

**TABLE OF CONTENT**

| **S. No.** | **Topic** | **Pg. No.** |
| --- | --- | --- |
|  | Objective | 2 |
|  | Introduction | 3 |
|  | Methodology | 5 |
|  | Analysis |  |
|  | Findings |  |
|  | Visualisation |  |
|  | Conclusion |  |

**OBJECTIVE**

The objective of this report is to explore and map the evolution of Corporate Social Responsibility (CSR) and Environmental, Social, and Governance (ESG) disclosure practices in corporate reporting of the company Shell over the past 10-12 years, focusing on the adoption and application of frameworks like GRI, IR, and TCFD.

**INTRODUCTION**

**ABOUT THE COMPANY**

Shell Oil Company, commonly known as Shell, is a global energy company headquartered in The Hague, Netherlands. Established in 1907, it has since become one of the largest oil and gas companies in the world, operating in over 70 countries and employing more than 93,000 people.

With a rich history spanning over a century, Shell has played a significant role in shaping the energy landscape through its exploration, production, refining, distribution, and marketing of oil and gas products. The company is also involved in renewable energy sources such as wind and solar power, as well as innovative technologies like carbon capture and storage.

Shell's operations encompass a wide range of activities, including upstream exploration and production, downstream refining and marketing, and integrated gas ventures. The company is known for its commitment to safety, environmental stewardship, and sustainable development, striving to meet the world's growing energy needs while minimizing its impact on the planet.

In recent years, Shell has embarked on a strategic shift towards cleaner energy solutions, investing in low-carbon technologies and transitioning towards a more sustainable energy portfolio. This includes initiatives to reduce greenhouse gas emissions, increase energy efficiency, and promote the use of renewable resources.

As a leading player in the global energy industry, Shell continues to innovate and adapt to changing market dynamics, driving progress towards a more sustainable and energy-efficient future.

**INTRODUCTION**

**CSR & ESG in Reporting**

Corporate Social Responsibility (CSR) and Environmental, Social, and Governance (ESG) disclosure practices have become increasingly important in corporate reporting as businesses recognize the significance of their impact on society and the environment. These practices entail companies voluntarily integrating social and environmental concerns into their business operations and interactions with stakeholders. CSR and ESG reporting serve as mechanisms for companies to communicate their commitments, performance, and impacts in these areas to stakeholders, including investors, customers, employees, regulators, and communities.

Frameworks such as the Global Reporting Initiative (GRI), Integrated Reporting (IR), and Task Force on Climate-related Financial Disclosures (TCFD) provide guidelines and standards for companies to structure and report their CSR and ESG activities. The GRI framework, for instance, offers a comprehensive set of principles and indicators for sustainability reporting, covering economic, environmental, and social dimensions. Integrated Reporting, on the other hand, encourages companies to provide a holistic view of their value creation process by integrating financial and non-financial information into a single report. The TCFD framework focuses specifically on climate-related risks and opportunities, helping companies assess and disclose their climate-related financial information.

These frameworks not only help companies improve transparency and accountability but also enable stakeholders to make informed decisions and evaluate companies' sustainability performance. As businesses face growing pressure to address societal and environmental challenges, CSR and ESG reporting have become integral components of corporate governance and risk management. In this report, we will explore the significance of CSR and ESG disclosure practices, examine the key frameworks and standards used in corporate reporting, and discuss their implications for businesses and stakeholders alike.

**INTRODUCTION**

**Shell & Sustainability**

Shell, a multinational energy company renowned for its global footprint and diverse operations, has increasingly focused on sustainability practices as an integral aspect of its corporate strategy. In response to growing environmental concerns, regulatory pressures, and shifting consumer expectations, Shell has embarked on a multifaceted approach to address sustainability challenges across its entire value chain.

At the core of Shell's sustainability efforts is a commitment to reducing greenhouse gas emissions and mitigating climate change impacts. Recognizing the urgent need for action, Shell has set ambitious targets to significantly decrease its carbon footprint. These targets include reducing the carbon intensity of its energy products, investing in low-carbon technologies such as renewable energy and electric vehicles, and implementing carbon capture and storage initiatives.

Moreover, Shell has demonstrated a commitment to fostering innovation and collaboration to accelerate the transition to a low-carbon future. Through partnerships with governments, research institutions, and other industry players, Shell seeks to develop and deploy innovative solutions that promote sustainable energy production and consumption.

In addition to addressing climate change, Shell is actively engaged in promoting biodiversity conservation, water stewardship, and responsible resource management. By integrating environmental considerations into its operations, Shell strives to minimize its impact on ecosystems and natural resources while maximizing positive outcomes for biodiversity and local communities.

Furthermore, Shell recognizes the importance of social responsibility and community engagement in its sustainability agenda. The company works closely with stakeholders, including local communities, indigenous groups, and non-governmental organizations, to ensure that its activities contribute to the well-being and prosperity of society at large.

Overall, Shell's sustainability practices reflect a holistic approach to addressing the complex challenges facing the energy industry. By embracing innovation, collaboration, and responsible business practices, Shell aims to create long-term value for its stakeholders while contributing to a more sustainable and resilient future for the planet.

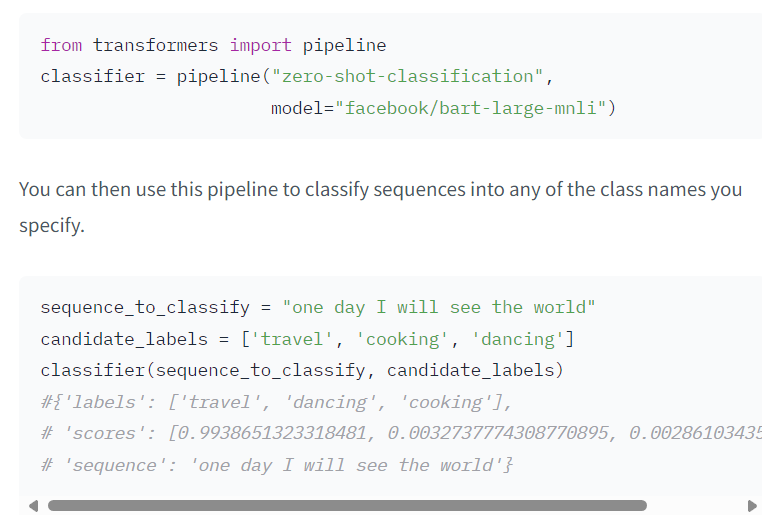
Shell utilizes CSR reporting to communicate its commitment to sustainability and responsible business practices. The company integrates social and environmental considerations into its operations and engages with stakeholders to address their concerns. Shell's CSR reports provide transparency on its performance, initiatives, and impacts in areas such as climate change, environmental stewardship, community engagement, and human rights. The company adheres to frameworks like the Global Reporting Initiative (GRI) to structure and report its CSR activities, enhancing accountability and enabling stakeholders to evaluate its sustainability efforts. Through CSR reporting, Shell demonstrates its dedication to mitigating risks, promoting long-term value creation, and contributing to a more sustainable future.

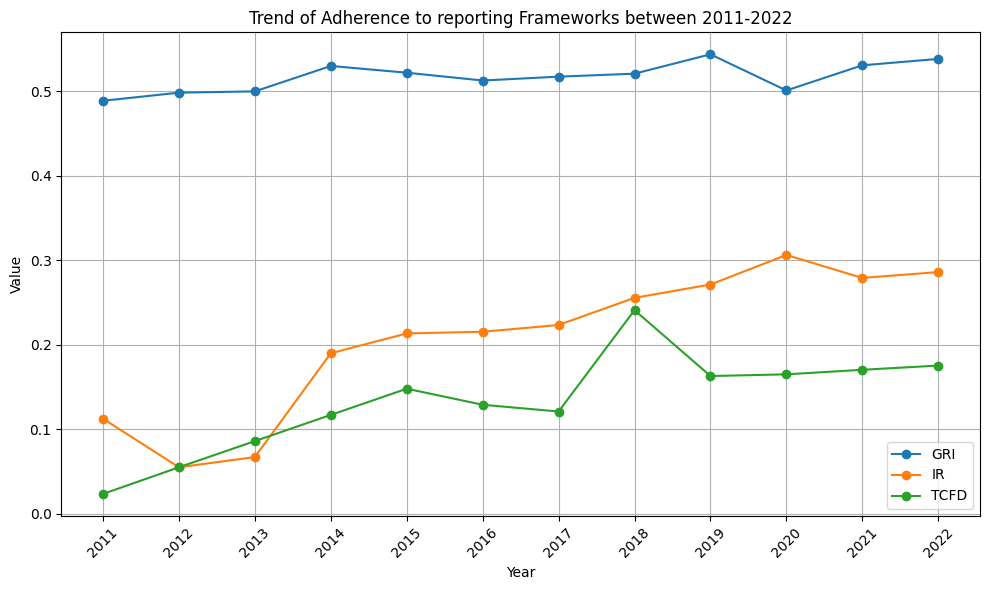
**METHODOLOGY**

**Data Pre Processing**

* The Sustainability reports were downloaded from the company website in the form of a PDF file.
* Next, the text was extracted from the PDF file, and each sentence of the PDF file was stored in a row of a CSV file.
* The sentences had to be split into shorter sentences since the ML models that will be used later for analysis only accepts sentences of length less than 512 tokens.
* After splitting the sentences, they are stored in a new CSV file. The data of this CSV file will be used as input for further analysis of the SUstainability report.

**0) Framework Identification**

* The first step is framework identification, i.e. to identify to what extent does the report follow various ESG reporting frameworks.
* For this project, we analyzed the reports to check to what extent the Sustainability reports of the company follows the GRI, IR and TCFD framework.
* For this, we used a model from Hugging face named “eleldar/theme-classification”
* This model uses a concept called “zero-shot learning” to predict the class to which a given input sentence belongs
* In zero-shot learning a model is trained on one set of data, but it is used to make predictions in a completely different area.
* The “eleldar/theme-classification” model wasn’t explicitly trained on Sustainability reports to classify which framework a report is adhering to.
* Instead, it was trained on MultiNLI dataset. The Multi-Genre Natural Language Inference (MultiNLI) corpus is a crowd-sourced collection of 433k sentence pairs annotated with textual entailment information.
* The model can be used as follows: 
* The second code block, it can be observed that we provide two input parameters to the classifier() function, a input sentence(that needs to classifier) and a list of labels. We can give our own choice of words in the list of labels.
* The model will then provide a “Confidence Score” between 0 to 1 for each label in the list, indicating how strongly the word define the theme/topic of the sentence.
* For our project, we iterated through each sentence of a report (this was the input sentence)
* For the “candidate\_labels”, we chose the words []
* Then we calculated the average score for all the sentences to get the overall score for a single report of a particular year.
* This procedure was repeated for all the years (2011-2022)
* The graph below shows the trend of scores

****

* It can be observed that the model classified the reports to be following the GRI framework with a higher Confidence score as compared to the other frameworks.

1. **ESG Score through Text mining**
2. Keyword extraction in text mining is an automated process that identifies the most relevant words and phrases within a text. It plays a vital role in distilling crucial information from paragraphs or documents, aiding in content summarization and issue identification. This technique is particularly useful when you need to analyse large amounts of text data, such as numerous product reviews online.

Keyword extraction is used to summarize the content of texts and recognize the main topics discussed. It uses machine learning artificial intelligence (AI) with natural language processing (NLP) to break down human language so that it can be understood and analysed by machines.

1. In our Project we have utilized the nltk library to Tokenize the text and arrange it for further sentimental analysis. We have used Finbert-esg pretrained ESG model to calculate the ESG Score from the Shell sustainability reports from 2011 to 2022. We can see that overall the scores generated for ESG are constant except a dip in 2021, which can be due to the impact of covid-19. There is also a dip in Governance score in 2014. This can be due to the $1.1 Billion scam by Shell and ex-Nigerian ministers who owned Malabu oil firm. We can see this phenomenon in Graph1 in Trends Analysis

Sentiment Analysis: Analyse the tone and sentiment of the sustainability disclosures.

2)Topic Modelling: Identifying main topics using LDA. (rohan)

* Trend Analysis: Analyse how reporting trends have evolved over time. (rohan)
* Categorization: Code the data based on the type of framework used and sustainability aspects reported. (rohan)

3)Theme Identification: Conduct thematic analysis to identify common themes and variations in reporting practices. (pramod)

1. **Reporting and Visualisation:**

* Use of Trend analysis and Graphs and Charts to compile the findings.

2) TOPIC Modelling and Categorization:

* For topic modelling, we used BERTopic.
* It was developed by Marten Groothendorst. BERTopic is a flexible and modular topic modeling framework that allows for the generation of easily interpretable topics from large datasets.
* It is trained on over 1,000,000 Wikipedia pages (first paragraph of each page). Specifically, we used the BERTopic\_Wikipedia model from Hugging Face.
* Next, we briefly explain how the BERTopic\_Wikipedia model works in terms of taking an input and generating an output. The model has 2378 categories. The model takes a sentence/paragraph as an input. Then, as an output, it assigns the most relevant category name (each category name contains 4 keywords) to the sentence along with a confidence score.
* We input the extracted textual data from the Sustainability PDF report.
* Then, for each sentence, it assigns the most relevant category along with a confidence score.
* Next we calculate the frequency of each category and select the top two categories as the ones that most aptly describe the main topic/theme of the report.
* We repeat this process for all the years and find the keywords that most aptly describe the topic of the report.
* Since each category name contains 4 keywords, and we selected the top two category names in terms of frequency, we get a total of eight keywords.
* These eight keywords serves as the categories of topics that most aptly describe the theme of the reports, and these 8 keywords are further ranked in the next step, thematic analysis.

3) Thematic Analysis

**Model:**

The pre-trained model used in this code is facebook/bart-large-mnli. It's a large language model (LLM) fine-tuned for a specific task called "natural language inference" (NLI). This task involves determining the relationship between two sentences (e.g., entailment, contradiction, neutral). Here, the model is cleverly adapted to classify sentences into user-defined categories (keywords) based on their semantic similarity.

**Methodology**

1. **Data Preparation:** The code starts by reading a CSV file containing text data (sentences) you want to classify. This file is assumed to be located at a specific path (input\_csv\_file).
2. **Model Loading:** The code. This model is specifically designed for zero-shot classification, which means it can classify text data into categories it hasn't seen before during training.
3. **Keyword List:** The code defines a list of keywords (categories) you're interested in classifying the sentences into. These keywords are not explicitly mentioned in the code snippet you provided, but they are likely defined elsewhere in your project.
4. **Sentence-by-Sentence Classification:** The code iterates through each sentence in the CSV file:
   * It extracts the sentence text.
   * It uses the loaded model to predict how well the sentence matches each keyword in the list. The model outputs a score between 0 (no match) and 1 (perfect match) for each keyword.
5. **Result Storage:** The code stores the sentence text and the corresponding scores for each keyword in a dictionary. These dictionaries are then collected into a list.
6. **Output Generation:** The code creates a new CSV file (output\_csv\_file) and writes the list of dictionaries (containing sentences and their classified scores) into this file.

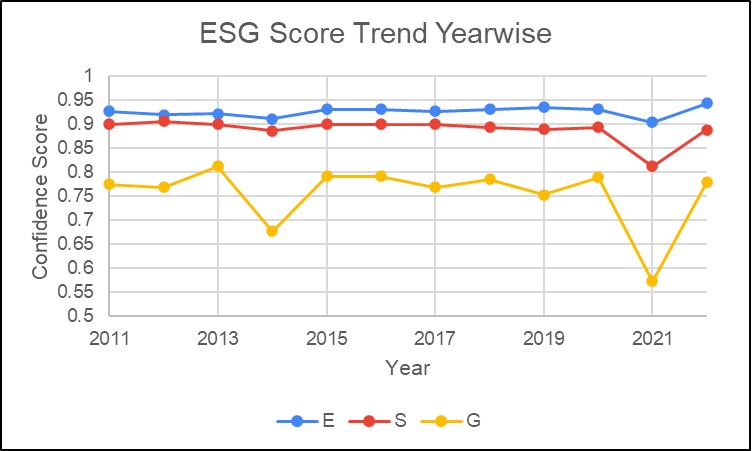
**How the Code Works Step-by-Step:**

1. **Imports:** The code imports necessary libraries (pandas for data manipulation and transformers for the pre-trained model).
2. **Model Loading:** It loads the pre-trained model (facebook/bart-large-mnli) for zero-shot classification.
3. **File Paths:** It defines the paths for the input and output CSV files.
4. **(Missing) Keyword List:** The keyword list is likely defined elsewhere in your project.
5. **Empty List:** An empty list is created to store the classification results.
6. **Read CSV:** The code reads the input CSV file using pandas.
7. **Iteration:** It iterates through each row (sentence) in the DataFrame.
8. **Extract Sentence:** The sentence text is extracted from the current row.
9. **Classification:** The model predicts scores for each keyword based on how well it matches the sentence.
10. **Store Results:** A dictionary is created to store the sentence

**ANALYSIS**

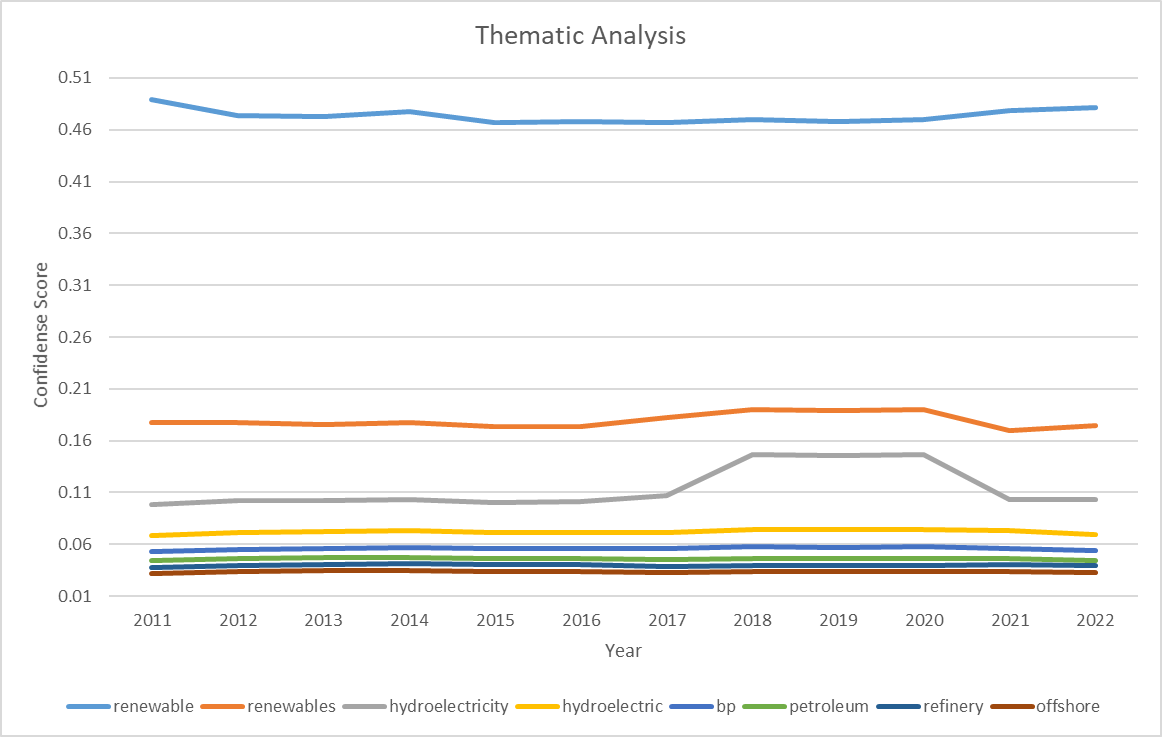
Trend Analysis:

The graph of the trend of ESG Scores from 2011 to 2022 is given below. The company has the highest score in the “Environmental” category, followed by “Social” Category and finally “Governance” category. This means that the company has comparatively mentioned most about Environmental related activities followed by Social related activities and finally Governance related activities. If we look at the trend, the scores are fairly consistent and slightly increasing over the years, apart from a dip in 2021.

****

**Graph1: ESG Score Trend Year Wise**

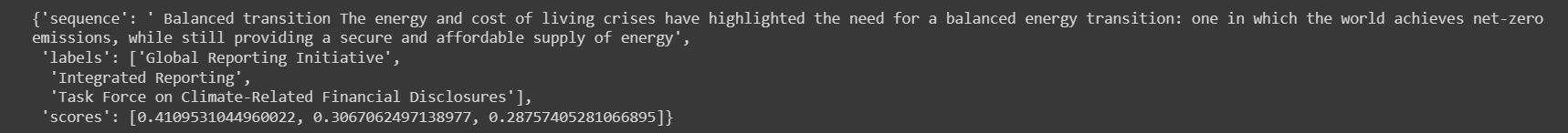
**Thematic Trend**

****

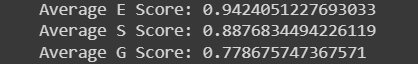
**Appendix**

**Output**

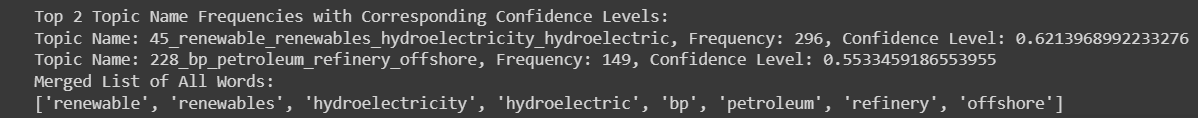
1. **Framework**

****

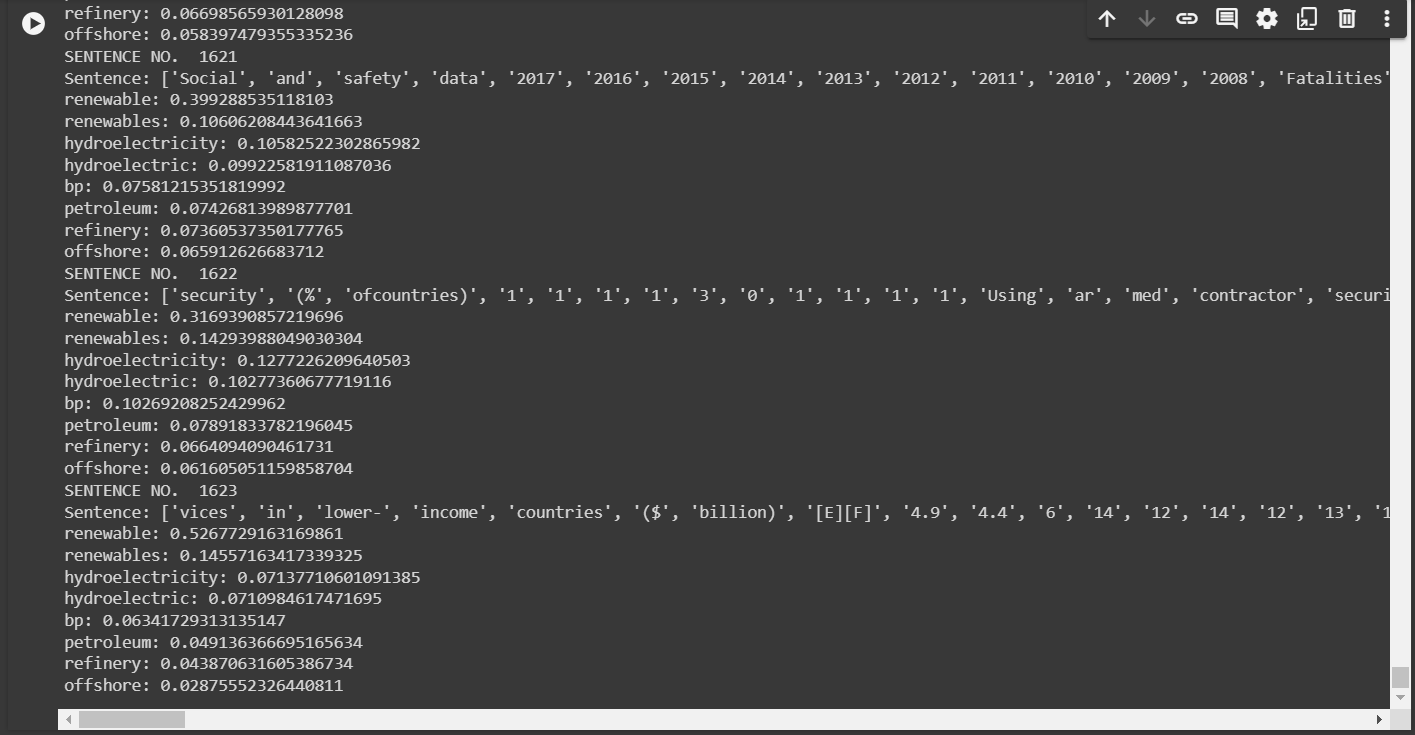
1. **Esg Score**

****

1. **LDA**

****

1. **Thematic Analysis**

****

**Model Link**

1. **Framework**

https://huggingface.co/eleldar/theme-classification?candidate\_labels=playing+sports%2C+music&multi\_class=true&text=I+train+football%20-

1. **ESG**

[**yiyanghkust/finbert-esg · Hugging Face**](https://huggingface.co/yiyanghkust/finbert-esg)

1. **Thematic**

<https://huggingface.co/eleldar/theme-classification?candidate_labels=playing+sports%2C+music&multi_class=true&text=I+train+football%20->

1. **Code repository of the project**

[**rohanPatgiri/Sustainability-Report-Analysis-using-various-ML-Models (github.com)**](https://github.com/rohanPatgiri/Sustainability-Report-Analysis-using-various-ML-Models/tree/main)

**FINDINGS**

**VISUALISATION**

**CONCLUSION**