International Journal of Cultural Inheritance & Social Sciences (IJCISS)
Vol. 8 Issue 14, September 2025, ISSN: 2632-7597 Impact Factor: 8.985
Journal Homepage: http://ijciss.com/, Email: journal.ijciss@gmail.com
Double-Blind, Peer Reviewed, Refereed & Open Access International Journal

NLP-DRIVEN ESG DISCLOSURE ANALYSIS IN BANKING ANNUAL REPORTS: STRENGTHENING REGULATORY AND SUPERVISORY TECHNOLOGIES

Nadir Hammadi
MSc Artificial Intelligence and Computer Science
Position: Supervisor - IT Support and Operations
in Arab Monetary Fund - Abu Dhabi - UAE
Email: nadir.hammadi@hotmail.com

Abstract

The objective of this project is to improve the state of Environmental, Social, and Governance (ESG) reporting by the application of machine learning and natural language processing, two forms of artificial intelligence (AI). Improvements to the summary and clarity of ESG reports, the use of RegTech solutions to expedite regulatory compliance, the use of SupTech solutions to facilitate the more effective and efficient delivery of regulatory requirements, and the promotion of transparency in business operations to support sustainable development are the biggest objectives.

To achieve these objectives, a methodology was followed to collect data from the relevant financial institutions (ADIB published on its website), that helps to include as many types of data as possible so that the study is comprehensive and does not exclude any type of data, followed by accurate pre-processing of this data, (with the help of artificial intelligence tools) to ensure the quality of the resulting summary content starting from the report index in the first attributes to the rest of the text.

To extract and analyze the ESG dataset, machine learning techniques were used, especially natural language processing (NLP, such as entity recognition, symbolism, sentiment analysis, and topic modeling) to identify and predict trends.

In addition, the integration of RegTech and SupTech tools will help in compliance checks and provide supervisory bodies with real-time insights, which is a challenge for effective supervision, and continuous performance evaluation, tuning and monitoring of AI models will ensure their accuracy and effectiveness.

The expected outcomes are the provision of accurate and human-like ESG report summary, improved regulatory compliance through automation, and accuracy of summary and clarity

of ESG reports. Furthermore, these outcomes will increase business transparency and trust among stakeholders, thus improving strategic decision-making in line with sustainability goals on the one hand and governance recommendations on the other. So, my project highlights the potential use of AI, not only to simplify and enhance regulatory processes but also to enhance accountability and support long-term sustainable business practices.

Keywords: Natural Language Processing (NLP), Environmental, Social, and Governance (ESG), Machine Learning, Artificial Intelligence (AI), RegTech, SupTech, Regulatory Compliance.

Table of Acronyms

AI	Artificial Intelligent
ML	Machine Learning
NLP	Naturel language Processing
ESG	Environmental, Social, Governance
ADIB	Abu Dhabi Islamic Bank
NER	Named entity recognition
LDA	Latent Dirichlet Allocation

1. Introduction

1.1 Background and Motivation

In the current rapidly evolving financial landscape, there is an increasing demand for greater attention to the quality, credibility and reliability of reports issued by financial institutions. This type of reporting has become mandatory for these institutions due to mandates from financial authorities and governments on the one hand, and to attract more capital and investors on the other hand.

In addition, I have included modern financial technologies that I believe help in shaping accurate outcomes that are in line with expectations from various responsible parties. regulatory technology and supervisory technology to change the way institutions meet any regulatory obligations and requirements in addition to how regulatory bodies monitor compliance.

"The components of ESG reporting are Environmental, Social, and Governance disclosures. Corporations are expected to maintain sustainable practices and consider the civil impact of their business decisions; beyond that, they are increasingly expected to report on these

efforts. Reporting regulations through the U.S. Securities and Exchange Committee (SEC) will release formal guidance for ESG disclosures for publicly traded companies in the US, but these companies may already be subject to international reporting directives such as the EU's CSRD.. Currently, the winds are blowing towards an integrated approach to ESG reporting, treating this information as part of an organization's financial reporting, and subjecting ESG disclosures to the same rigor of audit as financial disclosures. Though companies can define their own ESG strategy through a materiality assessment, there are some common threads that fall into each category of the E, S, and G." [1]

Today, regulatory compliance and supervisory oversight technologies are of paramount importance to maintaining the efficiency and transparency of financial results, and the impact of this series of actions and developments is reflected in upcoming annual reports.

"Regulatory and supervisory technologies are developing in response to various demand and supply drivers. On the demand side, regulatory pressure and budget limitations are pushing the market towards an increased use of automated software to replace human decision-making activities. This trend is reinforced by supply drivers such as increasing computing capacity and improved data architecture." [2]

RegTech leverages available technological tools to streamline and enhance the efficiency of regulatory compliance processes more effectively, including automating compliance reporting, leveraging predictive analytics for risk management, implementing machine learning algorithms for fraud detection, and improving anti-money laundering and customer knowledge procedures.

SupTech, on the other hand, empowers regulatory and supervisory bodies with these advanced technological tools to supervise and monitor financial institutions and markets, including real-time market monitoring, behavioral analytics, efficient data collection and analysis, and predictive risk assessment.[3]

"Imagine a financial world where compliance with complex regulations is streamlined, efficient and automated. Where financial institutions effectively manage risks, detect and prevent financial crimes and report accurate data to regulatory authorities, all while reducing costs and increasing operational effectiveness. This is the reality of Regtech and Suptech, the cutting-edge technologies transforming the compliance landscape in the UK." [4]

1.2 Limitations

Despite the existence of basic software tools for data collection and management that typically provide essential support for regulatory and supervisory functions through, as well as traditional IT systems for compliance and risk monitoring and automation, these traditional tools have become limited in capabilities and thus outputs, and often consume more time and incur higher costs, with a significant risk of human error. Therefore, producing accurate and high-quality reports or report summaries, with the above-mentioned capabilities, has become very expensive.

1.3 Project Objectives

The integration of artificial intelligence into these technologies greatly enhances their capabilities and eliminates the limitations of their capabilities, as artificial intelligence allows for more accurate analysis of big data, predictive analytics and real-time monitoring, which further improves the efficiency and effectiveness of financial institutions in preparing reports and their summaries, especially annual reports, which are relied upon for several reasons, including governance compliance and attracting investors.

"Stakeholders – from regulators and investors to customers and the public – are putting your ESG metrics and disclosures under increasing scrutiny. They want access to credible, verifiable and comparable ESG metrics so that they can make decisions on areas that matter most to them. Meeting their expectations can mean going beyond mandatory requirements." [5]

My project aims to create a flexible tool that provides a way to obtain a comprehensive, clear and understandable summary of financial reports on the one hand, and on the other hand a tool that plays a role in mitigating the fundamental challenges in creating report summaries, specifically eliminating inefficiencies and high error rates that may occur due to traditional manual processes carried out by humans.

1.4 Thesis Overview

By developing a tool that generates comprehensive reports using environmental, social and governance (ESG) aspects, these important aspects that every financial and economic entity looking for as a main target, the project also seeks to facilitate more effective oversight and ensure compliance with known advanced regulatory standards, both current and future.

At the end of this project, we aim to create a tool that harnesses the power of Natural Language Processing (NLP) "using supervised machine learning: to help classify texts, or

unsupervised learning algorithms, to help cluster similar documents or topics" or both (depending on the chosen algorithms and expected outputs), to analyze and summarize ESG report disclosures, with the aim of providing supervisory authorities and stakeholders with clear and actionable insights. The aim is to improve the accuracy, efficiency and effectiveness of ESG reporting summarization methods, thus contributing to a more transparent and reliable financial system supported by RegTech and SupTech.

2. Literature Review

2.1 Overview of ESG report

Environmental, Social, and Governance (ESG) reporting has become key for corporate transparency and liability. And helps in exploring the ESG reporting the use of Regulatory Technology (RegTech), Supervisory Technology (SupTech), and Natural Language Processing (NLP) to enhance the analysis and compliance of these reports.

"ESG reporting is the new frontier for enterprise disclosures, becoming another avenue through which investors and stakeholders make critical decisions. Over 90% of the S&P 500 already recognize the importance of environmental, social and governance matters, demonstrated by their ESG reporting practices already in place." [3]

2.2 Role of Artificial Intelligence in Regulatory Compliance

NLP technologies are essential for processing and analyzing the unstructured text data found in ESG reports. Tools like SpaCy, NLTK, allow for comprehensive text analysis, including tokenization, named entity recognition, and sentiment analysis. By leveraging NLP, we can extract and categorize information from ESG reports, revealing insights into corporate behaviors and compliance. Although NLP is widely used in various text analysis applications, its specific application to ESG reporting needs further exploration.

"Artificial Intelligence (AI) and Machine Learning (ML) have emerged as game-changing technologies that are reshaping various industries, including finance. These technologies have particularly revolutionized the field of regulatory compliance, introducing unprecedented efficiency, accuracy, and predictive capabilities." [6]

"Compliance officers can boost efficiency with **AI tools**, automating routine tasks like data analysis. This allows them to focus on complex compliance issues. AI-driven analytics offer risk insights, improving report accuracy and reducing time/resource requirements. Furthermore, it enhances transparency for regulators and aids employee compliance training through interactive programs. [7]

2.3 Natural Language Processing (NLP) in Financial Analysis

Combining RegTech, SupTech, and NLP offers a transformative approach to ESG reporting and monitoring. This integration can automate data collection, enhance compliance checks, and provide real-time insights into ESG performance [8]. The project aims to develop a framework that leverages these technologies to address the complexities of ESG disclosures, offering a more robust and transparent analysis of corporate sustainability practices.

"Financial institutions deal with vast amounts of unstructured data daily, from news articles and reports to social media feeds and regulatory filings. Manually processing this data is time-consuming and prone to errors. NLP automates this process, enabling institutions to quickly and accurately sift through data, identify relevant information, and generate actionable insights.

One significant application is in the automation of routine document processing tasks. For example, legal and financial documents can be reviewed and analyzed using NLP algorithms, extracting key information and identifying any potential issues. This not only saves time but also reduces the risk of human error." [9]

2.4 Summarize Existing Research

Nowadays, despite highlighting the potential of integrating RegTech, SupTech, with NLP to improve ESG reporting. While significant advancements have been made, gaps remain in effectively applying these technologies to address the qualitative aspects of ESG disclosures. The proposed project seeks to bridge these gaps by creating an AI-driven framework that automates and enhances the analysis of ESG reports, promoting a more sustainable and transparent financial ecosystem.

Despite advancements, there remain challenges in capturing the qualitative Analysis of ESG disclosures. Current methodologies often focus on quantitative metrics, leaving gaps in understanding the broader context and implications of ESG practices. [10]

The project mentioned below aims to bridge these gaps by developing an AI-driven framework that automates and enhances the analysis of ESG reports. This framework will leverage advanced NLP techniques to extract meaningful insights, identify key themes, and analyze sentiment but uses a tool can harm the performance as Handling very large datasets or multiple users simultaneously might impact its performance and require technical

expertise to fully customize it for specific use cases, moreover is a Java-based application, it requires a Java environment, which might not be ideal for all users. [11]

The exist studies are providing a thorough examination of how NLP can transform ESG reporting, while also addressing existing challenges and proposing avenues for future research, the studies focused on particular area which may limit the generalizability of the results to other industries or languages, and temporal bias**.** [12]

Moreover, the Small and Medium-sized Enterprises SMEs struggle with data availability, quality, and standardization across various sources. Addressing these challenges can enhance their sustainability practices and boost their corporate reputation. [13]

However, the research also highlights the challenges in integrating these technologies, such as ensuring data security, addressing algorithmic bias, and maintaining human oversight, and most of previous studies focused on streamline of the ESG reporting process there is a need for more research and practical applications to fully understand and harness the synergies between these technologies. [14]

While the potential benefits of integrating RegTech, SupTech, and NLP for ESG report summarization have been recognized, there is a need for more in-depth research and practical applications to fully understand and harness the synergies between these technologies.

NLP technologies enable more efficient and accurate assessments, including a variety of modules and components that work together to enable computers to understand, interpret, and generate human language, these models, driven by advancements in deep learning and natural language processing (NLP).

2.5 Case Study

2.5.1 Background and Context

“ADIB is a leading bank in the UAE and the 4th largest Islamic bank globally by assets. Headquartered and listed in Abu Dhabi, ADIB was incorporated in 1997 to serve as the first Islamic bank in the Emirate of Abu Dhabi.

Over the last two decades, ADIB has demonstrated a consistent track record of growth with assets now totaling USD 39 billion. The bank currently serves more than 1 million customers through a balanced proposition that combines a highly personalized customer experience with world-class digital banking services”. [15]

2.5.2 ESG Reporting Practices:

ADIB's annual reports detail the bank's ongoing efforts to address social challenges and build more sustainable, diverse, and equal societies, as well as the criteria that the bank adopts to build these reports. Accordingly, with these specifications, the bank was chosen as a case study in this project, and I chose its annual reports to be a source of data to bring these annual reports, which will be sufficient to obtain understandable and scalable outputs, as well as to introduce artificial intelligence technologies.

"ADIB's Environmental, Social and Governance Risk Management (ESG Risk Management Policy or the "Policy") is designed to integrate ESG risks, emanating from credit customers and investees, within its Enterprise Risk Management framework to ensure ADIB Group remains within the risk appetite whilst being fully aligned with ADIB Group's ESG strategy and

targets and the strategic business objectives of ADIB Group. The Policy is aligned with the relevant UAE laws and regulations, Shari'a principles (as determined by ADIB's Internal Shari'a Supervisory Committee and CBUAE Higher Shari'ah Authority), as well as international guidelines, standards and industry best practices such as Equator Principles, International Finance Corporation (IFC), Basel Committee on Banking Supervision (BCBS) and European Bank for Reconstruction and Development (EBRD)." [16]

3. Methodology

The introduction of methodology outlines a systematic approach to integrating NLP with RegTech and SupTech for enhancing ESG reporting analysis. It involves stages like data collection, processing, and analysis to create a framework that automates the extraction and evaluation of ESG information. The goal is to address the complexities of ESG disclosures, providing insights into corporate sustainability and regulatory compliance through a combination of quantitative and qualitative techniques. Figure(1)

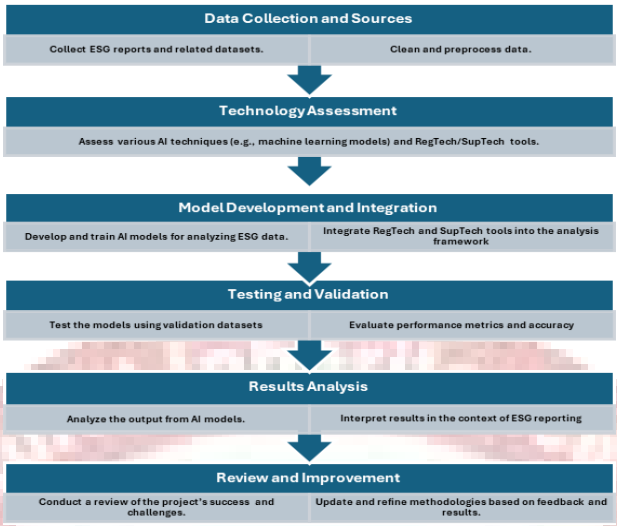


Figure (1): Methodology streamline

"In the data structure process, data mining and process mining technologies could find anomalies, associative pat-

terns, and outliers in collected data. Using sentiment and semantic analysis, NLP techniques could identify meaningful

patterns and new insights in textual data. Anomalies could indicate irregularities in environmental impact, social performance, or governance practices. For example, data mining can flag unusually high emissions levels for a specific company or industry, indicating a potential environmental concern ". [17]

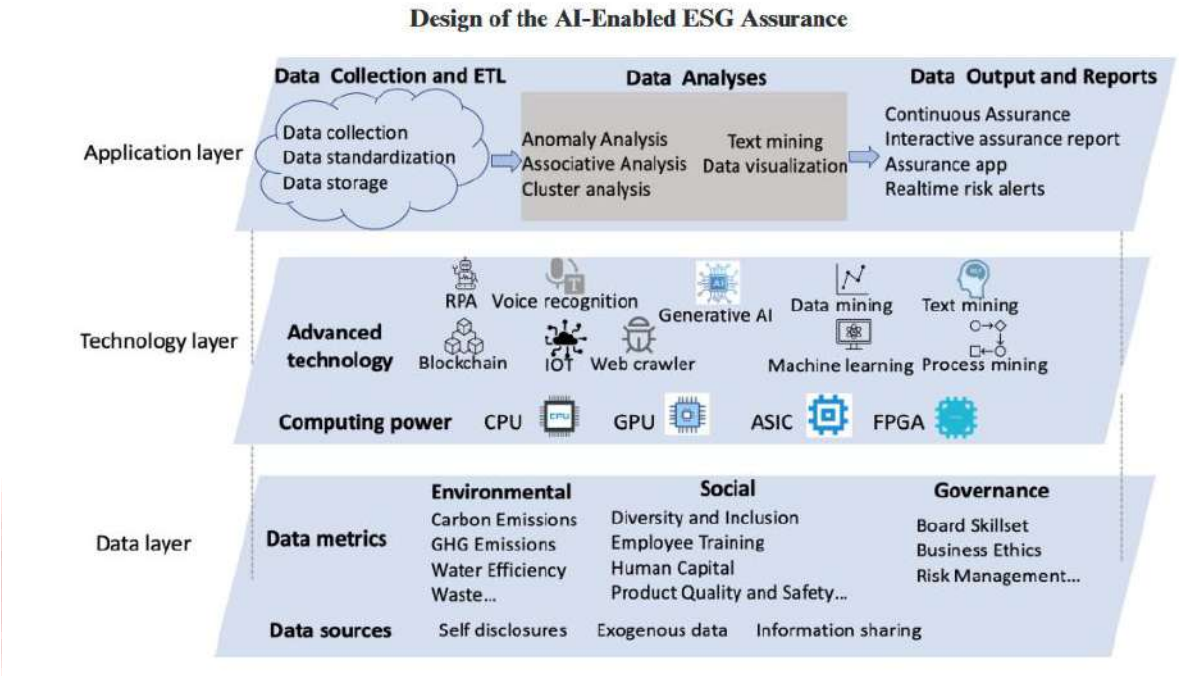
3.1. Data Collection and Sources

I carefully selected the data, ensuring that even if it wasn't fully organized or prepared, it was still suitable for my project in terms of outputs and its structural nature, including texts, tables, and other formats. The result is a comprehensive, well-prepared dataset ready for analysis using AI models and technology solutions, ensuring the data is reliable, relevant, and aligned with the project's objectives.

As a financial institution, So I chose ADIB Bank "Abu Dhabi Islamic Bank" as an example, "the leading Islamic financial institution and continued to lead in sustainable finance for these last few years", as it has reports spanning several years, making it easier to compare and examine the NLP module we plan to use. [18]

The ADIB ESG reports are : 2021 [19], 2022 [20], and 2023 [21], which we can compare between them in streamline manner that can help identify trends, improvements, or inconsistencies over the years.

Below in **Figure (2)** show how the AI Enabled ESG assured output with different levels of data processing:



3.2.1. Google Pegasus Model:

Role: Google Pegasus is employed for its advanced text summarization capabilities. Its design focuses on generating high-quality summaries of lengthy and complex documents, making it particularly suitable for condensing and analyzing the extensive and detailed content of ESG reports.

Google Pegasus "Pre-training with Extracted Gap-sentences for Abstractive Summarization" [12] was chosen for its advanced capabilities **Figure (3)**, especially in text summarization, unlike traditional models, Pegasus is specifically optimized for producing high-quality summaries of lengthy documents, making it ideal for extracting and condensing critical insights from extensive ESG reports. Its pre-training on large datasets and subsequent fine-tuning for summarization tasks ensure that it can handle the nuanced and detailed language found in financial and governance texts.

Pegasus Model offers superior performance in generating coherent and contextually accurate summaries, which is essential for analyzing complex ESG reports.

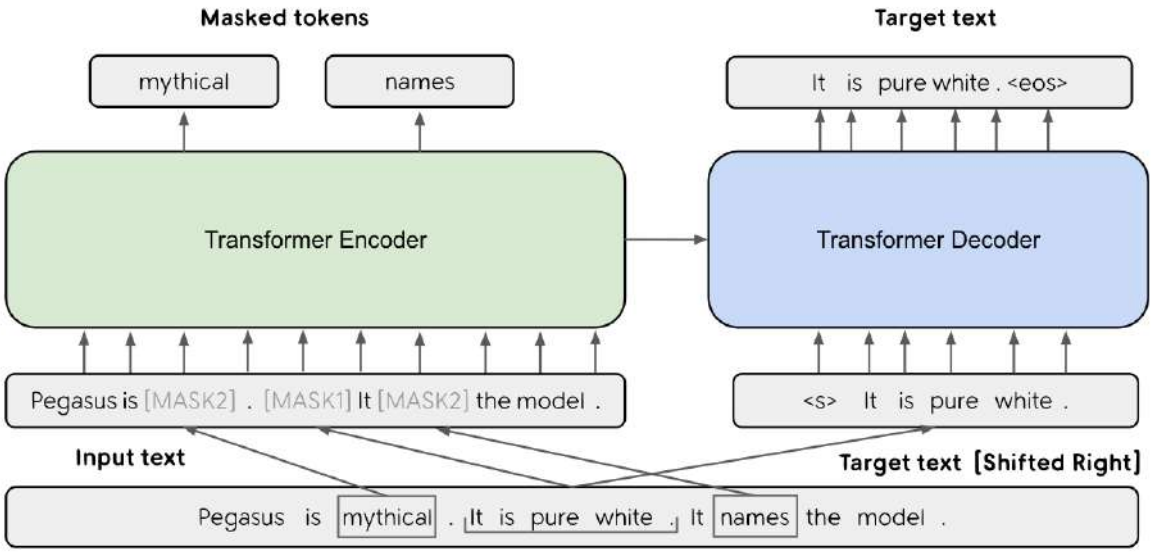


Figure (3): The base architecture of PEGASUS [22]

Strengths: Pegasus excels in producing coherent and contextually relevant summaries, which is crucial for extracting actionable insights from ESG reports. Its pre-training on large datasets enhances its ability to handle diverse and complex text structures.

Weaknesses: The model may require additional fine-tuning to adapt to the specific terminology and nuances of ESG language, which can be a limitation if not addressed.

3.2.2. Python:

Python is the primary programming language for implementing and managing an analysis pipeline. Its versatility and comprehensive support for data science and natural language processing tasks makes it an ideal choice for developing and implementing project algorithms.

Python's readability and comprehensive libraries that can be called and used from within the Python environment provide support, simplify the development process, and facilitate efficient data processing and analysis, that's why was my option.

Including Python tools and libraries such as "Panda, torch, matplotlib, nltk, PyMuPDF" collectively provide a robust framework for analyzing ESG reports, each contributing its strengths to ensure a comprehensive and effective analysis process and understandable output.

3.3. NLP Techniques for Report Analysis:

Analyzing the ESG report is essential for extracting meaningful understanding from large volumes of text by Natural Language Processing (NLP) techniques. These techniques enable the automatic processing and interpretation of unstructured dataset, such as the textual content of ESG reports that we are focused on it, to identify key themes, summarize information, and identify areas for improvement. Below are steps of the primary NLP techniques utilized for report analysis.

3.3.1 Data (text) preprocessing

"Before a model processes text for a specific task, the text often needs to be preprocessed to improve model performance or to turn words and characters into a format the model can understand". [23]. Below we will find the processes and tasks that we need and will use in my project:

Named entity recognition "NER": is an NLP technique to classify and recognize entities, that identifies and classifies named entities into predefined categories such as the individual names, organizations, locations. quantities, money. **Figure (1)**

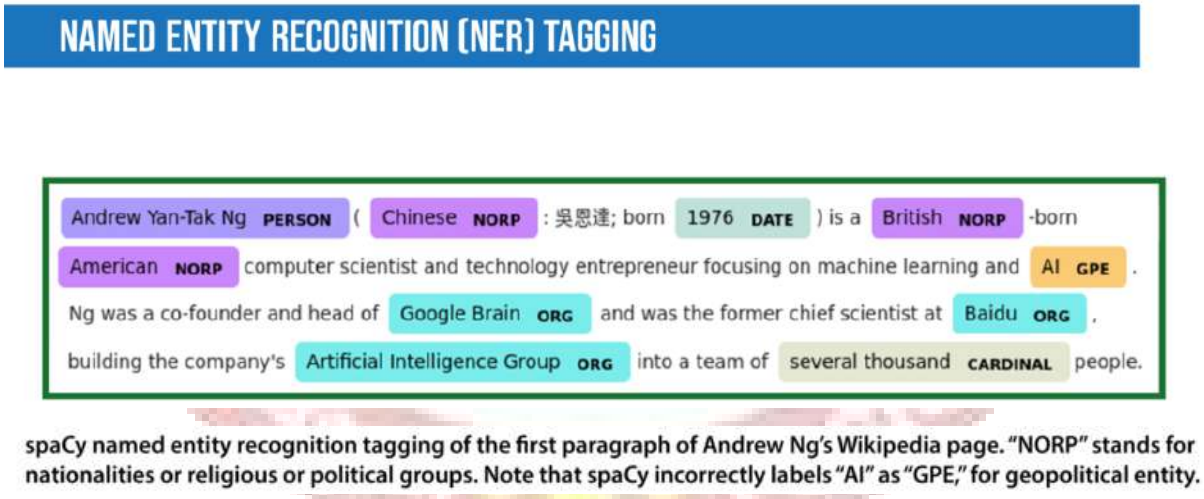


Figure (1) [24]

Topic Modeling: is an unsupervised NLP technique working by co-occurrence of common words, keyword or even phrases in the text to be extract and identify the main topics from whole document, "Most topic modeling approaches begin by generating a document-term

matrix. This matrix models the text dataset with documents as rows and individual words as columns, or vice-versa." [25]

Stemming: is a text processing, converts words to their original form using implication rules, where reduce and cut off prefixes and suffixes the additional form of the word to convert it its origin where we cannot process it again. The output of this technique is not always correct or has a meaning. [26]

Lemmatization: Is a more sophisticated technique reducing the word to root form with always considering the meaning of the word

(e.g.: ""better" → "good"), mean the output should always has a meaning. The popular lemmatization tool in python and I use it in my code in NLTK (Natural Language Toolkit). [27]

Stop word removal: Is preprocessing technique that filtering the text from common stop word, such as "articles, prepositions, pronouns and auxiliar verbs" (e.g. "is", "the," "are"), that do not significant much meaning in the text. As in ESG report we don't not give interest to this kind of stop word so this technique really will be efficacy and reduce time consumption in text processing.

Tokenization: Is one phase of Lemmatization technique, where breaking down the text into separate: word, phrase or sentence, this output called individual tokens, to get a better understating and enhance the search and text retrieval method which is consider as beneficial for NLP modules **Figure (2)**. [28]

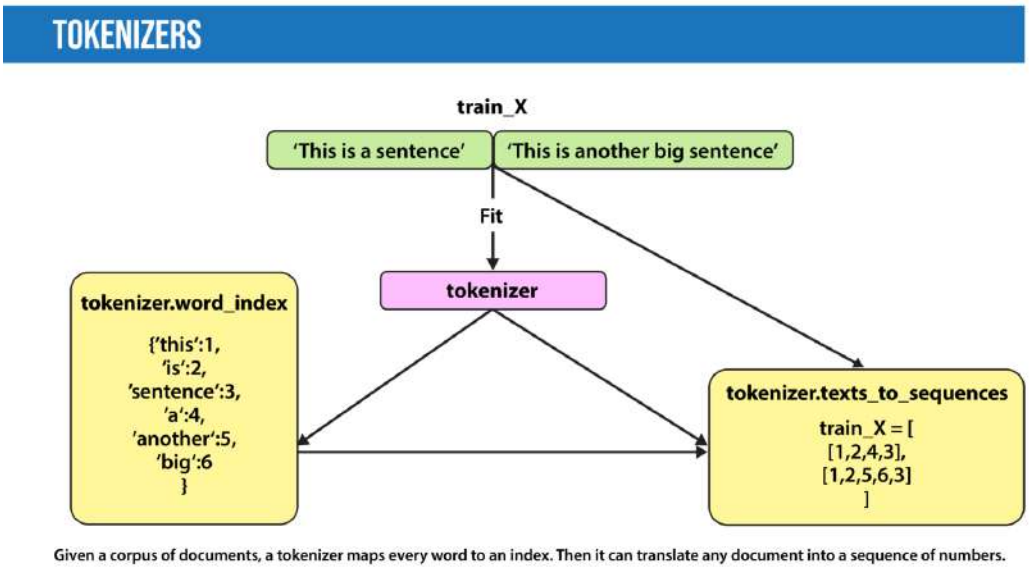


Figure (2) [24]

Latent Dirichlet Allocation (LDA): is one example of topic modeling technique base in clustering method, "LDA is a Bayesian network, meaning it's a generative statistical model that assumes documents are made up of words that aid in determining the topics". [29]

3.3.2 Algorithms for Data Analysis:

Phase 1: My code starts by uploading the ESG reports (**ADIB annual ESG reports**), where string by preparing the text environment using upload the pdf report file that contains around hundred pages and initialize the files output storage, including error handling. Later convert it to "txt" format to extract a clean text that allow me to analyze, identify and categorize the text content based on ESG defined categories.

Phase 2: The function in the code uses the "**fitz**" library to open the read the PDF files, as well as "**PyPDF2"** library helping in splitting and extracting information from PDF files.
The idea here is to catch the table content on the pdf file (I make sure that will be in the first five "5" pages), here we can split this section from the file into headlines and stored in new output text file in form of section groups, with attempts to remove duplicates words.
In this phase the code starts to create a dictionary category a list of predefined keywords associated with the three related to ESG categories "Environment, Social and Governance", which are considered as main ESG topics.

Phase 3: now the role of function named by (**append_content_to_titles)**, start to identify and categorize these last sections I mentioned in (**Phase2)** based on titles, this will be useful for analyzing the document with structured sections.

Phase 4: Install the necessary libraries such as (**PyTorch, Transformers**), then staring to wrapping the text is split into chunks base on given maximum length (**max_length = 1024**), with some overlap "**overlap = 100**" between chunks to ensure the continuity of the context, then save the return results.

Phase 5: Here starting by Install the required libraries (**sum_google_peg**) that uses the Pegasus model to generate a summary of the text file based on the list of topics extracted from the document, means the code here prepares the uses Pegasus model to summarize text.

Moreover, here start to tokenize the topics and feeds them into the model, which generates a summary of the ESG reports "or even similar other documents", includes flexibility to adjust the length output summary.

Phase 6: I developed a summary report that concerned with displaying text classifications from ESG documents using different NLP techniques and libraries for text and visual representation such as reportlab, matplotlib, numpy, pillow, and wordcloud. This is similar to "general" classification because it starts from the search and then the classification is based on the three main and important keywords "Environmental, Social, and Governance" and their derivatives as well. This task goes through several processes, the most important of which are:

Text processing: It is divided into lines, then the text is processed, summarized, and classified to facilitate visual representation.

Visualization: Column charts and word cloud technology are created for the above-mentioned words and their derivatives, to visually represent the results of the analysis.

Phase 7: The script here gathers several functions to re-summarize a text document using Google Peagasus Model, So mainly here is re-summarizing the input file and generate a summary and save it to a new output file.

The technique here is splitting the text into lines to be manageable chunks. This approach will help for summarizing a huge document by breaking it down into smaller parts to make it more manageable sections.

Phase 8: Instead of uploading the output file into third party World Cloud tool, I used here a world_cloud function, in order to bring up a visual representation of the most frequents words exist in output files and save this as PNG Image and returned it to save it.

At this advanced stage, after analyzing the data and conclusions of the ESG Report Analysis Project, it can be confirmed that the applied algorithms were effective in extracting useful insights from the reports. Natural Language Processing (NLP) techniques were proven to be able to identify key entities, themes and sentiment trends across different reports, which

helped in better understanding ESG practices in the financial sector, The results may also reveal weaknesses or gaps in reporting, indicating opportunities for improvement.

The results may also reveal weaknesses or gaps in reporting, indicating opportunities for improvement and development.

4.Data Analysis and Findings

At this stage I would Presenting the key insights gained from applying various algorithms to the ESG reports.

4.1 Entity extraction and topic identification

Key entities: Named Entity Recognition (N.E.R), or some similar tool, assists in primary identification of certain important entities like company titles, geographical locations **Figure (1)**, and other terms associated with ESG analysis thereby preparing us on which areas to concentrate in our reports and also in creating a database that assists the tool in comprehending the text and recognizing entities in less time and with more ease. **Figure (2)**

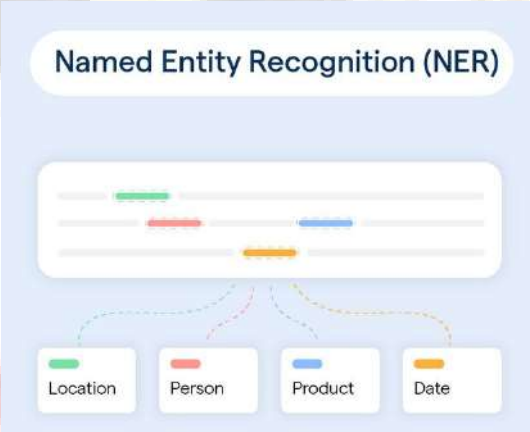


Figure (1) [30]

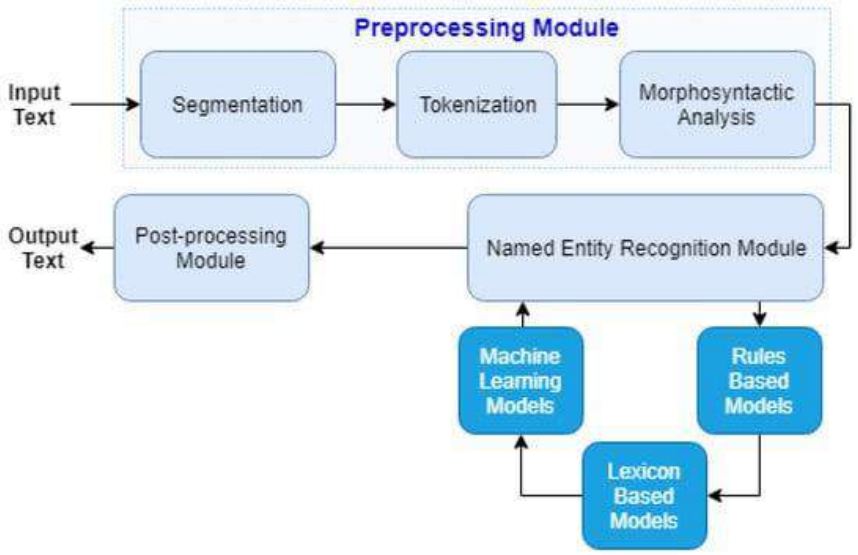


Figure (2) [31]

Topic modeling: This tool assists me in conducting "Latent Dirichlet Allocation" LDA (**Figure (3)**) topic modeling algorithms in more effective ways, mainly I mean discovering the existing topics like environmental sustainability, social activities, and governance or organization management practices which I mean the three words (ESG) and their derivatives. **Figure (4)**

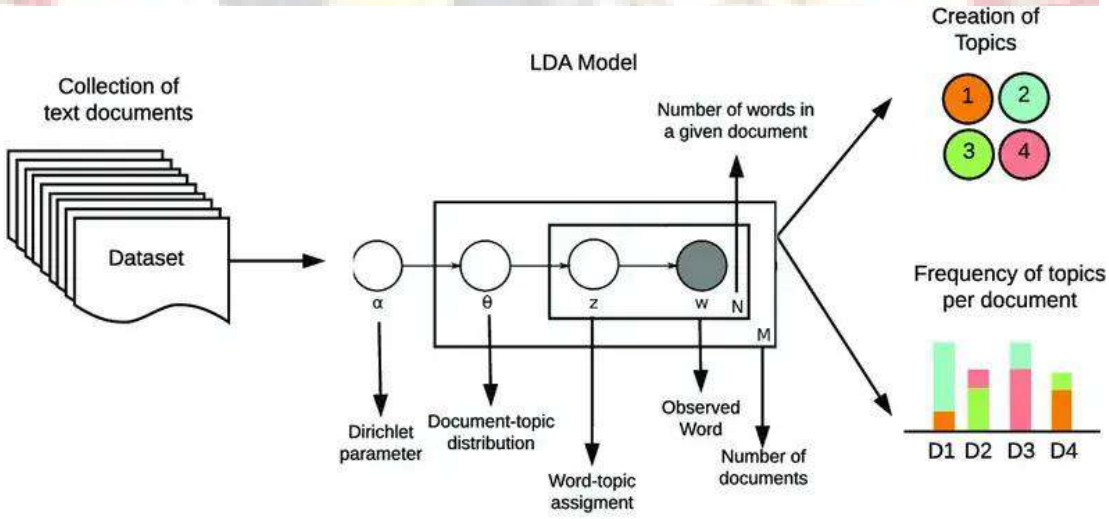


Figure (3) [32]

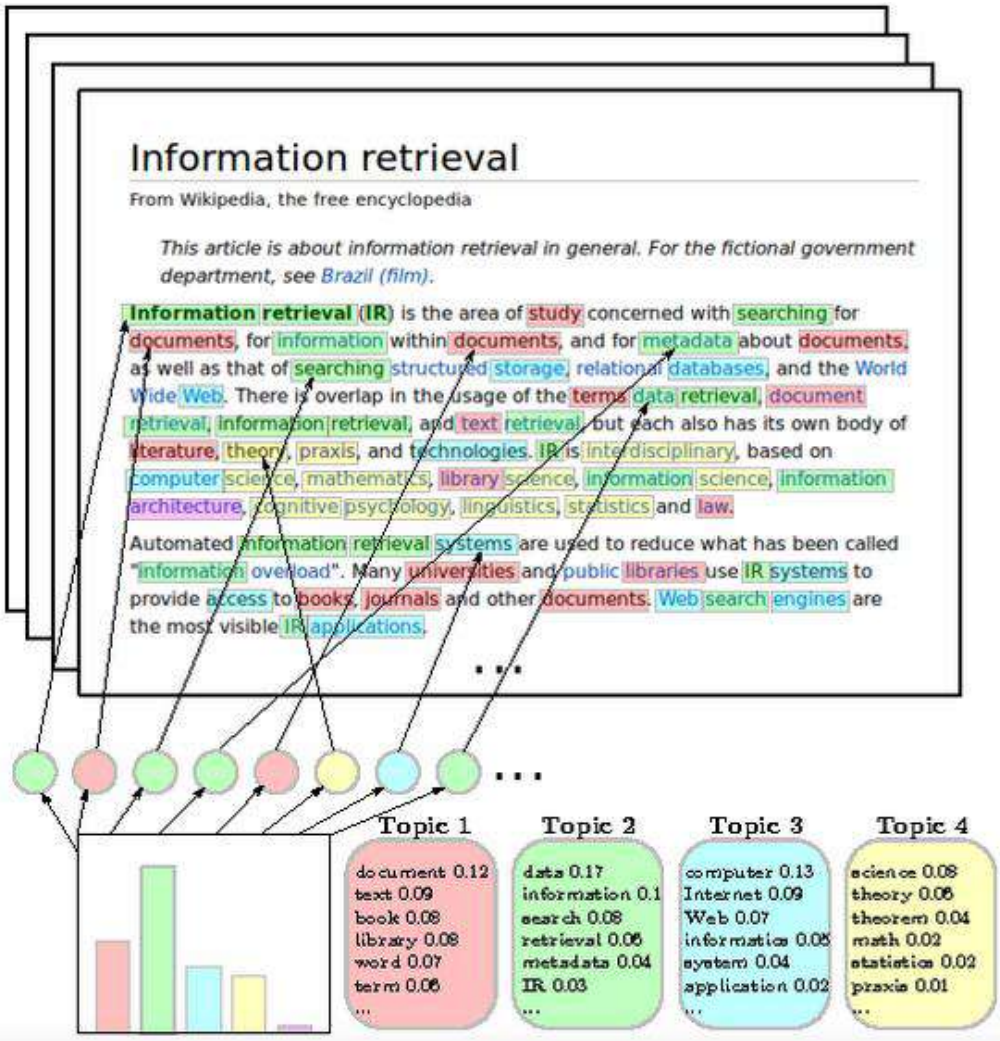


Figure (4) [33]

4.2 Sentiment Analysis

Conducting sentiment analysis for the type of reports targeted in my project is very important in analyzing and summarizing financial and technical reports, with the aim of determining whether they are generally positive, negative, or neutral towards certain aspects, opinion, or a pre-listed goal. For example, it might be said that social responsibility is viewed more positively than environmental sustainability. **Figure (5)**

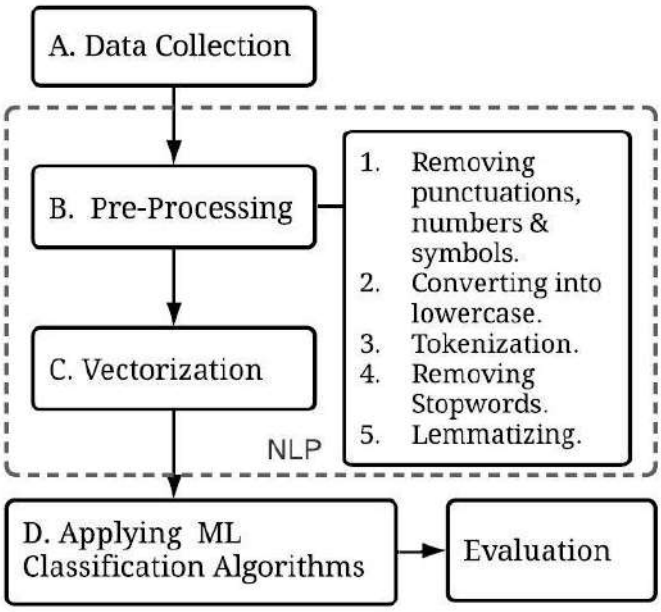


Figure (5) [34]

Aspect-based analysis: A more sophisticated sentiment analysis tool based on opinion can identify specific aspects or features of a product or service that are praised or criticized in a report.

4.3 Text Classification

Text classification, which divides text into predetermined categories of unstructured text data, is a basic component of natural language processing (NLP). The data is more organized and structured as a result of this process, which additionally makes it easier to analyze and extract insights and prepares it for further in-depth study. Putting text or reports into "core," "social," or "promotions" categories is a common use case.

A variety of algorithms are used, including **transformer** along with different deep learning models. **Figure (6)**

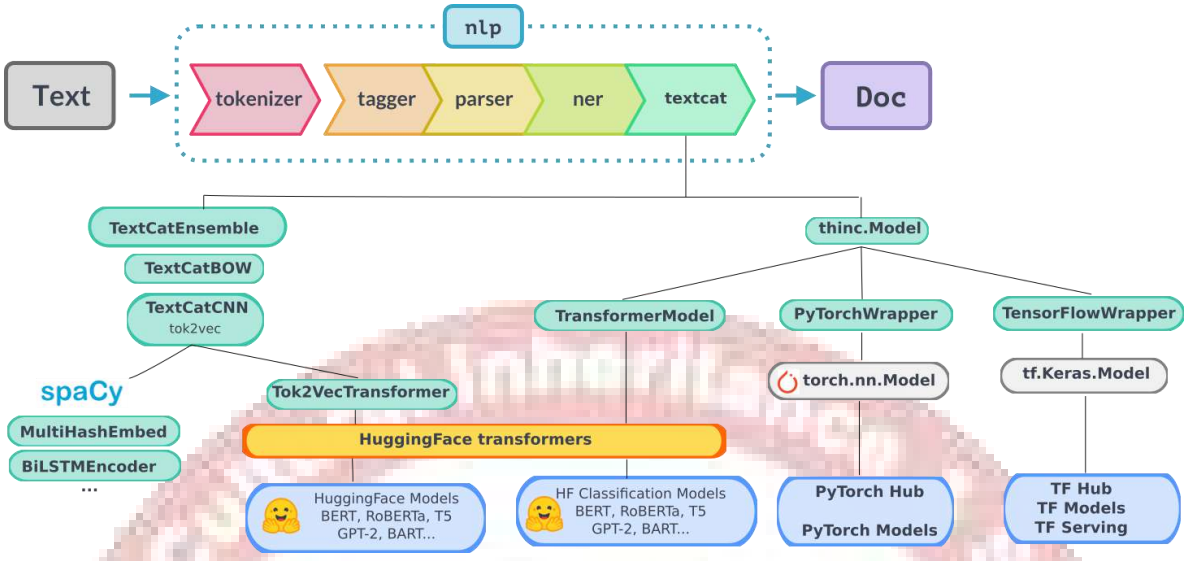


Figure (6) [35]

4.4 ESG Reports of Abu Dhabi Islamic Bank (ADIB)

"ESG *report* contains *statements* that may be deemed as. "forward-looking *statements*" that express how ADIB intends to conduct its activities.

This report has been prepared for key stakeholders, including employees, suppliers, customers, partners, and communities, to provide an in-depth overview of our performance on material ESG issues". [21]

4.5 Suggestions

The recommendations I can make based on the findings and outcomes are how well the tool works to, first, help improve the standard and caliber of ESG reporting and its summaries, which is essential, secondly, suggest areas where businesses should concentrate, where more work is required, or where the sector could benefit from standardized reporting guidelines. These results will offer an in-depth overview of the current state of ESG reporting and summary in the financial industry and serve as a foundation for further improvements in the future.

5. Conclusion

In my project I developed a tool that employs sophisticated NLP methodologies, or rather the model's functionalities such as text categorization, emotion detection, and summation to

interpret ESG documents from the finance industry and provide a concise clear cut summary using NLP Model specifically Google Pegasus summarization, NER to extract the most relevant key players, and topic modeling to find mostly used topics.

My project offered an overall evaluation of how companies are dealing with ESG issues. It demonstrated the benefits of such tools and areas and the markets responsible for enhancing ESG reports and summaries reporting, thereby aiding professionally sensible and timely sustainability reports/ disclosures from the companies.

These results also allow for better decision making for all stakeholders and increase the knowledge and comprehension of ESG in the financial industry.

This project is a chance for expanding and enhancing the products and products working on ESG reporting through the processes of NLP reporting.

My project has given the chance for other researchers to research and carry out studies to improve the products and outputs, and to employ other methods that can be applicable to meet future needs.

6. Future Directions

6.1 User Feedback: This point is not included and is also considered blind. Feeding back occupant or stakeholder's comments of any aspect further recommendations or proposals for future research.

6.2 Incorporate Visual Analytics: Adding Interactive tools like dashboards can be provided to the stakeholders to make reasonable interpretations of the analysis without much strain.

6.3 Expand Comparison: Conduct a thorough comparison of the ESG reports of different industries or regions for better understanding.

6. APPENDIX:

1.Githlab repository:

The links below shows the code in the Githlab repository:

2.GitLab project repository:

https://git.cs.bham.ac.uk/nxh263/myprpoject/-/blob/main/document_summarizer.ipynb

3.Code Structure:

The code was developed in Python and run on the Google Colab platform. My software takes around 5 to 10 minutes for execution time to create the final output as a PDF file named report.pdf.

4.Execution Steps:

4.1 Download any pdf file report we want to summarize, below the link from its source:

adib-esg-report-2023.pdf

Adib-esg-report-2022_en.pdf

adib-esg-report-2021_en.pdf

4.2 Run the code using the GitLab URL.

4.3 From runtime in the navbar, press run all.

4.4 During execution may find:

1. Runtime Disconnected: Press reconnect and press running all

2. Cannot connect to GPU backend: press connect without GPU and press running all

4.5 After you press run all, you should upload one of the pdf files mentioned above. After the file got uploaded, we can see it in the Table of Contents part in Colab in the link (/content/drive_2/MyDrive/DocumentSummarizer) we can see the text files start to appear. The first file is the report converted file from pdf to txt. (all_text_path.txt), and then successively we can the text files, (In_between_output) that contain the separate paragraphs. After this point, the software will start to summarize this text using the model used in the code and create a new file for this name (adib_summary). All these files we can see them in the Table of Contents part in colab left side.

4.6 After finishing the execution, the report.pdf file will be created at the same location in Table of contents part in colab.

7. References

[1] "ESG Reporting: A Complete Guide," AuditBoard. Accessed: Sep. 01, 2024. [Online]. Available: https://www.auditboard.com/blog/esg-reporting

[2] "trv_2019_1-regtech_and_suptech_change_for_markets_and_authorities.pdf." Accessed: Sep. 02, 2024. [Online]. Available: https://www.esma.europa.eu/sites/default/files/trv_2019_1-regtech_and_suptech_change_for_markets_and_authorities.pdf

[3] "RegTech, FinTech, SupTech: Key Differences - Sanction Scanner." Accessed: Sep. 02, 2024. [Online]. Available: https://www.sanctionscanner.com/blog/whats-the-difference-between-regtech-fintech-and-suptech-519

[4] SEO, "What Is Suptech?," FinregE. Accessed: Sep. 02, 2024. [Online]. Available: https://finreg-e.com/what-is-suptech/

[5] "ESG reporting," KPMG. Accessed: Sep. 02, 2024. [Online]. Available: https://kpmg.com/dp/en/home/industries/environmental-social-governance/esg-reporting.html

[6] "AI & Machine Learning for Regulatory Compliance," Akkio. Accessed: Sep. 02, 2024. [Online]. Available: https://www.akkio.com/post/compliance-artificial-intelligence

[7] "Challenges of AI Adoption in Compliance | MEGA." Accessed: Sep. 02, 2024. [Online]. Available: https://www.mega.com/blog/how-artificial-intelligence-can-be-used-compliance

[8] "Harnessing NLP for Effective ESG Reporting: Insights for SMEs." Accessed: Sep. 02, 2024. [Online]. Available: https://positivethinking.tech/insights/nlp-esg-reporting-smes/

[9] Finextra, "Why Finance is Deploying Natural Language Processing," Finextra Research. Accessed: Sep. 02, 2024. [Online]. Available: https://www.finextra.com/blogposting/26254/why-finance-is-deploying-natural-language-processing

[10] "How to Use Data and AI to Measure ESG Performance." Accessed: Sep. 02, 2024. [Online]. Available: https://roundtable.datascience.salon/how-to-use-data-and-ai-to-measure-esg-performance

[11] "file.pdf." Accessed: Sep. 02, 2024. [Online]. Available: https://pergamos.lib.uoa.gr/uoa/dl/object/3363684/file.pdf

[12] C. P. Soares, "Leveraging Natural Language and Item Response Theory Models for ESG Scoring," Jul. 29, 2024, *arXiv*: arXiv:2407.20377. Accessed: Sep. 02, 2024. [Online]. Available: http://arxiv.org/abs/2407.20377

[13] "Sustainability | Free Full-Text | Natural Language Processing Methods for Scoring Sustainability Reports—A Study of Nordic Listed Companies." Accessed: Sep. 02, 2024. [Online]. Available: https://www.mdpi.com/2071-1050/14/15/9165

[14] S. Sharma, "Automate and Streamline the ESG Reporting Process." Accessed: Sep. 02, 2024. [Online]. Available: https://www.lythouse.com/blog/automate-and-streamline-the-esg-reporting-process

[15] "ADIB | ADIB Vision and Mission | Abu Dhabi Islamic Bank." Accessed: Sep. 02, 2024. [Online]. Available: https://www.adib.ae/mission-objectives

[16] "esg_rsk_policy_r05.pdf." Accessed: Sep. 02, 2024. [Online]. Available: https://www.adib.ae/en/siteassets/sustainability/esg_rsk_policy_r05.pdf

[17] "Start with Silverchair & Go Anywhere," Silverchair. Accessed: Sep. 02, 2024. [Online]. Available: https://www.silverchair.com/

[18] "ADIB Releases Its Annual Environmental, Social And Governance Report And Sets Its Net Zero Reduction Target For 2030 | Go Green - A Cyber Gear initiative for living a green lifestyle." Accessed: Sep. 02, 2024. [Online]. Available: https://www.go-green.ae/adib-releases-its-annual-environmental-social-and-governance-report-and-sets-its-net-zero-reduction-target-for-2030/

[19] "adib-esg-report-2021_en.pdf." Accessed: Sep. 02, 2024. [Online]. Available: https://www.adib.ae/en/siteassets/investor-relations/esg-reports/adib-esg-report-2021_en.pdf

[20] "adib-esg-report-2022_en.pdf." Accessed: Sep. 02, 2024. [Online]. Available: https://www.adib.ae/en/siteassets/investor-relations/esg-reports/adib-esg-report-2022_en.pdf

[21] "adib-esg-report-2023.pdf." Accessed: Sep. 02, 2024. [Online]. Available: https://www.adib.ae/en/siteassets/investor-relations/esg-reports/adib-esg-report-2023.pdf

[22] J. Zhang, Y. Zhao, M. Saleh, and P. J. Liu, "PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization," Jul. 10, 2020, *arXiv*: arXiv:1912.08777. Accessed: Sep. 02, 2024. [Online]. Available: http://arxiv.org/abs/1912.08777

[23] "Natural Language Processing (NLP) [A Complete Guide]." Accessed: Sep. 02, 2024. [Online]. Available: https://www.deeplearning.ai/resources/natural-language-processing/

[24] "Natural Language Processing (NLP) - A Complete Guide." Accessed: Sep. 02, 2024. [Online]. Available: https://www.deeplearning.ai/resources/natural-language-processing/

[25] "What is topic modeling? | IBM." Accessed: Sep. 02, 2024. [Online]. Available: https://www.ibm.com/topics/topic-modeling

[26] "Stemming in NLP- A Beginner's Guide to NLP Mastery," ProjectPro. Accessed: Sep. 02, 2024. [Online]. Available: https://www.projectpro.io/article/stemming-in-nlp/780

[27] "What is Lemmatization? | Definition from TechTarget," Enterprise AI. Accessed: Sep. 02, 2024. [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/lemmatization

[28] "What is Tokenization in NLP & Machine Learning? | Deepgram." Accessed: Sep. 02, 2024. [Online]. Available: https://deepgram.com/ai-glossary/tokenization

[29] "What is Topic Modeling? An Introduction With Examples." Accessed: Sep. 02, 2024. [Online]. Available: https://www.datacamp.com/tutorial/what-is-topic-modeling

[30] "Named Entity Recognition: Types & Challenges | BotPenguin." Accessed: Sep. 02, 2024. [Online]. Available: https://botpenguin.com/glossary/named-entity-recognition, https://botpenguin.com/glossary/named-entity-recognition

[31] "Applied Sciences | Free Full-Text | Named Entity Recognition for Sensitive Data Discovery in Portuguese." Accessed: Sep. 02, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/10/7/2303

[32] "Schematic-of-LDA-algorithm.png (PNG Image, 850 × 373 pixels)." Accessed: Sep. 02, 2024. [Online]. Available: https://www.researchgate.net/profile/Diego-Buenano-Fernandez/publication/339368709/figure/fig1/AS:860489982689280@1582168207260/Schematic-of-LDA-algorithm.png

[33] I. Vulic, W. Smet, J. Tang, and M.-F. Moens, "Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications," *Inf. Process. Manag.*, 2015, Accessed: Sep. 02, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Probabilistic-topic-modeling-in-multilingual-An-of-Vulic-Smet/a2d1bed6ada7cc7bb255016d12c68d5cc1146f3e

[34] "Figure 1. Workflow of Sentiment Analysis using NLP and Machine Learning," ResearchGate. Accessed: Sep. 02, 2024. [Online]. Available: https://www.researchgate.net/figure/Workflow-of-Sentiment-Analysis-using-NLP-and-Machine-Learning_fig1_357236562

[35] "Building Production-Grade spaCy Text Classification Pipelines for Business Data | Width.ai." Accessed: Sep. 02, 2024. [Online]. Available: https://www.width.ai/post/spacy-text-classification