Project Report

On

ESG Report Analysis using NLP

Submitted by

Harsh Thakkar (IU2041230183)
Jaynil Zala (IU2041230200)
Rikin Zala (IU2041230201)

In the fulfilment for the award of the degree

Of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE & ENGINEERING

INSTITUTE OF TECHNOLOGY AND ENGINEERING

INDUS UNIVERSITY CAMPUS, RANCHARDA, VIA-THALTEJ

AHMEDABAD-382115, GUJARAT, INDIA

WEB: www.indusuni.ac.in

OCTOBER, 2023

Project Report

ON

ESG Report Analysis using NLP

AT

In the partial fulfilment of the requirement

for the degree of

Bachelor of Technology in

Computer Science & Engineering

PREPARED BY

Harsh Thakkar (IU2041230183)

Jaynil Zala (IU2041230200)

Rikin Zala (IU2041230201)

UNDER GUIDANCE OF

Internal Guide
Prof. Darshan Solanki,
Assistant Professor,
Department of CSE, IITE,
Indus University, Ahmedabad

SUBMITTED TO

INSTITUTE OF TECHNOLOGY AND ENGINEERING

INDUS UNIVERSITY CAMPUS, RANCHARDA, VIA-THALTEJ

AHMEDABAD-382115, GUJARAT, INDIA

WEB: www.indusuni.ac.in

OCTOBER, 2023

CANDIDATE'S DECLARATION

We declare that final semester report entitled "ESG Report Analysis using NLP" is our own work conducted under the supervision of the guide Prof. Darshan Solanki.

We further declare that to the best of our knowledge, the report for B. Tech final semester does not contain part of the work which has been submitted for the award of B. Tech Degree either in this university or any other university without proper citation.

Candidate's Signature

Rikin Zala – IU2041230201

Candidate's Signature

Harsh Thakkar – IU2041230183

Candidate's Signature

Jaynil Zala – IU2041230200

Guide: **Prof. Darshan Solanki**

Assistant Professor,
Department of CSE, IITE,
Indus University - Ahmedabad,
State: Gujarat

INSTITUTE OF TECHNOLOGY AND ENGINEERING

COMPUTER SCIENCE AND ENGINEERING

2023-2024

CERTIFICATE

Date: 09-10-2023

This is to certify that the project work entitled "**ESG Report Analysis using NLP**" has been carried out by **RIKIN VISHALKUMAR ZALA** under my guidance in partial fulfillment of degree of Bachelor of Technology in **COMPUTER SCIENCE & ENGINEERING (Final Year)** of Indus University, Ahmedabad during the academic year 2023 - 2024.

_____
PROF. DARSHAN SOLANKI
Assistant Professor,
Department of CSE,
I.I.T.E, Indus University,
Ahmedabad

_____
PROF. ZALAK TRIVEDI
Head of the Department,
Department of CSE,
I.I.T.E, Indus University,
Ahmedabad

INSTITUTE OF TECHNOLOGY AND ENGINEERING

COMPUTER SCIENCE AND ENGINEERING

2023-2024

CERTIFICATE

Date: 09-10-2023

This is to certify that the project work entitled "**ESG Report Analysis using NLP**" has been carried out by **HARSH JAGDISH THAKKAR** under my guidance in partial fulfillment of degree of Bachelor of Technology in **COMPUTER SCIENCE & ENGINEERING (Final Year)** of Indus University, Ahmedabad during the academic year 2023 - 2024.

_____	_____
PROF. DARSHAN SOLANKI Assistant Professor, Department of CSE, I.I.T.E, Indus University, Ahmedabad	PROF. ZALAK TRIVEDI Head of the Department, Department of CSE, I.I.T.E, Indus University, Ahmedabad

INSTITUTE OF TECHNOLOGY AND ENGINEERING

COMPUTER SCIENCE AND ENGINEERING

2023-2024

CERTIFICATE

Date: 09-10-2023

This is to certify that the project work entitled "**ESG Report Analysis using NLP**" has been carried out by **JAYNIL BHAVESHBHAI ZALA** under my guidance in partial fulfillment of degree of Bachelor of Technology in **COMPUTER SCIENCE & ENGINEERING (Final Year)** of Indus University, Ahmedabad during the academic year 2023 - 2024.

_____	_____
PROF. DARSHAN SOLANKI Assistant Professor, Department of CSE, I.I.T.E, Indus University, Ahmedabad	PROF. ZALAK TRIVEDI Head of the Department, Department of CSE, I.I.T.E, Indus University, Ahmedabad

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all those who contributed to the successful completion of this college project on "ESG Report Analysis using Natural Language Processing." It has been a collaborative effort, and we are thankful for the support, guidance, and encouragement we received throughout this endeavour.

First and foremost, we extend our heartfelt appreciation to our college faculty and project supervisor **Prof. Darshan Solanki** [Assistant Professor in the Department of Computer Science & Engineering, IITE, Indus University] for their invaluable guidance and mentorship. Their expertise and unwavering support were instrumental in shaping the project and ensuring its successful execution.

Thank you all for being an integral part of this journey.

Sincerely,

-Harsh Thakkar

IU2041230183

Computer Science & Engineering

-Jaynil Zala

IU2041230200

Computer Science & Engineering

-Rikin Zala

IU2041230201

Computer Science & Engineering

TABLE OF CONTENT

Title **Page No**

ABSTRACT

This abstract outline a research study that utilizes data-driven methods and natural language processing (NLP) techniques to analyse Environmental, Social, and Governance (ESG) reports. The study's primary objective is to unravel the language employed within these reports, pinpoint prevalent concerns, and discern variations among industries and regions.

The researchers employ a robust dataset of over 10,000 ESG reports from global companies and harness various NLP techniques to extract meaningful insights. NLP, a branch of artificial intelligence, is adept at comprehending and processing human language, making it an ideal tool for parsing textual data such as ESG reports.

The study's findings shed light on substantial disparities in the language used by companies operating in different industries and regions. This insight underscores the significance of tailoring ESG assessments to the unique characteristics of each sector and geographic area. Notably, the research highlights certain ESG topics, notably climate change and human rights, as recurring areas of emphasis.

The implications of this study extend to a diverse set of stakeholders. Investors stand to gain valuable insights to inform their investment strategies by understanding how companies communicate their ESG endeavours. Regulators tasked with enforcing ESG reporting standards can better comprehend reporting trends. Moreover, advocacy groups and the broader public can use this research to hold companies accountable for their ESG commitments.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ESG - Environmental, Social and Governance

NLP - Natural Language Processing

GDPR - General Data Protection Regulation

CCPA - California Consumer Privacy Act

PDF - Portable Document Format

CSV - Comma Separated Value

AI - Artificial Intelligence

ML - Machine Learning

RMSE - Root Mean Square Error

MAE - Mean Absolute Error

S&P 500 - Standard and Poor's 500

ESG Bert - NLP Model

AJAX - Asynchronous JavaScript and XML

BSON - Binary JSON

JSON - JavaScript Object Notation

WSGI - Web Server Gateway Interface

BERT - Bidirectional Encoder Representations from Transformers

TIKA - Java Library

HTML - Hyper Text Transfer Protocol

URL - Uniform Resource Locator

DB – Data Base

CSS - Cascading Style Sheets

API - Application Programming Interface

CHAPTER 1
INTRODUCTION

- PROBLEM STATEMENT
- OBJECTIVES AND SCOPE OF THE PROJECT
- SIGNIFICANCE OF THE PROJECT

1.1 PROBLEM STATEMENT

In today's rapidly evolving business landscape, the Environmental, Social, and Governance (ESG) factors have become central to investment decisions, regulatory compliance, and corporate sustainability initiatives. The volume and complexity of ESG data generated by companies worldwide have surged, making manual analysis impractical and error-prone. As a result, there is a critical need for an advanced technological solution that can harness the power of Natural Language Processing (NLP) to efficiently and accurately analyse ESG reports.

The problem at hand is twofold:

1. **Data Overload:** ESG reports produced by companies encompass a vast array of information, often presented in unstructured textual formats. Parsing, understanding, and extracting valuable insights from these reports in a timely manner have become increasingly challenging, particularly for stakeholders such as investors, regulators, and sustainability analysts.

2. **Complexity and Variability**: ESG reporting standards and practices vary widely across industries, regions, and organizations. Analysing these reports requires not only the ability to process textual data but also the capacity to recognize nuanced ESG concepts, sentiments, and trends specific to each context.

Hence, our project seeks to address these challenges by developing an ESG report analysis system using NLP. This system will serve as a comprehensive and efficient tool to automatically process, interpret, and derive meaningful insights from ESG reports, facilitating informed decision-making and promoting responsible business practices.

1.2 OBJECTIVES AND SCOPE OF THE PROJECT

Project Objectives:

1. Automated ESG Report Analysis:
   Develop a system that can automatically ingest, process, and analyze ESG reports from various sources using Natural Language Processing (NLP) techniques.

2. Data Extraction and Classification:
   Implement NLP algorithms to extract relevant information, categorize reports based on industry or region, and identify key ESG concepts, trends, and sentiments within the textual data.

3. Trend Identification:
   Utilize NLP-driven topic modeling to identify recurring themes, emerging trends, and shifts in ESG reporting across different industries and regions.

4. Anomaly Detection:
   Implement machine learning algorithms to flag anomalies or irregularities in ESG reports, alerting users to potential risks or non-compliance issues.

5. User-Friendly Interface:
   Create an intuitive and user-friendly web-based interface that allows stakeholders, including investors, regulators, and sustainability analysts, to access and interact with analysis results and visualizations.

6. Customization and Reporting:
   Enable users to customize their analysis criteria and generate detailed ESG analysis reports, which can be exported in various formats (e.g., PDF, CSV).

7. Scalability:
   Design the system architecture to scale efficiently, accommodating a growing volume of ESG reports and users over time.

8. Compliance:
   Ensure compliance with relevant data privacy regulations (e.g., GDPR, CCPA) and industry standards governing ESG reporting.

9. Performance Optimization:
   Optimize system performance to provide real-time or near-real-time analysis results, with a focus on responsiveness and low latency.

10. Documentation and Training:
    Provide comprehensive documentation and user training materials to facilitate the effective use of the ESG report analysis system.

By achieving these objectives, the project aims to empower stakeholders with a powerful tool for informed decision-making, risk assessment, and promoting sustainable and ethical business practices based on ESG reporting and analysis.

Scope of the Project:

The scope of this project encompasses the development of a specialized ESG (Environmental, Social, and Governance) report analysis system. The system will allow users to manually upload ESG reports in PDF. The textual data from uploaded reports will undergo preprocessing, including tokenization, stemming, and stop word removal, to prepare it for analysis.

The primary focus will be on the application of Natural Language Processing (NLP) techniques to analyse the content of ESG reports. This analysis will involve the identification and extraction of relevant keywords, phrases, and concepts related to environmental, social, and governance factors. To aid user understanding, the system will generate basic visualizations, such as word clouds and bar charts, to represent the most frequently mentioned ESG topics in the reports. Users will interact with the system through a simple and user-friendly web-based interface, facilitating report uploads, access to analysis results, and the download of basic visualizations.

Additionally, users will have the flexibility to customize their analysis criteria and generate summary reports, containing key insights extracted from the ESG reports. Ensuring system functionality and accuracy, testing will be conducted, focusing on achieving acceptable performance for small to medium-sized ESG reports with minimal response times. Compliance with relevant data privacy regulations, particularly concerning the handling of uploaded reports, is a critical aspect of the project's scope. Lastly, comprehensive user documentation will be provided, guiding users in effectively utilizing the system for ESG report analysis.

1.3 SIGNIFICANCE OF THE PROJECT

1. Informed Decision-Making:
   The project addresses the growing need for informed decision-making in today's business landscape. By automating ESG report analysis, it empowers stakeholders, including investors, regulators, and sustainability analysts, to make data-driven choices regarding investments, risk assessment, and ethical business practices.

2. Enhancing Transparency:
   Transparency is a cornerstone of responsible corporate governance. This project contributes to enhanced transparency by offering a systematic approach to ESG reporting, enabling companies to showcase their commitment to sustainability and ethics.

3. Risk Mitigation:
   Effective ESG report analysis helps identify potential risks and non-compliance issues, allowing proactive risk mitigation. This is particularly vital in industries where regulatory changes and public perception can significantly impact business operations.

4. Promoting Sustainable Practices:
   By highlighting the importance of ESG factors, the project encourages organizations to adopt sustainable and ethical business practices. It aligns with global efforts to address environmental and social challenges, such as climate change and inequality.

5. Efficiency and Accuracy:
   Manual analysis of ESG reports is time-consuming and prone to errors. This project streamlines the process with NLP and machine learning, improving the efficiency and accuracy of extracting insights from large volumes of textual data.

6. Adapting to Market Trends:
   ESG reporting is evolving rapidly, with new topics and trends emerging. The project's trend identification feature helps users stay up-to-date with the latest developments, providing a competitive edge in adapting to changing market expectations.

7. User-Friendly Accessibility:
   The user-friendly web interface and customizable reporting options ensure that both experts and non-experts can access and interpret ESG analysis results. This accessibility broadens the project's impact across various user groups.

8. Scalability:
   While initially designed for small to medium-sized ESG reports, the project's scalability features ensure it can grow to accommodate larger datasets and user bases, expanding its utility over time.

In summary, this project significantly contributes to promoting responsible and sustainable business practices, facilitates data-driven decision-making, and offers efficiency gains in the analysis of ESG reports. It aligns with the global trend towards greater transparency, ethical considerations, and sustainability in the corporate world.

CHAPTER 2
LITERATURE REVIEW

LITERATURE REVIEW

The literature on ESG (Environmental, Social, and Governance) analysis utilizing Natural Language Processing (NLP) techniques is burgeoning, reflecting the increasing recognition of ESG factors in investment decisions and corporate sustainability practices. Researchers have explored various dimensions of this topic, ranging from sentiment analysis to topic modeling and entity recognition.

Topic Modeling and Trend Identification:

As we investigated the utility of NLP techniques for identifying recurring themes, emerging trends, and shifts in ESG reporting across diverse industries and regions. Their findings underscored the significance of NLP-driven topic modeling in keeping investors well-informed about evolving ESG dynamics. This research accentuates the potential of NLP not only to extract insights from reports but also to serve as a vigilant tool for monitoring and adapting to changing market expectations and regulatory landscapes.

Overall Significance of NLP in ESG Analysis:

Collectively, the existing literature underscores the pivotal role of NLP in ESG analysis. It affirms that NLP techniques can serve as invaluable tools for stakeholders, including investors, regulators, and sustainability analysts. NLP not only facilitates the extraction of critical insights from textual data but also equips users with the means to assess sentiment, identify relevant entities, and detect emerging trends. By doing so, NLP empowers stakeholders to make more informed decisions, mitigate risks, and promote sustainable and ethical business practices. This growing body of research signals the continued importance of advancing NLP capabilities in the context of ESG analysis, aligning with global efforts to enhance transparency and corporate responsibility.

Challenges and Opportunities:

While the literature demonstrates the promise of NLP in ESG analysis, it also sheds light on some of the challenges and opportunities within this evolving field. Researchers have noted the need for robust algorithms capable of handling the nuances of ESG reporting across various industries and regions. Achieving a balance between precision and recall in sentiment analysis and entity recognition

remains a challenge, as the context and language used in ESG reports can vary significantly.

Moreover, as ESG reporting standards continue to evolve, there is an opportunity for researchers and practitioners to enhance NLP models to adapt to these changes seamlessly. This adaptability will be critical for keeping pace with regulatory shifts and emerging ESG priorities.

Integration with Stakeholder Needs:

One of the notable findings in the literature is the alignment between NLP-based ESG analysis and the evolving needs of stakeholders. Investors increasingly seek comprehensive insights beyond traditional financial metrics, making NLP a pertinent tool for assessing companies' sustainability and ethical practices. Regulators, too, can benefit from automated ESG analysis to enhance compliance monitoring and oversight.

Holistic ESG Assessment:

The literature review highlights the potential for NLP to contribute to holistic ESG assessment. By combining sentiment analysis, entity recognition, and trend identification, NLP can offer a more complete picture of a company's sustainability practices. This comprehensive understanding enables stakeholders to make more informed and responsible investment decisions.

Future Directions:

Looking ahead, the literature suggests several avenues for further research and development. These include refining NLP models to handle industry-specific ESG reporting nuances, enhancing the accuracy of sentiment analysis, and expanding the scope of entity recognition to encompass a broader range of ESG-related entities and concepts.

Additionally, there is a growing interest in the integration of machine learning and artificial intelligence (AI) techniques to augment NLP capabilities in ESG analysis. Advanced machine learning models can help uncover intricate patterns and insights within ESG data, further enhancing decision-making processes.

In conclusion, the literature on ESG analysis using NLP underscores the transformative potential of natural language processing in deciphering the complexities of environmental, social, and governance reporting. It not only addresses existing challenges but also opens-up new horizons for stakeholders to harness textual data for ethical investment, risk mitigation, and sustainable business practices. As NLP technology continues to advance, it is likely to play an increasingly integral role in shaping the future of ESG analysis and corporate responsibility.

CHAPTER 3

METHODLOGY

- DETAILED EXPLANATION OF THE METHODS
- DESCRIPTION OF TOOLS AND TECHNOLOGIES
- FLOWCHARTS & DIAGRAM

3.1 DETAILED EXPLANATION OF THE METHODS

Sector 1: S&P 500 List - ML-Based ESG Analysis

Data Collection and Preprocessing:

- Collect historical stock data for all S&P 500 companies from 2015 to 2022.
- Acquire ESG scores for each company. This data can be obtained from ESG rating agencies or sources like MSCI, Sustainalytics, or Bloomberg.
- Preprocess the data to handle missing values, outliers, and standardize data formats.

Machine Learning Model:

- Build a machine learning model to predict ESG scores based on historical data and other relevant features (e.g., financial metrics, sector-specific data).
- Train the model using historical data and evaluate its performance using appropriate metrics (e.g., RMSE, MAE).
- Use the trained model to predict ESG scores for companies in the S&P 500 list for the desired period.

Sector 2: Dynamic ESG Approach with Stock Prediction

Data Collection:

- Collect monthly stock data for companies of interest.
- Obtain ESG scores as described in Sector 1.

Dynamic ESG Allocation:

- Develop a backend Python system that dynamically allocates investments based on ESG scores and stock performance.
- Define a strategy for dynamic allocation (e.g., monthly rebalancing) and implement it.
- Monitor the performance of the dynamic ESG allocation strategy over time.

Stock Prediction:

- Use linear regression or other suitable models to predict stock prices.
- Train the stock prediction model on historical data.
- Use the model to predict stock prices for the selected companies.

Sector 3: NLP-Based ESG Analysis

Data Collection:

- Collect sustainability reports for the companies of interest.
- Preprocess the text data, including cleaning, tokenization, and removing stop words.

ESG-BERT Model:

- Utilize a pre-trained ESG-BERT model to extract relevant information from the sustainability reports.

- Fine-tune the ESG-BERT model on your specific dataset if necessary.

- Classify and categorize the extracted information into ESG criteria (e.g., Environmental, Social, Governance).

ESG Scoring and Reporting:

- Develop a scoring system based on the classified ESG criteria to generate ESG scores for each company.

- Create reports summarizing the ESG performance of each company, highlighting strengths and weaknesses.

Distribution and Use:

- Use the generated ESG scores and reports for decision-making, such as investment analysis, portfolio management, or corporate governance assessments.

3.2 DESCRIPTION OF TOOLS AND TECHNOLOGIES

HTML (Hyper-Text Markup Language):

- HTML is a standard markup language used to create the structure and content of web pages.
- It provides a way to define the elements and layout of a web page, including headings, paragraphs, images, links, forms, and more.
- HTML tags are used to mark-up content, and browsers interpret these tags to render web pages.

CSS (Cascading Style Sheets):

- CSS is a stylesheet language used for controlling the presentation and layout of web pages.
- It allows web developers to apply styles such as colors, fonts, spacing, and positioning to HTML elements.
- CSS can be used to create responsive and visually appealing web designs.

JavaScript:

- JavaScript is a versatile programming language primarily used for adding interactivity to web pages.
- It enables dynamic behavior in web applications, such as handling user interactions, updating content without page refreshes, and making asynchronous requests to servers (AJAX).

MongoDB:

- MongoDB is a NoSQL database management system known for its flexibility and scalability.

- It stores data in a document-oriented format, using BSON (Binary JSON), making it suitable for handling unstructured or semi-structured data.
- MongoDB is often used for web applications, including those involving large volumes of data.

Python:

- Python is a high-level, general-purpose programming language known for its readability and ease of use.
- It is widely used in web development, data analysis, machine learning, and various other fields.
- Python's extensive libraries and frameworks make it suitable for a wide range of applications.

Flask:

- Flask is a lightweight and flexible web framework for building web applications in Python.
- It provides the essential tools and libraries needed to create web-based applications, making it well-suited for small to medium-sized projects.
- Flask follows the WSGI (Web Server Gateway Interface) standard and is known for its simplicity and extensibility.

ESG-Bert Algorithm:

- ESG-Bert is likely a specialized algorithm or model for analyzing ESG (Environmental, Social, and Governance) reports.
- It could be based on the BERT (Bidirectional Encoder Representations from Transformers) model, which is a popular NLP model for various text analysis tasks.
- The ESG-Bert algorithm likely focuses on sentiment analysis, keyword extraction, and entity recognition within ESG reports.

TIKA (Java Library):

- Apache TIKA is a Java library and toolkit for extracting text and metadata from various document formats, including PDF, Word, HTML, and more.
- It can be used to parse and extract content from documents, making it valuable for text analysis and data extraction tasks.
- TIKA simplifies the process of working with diverse document formats in Java-based applications.

These technologies and components are essential building blocks for ESG report analysis project, enabling web-based interfaces, data storage, data extraction, text analysis, and more.

3.3 FLOWCHARTS & DIAGRAM

(1) Sequence Diagram:

A sequence diagram is a type of interaction diagram that displays the interactions between objects in a system in a particular sequence. In the context of an ESG Report Analysis using NLP, the sequence diagram would illustrate the flow of actions and events that occur between the user, the server and the database. A Sequence Diagram for ESG (Environmental, Social, and Governance) Report Analysis using Natural Language Processing (NLP) visually represents the dynamic interactions and processes involved in the analysis of ESG reports through the application of NLP techniques. It provides a comprehensive view of how different system components and actors collaborate to extract valuable insights from ESG reports. It also provides the feedback loop from the database to the user based on his/her request.

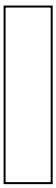
	Object
	Message
	Feedback / Response
	Activation Box

Table-3.3.1 Symbols used in Sequence Diagram.

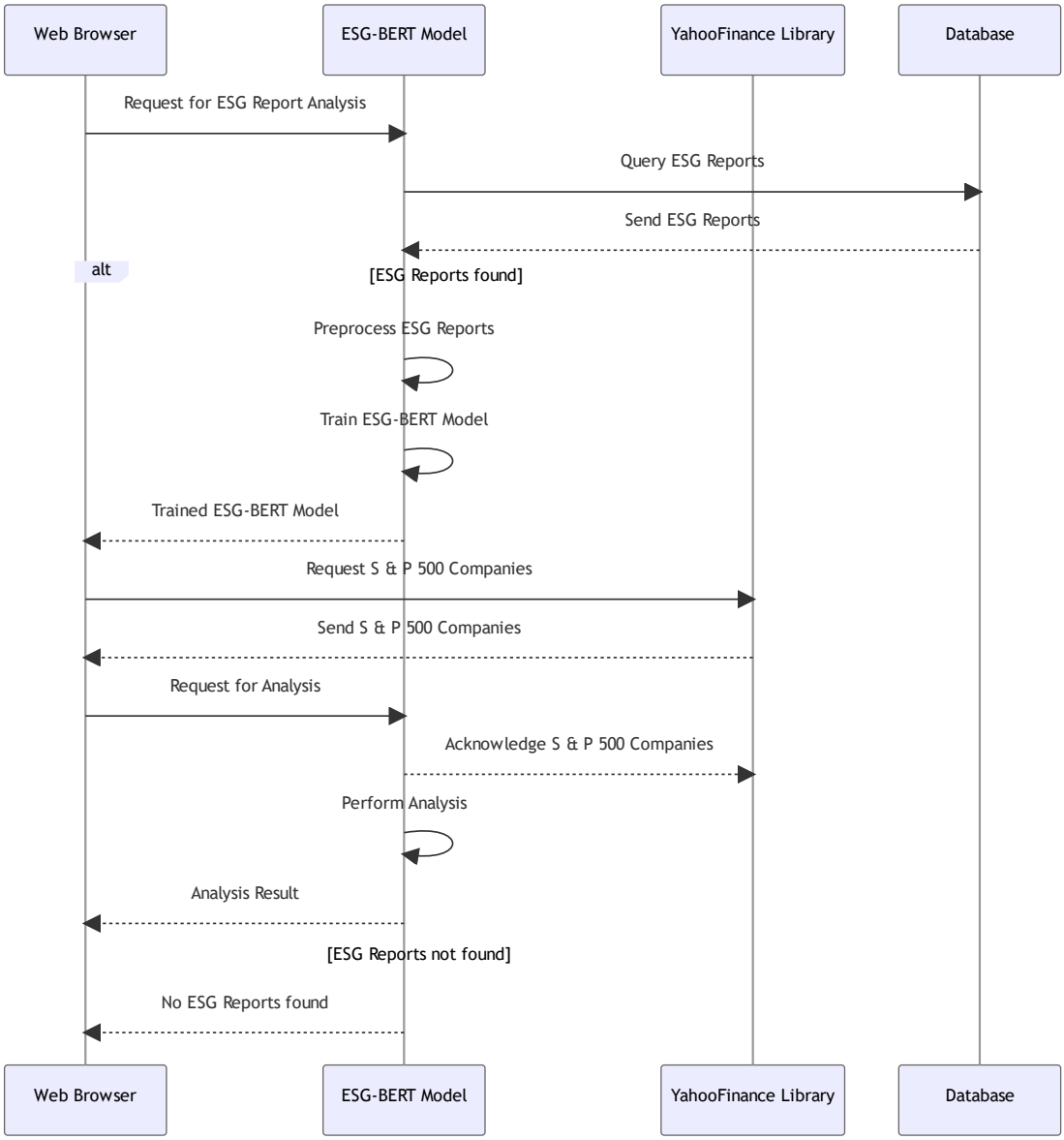


Figure-3.3.1 Sequence Diagram.

(2) Class Diagram:

A class diagram is a type of diagram in object-oriented programming (OOP) that displays the classes and their relationships to each other. For our project topic-ESG Report Analysis using NLP, the class diagram will show the various classes such as ESG Report Analysis, ESG BERT Model, Yahoo Finance Library, Yahoo Query Library, ESG Report. These classes will have attributes and methods that define their behaviour within the system. The class diagram will also show the relationships between these classes, such as inheritance, aggregation, and association. This diagram is an essential tool for developers as it provides a visual representation of the system's architecture, making it easier to understand and modify the code.

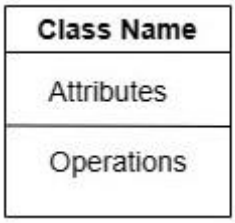
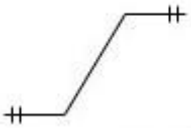
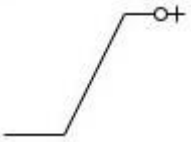
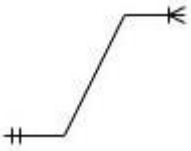
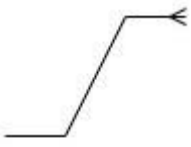
	Class
	One to One
	Zero to One
	One Mandatory to Many Optional
	One to Many

Table-3.3.2 Symbols used in Class Diagram.

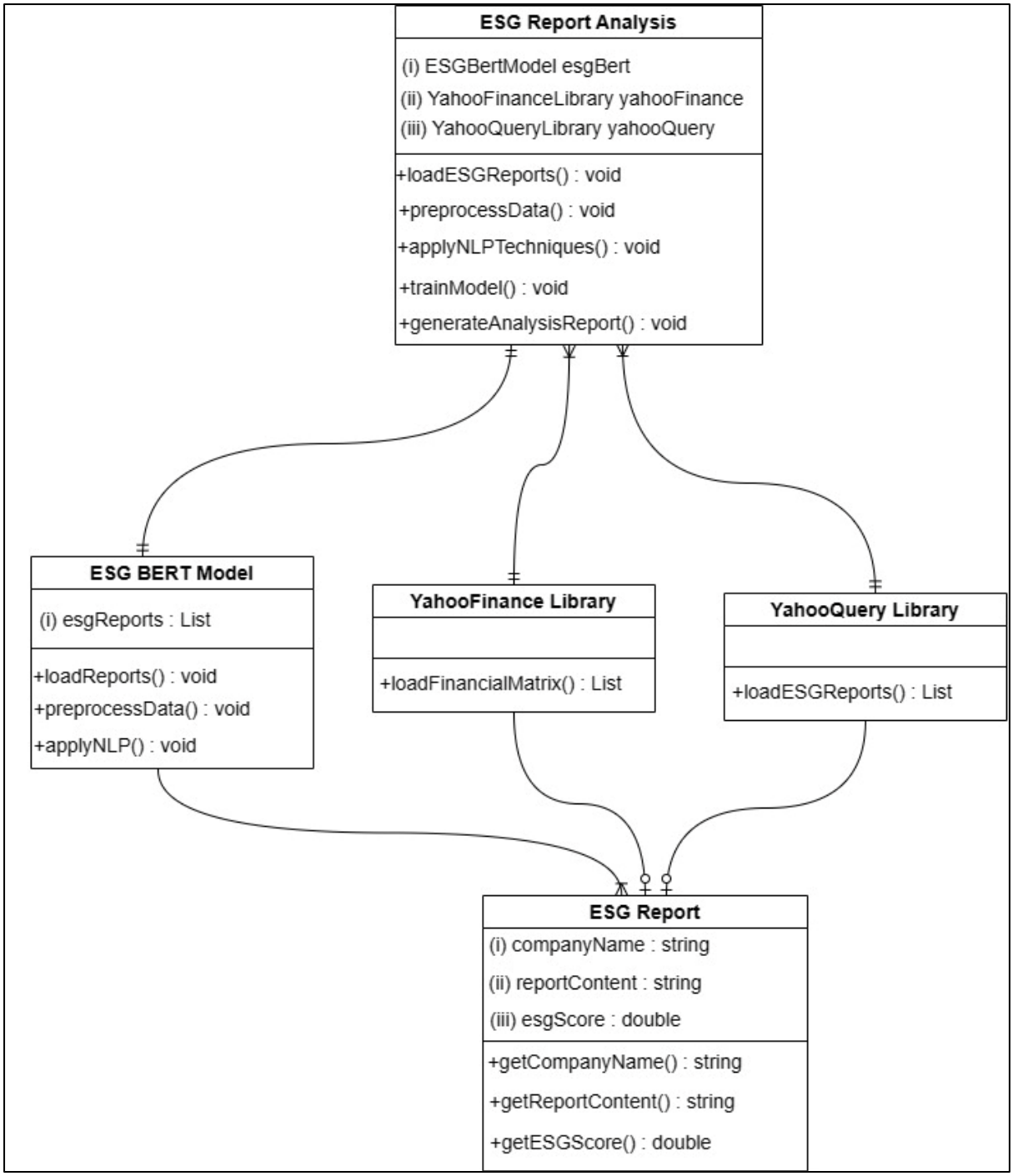


Figure-3.3.2 Class Diagram.

(3) Flowchart Diagram:

A flowchart diagram is a visual representation of a process or system using symbols, lines, and arrows to depict the sequence of steps or activities involved. Flowcharts are a powerful tool for illustrating complex processes, making them easier to understand and analyze. For our project topic-ESG Report Analysis using NLP, our project comprises of four different Flowchart Diagrams: S & P 500 Companies Flowchart, Dynamic Allocation Flowchart, NLP Approach Flowchart and Compare two ESG's Flowchart.

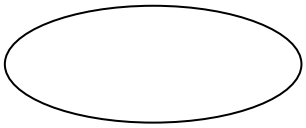
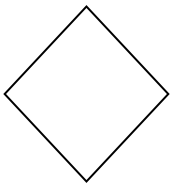
	Start/End
	Processing
	Decision
	Procedure/Method
	Flowlines/Arrows

Table-3.3.3 Symbols used in Flowchart Diagram.

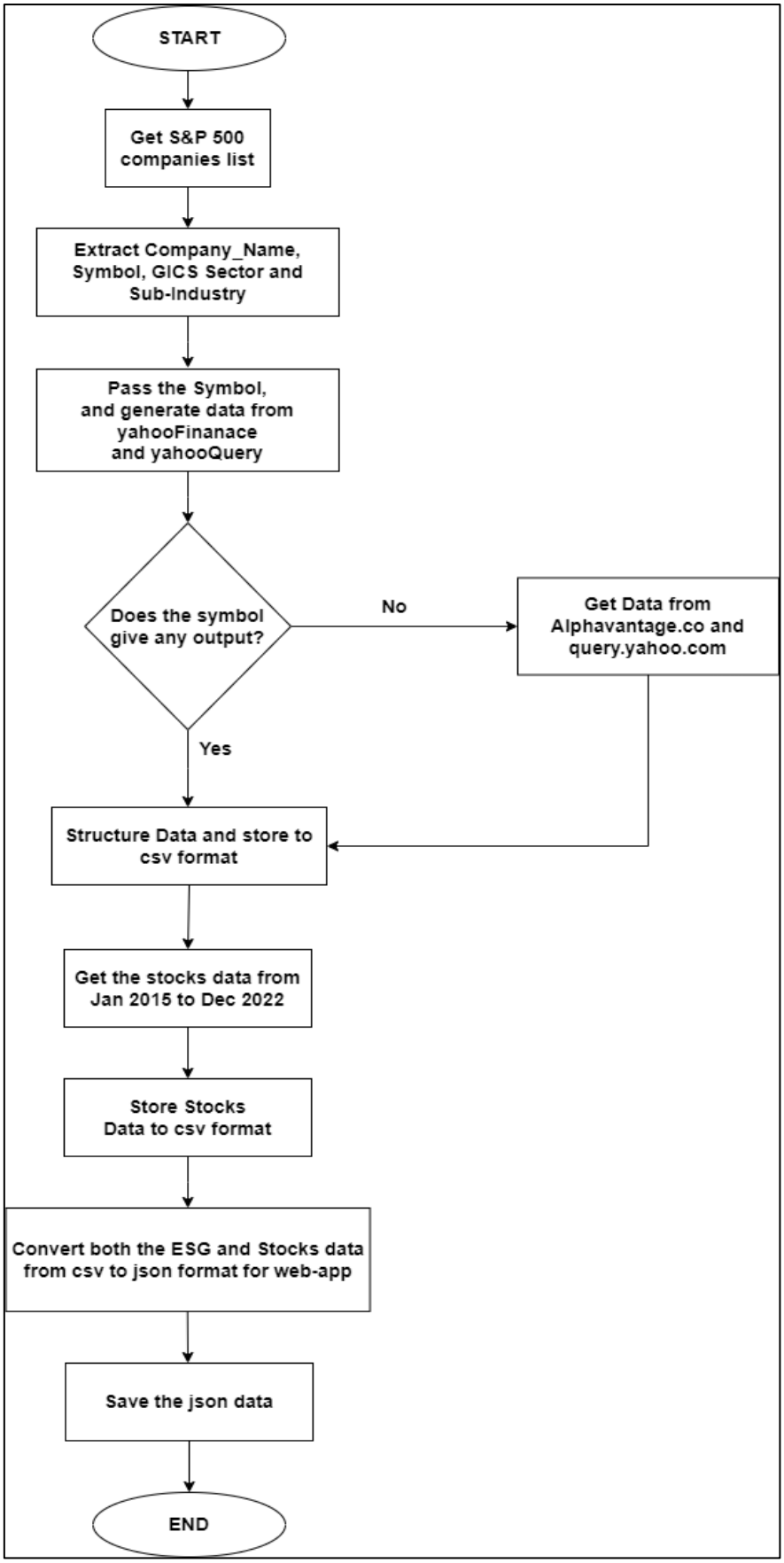


Figure-3.3.31 S & P 500 Constituents Flowchart Diagram.

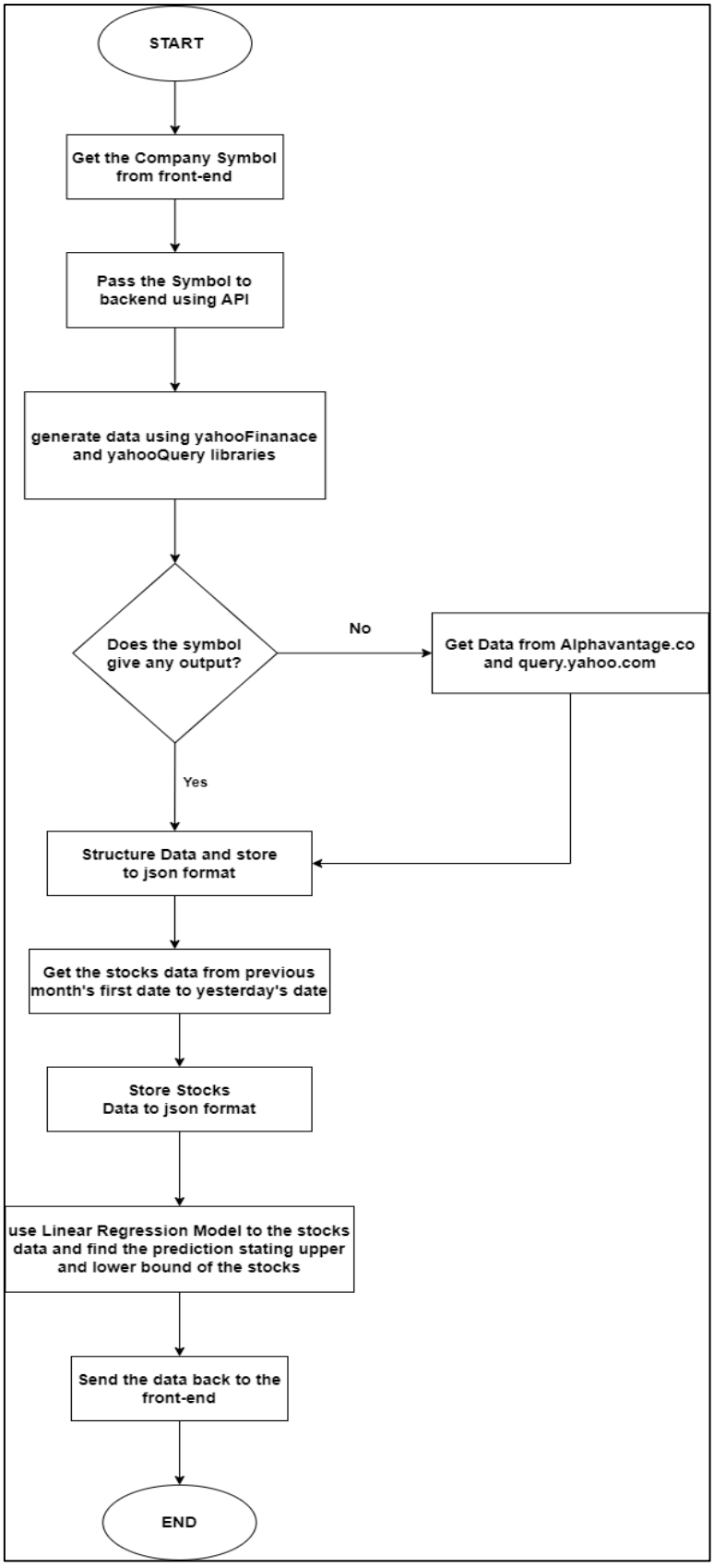


Figure-3.3.32 Dynamic Allocation Flowchart Diagram.

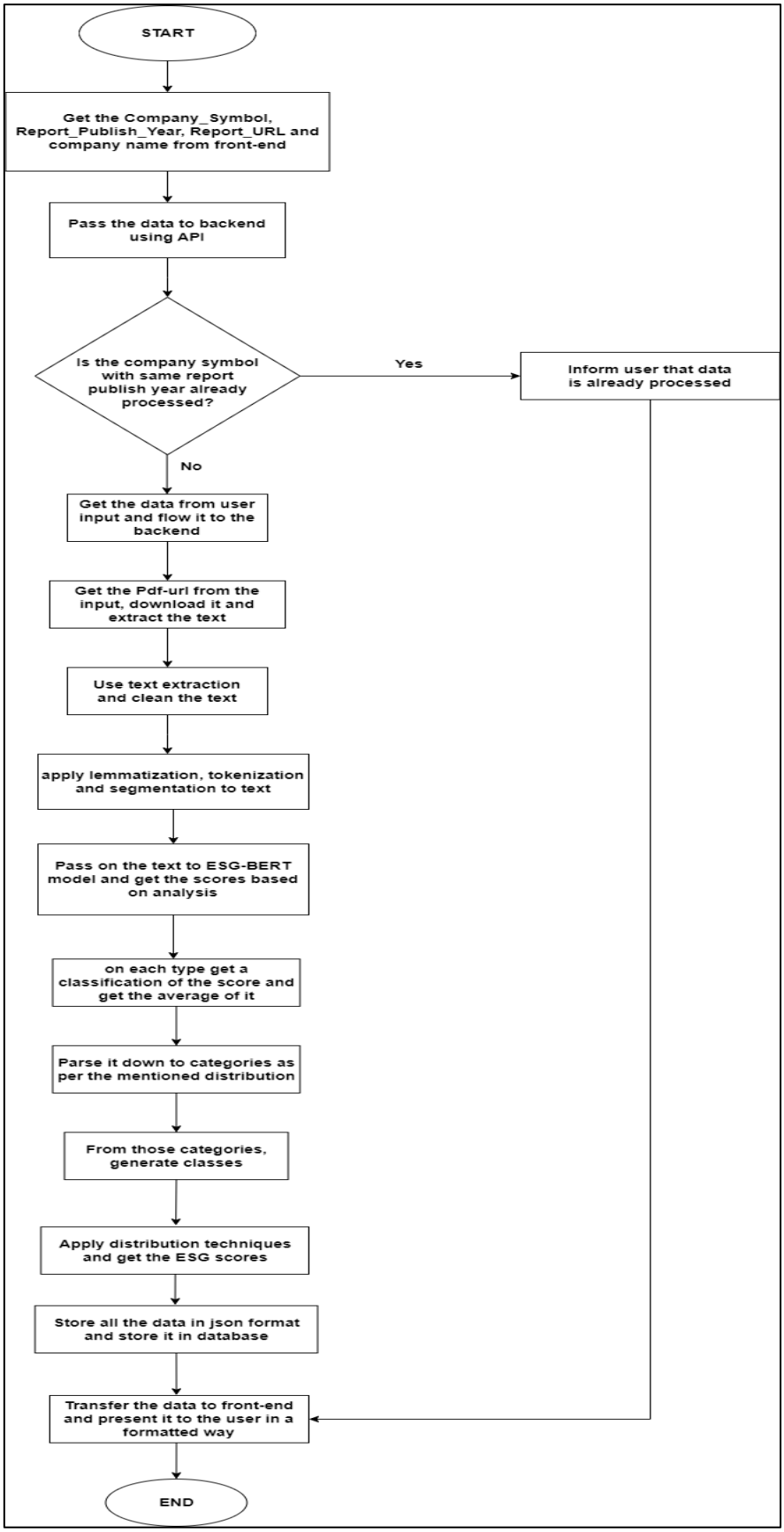


Figure-3.3.33 NLP Approach Flowchart Diagram.

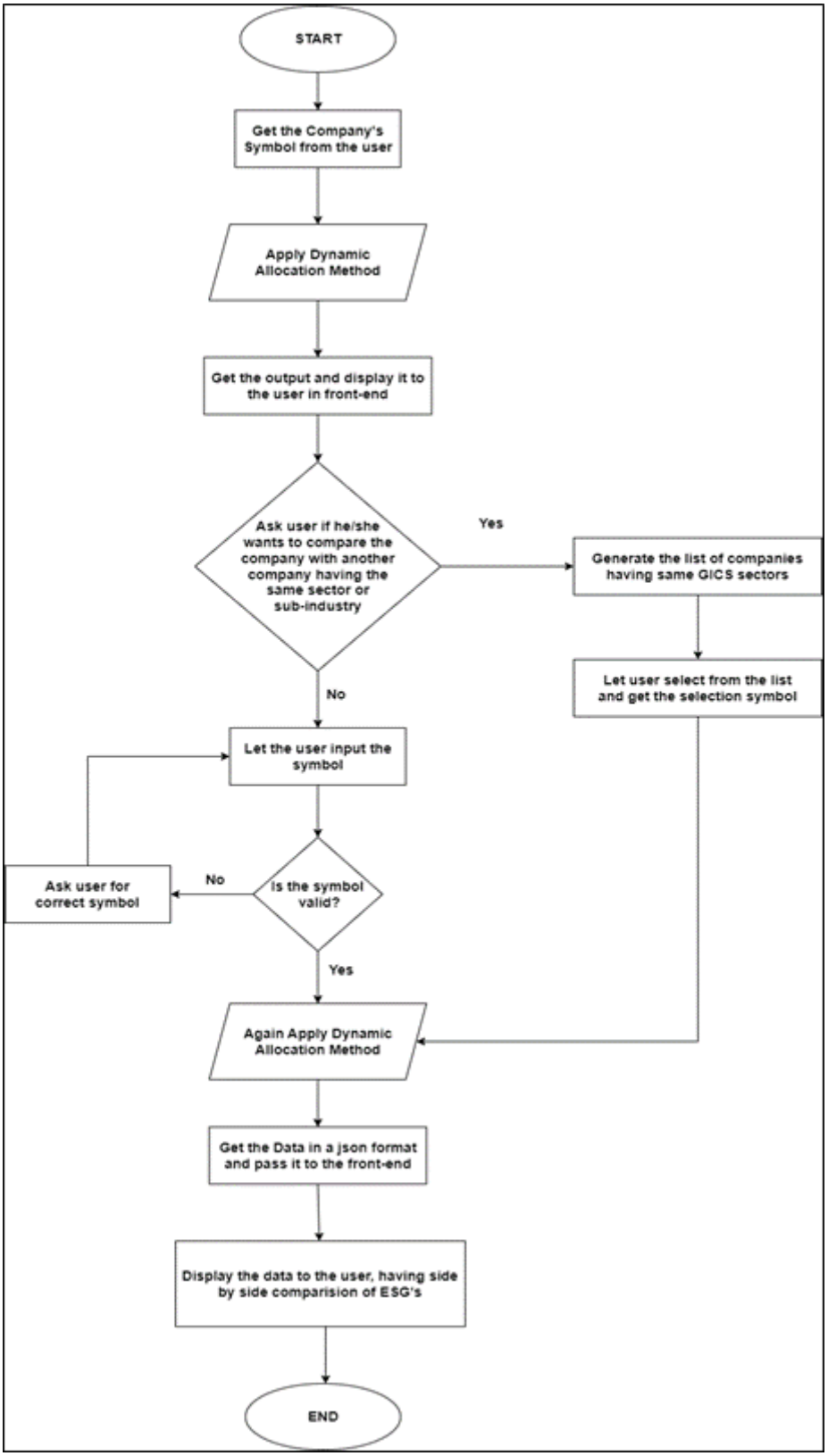


Figure-3.3.34 Compare two ESG's Flowchart Diagram.

CHAPTER 4

SYSTEM DESIGN

- ARCHITECTURE AND SYSTEM OVERVIEW
- BLOCK DIAGRAM
- DATA FLOW DIAGRAM (DFD)
- USER INTERFACE DESIGN

4.1 ARCHITECTURE AND SYSTEM OVERVIEW

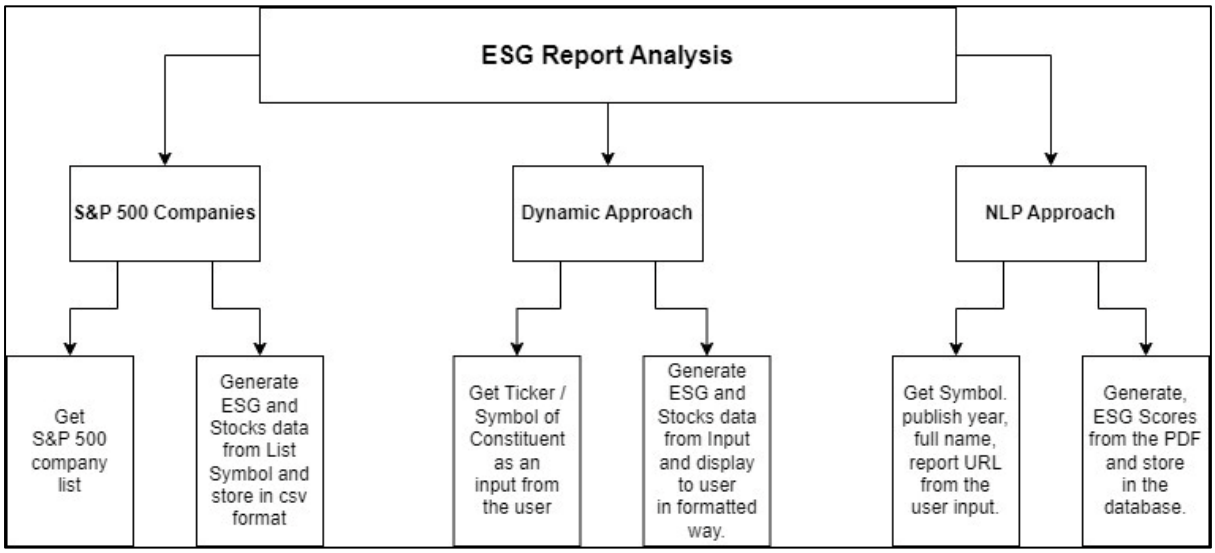


Figure-4.1 Architecture and System Overview Diagram.

An Architecture and System Overview Diagram for ESG (Environmental, Social, and Governance) Report Analysis Using Natural Language Processing (NLP) provides a comprehensive visual representation of the underlying framework and components of a sophisticated system designed to assess and interpret ESG reports, thereby aiding in sustainable investment decision-making and corporate responsibility evaluation. This diagram showcases the essential elements and interactions within the system, offering a clear insight into its operational mechanisms.

Key components and features of this Architecture and System Overview Diagram may include:

1. Data Ingestion: At the core of the system, data from various sources, such as ESG reports, financial data, news articles, and social media, are ingested into the system. This may involve data extraction, transformation, and loading (ETL) processes.

2. NLP Processing Engine: This represents the heart of the system. NLP algorithms and models are applied to the ingested data to analyze text,

extract relevant information, and comprehend the sentiments and context of ESG-related content.

3. Data Storage: Processed data is stored in databases or data lakes for easy retrieval and analysis. This may include structured and unstructured data storage solutions.

4. Machine Learning Models: Trained machine learning models are employed to identify trends, patterns, and correlations within the ESG data. These models help in scoring and classifying ESG performance metrics.

5. User Interface: A user-friendly interface, such as a web application or dashboard, allows users to interact with the system. Users can query, visualize, and explore ESG data insights in an intuitive manner.

6. Reporting and Visualization: The system generates reports and visual representations of ESG performance, which can include charts, graphs, and other data visualization tools to facilitate decision-making.

7. Integration with External Systems: The system may integrate with external data sources, analytics platforms, and investment tools to enhance its capabilities and provide a more holistic view of ESG-related information.

8. Compliance and Governance: To ensure that the system adheres to regulatory requirements, a governance layer is incorporated to monitor data handling, model accuracy, and maintain data security.

9. Scalability and Performance: The diagram may highlight the system's scalability features, including cloud-based infrastructure for handling large volumes of data and high-performance computing capabilities.

10. Feedback Loop: Continuous learning and improvement are achieved through feedback loops, where the system learns from user interactions, adjusts models, and refines its analysis over time.

11. Security and Data Privacy: Security measures to protect sensitive ESG data and user privacy are fundamental to the system's design.

12. APIs and Data Export: The system may offer APIs for integration with other applications and provide data export options for users to leverage the insights in their decision-making processes.

This Architecture and System Overview Diagram serves as a valuable tool for stakeholders, developers, and decision-makers to understand the intricate workings of the ESG Report Analysis system. It demonstrates how NLP and data science techniques are leveraged to transform raw ESG data into actionable insights, ultimately contributing to more informed and responsible investment decisions in line with sustainable and ethical practices.

4.2 BLOCK DIAGRAM

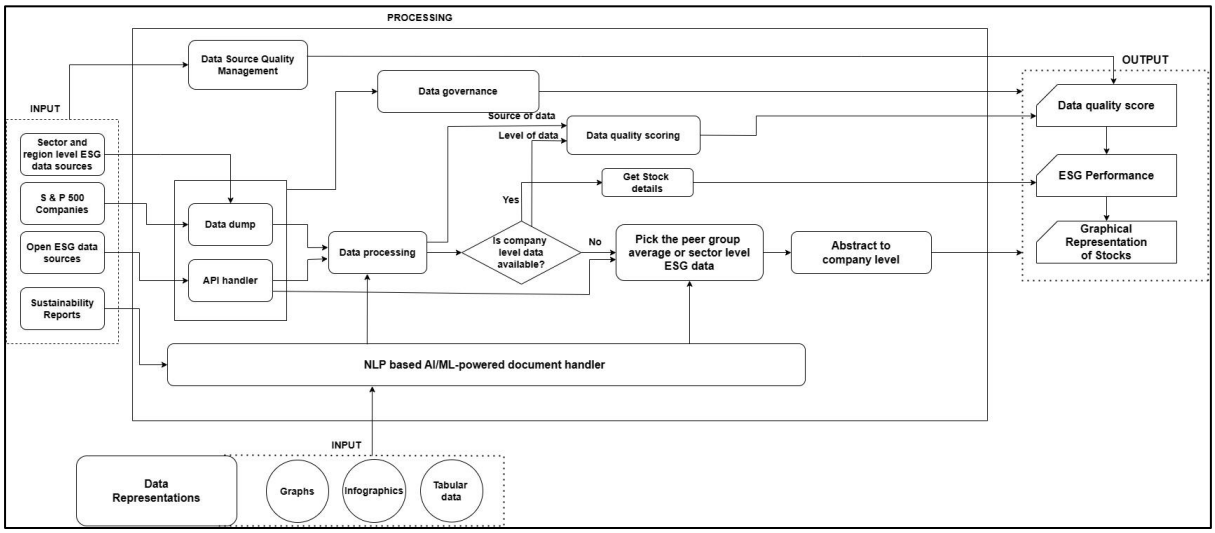


Figure-4.2 Functionality Block Diagram.

Here, for our project topic: ESG Report Analysis using NLP, there are 4 parts, specifically, two different ways of Input, one main processing unit and one output unit. We take S & P 500 Companies Reports and Sustainability Reports as one Input type and other input taken is different types of Data Representations, i.e., Graphs, Infographics, Tabular Data. We use NLP-based ESG-Bert Model to generate the ESG Score based on various risk factors. We get an output in form of Data quality Score, ESG Performance and Graphical Representation of Stocks.

4.3 DATA FLOW DIAGRAM (DFD)

A data flow diagram (DFD) is a graphical representation of how data flows through a system. It shows how information is processed and exchanged between different components of a system. In the case of our project topic: ESG Report Analysis using NLP, a DFD can be used to illustrate the flow of data between the user, the web application, and the database. The DFD can help identify potential bottlenecks or inefficiencies in the data flow, and can be used to optimize the system design for improved performance. It can also help identify security risks and suggest ways to mitigate them. The DFD for ESG Report Analysis Using NLP could include different levels of detail, ranging from a high-level overview of the system to a more detailed breakdown of specific data flows and processes.

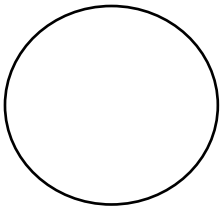
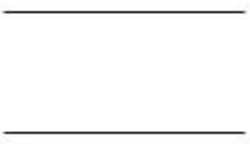
	Entity
	Process
	Data Flow
	Database/Data Store

Table-4.3.1 Symbols used in Data Flow Diagram.

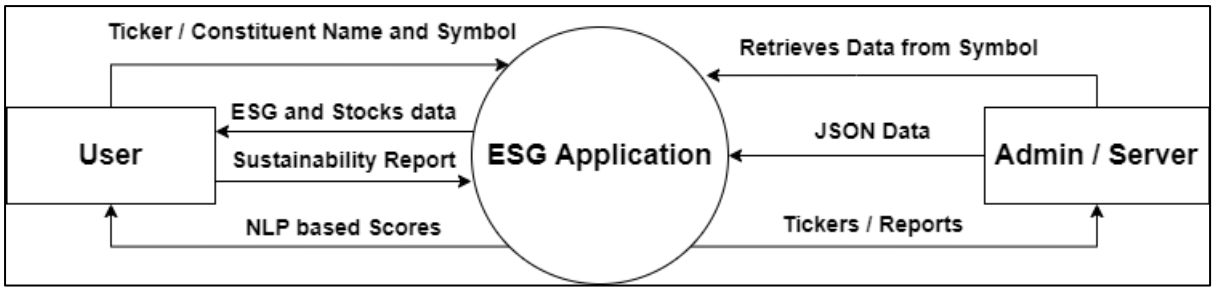


Figure-4.3.1 Data Flow Diagram Level - 0.

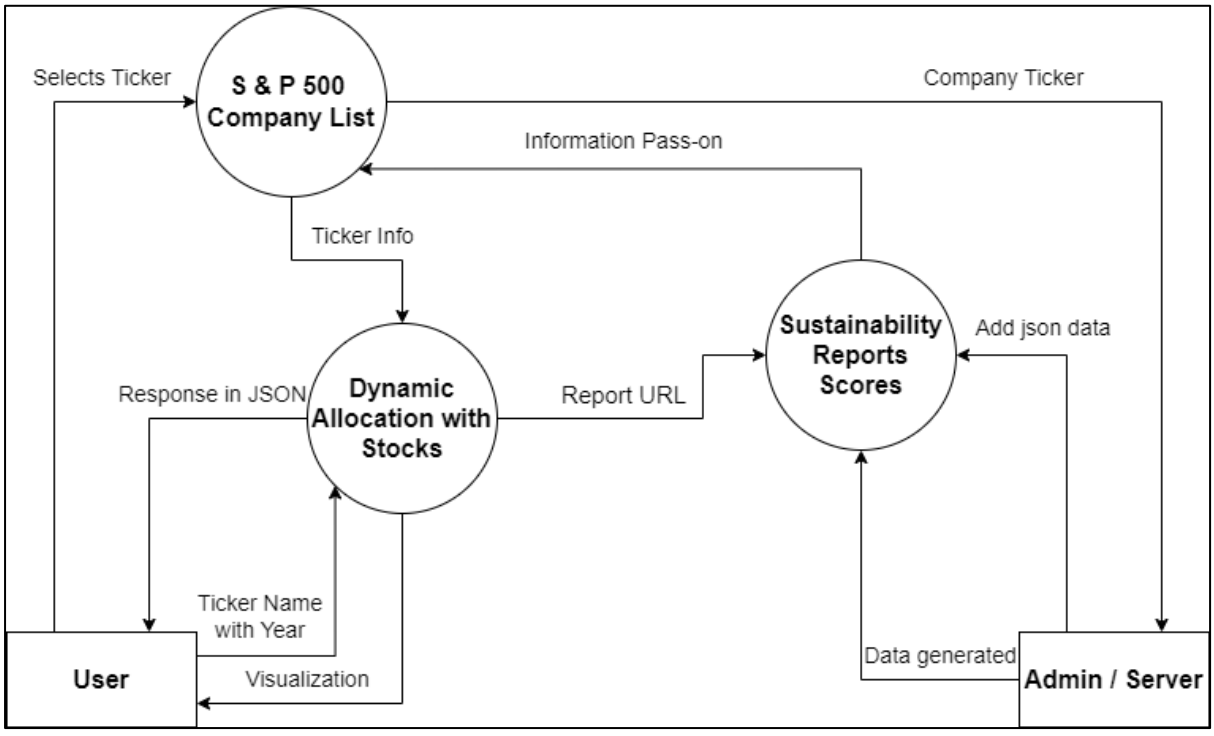


Figure-4.3.2 Data Flow Diagram Level - 1.

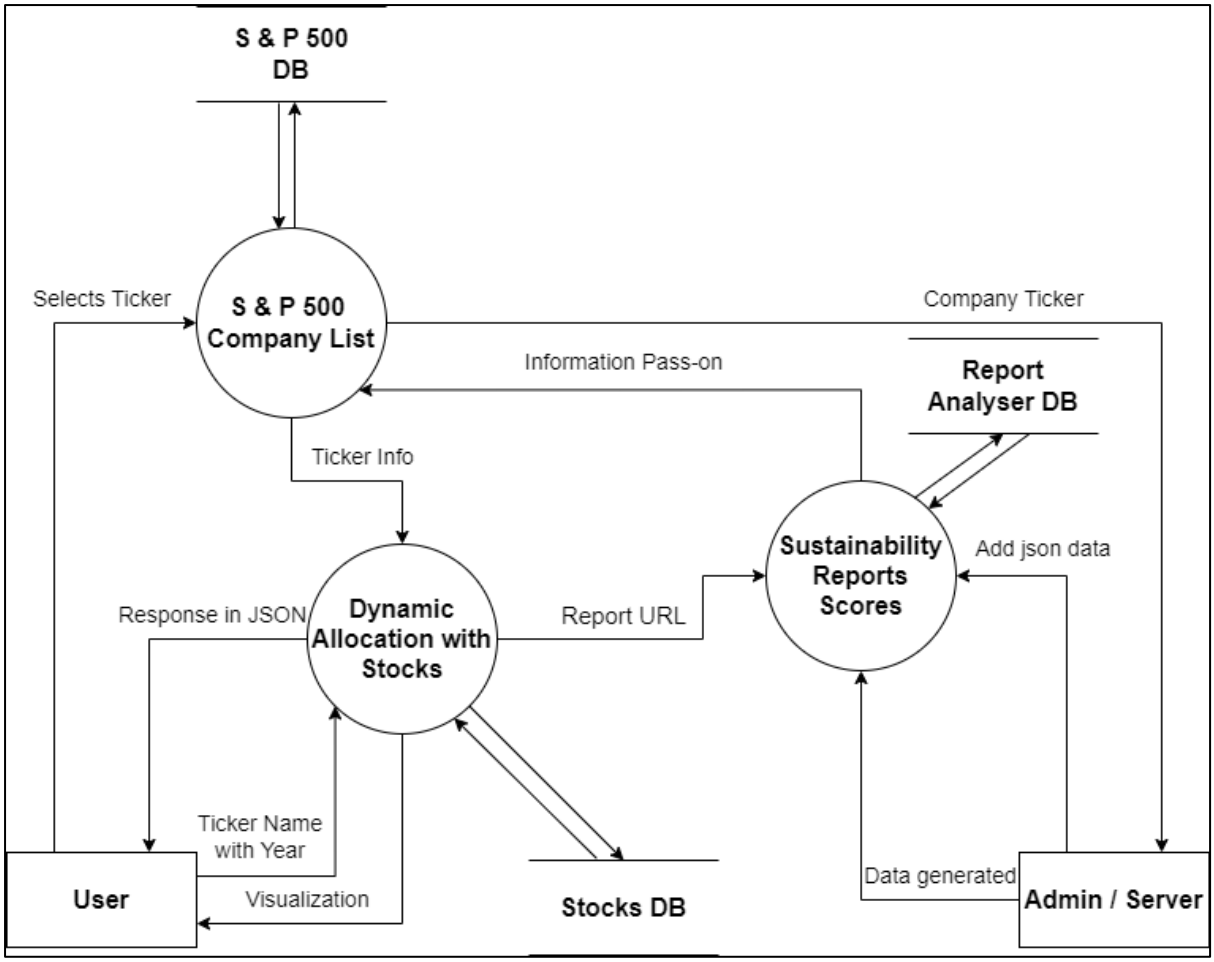


Figure-4.3.3 Data Flow Diagram Level - 2.

4.4 USER INTERFACE DESIGN

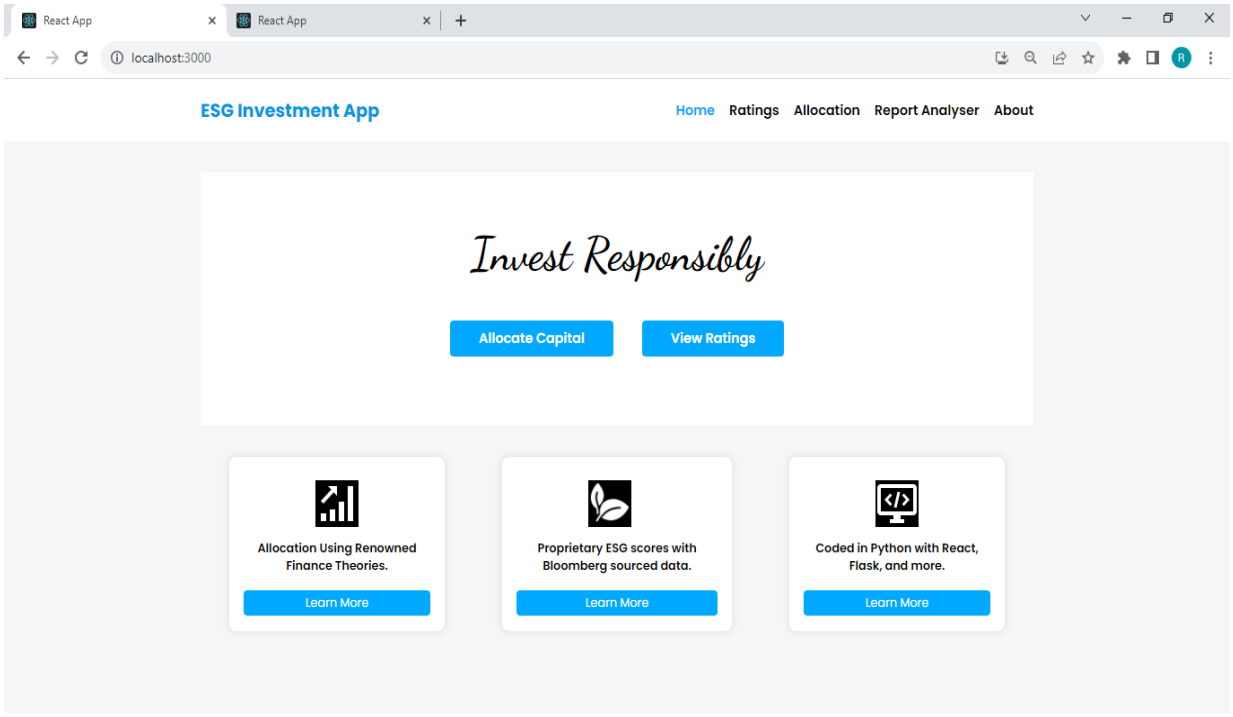


Figure-4.4.1 Home Page Design.

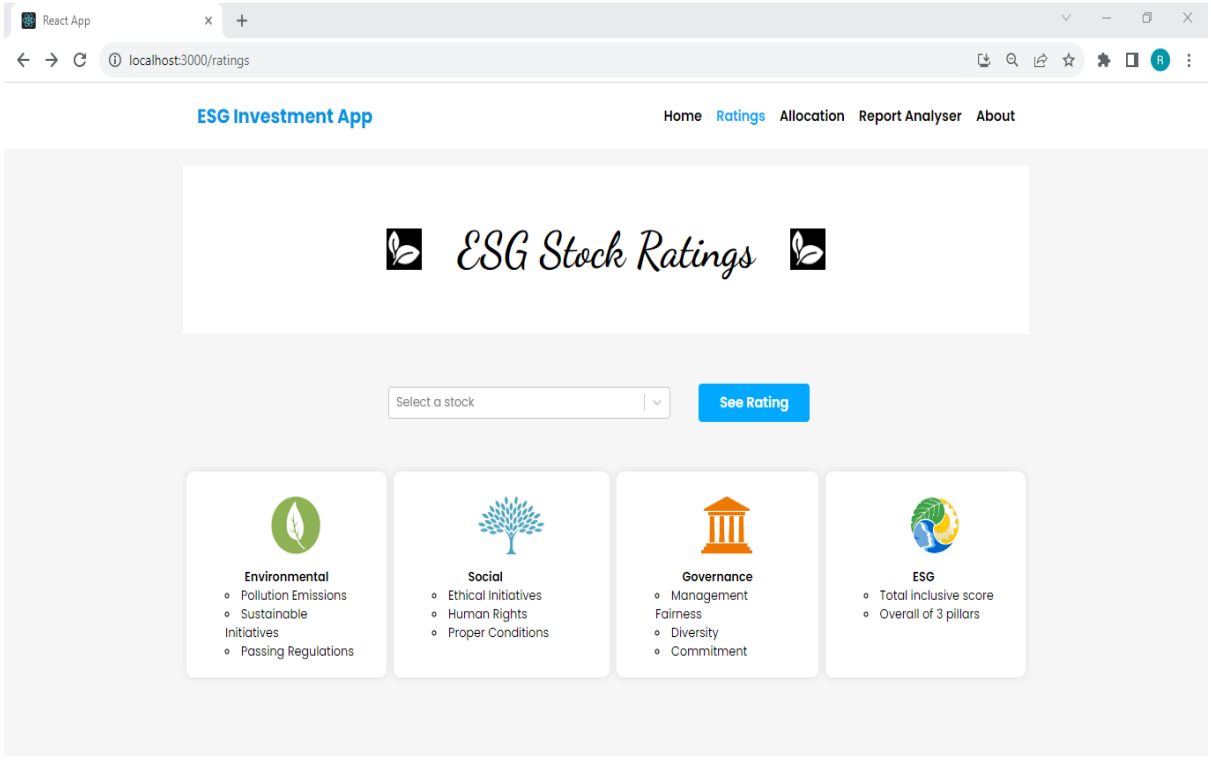


Figure-4.4.2 Ratings Page Design.

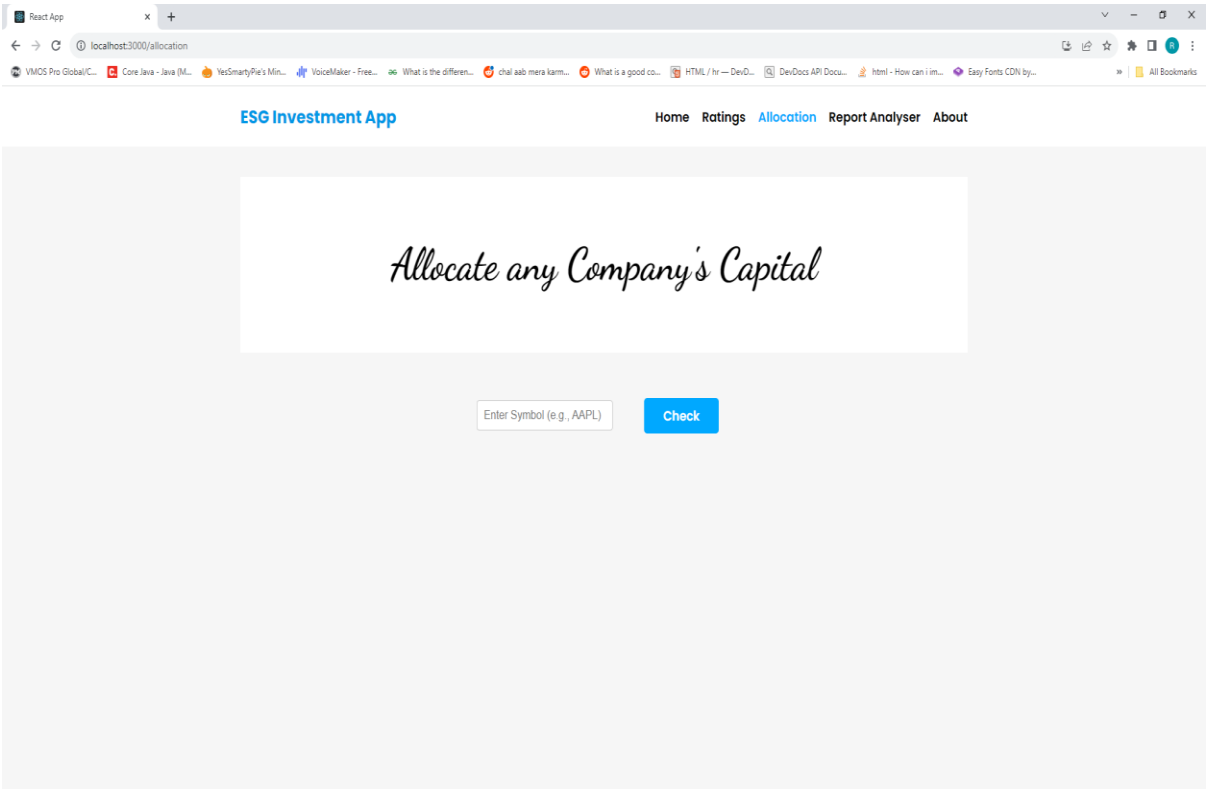


Figure-4.4.3 Allocation Page Design.

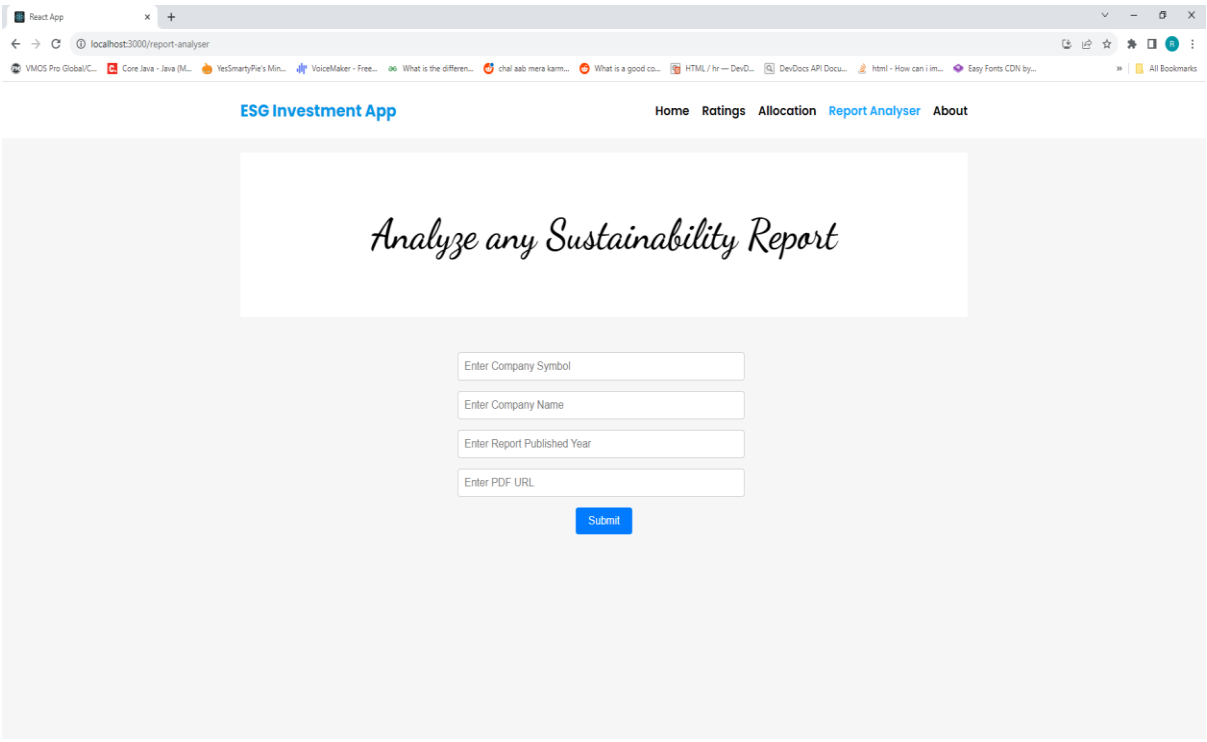


Figure-4.4.4 Reports Analyser Page Design.

CHAPTER 5

IMPLEMENTATION

- CODE SNIPPETS
- TESTING PROCEDURES AND RESULTS
- CHALLENGES DURING IMPLEMENTATION

5.1 CODE SNIPPETS

Libraries

```
# libraries
import re
import numpy as np
import pandas as pd
import yahooquery as yq
import yfinance as yf
from transformers import AutoTokenizer, AutoModelForSequenceClassification
from transformers import pipeline
from tika import parser
```

S & P 500 Company List

```
# S&P 500 Company List
tickers = pd.read_html('https://en.wikipedia.org/wiki/List_of_S%26P_500_companies')[0]
columns_to_remove = ["Headquarters Location", "Date added", "CIK", "Founded"]
tickers = tickers.drop(columns=columns_to_remove)
tickers = tickers.rename(columns={"Security": "Full Name"})
```

Yahoo Query – Get ESG Data

```
# Yahoo Query - get ESG Data
yq_data = pd.DataFrame([])

for ticker in tickers['Symbol']:
    try:

      ticker_yq = yq.Ticker(ticker).esg_scores

      ticker_yq[ticker]["Symbol"] = ticker

      ticker_yq = pd.DataFrame([ticker_yq[ticker]])
      yq_data = pd.concat([yq_data, ticker_yq], ignore_index=True)

    except Exception as e:
      continue
```

Yahoo Finance – Get Financial Metrics

```
# Yahoo Finanace - get Financial Metrics
yf_data = pd.DataFrame([])

for ticker in tickers['Symbol']:
    try:

      ticker_info = yf.Ticker(ticker).info

      ticker_data = pd.DataFrame({
        'Symbol': [ticker],
        'Market Cap': [ticker_info.get('marketCap')],
        'Beta': [ticker_info.get('beta')],
        'Overall Risk': [ticker_info.get('overallRisk')]
      })

      yf_data = pd.concat([yf_data, ticker_data], ignore_index=True)

    except Exception as e:
      continue
```

Generate Stocks Data

```
# Generate Stocks Data
start_date = "2015-01-01"
end_date = "2022-12-31"

ticker_name = tickers["Symbol"].to_list()
price_data = yf.download(ticker_name, start=start_date, end=end_date)
prices = price_data["Adj Close"][ticker_name]
```

NLP Approach

```
# NLP Approach
tokenizer = AutoTokenizer.from_pretrained("nbroad/ESG-BERT")
model = AutoModelForSequenceClassification.from_pretrained("nbroad/ESG-BERT")
classifier = pipeline('text-classification', model=model, tokenizer=tokenizer)

class PDFParser:
    def __init__(self, file_path):
        self.file_path = file_path
        self.raw = parser.from_file(self.file_path)
        self.text = self.raw['content']

    def get_text(self):
        return self.text

    def get_text_clean(self):
        text = self.text
        text = re.sub(r'\n', ' ', text)
        text = re.sub(r'\s+', ' ', text)
        return text

    def get_text_clean_list(self):
        text = self.get_text_clean()
        text_list = text.split('.')
        return text_list

def run_classifier(url):
    pp = PDFParser(url)
    sentences = pp.get_text_clean_list()
    print(f"The CSR report has {len(sentences):,d} sentences")
    result = classifier(sentences)
    df = pd.DataFrame(result)
    return(df)

run_classifier("link")
```

5.2 TESTING PROCEDURES AND RESULTS

The results are displayed here below taking into account all of the strategies stated in the procedures section.

A. The results of information received from the S&P 500 list are displayed in the table.

Input: S&P 500 Constituents list.

Output: CSV. *[From Jupyter Notebook] [Financial Metrics of Constituents]*

Symbol	Full Name	GICS Sector	GICS Sub-Industry	Environment Score	Social Score	Governance Score
MMM	3M	Industrials	Industrial Conglomerates	12.33	13.56	7.72
AOS	A. O. Smith	Industrials	Building Products	7.29	11.87	6.27
ABT	Abbott	Health Care	Health Care Equipment	3.03	13.59	8.36
ABBV	AbbVie	Health Care	Pharmaceuticals	1.12	16.77	9.95
ACN	Accenture	Information Technology	IT Consulting & Other Services	0.29	4.58	4.83

Table-5.2.1 S&P 500 Company list – generated scores from yfinance and yquery.

Symbol	Total Esg	Highest Controversy	Percentile	Market Cap	Beta	Overall Risk
MMM	33.61	3	74.68	55784308736.0	0.968188	7
AOS	25.43	0	46.64	10048283648.0	1.286998	10
ABT	24.98	3	44.01	176555524096.0	0.673236	8
ABBV	27.84	3	55.08	268499402752.0	0.547026	9
ACN	9.71	2	2.21	197301125120.0	1.215421	1

Table-5.2.2 Continuation of Table-5.2.1.

The ESG – ratings are available on the page in a systematic way followed by stock prices from 1st January 2015 to 31st December 2022.

B. The results of information retrieval for companies not on the S&P 500 list are displayed on the Allocation page.

Input: GGG *[not in the S&P 500 list]*

Output: *[From Server] [JSON Format]*


```
{
  "beta": 0.781195,
  "environment_score": 8.89,
  "full_name": "Graco Inc.",
  "gics_sector": "Industrials",
  "gics_sub_industry": "Machinery",
  "governance_score": 7.51,
  "highest_controversy": 0.0,
  "market_cap": 12633916416,
  "overall_risk": 2,
  "percentile": 64.99,
  "prediction_lower_bound": 73.9269550917081,
  "prediction_upper_bound": 75.06709973831977,
  "social_score": 13.51,
  "stocks_date_list": [
    "2023-09-01",
            ...
    "2023-10-16"
  ],
  "stocks_end_date": "2023-10-17",
  "stocks_lower_bound": [
    78.6316175866739,
            ...
    74.0
  ],
  "stocks_start_date": "2023-9-1",
  "stocks_symbol": "GGG",
  "stocks_upper_bound": [
    79.46897652463954,
            ...
    75.04000091552734
  ],
  "symbol": "GGG",
  "total_esg": 29.9
}
```

The result for GGG – Graco Inc. is dynamically generated from the same procedures, which will be visible on-click check button and stocks will be listed with predictions followed as well.

C. The ESG ratings are displayed on the Report Analyzer page based on the uploaded document URL.

Input: MCD, McDonalds, 2021, https://www.responsibilityreports.com/Click/2534

Output: *[From Server] [JSON Format]*


```
"MCD": {
    "tickerName": "MCD",
    "longName": "McDonald's Corporation",
    "publishYear": 2021,
    "processedDate": "2023-10-08",
    "categories": {
      "Access and Affordability": 0.5078979084889094,
      "Business Ethics": 0.5254482686519623,
      "Business Model Resilience": 0.4128567725419998,
      "Competitive Behavior": 0.2863025837517404,
      "Critical Incident Risk Management": 0.9350666205088297,
      "Customer Privacy": 0.445619385689497,
      "Customer Welfare": 0.40044992083105546,
      "Data Security": 0.13562656566500664,
      "Director Removal": 0.10258898884057999,
      "Ecological Impacts": 0.5955852940678596,
      "Employee Engagement Inclusion and Diversity": 0.6404755349670138,
      "Employee Health and Safety": 0.8497115423281988,
      "Energy Management": 0.5377880972292688,
      "GHG Emissions": 0.5400322700540224,
      "Human Rights and Community Relations": 0.5719779746399986,
      "Labor Practices": 0.7389356621674129,
      "Management of Legal and Regulatory Framework": 0.35023836963451827,
      "Physical Impacts of Climate Change": 0.8934897383054098,
      "Product Design and Lifecycle Management": 0.681599784642458,
      "Product Quality and Safety": 0.716889907916387,
      "Selling Practices and Product Labeling": 0.6863646686077118,
      "Supply Chain Management": 0.8424439405401548,
      "Systemic Risk Management": 0.3190341554582119,
      "Waste and Hazardous Materials Management": 0.6823649227619171,
      "Water and Wastewater Management": 0.8415388464927673
    },
    "classes": {
      "Emission": 0.6417238499141402,
```

```
        "Resource Use": 0.7619518846273422,
        "Innovation": 0.681599784642458,
        "Workforce": 0.7450935386476063,
        "Community": 0.4230346532602762,
        "Human Rights": 0.5719779746399986,
        "Product Responsibility": 0.6573968736661806,
        "Management": 0.22641367923754913,
        "Shareholders": 0.5247782895134555
    },
    "distribution": {
        "E": 0.6950918397279802,
        "S": 0.5993757600535154,
        "G": 0.37559598437550235
    },
    "executionTime": "43 Seconds"
}
```


Average Time for Processing: 3-4 minutes.

5.3 CHALLENGES DURING IMPLEMENTATION

1. Integration of Diverse Technologies:
   Combining multiple technologies and languages, including Python, Flask, Java libraries, HTML, CSS, JavaScript, MongoDB, Yahoo Finance API, and Hugging Face models, can be complex. Ensuring seamless integration and communication between these components is a significant challenge.

2. Data Retrieval and Parsing:
   Obtaining ESG reports from various sources and parsing them effectively is a challenge, as these reports may come in different formats, structures, and languages. Consistency in data retrieval and handling is crucial for accurate analysis.

3. Natural Language Processing Complexity:
   Implementing NLP techniques for text analysis, including sentiment analysis, keyword extraction, and entity recognition, can be challenging due to the complexity of the algorithms involved. Ensuring accurate and meaningful results requires a deep understanding of NLP.

4. Scalability:
   As the project may involve processing a large volume of ESG reports and user requests, ensuring scalability is vital. The system should handle increased loads efficiently and effectively.

5. Real-Time Analysis:
   Providing real-time or near-real-time analysis results while maintaining low latency can be a significant technical challenge, especially when handling large datasets and complex NLP models.

6. User Interface Design:
   Creating an intuitive and user-friendly web interface (HTML, CSS, JavaScript) that accommodates various user preferences and customizations can be challenging. Balancing functionality with usability is crucial.

7. Error Handling and Anomaly Detection:
   Developing mechanisms for detecting anomalies in ESG reports and implementing effective error handling procedures to maintain system

reliability is challenging, as it involves both NLP and machine learning components.

8. Testing and Validation:
   Comprehensive testing, including unit testing, integration testing, and validation against real-world ESG data, is essential to ensure the accuracy and reliability of the analysis results.

9. Documentation and User Training:
   Creating comprehensive documentation and user training materials is time-consuming but crucial for enabling users to effectively utilize the system.

10. External API Integration:
    Integrating external APIs like Yahoo Finance for financial data retrieval may introduce challenges related to data consistency, availability, and API limitations.

11. Model Updates and Maintenance:
    NLP models provided by Hugging Face may require regular updates and maintenance to ensure they remain effective and accurate over time.

12. Cross-Browser Compatibility:
    Ensuring that the web-based interface works seamlessly across different web browsers can be challenging due to varying browser standards and behaviours.

Addressing these challenges during the implementation phase requires careful planning, collaboration among team members, and continuous monitoring and testing to ensure the project's success.

CHAPTER 6

RESULTS AND DISCUSSIONS

- PROJECT RESULTS AND ANALYSIS
- COMPARISON WITH PROJECT OBJECTIVES

6.1 PROJECT RESULTS AND ANALYSIS

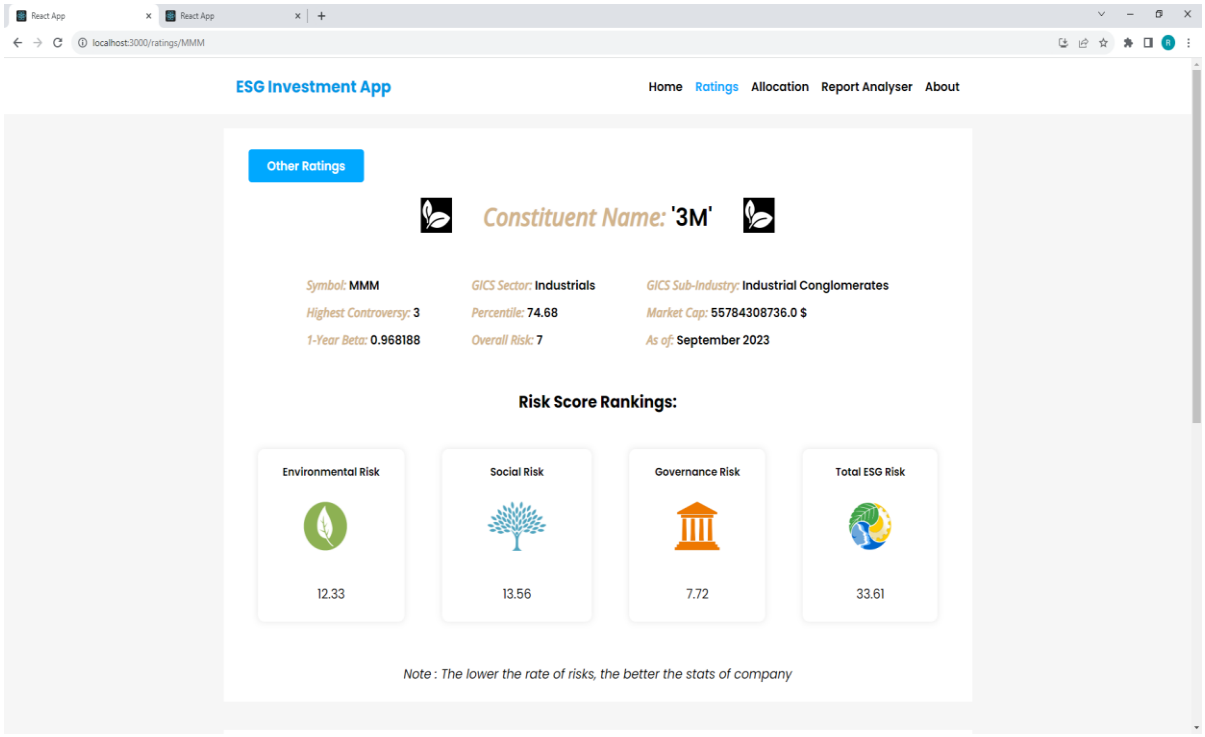


Figure-6.1.1 S & P 500 Constituents Page Output

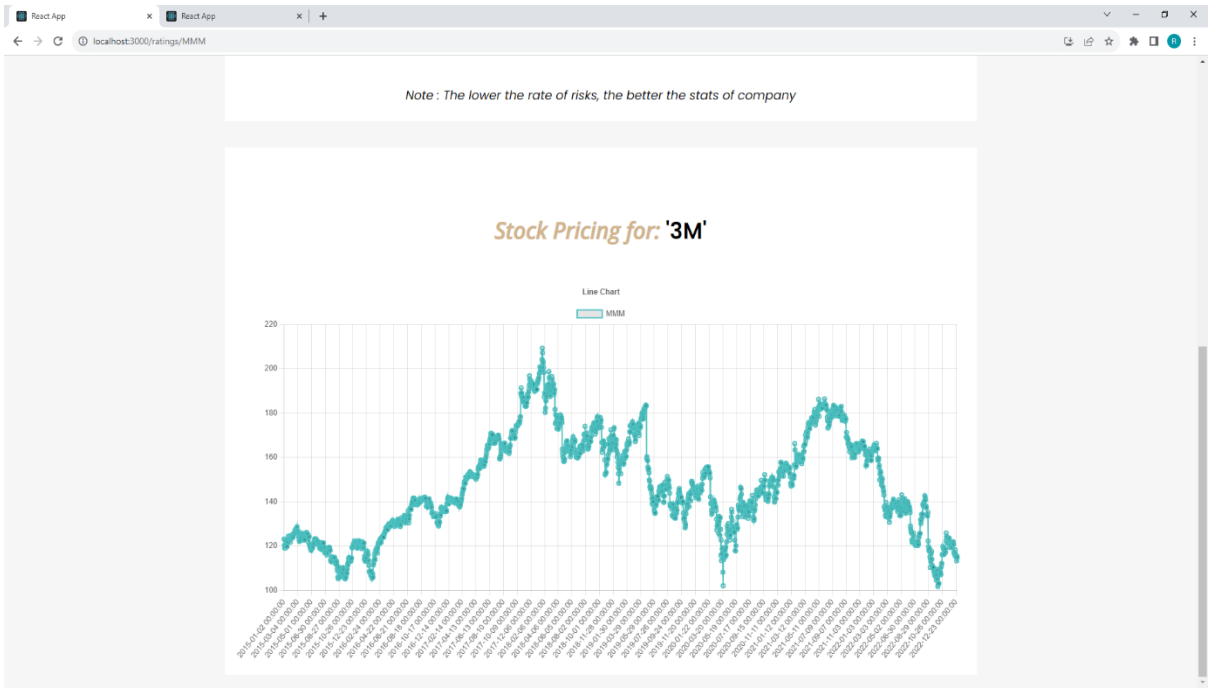


Figure-6.1.2 S & P 500 Constituents Page Output

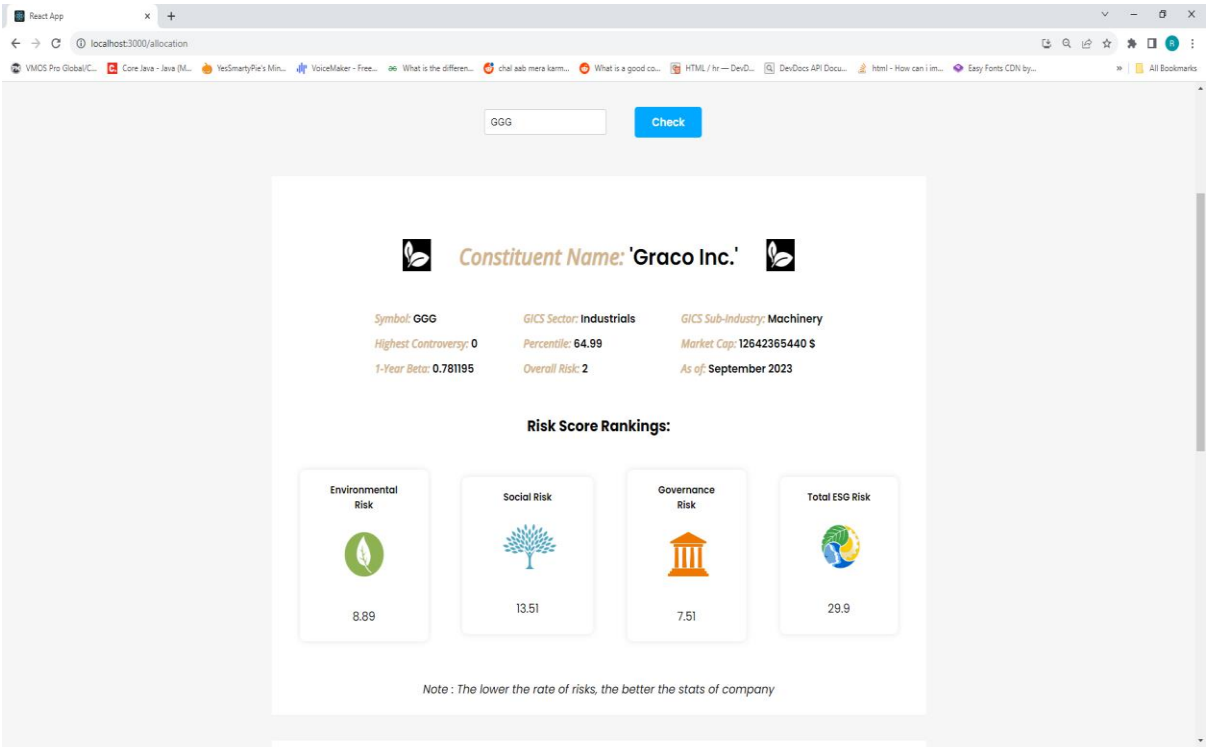


Figure-6.1.3 Dynamic Ratings Page Output

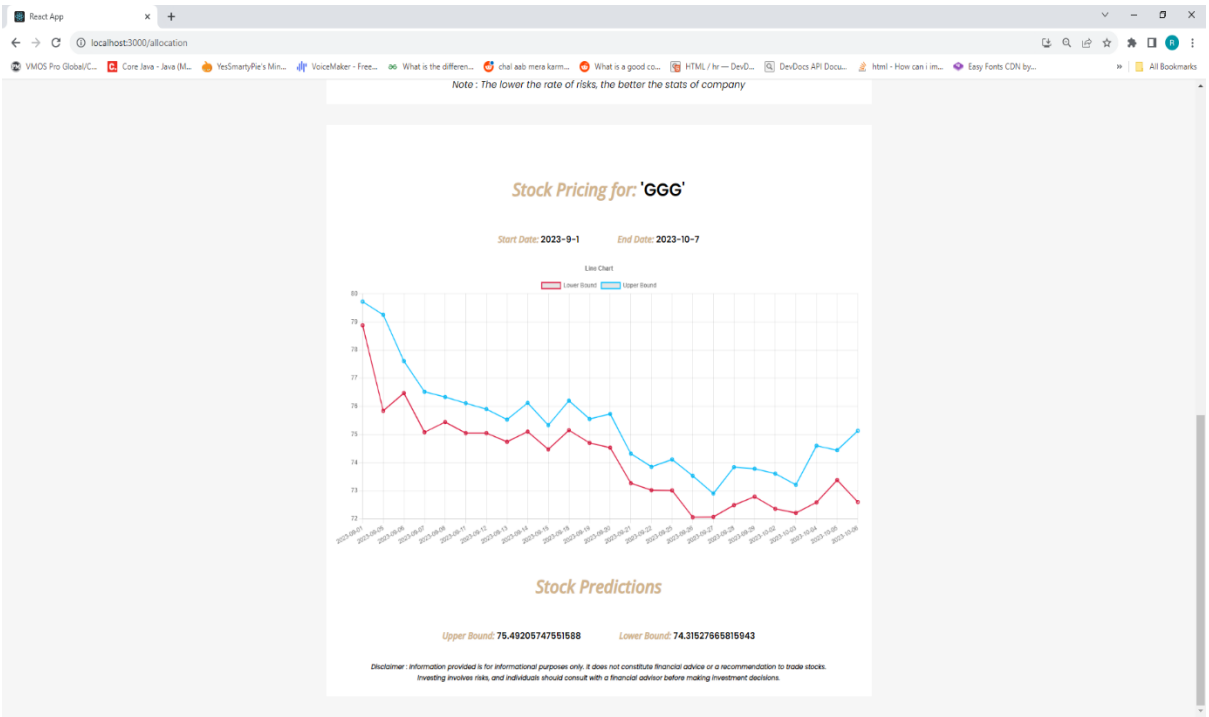


Figure-6.1.4 NLP Approach Page Output

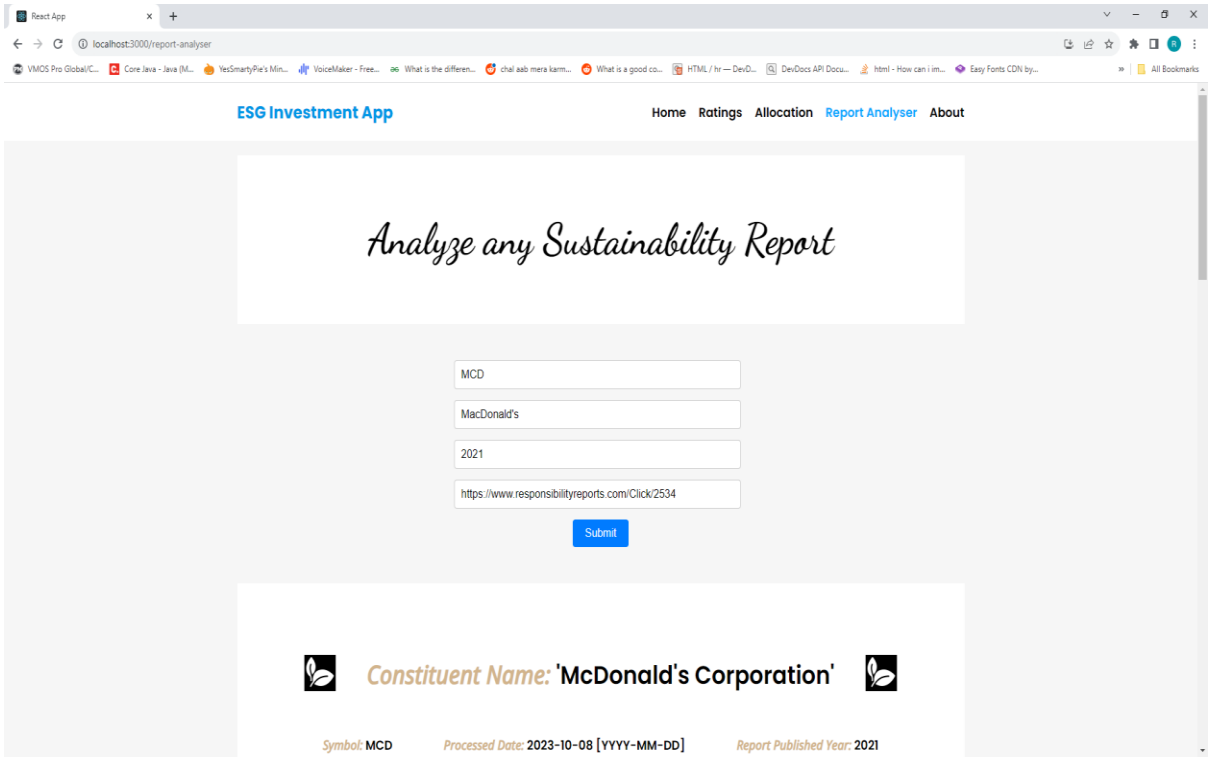


Figure-6.1.5 Report Analyser Page Output

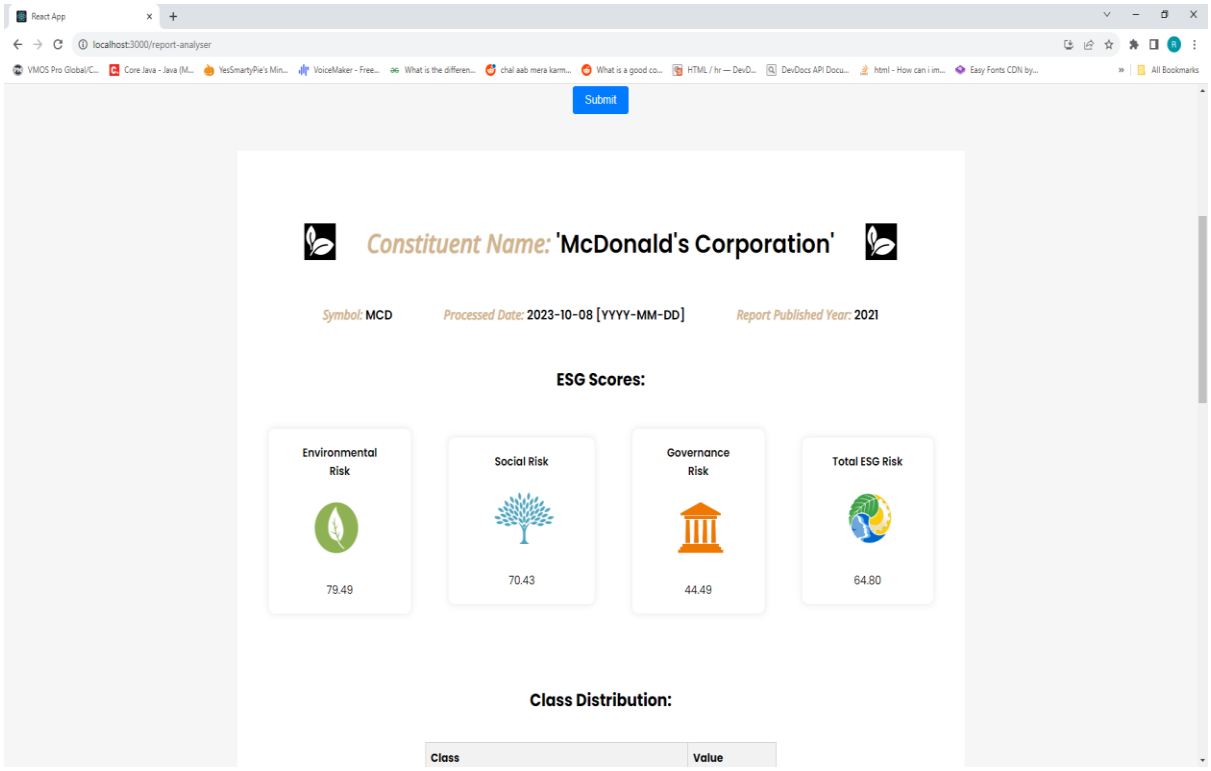
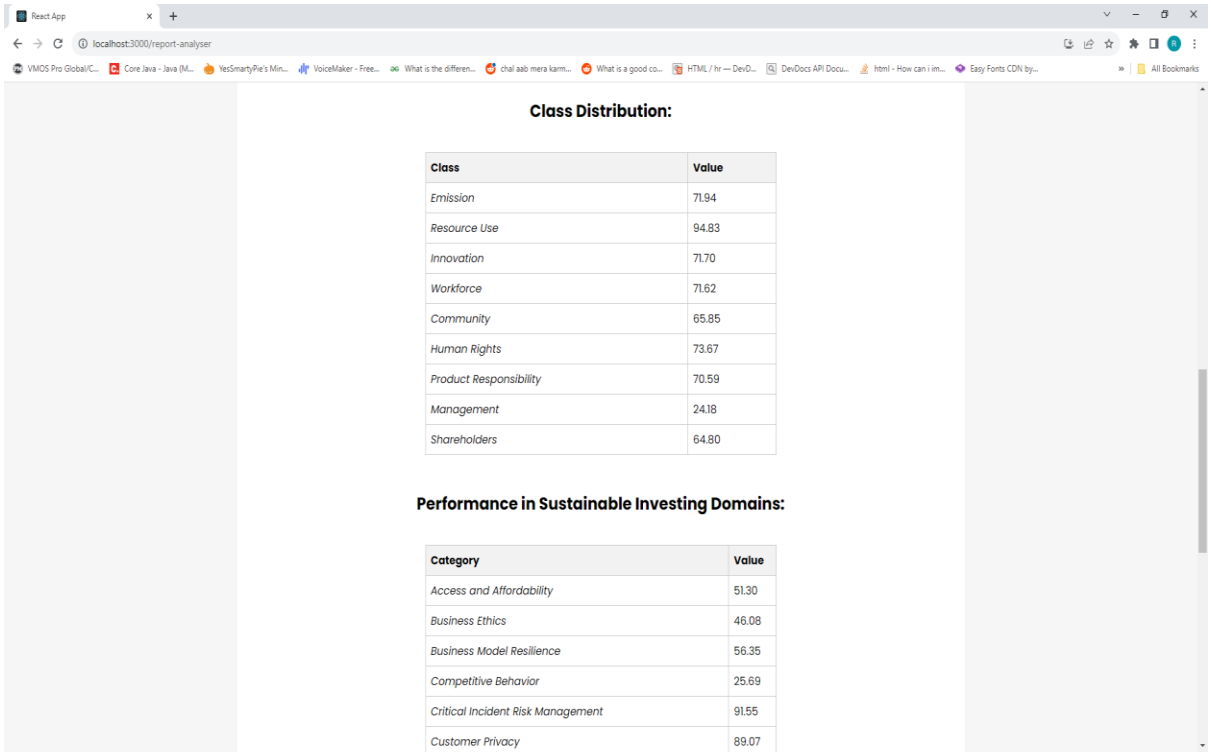


Figure-6.1.6 Report Analyser Page Output



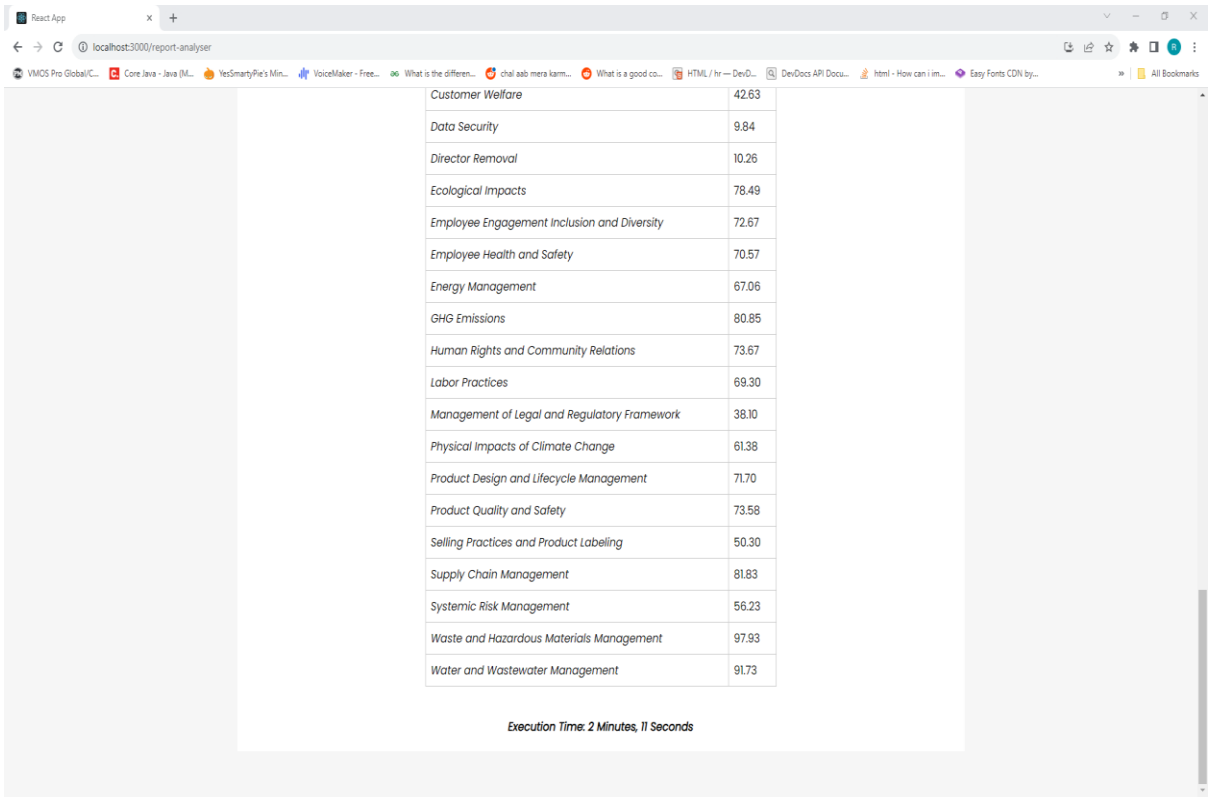
Class Distribution:

Class	Value
Emission	71.94
Resource Use	94.83
Innovation	71.70
Workforce	71.62
Community	65.85
Human Rights	73.67
Product Responsibility	70.59
Management	24.18
Shareholders	64.80

Performance in Sustainable Investing Domains:

Category	Value
Access and Affordability	51.30
Business Ethics	46.08
Business Model Resilience	56.35
Competitive Behavior	25.69
Critical Incident Risk Management	91.55
Customer Privacy	89.07

Figure-6.1.7 Report Analyser Page Output



Customer Welfare	42.63
Data Security	9.84
Director Removal	10.26
Ecological Impacts	78.49
Employee Engagement Inclusion and Diversity	72.67
Employee Health and Safety	70.57
Energy Management	67.06
GHG Emissions	80.85
Human Rights and Community Relations	73.67
Labor Practices	69.30
Management of Legal and Regulatory Framework	38.10
Physical Impacts of Climate Change	61.38
Product Design and Lifecycle Management	71.70
Product Quality and Safety	73.58
Selling Practices and Product Labeling	50.30
Supply Chain Management	81.83
Systemic Risk Management	56.23
Waste and Hazardous Materials Management	97.93
Water and Wastewater Management	91.73

Execution Time: 2 Minutes, 11 Seconds

Figure-6.1.8 Report Analyser Page Output

6.2 COMPARISON WITH PROJECT OBJECTIVES

After successfully completing ESG (Environmental, Social, and Governance) report analysis project, we can compare the achieved outcomes with the initial project objectives to assess its success and measure the degree to which each objective has been met. Here's how we can make the comparisons:

- ✓ Objective 1: Automated ESG Report Analysis
  Comparison: Evaluate the extent to which the system automates ESG report analysis. Measure the reduction in manual effort required to process and analyze reports.

- ✓ Objective 2: Data Extraction and Classification
  Comparison: Analyze the accuracy and efficiency of NLP algorithms in extracting relevant information and categorizing reports by industry or region.

- ✓ Objective 3: Trend Identification
  Comparison: Assess the system's effectiveness in identifying recurring themes and emerging trends in ESG reporting across different industries and regions.

- ✓ Objective 4: Anomaly Detection
  Comparison: Measure the system's success in detecting anomalies or irregularities in ESG reports and how well it alerts users to potential risks or non-compliance issues.

- ✓ Objective 5: User-Friendly Interface
  Comparison: Gather user feedback on the web-based interface's ease of use, accessibility, and functionality compared to the project's user interface objectives.

- ✓ Objective 6: Customization and Reporting
  Comparison: Evaluate the extent to which users can customize analysis criteria and generate detailed ESG analysis reports as specified in the objectives.

- ✓ Objective 7: Scalability
  Comparison: Measure the system's capacity to scale efficiently, accommodating a growing volume of ESG reports and users, and compare it to the scalability objectives.

- ✓ Objective 8: Compliance
  Comparison: Assess the system's compliance with relevant data privacy regulations (e.g., GDPR, CCPA) and industry standards governing ESG reporting, comparing it to compliance objectives.

- ✓ Objective 9: Performance Optimization
  Comparison: Analyze system performance to ensure it provides real-time or near-real-time analysis results with a focus on responsiveness and low latency, comparing it to performance optimization objectives.

- ✓ Objective 10: Documentation and Training
  Comparison: Evaluate the comprehensiveness and effectiveness of user documentation and training materials in facilitating the use of the ESG report analysis system.

By comparing the project outcomes to these objectives, we can provide a clear assessment of the project's success and identify areas where it has excelled or where improvements may be needed. This comparison will also help in demonstrating the project's value in meeting its intended goals and benefiting stakeholders.

CHAPTER 7

CONCLUSION

- SUMMARY OF THE PROJECT
- FUTURE WORK AND RECOMMENDATIONS

7.1 SUMMARY OF THE PROJECT

This project is dedicated to the development of a robust ESG (Environmental, Social, and Governance) report analysis system, leveraging a diverse stack of technologies. It combines HTML, CSS, and JavaScript for creating a user-friendly web interface. Python, alongside the Flask framework, serves as the backbone for server-side logic and data handling. MongoDB acts as the database for efficient data storage.

At the heart of the project lies the ESG-Bert algorithm, a specialized NLP model built upon the BERT architecture. This algorithm powers the analysis of ESG reports, enabling sentiment analysis, keyword extraction, and entity recognition, thus providing meaningful insights into sustainability and ethical practices of companies.

Furthermore, the project integrates the TIKA Java library to effectively extract textual data from various document formats, ensuring compatibility with diverse ESG report sources.

The significance of this project extends to promoting informed decision-making and sustainable business practices. It empowers users, including investors, regulators, and sustainability analysts, to assess companies' ESG performance. The system's capabilities encompass data ingestion, preprocessing, analysis, and customizable reporting, delivering actionable information in real-time. By addressing the challenges associated with ESG report analysis, this project contributes to transparency, ethical investment, and corporate responsibility.

In summary, this project's comprehensive approach leverages a range of technologies to enable efficient, accurate, and user-friendly analysis of ESG reports, supporting stakeholders in making informed decisions and fostering sustainable business practices.

7.2 FUTURE WORK AND RECOMMENDATIONS

Future Work:

- Enhanced NLP Models:
  Explore the use of more advanced NLP models beyond ESG-Bert, such as GPT-4 or RoBERTa, to improve the accuracy and depth of textual analysis in ESG reports.

- Machine Learning Integration:
  Incorporate machine learning techniques for advanced anomaly detection, predictive analysis, and trend forecasting based on historical ESG data.

- Real-Time Data Streaming:
  Implement real-time data streaming capabilities to enable continuous analysis of ESG reports as they are published, providing stakeholders with up-to-the-minute insights.

- Multi-Language Support:
  Extend language support for ESG report analysis to accommodate reports in multiple languages, enhancing the system's global applicability.

- User Collaboration:
  Introduce collaboration features that allow multiple users to collaborate on ESG analysis projects and share findings and reports.

- API Integrations:
  Integrate with more data sources and APIs beyond Yahoo Finance, expanding the range of financial and ESG-related data available for analysis.

- Advanced Visualization:
  Enhance data visualization capabilities with interactive and customizable charts, graphs, and dashboards for in-depth analysis.

- Scalability and Cloud Deployment:
  Prepare the system for cloud deployment, ensuring scalability to handle larger datasets and a growing user base.

Recommendations:

- User Training and Support:
  Offer comprehensive user training and support materials to assist users in navigating and maximizing the system's capabilities effectively.

- Feedback Mechanism:
  Implement a feedback mechanism to collect user feedback and feature requests, allowing for continuous improvement of the system based on user needs.

- Documentation Updates:
  Keep user documentation and system documentation up-to-date to reflect any changes, enhancements, or new features in the system.

- Community Engagement:
  Foster a community of users and developers interested in ESG analysis to share insights, best practices, and collaborative projects.

- Collaboration with ESG Initiatives:
  Collaborate with ESG-focused organizations, research institutions, and industry bodies to align the system with emerging ESG standards and reporting requirements.

- Compliance Monitoring:
  Continuously monitor and adapt to changes in data privacy regulations and ESG reporting standards to ensure ongoing compliance.

- Performance Optimization:
  Regularly assess and optimize system performance to maintain responsiveness and low latency, especially as the user base grows.

- Market Analysis:
  Stay informed about the evolving landscape of ESG reporting and analysis tools to remain competitive and innovative in the field.

By considering these future work possibilities and recommendations, our project can evolve to meet the changing needs of users and the dynamic ESG analysis landscape, ensuring its continued relevance and impact.

REFERENCES

REFERENCES

1. https://www.bloomberg.com/professional/blog/esg-assets-may-hit-53-trillion-by-2025-a-third-of-global-aum/#:~:text=GSIA%2C%20Bloomberg%20Intelligence
2. https://www.sciencedirect.com/science/article/pii/S187770581400349X
3. https://onlinelibrary.wiley.com/doi/10.1002/bse.1765
4. https://www.sciencedirect.com/science/article/abs/pii/S0264999315002667
5. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2471514
6. https://link.springer.com/article/10.1007/s10551-006-9253-8
7. https://en.wikipedia.org/wiki/List_of_S%26P_500_companies
8. https://www.responsibilityreports.com/
9. https://doi.org/10.5539/ijbm.v12n9p53
10. https://www.symanto.com/blog/the-emerging-role-of-natural-language-processing-in-esg-analytics/
11. https://deepai.org/publication/detecting-esg-topics-using-domain-specific-language-models-and-data-augmentation-approaches
12. https://www.amenityanalytics.com/blog-articles/5-use-cases-for-nlp-in-esg
13. https://www.refinitiv.com/perspectives/future-of-investing-trading/how-can-nlp-be-used-to-quantify-esg-analytics/
14. https://arxiv.org/ftp/arxiv/papers/2203/2203.16788.pdf
15. https://github.com/RikinZala25/ESG-Report-Analysis/blob/main/Jupyter%20Notebook/outputs_csv/s%26p500_esg_data.csv