

# Data Mining for Sustainability

## Tracking Environmental Progress in the Oil and Gas Sector Using AI

For ESG funds, distinguishing between companies genuinely committed to sustainability and those engaging in superficial greenwashing is critical. Current reporting inconsistencies make this task challenging, leaving ESG funds reliant on incomplete or unstandardized data. This project tackles these issues by focusing on two primary objectives:

1. **Tracking Progress Toward Environmental Goals:** Automate the extraction and analysis of metrics such as Scope 1-3 emissions, green investments, and methane intensity from annual reports.
2. **Analyzing Impact of Financial Performance:** Evaluate the relationship between financial metrics (e.g., operating expenses, capital expenditures) and sustainability efforts to understand the economic trade-offs.

## Text Extraction

To mitigate token limitations in the ChatGPT API, relevant sustainability-related sections were extracted from the reports using the PyMuPDF library. Regular expressions dynamically identified specific sections, while “capture flags” ensured targeted text extraction. To improve accuracy, the process filtered out irrelevant content, such as footnotes, using font size and cleaned non-printable characters. The results were structured into Excel for year-over-year comparisons across the selected companies.

The analysis focused on Scope 1, Scope 2, and Scope 3 emissions, investments in green initiatives, and methane emissions. These metrics were chosen as they provide a comprehensive view of a company’s environmental impact:

- **Scope 1, 2, and 3 Emissions:** Covering direct, indirect, and value-chain carbon footprints to evaluate decarbonization efforts.
- **Investments in Green Initiatives:** Highlighting commitment to transitioning to sustainable energy.
- **Methane Emissions:** Addressing their high global warming potential as a critical area for environmental impact reduction.

## Sentiment Analysis

Leveraging OpenAI’s API, the project evaluate the companies’ performance on their sustainability goals. The sentiment analysis provided the following outputs for each company and goal:

1. **Sentiment Classification:** Positive, negative, or neutral.
2. **Explanation:** A brief rationale for the sentiment classification.

### 3. Percentage Changes:

- From the Previous Year: Highlighting annual progress.
- From the 2019 Baseline: Benchmarking performance relative to 2019.

The extracted data and analysis were validated against BP's ESG report, which contained detailed breakdown of their sustainability performance since 2019.

Aim	Measure/coverage	2023 performance
① Net zero operations★	Scope 1 and 2★	41% <sup>a</sup>
② Net zero production★	Scope 3★	13% <sup>a</sup>
③ Net zero sales★	Average lifecycle carbon intensity	3% <sup>b</sup>

Bp Aims	Year	Sentiment	vs previous year	vs 2019
Aim 1 (Scope 1 & 2)	2023	+Positive	+2%	-41%
Aim 2 (Scope 3)	2023	+Neutral	+3%	-13%
Aim 3 (Carbon Intensity)	2023	+Positive	+0%	-3%

Figure 1 - Comparing the values from Text Extraction vs BP's ESG Report

### Local LLMs vs. ChatGPT

While initial attempts utilized local Large Language Models (LLMs) to maintain cost control and flexibility, these required extensive training to interpret the nuanced language of the annual reports. Transitioning to ChatGPT (GPT-3.5), pre-trained for diverse contexts, allowed for more accurate sentiment classification and performance interpretation. This integration was achieved through OpenAI's API.

## Financial Analysis

This section evaluates the influence of financial performance on a company's ability to meet its environmental goals. Financial metrics were extracted from the company's financial reports using an R script, and a multivariate regression model was employed to analyze relationships between these metrics and environmental objectives.

The analysis included eight key financial metrics for each company, spanning data from 2019 to 2023. Adjusted  $R^2$  values and p-values were used to identify the financial metrics most strongly associated with sustainability outcomes, providing critical insights into the financial implications of sustainability strategies.

# Key Findings

## Increased Operating Expenses to Achieve Sustainability Outcomes

As companies in the oil and gas sector invest more in sustainability initiatives, they often face higher operating expenses (OPEX). These increased costs are typically associated with efforts to reduce emissions, such as upgrading infrastructure, adopting new technologies, or investing in renewable energy. For instance, BP's operational costs increased as methane emissions declined from 2019 to 2023. This demonstrates the direct financial impact of emission-reduction projects.

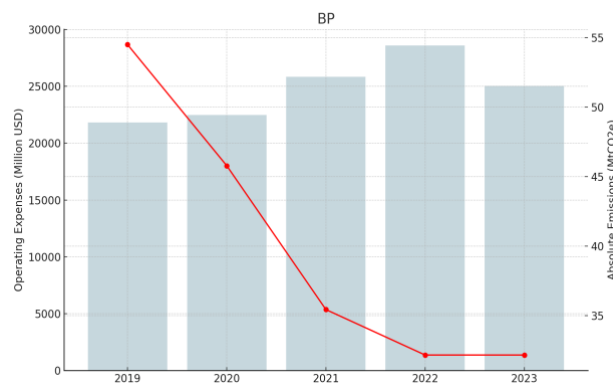


Figure 2 – BP's Operating Expenses & Methane Emissions (2019 – 2023)

## Increased Capital Expenditures for Emission Reductions

Capital expenditures (CapEx) are crucial for funding large-scale sustainability projects such as renewable energy infrastructure, carbon capture technologies, and methane reduction initiatives. These investments are typically capital-intensive and require substantial upfront spending. In Figure 3, we observe that Shell's increase in CapEx correlates with a significant reduction in Scope 1 and Scope 2 emissions. This suggests that the company's capital investment in green technologies has directly contributed to its environmental performance.

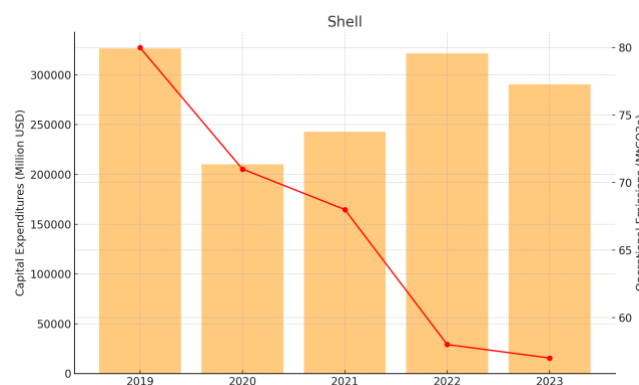


Figure 3 - Shell's Capital Expenditure & Scope 1 & 2 Emissions (2019 – 2023)

## Financial Disconnect

Despite environmental gains, companies often experience a disconnect between sustainability investments and short-term revenue. This challenge is particularly evident in the oil and gas sector, where long-term sustainability initiatives, such as renewable energy projects and carbon capture technologies, require substantial upfront costs without guaranteeing immediate financial returns. BP's pivot toward fossil fuels underscores the financial strain companies face when attempting to balance shareholder expectations for short-term profitability with the need for meaningful environmental progress (Ambrose, et al., 2024).

## Stakeholder Recommendations

To improve transparency and accountability in sustainability efforts, it is crucial to adopt a standardized ESG (Environmental, Social, and Governance) reporting framework. A unified structure, such as the European Sustainability Reporting Standards (ESRS), would ensure consistent and comparable disclosure of sustainability metrics across companies (Prewave, 2024). Currently, the project relies on annual reports to extract relevant data; however, not all companies provide this information, and when they do, it is often in varying formats, which hinders comparability.

In addition, introducing robust ESG performance benchmarking will allow stakeholders to effectively evaluate and compare sustainability efforts across companies. To further enhance credibility, mandatory third-party audits should be implemented. These audits will validate the accuracy of reported data, ensuring accountability and preventing greenwashing. Research shows that firms with external audits provide more accurate ESG reports, increasing stakeholder trust (Gelfand, 2024).

Combining standardized reporting, benchmarking, and independent audits will empower stakeholders - investors, regulators, and consumers - to assess companies' sustainability progress confidently and drive meaningful progress towards global environmental and social goals.

## Conclusion

The "Data Mining for Sustainability" project presents a transformative framework for evaluating the environmental progress of oil and gas companies. By integrating automated text extraction, AI-driven sentiment analysis, and financial modeling, this approach bridges the gap between corporate claims and measurable outcomes. The findings highlight the complex interplay between financial strategies and sustainability goals, underscoring the need for standardized ESG reporting and independent audits to ensure accountability.

This project equips stakeholders with actionable insights, fostering trust and accelerating progress toward achieving global environmental objectives.

## Bibliography

Ambrose, J., Kollewe, J. & Makortoff, K., 2024. *Very concerning': BP dilutes net zero targets as global retreat from green standards gathers pace*. [Online]

Available at: <https://www.theguardian.com/business/2024/oct/13/very-concerning-bp-dilutes-net-zero-targets-as-global-retreat-from-green-standards-gathers-pace>

[Accessed 15 November 2024].

Crookall, P., 2022. *EXPLORING KEY ESG CHALLENGES FOR THE OIL AND GAS INDUSTRY*. [Online]

Available at: <https://www.rpsgroup.com/imported-media/11073/challenges-of-esg-in-oil-and-gas.pdf>

[Accessed 13 December 2024].

Gelfand, A., 2024. *New Data on the Quality of ESG Audits Catches Regulators' Attention*. [Online]

Available at: [https://www.gsb.stanford.edu/insights/new-data-quality-esg-audits-catches-regulators-attention?utm\\_source=chatgpt.com](https://www.gsb.stanford.edu/insights/new-data-quality-esg-audits-catches-regulators-attention?utm_source=chatgpt.com)

[Accessed 13 December 2024].

International Energy Agency, 2023. *Oil and gas industry faces moment of truth – and opportunity to adapt – as clean energy transitions advance*. [Online]

Available at: <https://www.iea.org/news/oil-and-gas-industry-faces-moment-of-truth-and-opportunity-to-adapt-as-clean-energy-transitions-advance>

[Accessed 13 December 2024].

Prowave, 2024. *Breaking Down the ESRS Standards: A Practical Guide for Business Leaders*. [Online]

Available at: <https://www.prowave.com/blog/breaking-down-the-esrs-standards-a-practical-guide-for-business-leaders>

[Accessed 12 December 2024].