T. W., Seefloth, M., & Zülch, H. (2023). Managing sustainability—Does the integration of environmental, social and governance key performance indicators in the internal management systems contribute to companies' environmental, social and governance performance? *Business Strategy and the Environment*, *32*(4), 2175–2192. https://doi.org/10.1002/bse.3242

González-Serrano, M. H., Añó Sanz, V., & González-García, R. J. (2020). Sustainable sport entrepreneurship and innovation: A bibliometric analysis of this emerging field of research. *Sustainability*, *12*(12), 5209. https://doi.org/10.3390/su12125209

Grougiou, V., Anagnostopoulou, S., & Tingey-Holyoak, J. L. (2024). Sustainability literature review research: Advancing theory and practice. *Sustainability Accounting, Management and Policy Journal*, *15*(5), 1017–1037. https://doi.org/10.1108/SAMPJ-03-2024-0198

Iamandi, I. E., Constantin, L. G., Munteanu, S. M., & Cernat-Gruici, B. (2019). Mapping the ESG behavior of European companies. A holistic Kohonen approach. *Sustainability*, *11*(12), 3276. https://doi.org/10.3390/su11123276

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Janipour, Z., de Gooyert, V., Huijbregts, M., & de Coninck, H. (2022). Industrial clustering as a barrier and an enabler for deep emission reduction: A case study of a Dutch chemical cluster. *Climate Policy*, *22*(3), 320–338. https://doi.org/10.1080/14693062.2022.2025755

Jiménez, E., de la Cuesta-González, M., & Boronat-Navarro, M. (2021). How small and medium-sized enterprises can uptake the sustainable development goals through a cluster management organization: A case study. *Sustainability*, *13*(11), 5939. https://doi.org/10.3390/su13115939

Kanemoto, K., Hanaka, T., Kagawa, S., & Nansai, K. (2019). Industrial clusters with substantial carbon-reduction potential. *Economic Systems Research*, *31*(2), 248–266. https://doi.org/10.1080/09535314.2018.1492369

Khalil, M. A., Khalil, S., & Sinliamthong, P. (2024). From ratings to resilience: The role and implications of environmental, social, and governance (ESG) performance in corporate solvency. *Sustainable Futures*, *8*, 100304. https://doi.org/10.1016/j.sftr.2024.100304

Ko, B., Lee, K., Yoon, Y., & Park, S. (2022). Impact of ESG (Environmental, Social, Governance) on the Performance of Electric Utilities. *New & Renewable Energy*, *18*(2), 60–72. https://doi.org/10.7849/ksnre.2022.0010

Kocmanová, A., & Dočekalová, M. (2012). Construction of the economic indicators of performance in relation to environmental, social and corporate governance (ESG) factors. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *60*(4), 195–206.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means clustering. *International Journal*, *1*(6), 90–95.

Kotsantonis, S., & Serafeim, G. (2019). Four things no one will tell you about ESG data. *Journal of Applied Corporate Finance*, *31*(2), 50–58. https://doi.org/10.1111/jacf.12346

Kuo, K. C., Yu, H. Y., Lu, W. M., & Le, T. T. (2022). Sustainability and corporate performance: Moderating role of environmental, social, and governance investments in the transportation sector. *Sustainability*, *14*(7), 4095. https://doi.org/10.3390/su14074095

LaBella, M. J., Sullivan, L., Russell, J., & Novikov, D. (2019). The devil is in the details: The divergence in ESG data and implications for responsible investing. *New York: QS Investors*, *11*.

Li, T. T., Wang, K., Sueyoshi, T., & Wang, D. D. (2021). ESG: Research progress and future prospects. *Sustainability*, *13*(21), 11663. https://doi.org/10.3390/su132111663

Lin, A. J., Chang, H. Y., & Hung, B. (2022). Identifying key financial, environmental, social, governance (ESG), bond, and COVID-19 factors affecting global shipping companies—a hybrid multiple-criteria decision-making method. *Sustainability*, *14*(9), 5148. https://doi.org/10.3390/su14095148

Ma, B., Lin, S., Bashir, M. F., Sun, H., & Zafar, M. (2023). Revisiting the role of firm-level carbon disclosure in sustainable development goals: Research agenda and policy implications. *Gondwana Research*, *117*, 230–242. https://doi.org/10.1016/j.gr.2023.02.002

MacNeil, I., & Esser, I. M. (2022). From a financial to an entity model of ESG. *European Business Organization Law Review*, *23*(1), 9–45. https://doi.org/10.1007/s40804-021-00234-y

Manduchi, L., Chin-Cheong, K., Michel, H., Wellmann, S., & Vogt, J. (2021). Deep conditional Gaussian mixture model for constrained clustering. *Advances in Neural Information Processing Systems*, *34*, 11303–11314.

Marie, M., Qi, B., Gerged, A. M., & Nobanee, H. (2024). Exploring environmental, social and governance research in the wake of COVID-19: A bibliometric analysis of current trends and recommendations for future research. *Corporate Social Responsibility and Environmental Management*, *31*(6), 6131–6149. https://doi.org/10.1002/csr.2909

Nakielski, M. L. (2023). Moving forward with ESG, sustainability, and Corporate Responsibility. *Frontiers of Health Services Management*, *40*(1), 33–39. https://doi.org/10.1097/HAP.0000000000000178

Nielsen, H., & Villadsen, K. (2023). The ESG discourse is neither timeless nor stable: How Danish companies ‘tactically ’embrace ESG concepts. *Sustainability*, *15*(3), 2766. https://doi.org/10.3390/su15032766

Orchard, T., & Tasiemski, L. (2023). The rise of generative AI and possible effects on the economy. *Economics and Business Review*, *9*(2), 9–26. https://doi.org/10.18559/ebr.2023.2.732

Paolone, F., Cucari, N., Wu, J., & Tiscini, R. (2022). How do ESG pillars impact firms’ marketing performance? A configurational analysis in the pharmaceutical sector. *Journal of Business & Industrial Marketing*, *37*(8), 1594–1606. https://doi.org/10.1108/JBIM-07-2020-0356

Papagiannidis, S., See-To, E. W., Assimakopoulos, D. G., & Yang, Y. (2018). Identifying industrial clusters with a novel big-data methodology: Are SIC codes (not) fit for purpose in the Internet age? *Computers & Operations Research*, *98*, 355–366. https://doi.org/10.1016/j.cor.2017.06.010

Park, S. R., & Jang, J. Y. (2021). The impact of ESG management on investment decision: Institutional investors’ perceptions of country-specific ESG criteria. *International Journal of Financial Studies*, *9*(3), 48. https://doi.org/10.3390/ijfs9030048

Park, A., & Ravenel, C. (2013). Integrating sustainability into capital markets: Bloomberg LP And ESG's quantitative legitimacy. *Journal of Applied Corporate Finance*, *25*(3), 62-67. https://doi.org/10.1111/jacf.12030

Radu, C., & Smaili, N. (2021). Corporate performance patterns of Canadian listed firms: Balancing financial and Corporate Social Responsibility outcomes. *Business Strategy and the Environment*, *30*(7), 3344–3359. https://doi.org/10.1002/bse.2806

Ratnam, K. A., & Dominic, P. D. D. (2011, September 19–20). *A study of technology sustainability on hospital information management system (HIMS) Governance in Malaysia*. 2011 National Postgraduate Conference, IEEE. Perak. Malaysia. https://doi.org/10.1109/NatPC.2011.6136309

Ronalter, L. M., Bernardo, M., & Romaní, J. M. (2023). Quality and environmental management systems as business tools to enhance ESG performance: A cross-regional empirical study. *Environment, Development and Sustainability*, *25*(9), 9067–9109. https://doi.org/10.1007/s10668-022-02425-0

Rusu, Ș., Boloș, M. I., & Leordeanu, M. (2023). K-means and agglomerative hierarchical clustering analysis of ESG scores, yearly variations, and stock returns: Insights from the energy sector in Europe and the United States. *Journal of Financial Studies*, *8*(Special-June), 166–180.

Saini, N., Singhania, M., Hasan, M., Yadav, M. P., & Abedin, M. Z. (2022). Non-financial disclosures and sustainable development: A scientometric analysis. *Journal of Cleaner Production*, *381*, 135173. https://doi.org/10.1016/j.jclepro.2022.135173

Saraswati, E., Ghofar, A., Atmini, S., & Dewi, A. A. (2024). Clustering of companies based on sustainability performance using ESG materiality approach: evidence from Indonesia. *International Journal of Energy Economics and Policy*, *14*(2), 112–125. https://doi.org/10.32479/ijeep.15393

Sariyer, G., Mangla, S. K., Chowdhury, S., Sozen, M. E., & Kazancoglu, Y. (2024). Predictive and prescriptive analytics for ESG performance evaluation: A case of Fortune 500 companies. *Journal of Business Research*, *181*, 114742. https://doi.org/10.1016/j.jbusres.2024.114742

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, *8*(1), 289.

Spence, M. (1978). Job market signaling. In *Uncertainty in economics* (pp. 281–306). Academic Press.

Sultana, S., Zulkifli, N., & Zainal, D. (2018). Environmental, social and governance (ESG) and investment decision in Bangladesh. *Sustainability*, *10*(6), 1831. https://doi.org/10.3390/su10061831

Sun, Y., Chen, M., Yang, J., Ying, L., & Niu, Y. (2022). Understanding technological input and low-carbon innovation from multiple perspectives: Focusing on sustainable building energy in China. *Sustainable Energy Technologies and Assessments*, *53*, 102474. https://doi.org/10.1016/j.seta.2022.102474

Suproń, B. (2025). Environmental pollution and economic growth in the European Union countries: A systematic literature review. Economics and Business Review, 11(1), 7-30. https://doi.org/10.18559/ebr.2025.1.1777

Surroca, J., Tribó, J. A., & Waddock, S. (2010). Corporate responsibility and financial performance: The role of intangible resources. *Strategic Management Journal*, *31*(5), 463–490. https://doi.org/10.1002/smj.820

Van Holt, T., & Whelan, T. (2021). Research frontiers in the era of embedding sustainability: Bringing social and environmental systems to the forefront. *Journal of Sustainability Research*, *3*(2).

Vichi, M., Cavicchia, C., & Groenen, P. J. (2022). Hierarchical means clustering. *Journal of Classification*, *39*(3), 553–577. https://doi.org/10.1007/s00357-022-09419-7

Vilas, P., Andreu, L., & Sarto, J. L. (2022). Cluster analysis to validate the sustainability label of stock indices: An analysis of the inclusion and exclusion processes in terms of size and ESG ratings. *Journal of Cleaner Production*, *330*, 129862. https://doi.org/10.1016/j.jclepro.2021.129862

Vinayavekhin, S., Li, F., Banerjee, A., & Caputo, A. (2023). The academic landscape of sustainability in management literature: Towards a more interdisciplinary research agenda. *Business Strategy and the Environment*, *32*(8), 5748-5784. https://doi.org/10.1002/bse.3447

Wu, B., Gu, Q., Liu, Z., & Liu, J. (2023). Clustered institutional investors, shared ESG preferences and low-carbon innovation in family firm. *Technological Forecasting and Social Change*, *194*, 122676. https://doi.org/10.1016/j.techfore.2023.122676

Xie, J., Nozawa, W., Yagi, M., Fujii, H., & Managi, S. (2019). Do environmental, social, and governance activities improve corporate financial performance? *Business Strategy and the Environment*, *28*(2), 286–300. https://doi.org/10.1002/bse.2224

Zhong, W. (2023). An analysis of focusing on ESG investment. *Highlights in Business, Economics and Management*, *17*(1), 8–13.