ESG

Cross-Industry ESG Analysis

Uncovering Sustainability Sentiments, Greenwashing, and Corporate Responsibility Insights Through Natural Language Processing

AN6002 – Analytics and Machine Learning in Business

Group 8:
Dhruva Kashyap Shastry, He Jingmei, Sonali Shripad Shanbhag, Svadhaa Agarwal, Simon Eppig

ESG is affecting company perception and sustainability efforts across industries

ESG

Stakeholder Expectations

E
Environmental

Impact on the physical environment and the climate

e.g. greenhouse gas emissions, water pollution

S
Social

Relationship with different relevant stakeholder groups

e.g. staff diversity, health & safety measures

G
Governance

Company organisation and management

e.g. risk management, compliance

Sustainability Reporting

Bias, inconsistencies and greenwashing undermine sustainability efforts

ESG

1 Situation

Current Sustainability Reporting Standards

Regional

OR

Non-mandatory

OR

Difficult to compare

2 Consequence

Corporate Sustainability Reports

Inconsistent

AND

Biased

AND

Filled with Greenwashing

3 Problem

Troubles for Stakeholders

Investors
Making sustainable investments

Legislators
Supporting sustainable businesses

Market Analysts
Advising clients on sustainability

Company Executives
Comparing with competitors

Stakeholder needs are addressed through precise objectives and advanced analytics

ESG

Project Objectives

Evaluate and compare ESG performance and public perception across various industries and companies

1. Identify the most important sustainability topics in different industries
2. Examine the potential for greenwashing in different industries
3. Measure greenwashing in different industries by analysing News Media

Analytics Methodology

Natural Language Processing on ESG reports and news articles

1. Word Cloud Analysis
2. WordNet Expansion
3. Term Frequency–Inverse Document Frequency
4. Sentiment Analysis
5. Greenwashing Detection

Comprehensive analytics approach maximises stability and reliability

ESG

Natural Language Processing (NLP) can help identify sentiment trends, detect potential greenwashing, and classify companies based on their emphasis on specific ESG pillars, offering a data-driven understanding of sustainability communication patterns.

Word Cloud Analysis

Conduct **WordCloud analysis** to identify common themes in each industry

WordNet Expansion

Enhance keyword list to account for variations and linguistic differentiation, ensuring accurate analysis

TF-IDF

Importance based metric to identify the ESG pillar being represented most by each company

Sentiment Analysis

Applied to sustainability reports and ESG-related news using **VADER** and **TextBlob** to measure tone, **polarity**, and **subjectivity**.

Greenwashing Detection

Measured the ratio of **aspirational** terms to **substantive** terms to identify cases where companies may be **overstating** sustainability claims

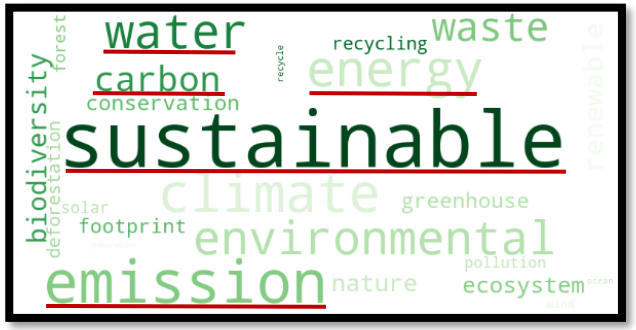
Word cloud analysis reveals unique set of important keywords for each ESG pillar

ESG

Strong emphasis on **climate change**, **resource management**, **renewable energy**, and **environmental restoration**.

Emphasis on **regulatory alignment**, **transparency**, and **ethical practices**.

Environmental



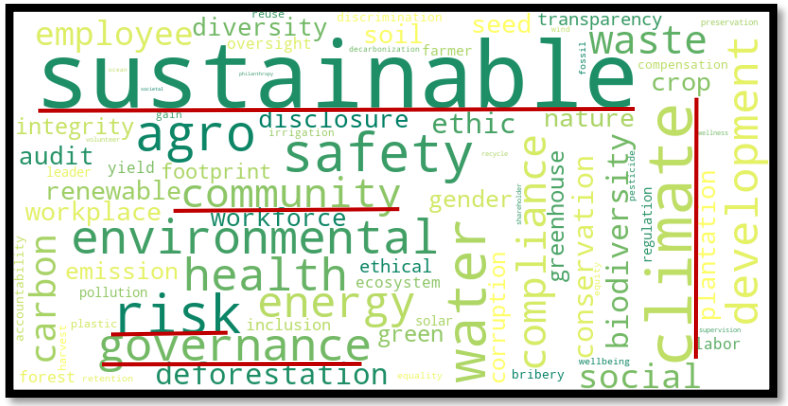
Social

Focus on **occupational safety**, **healthy and inclusive workplaces**, and **broad social value creation**.

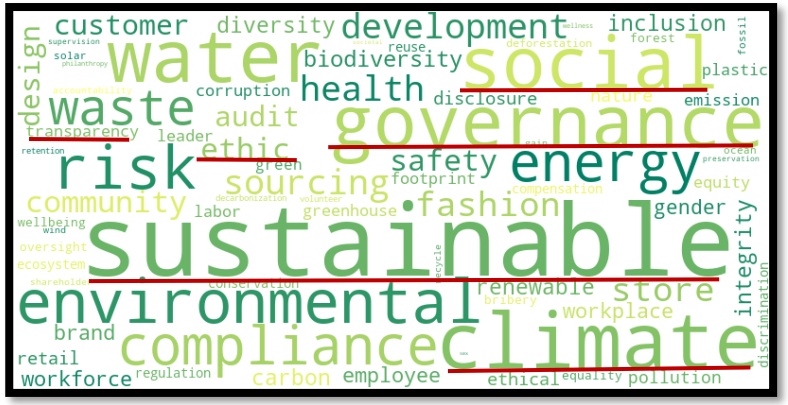
Governance

Industry-wise word cloud analysis reveals distinct ESG focus in different sectors

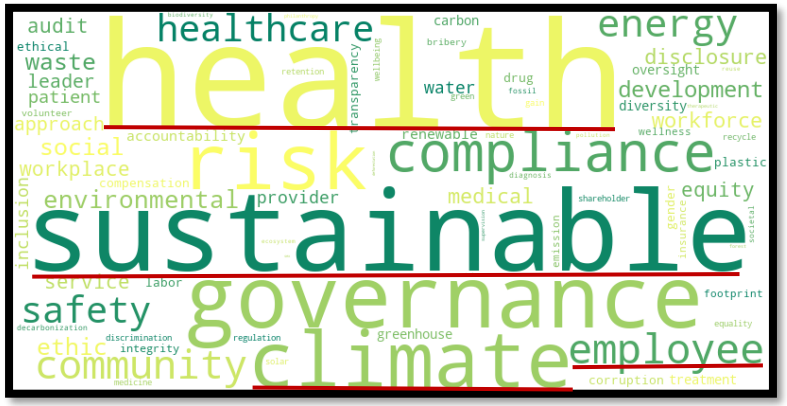
ESG



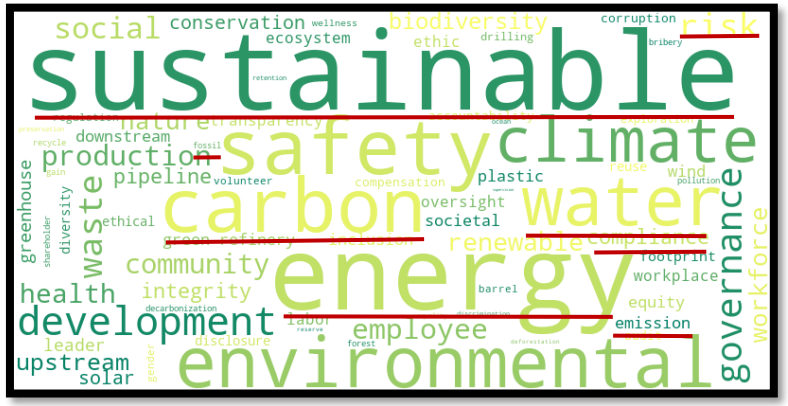
Agriculture – Environmental/Social



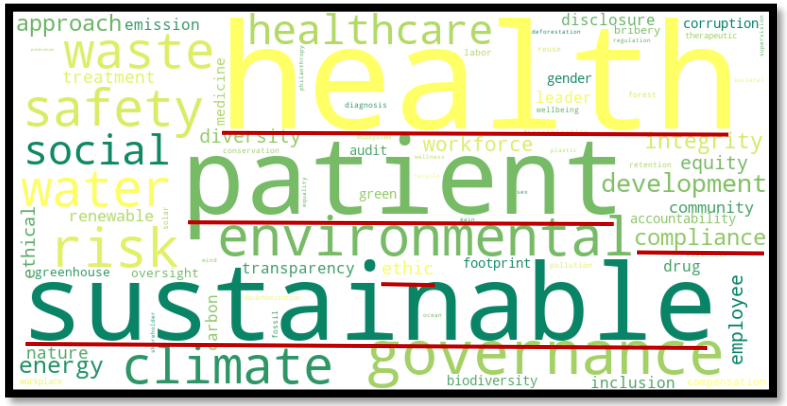
Fashion Retail – Environmental/Social



Healthcare – Social



Oil & Gas – Environmental



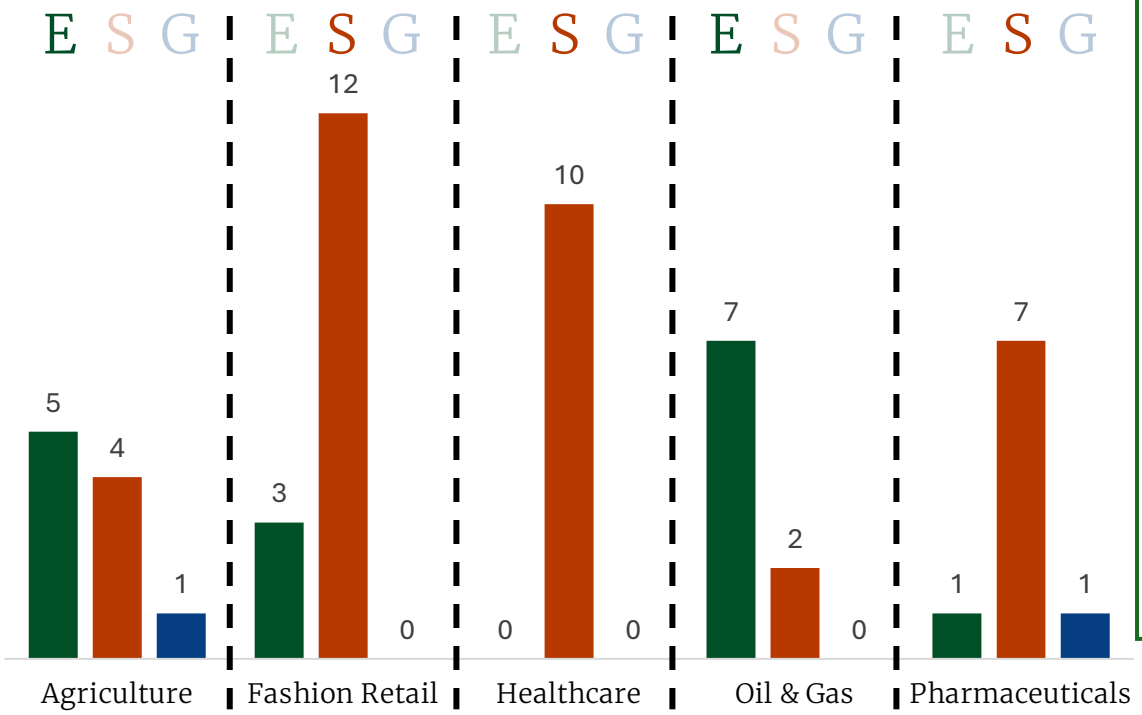
Pharmaceutical – Social

Quantifying the ESG Pillar Dominance Across Industries and Companies

ESG

Industry-wise Pillars

Dominant Pillars Among Each Industry



1. 🌽 Agricultural Industry: **food security** and **worker welfare,** managing **environmental impact** on **biodiversity**
2. 👗 **Fashion Retail:** labor-intensive supply chains with concerns around **worker rights, fair wages**, and **working conditions**
3. 🛢️ **Oil & Gas Industry: high carbon emissions, pollution risks**, and impact on **climate change** and **ecosystem degradation**
4. 🏥 **Healthcare: patient care, health equity, accessibility**, and **community wellbeing**
5. 💊 **Pharmaceutical Industry: drug affordability, patient access, ethical clinical trials**

Outliers arise when companies redefine their ESG focus:

- **Mitigate risk exposure or rebuild trust after controversies (KPK, Exxon)**
- **Differentiate through brand identity and consumer alignment (Fashion)**
- **Comply with evolving regulations or meet investor disclosure expectations (US Healthcare)**
- **Showcase internal governance strength to enhance transparency and control (KPK, US Healthcare)**

Assessing Greenwashing Risk Scores and Vader Scores using Corporate ESG Reports

VADER and TextBlob Scoring

High Sentiment Scores, companies consistently use **highly positive language** in their sustainability reports. Persists **even among companies with higher controversy levels**, indicating a **potential risk of greenwashing**.

Defining a Risk Measurement Metric

Greenwashing Ratio = Aspirational/ Substantive
Greenwashing Density = Aspirational Count / Total Words

Risk Score = 0.6* Ratio * 0.4* Density

to be considered high-risk, a report needs a greenwashing ratio of at least 2.5 and a density greater than 25

Industry-wise Risk Analysis

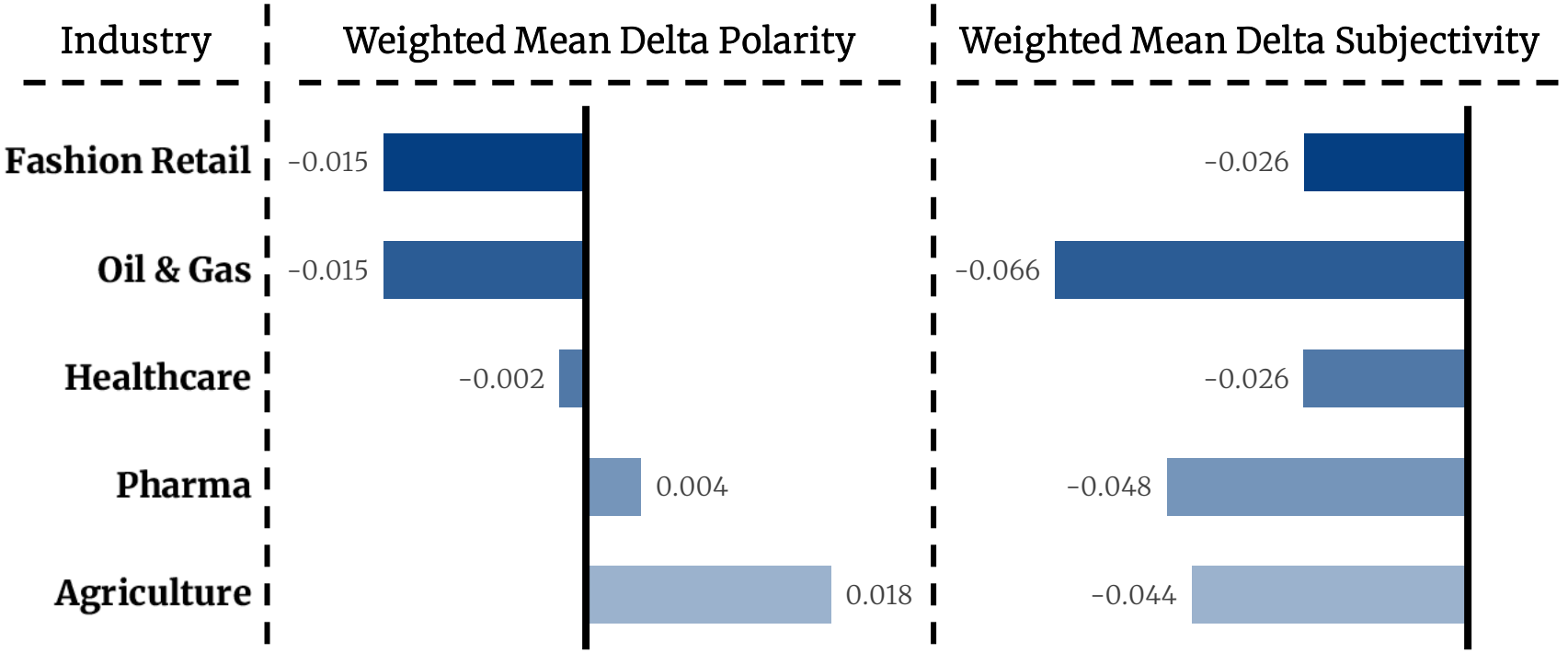
Industry	Average Risk Score	% Companies At High Risk
Agriculture	6.15	50%
Fashion Retail	5.94	53%
Healthcare	6.35	70%
Oil & Gas	6.38	78%
Pharmaceuticals	8.07	78%

Industry Sentiment Analysis and Greenwashing Detection Using News Media

ESG

Industry-wise Polarity and Subjectivity

Weighted Mean Measures across Industries



Polarity and Subjectivity Analysis

Polarity = measures positivity vs. negativity
Subjectivity = measures personal/emotional tone

Both values are small, which is expected in journalistic or formal text, as such writing tends to stay neutral and factual rather than emotional.

Agriculture: Highest positive delta polarity → Companies present themselves more positively.

Pharma: Slight positive gap → Mildly promotional tone.

Oil & Gas & Fashion Retail: Negative gaps → Reports are more restrained/cautious, reflecting credibility and trust-building efforts.

Healthcare: Neutral gap → Maintains balance between corporate and media sentiment.

The analysis results provide meaningful insights for all stakeholder groups

ESG

Dominant ESG Pillars

- Each industry shows a clear ESG pillar focus, aligned with its operational impact.

Aspirational Vs Substantive Language

- Aspirational language is common across sectors, with Pharma and Oil & Gas showing stronger greenwashing signals.

News Article Analysis

- News articles tend to be more neutral, which may mask the promotional tone found in some ESG reports.

Stakeholder Implications

1. **Investors**
   Can use assess aspirational claims to make better investment decisions
2. **Legislators**
   Can identify discrepancies in reporting to impose fines and penalties
3. **Market Analysts**
   Can Apply NLP techniques to assess performance and detect greenwashing
4. **Company Executives**
   Can gain insight into competitors and where changes can be made for accurate reporting

Limitations of the analysis should be addressed through further research

ESG

1 Limitations

Limited number of industries, companies and data types used

While generally representative, there could be bias in only including largest companies

Neutral tones in reports and news articles

Limits the ability of stakeholders to assess overall greenwashing risks and sentiment

2 Future Direction

Including more companies and quantitative data for further analysis

Combining company results or ESG scores with NLP results to achieve deeper insights

Including social media analysis for further sentiment analysis

Combining sentiment analysis with social listening to include public opinions