Exploring portfolio decarbonization using AI

Documentation

By Group 7

Team members:
Ho Jheng-Yuan
Ho Lok Sang Kelvin
Lu Xinyi
Soh Jun Heng
Toh Hui Shan Alicia

2021
Project Title: NUS Capstone Project
Project Supervisor: Mr. Puneet Gupta, Prof. Lee Boon Kee

Table of Content

1 Solution Description **3**

1.1 Solution Overview 3

1.2 User Flow 3

1.3 Data Preprocessing 4

1.4 Machine Learning Algorithms 5

1.4.1 Decision Tree and Linear Regression for Sentiment Analysis 5

1.4.2 LDA for Topic Modelling 5

1.4.3 Bigram Analysis 5

2 Solution Architecture **6**

2.1 Technology Stack 6

2.2 System Requirements 6

2.3 Functions 8

2.3.1 Data Preprocessing 8

2.3.2 Model Building 11

2.3.3 Data Analysis 12

2.3.4 Dashboard & Visualisation 16

3 User guide **18**

3.1 Step 1: Cloning GitHub 18

3.2 Step 2: Setup the Repository 18

3.3 Step 3: Updating Dataset (if necessary) 19

4 Dashboard Documentation **20**

4.1 Dashboard GUI 20

4.2 Dashboard Features 20

4.2.1 First Interface 20

4.2.2 Second Interface 21

5 Appendices **22**

5.1 Current List of Dataset 22

5.2 Full list of environment.yml 31

5.3 Screenshots of the Dashboard 35

5.3.1 First Interface 35

5.3.2 Second Interface 35

1 Solution Description

1.1 Solution Overview

Our group would like to combine machine learning algorithms and data visualisation using Python to solve this issue. Firstly, we preprocess all the reports from different companies and select key information from those reports using PyMuPDF. Secondly, with different machine learning algorithms such as Linear Regression, Decision Tree and Latent Dirichlet Allocation, we get the information of each firm's actions on decarbonization and give a score based on their effort. Lastly, we create a dashboard with several filters for users to view those data in a user-friendly way, hence assisting them on making decisions on portfolio decarbonization.

1.2 User Flow

The project is designed for users (i.e NatWest Markets and its clients), and their user flow is shown in Figure 1.

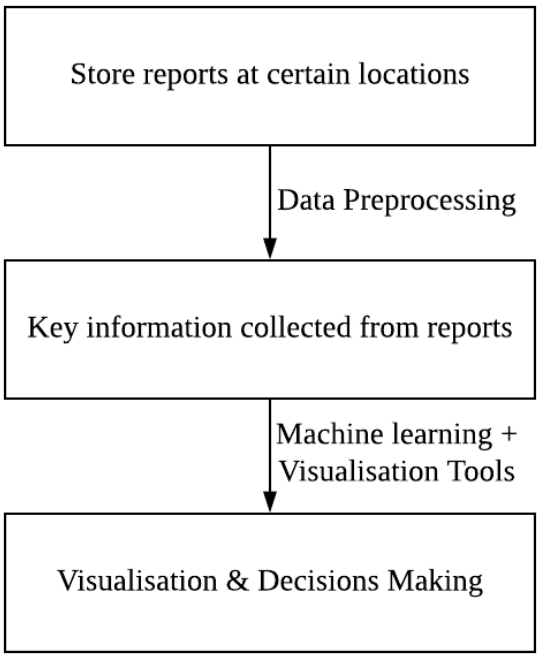


Figure 1. User flow for the project

Users would first have to store all companies' reports and/or choose to use the files we collected to a place where the framework can access. After settling all those files to the right repository, users can run the code to access all those reports to be analysed. The data preprocess part will drop stopwords(e.g. 'The', 'it', etc.) and filter out decarbonization-related keywords. It will also categorise those keywords based on its optimism on decarbonization. With all those different machine learning models, we will generate a score for each firm. Lastly, all information will be displayed in a dashboard, with several filters available for users to set. With those insights, we believe users can make better and more precise decisions.

1.3 Data Preprocessing

The data preprocessing workflow is shown in Figure 2.

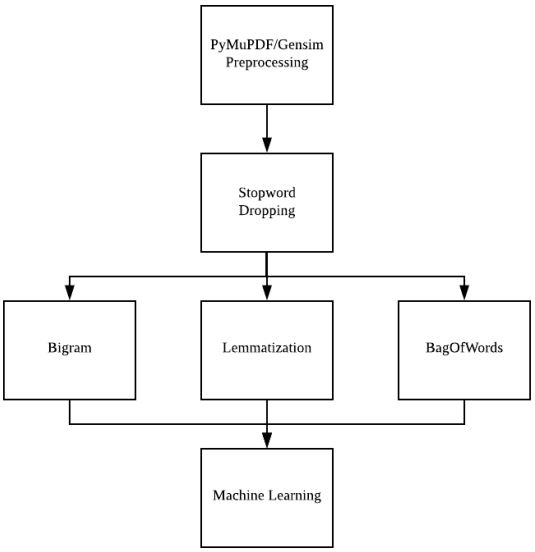


Figure 2. Data Preprocessing Workflow

First, the code will read sustainability reports in PDF format stored in a repository. Then, we will use both the PyMuPDF and Gensim package to preprocess these reports. PyMuPDF helps us parse the information in the report, while Gensim helps us tokenize texts and drop the punctuations from the text data. Then, we use the NLTK package to drop stopwords (e.g. I, we, etc.). Lastly, we get bigrams, bag of words, and conduct lemmatization on the data to finish the preprocessing. Those information generated by this step will be passed into our machine learning models.

1.4 Machine Learning Algorithms

1.4.1 Decision Tree and Linear Regression for Sentiment Analysis

For sentiment analysis, we built a regression model to predict decarbonization sentiments of a company using algorithms such as Decision Tree and Linear Regression. The decision tree model achieved a better performance and is being implemented in our solution. We imported the packages from sklearn. The accuracy of the decision tree is 0.823.

1.4.2 LDA for Topic Modelling

The Latent Dirichlet Allocation Mallet (LDAMallet) algorithm from the Gensim package was used in our solution. The chosen model with 12 topics attained the highest coherence score of 0.7024.

1.4.3 Bigram Analysis

Bigram Analysis is conducted to extract the top occurring decarbonization related word pairs. This is conducted using TfidfVectorizer from the sklearn library.

2 Solution Architecture

This project is a python-based code build for laptop / desktop users. It can run on any device with the code installed. There will be no database to store those data, but downloading the models or filtered data can be done to save time for preprocessing and processing if required.

2.1 Technology Stack

Architecture	**Technology Used**
Repository	GitHub
Backend	Python
External Database	N/A
Development & Deployment	Python
Frontend Visualisation	Python
Dashboard deployment	Heroku

2.2 System Requirements

The key system requirements are listed below, while users can install all the requirements using the environment.yml. Users can also view all of the requirements by opening the environment.yml file.

Stages	**Package name**[1]	**Version**
Data Preprocessing, Data Analysis	Gensim	3.6.0
	Matplotlib	3.4.3
	nltk	3.6.5

[1] There are 79 packages used in the environment.yml. Only the top 6 packages are listed in this table.

	seaborn	0.11.2
	sklearn	0.0
	spacy	3.2.0

The requirements below are not included in environment.yml as they can be installed using other commands when creating the dashboard (refer to deploy-dashboard-guide.md)

Stages	**Package name**	**Version**
Dashboard deployment (to be installed on virtual environment in dashboard repository)	Brotli	1.0.9
	click	8.0.3
	colorama	0.4.4
	dash	2.0.0
	dash-bootstrap-components	1.0.0
	dash-core-components	2.0.0
	dash-html-components	2.0.0
	dash-table	5.0.0
	Flask	2.0.2
	Flask-Compress	1.10.1
	gunicorn	20.1.0
	importlib-metadata	4.8.1
	itsdangerous	2.0.1
	Jinja2	3.0.2
	MarkupSafe	2.0.1
	numpy	1.21.4

	pandas	1.3.4
	plotly	5.3.1
	python-dateutil	2.8.2
	pytz	2021.3
	six	1.16.0
	tenacity	8.0.1
	typing-extensions	3.10.0.2
	Werkzeug	2.0.2
	zipp	3.6.0

2.3 Functions

2.3.1 Data Preprocessing

Function Name	**Description**
`read_pdf(fi)`	To extract text from each company's report for a specific type of financial institution. **Input:** A *string* that specifies the type of financial institution. Eg. 'asian_banks', 'asset_managers', 'insurance' or 'pension_funds' **Output:** A *dataframe* with 3 columns: name, sentence, year that contains the extracted text.
`convert_pdf_into_sentences (text)`	To convert a string of report into individual sentence strings **Input:** A *string* with all text from a single report

	Output: A *list* of sentences from the same report
`clean_text(df)`	To separate all extracted text into individual sentences, as well as remove unnecessary line breaks and spaces. **Input:** A *dataframe* containing all the raw extracted text of a given company's report. **Output:** A *dataframe* containing the individual sentences of a given company's report.
`extract_text(path)`	Combines all text extraction functions into a single function to be executed. **Input:** A *string* that specifies the type of financial institution. Eg. 'asian_banks', 'asset_managers', 'insurance' or 'pension_funds' **Output:** A *dataframe* with 3 columns: name, sentence, year containing the individual sentences of a given financial institution's report.
`sent_to_words(sentences)`	To use Gensim's default preprocess function to clean the given sentences. This lowercases, tokenizes, de-accents the input and the outputs are final tokens of unicode strings. **Input:** A *list* of sentences. **Output:** A *list* of tokens.
`create_stopwords()`	Compile all stop words which includes stopwords in NLTK, names of companies in our dataset and context specific stop words such as 'report'. **Input:** -

	Output: A *list* of stop words.
`remove_stopwords(texts)`	To remove stopwords from the list of text. **Input:** A *list* of text. **Output**: A new *list* of text with stopwords dropped
`create_bigram_mod(data_words)`	To create a bigram model with the created list of words. **Input:** List of tokens generated from sent_to_words() **Output**: A bigram model.
`make_bigrams(texts)`	Create bigram models. Bigrams are two words frequently occurring together in the document. **Input:** A *list* of text with stopwords dropped. **Output**: A new *list* of bigrams.
`lemmatization(texts)`	To lemmatize the list of text. In lemmatization, words are returned to their dictionary form. **Input:** A *list* of bigrams. **Output**: A new *list* of lemmatized text.
`bag_of_words(df)`	Create a bag of word model for the input dataframe and output is a continuous variable. **Input:** A *dataframe.* **Output**: A *dataframe* that shows the counts of features for each row.

`bag_of_words_discrete(df)`	Create a bag of word model for the input dataframe and output is a discrete variable. **Input:** A *dataframe.* **Output**: A *dataframe* that shows the counts of features for each row.

2.3.2 Model Building

Function Name	**Description**
`word2vec(df,company_name,*args)`	Given a keyword, the Word2Vec model will generate the top 50 words in a specific company's report that are most similar to it. **Input:** *dataframe* containing all extracted text, *string* company's name, *string* *args to specify the keyword that the user is interested in. **Output:** *list* of the top 50 most similar words.
`tsnescatterplot(model, word, list_names)`	To generate a visualisation that helps to give the user a visual representation on the similarity of words. Words that are more similar will be closer to each other in the scatterplot. **Input:** *model* used to determine word similarity, *string* to specify keyword, *list* of similar words. **Output:** Scatter plot that displays relationship between keywords and its similar words.
`word2vec_tsne(df,company_name,args)`	Combines both functions above. **Input:** *dataframe* containing all extracted text, *string* company's

	name, *string* *args to specify the keyword that the user is interested in. **Output:** *list* of the top 50 most similar words and a scatter plot that displays relationship between keywords and its similar words.
`compute_coherence_values(dictionary, corpus, texts, limit, start, step)`	To train multiple LDA models with varying numbers of topics and provide the models and their corresponding coherence scores as output. **Input:** Gensim dictionary, Gensim corpus, *list* of input text, *int* limit to specify the maximum number of topics, *int* start to specify minimum number of topics and *int* step to represent the number of topics to increase when training the next model. **Output:** *list* of LDA topic model and *list* of coherence values corresponding to the LDA model with the respective number of topics.

2.3.3 Data Analysis

Function Name	**Description**
`find_e_topics(topics)`	To determine which topics derived from the optimal LDA model are related to decarbonization. **Input:** *list* of topics and associated keywords. **Output:** *list* containing topic numbers that are related to decarbonization.
`format_topics_sentences(ldamodel, corpus=corpus, texts=data)`	To assign a dominant topic to each sentence of each company's report.

	Input: *model* to specify optimal LDA model, Gensim corpus and *list* of input text. **Output:** *dataframe* that has a column indicating which topic each sentence belongs to.
`assign_esg(df)`	To determine which sentences are related to decarbonization based on its assigned dominant topic. **Input:** *dataframe* with a column that indicates which topic each sentence belongs to. **Output:** *dataframe* with a column that indicates which category (E/SG) each sentence belongs to.
`extract_initiatives(data)`	To extract the global standards and initiatives relating to sustainable finance/decarbonization that a company is committed to. **Input:** *dataframe* containing all extracted text **Output:** *dataframe* with following columns: name, initiatives, count, type. Initiatives is an *array* containing all extracted initiatives name and count is an *int* representing total number of initiatives and standards of each company.
`bigram(sub_df)`	To extract top ten occurring bigrams (word pairs) from lemmatized sentences. **Input:** *dataframe* containing all extracted text which have undergone data preprocessing and lemmatization. **Output:** *list* containing ten tuples. Each tuple consists of a bigram and its corresponding TF-IDF score.

`generate_sentiment_score(df)`	To generate the sentiment score for each company using Textblob library and frequency of ESG related keywords. **Input:** *dataframe* containing all extracted text. **Output:** *dataframe* with two newly added columns: sentiment_score and optimism. Sentiment_score indicates the generated sentiment level and optimism indicates the label of the corresponding score (neutral, optimistic, pessimistic).
`decision_tree_final(df)`	To predict the sentiment score of each company using the decision tree algorithm. **Input:** *dataframe* that is generated by the bag of words model. **Output:** *dataframe* with a newly added column: predicted_sentiment_tree which indicates the sentiment score predicted by the decision tree and a *float* that displays the Root Mean Square Error of the model.
`linear_regression(df)`	To predict the sentiment score of each company using linear regression. **Input:** *dataframe* that is generated by the bag of words model. **Output:** *dataframe* with a newly added column: predicted_sentiment_lr which indicates the sentiment score predicted by linear regression and a *float* that displays the Root Mean Square Error of the model.
`rmse_lr(data, type_of_fi)`	To calculate the Root Mean Square Error (RMSE) of the linear regression model used in sentiment analysis. **Input:** *dataframe* that contains the predicted sentiment scores and *string* type_of_fi indicating which financial institution the

	RMSE is applicable to. **Output:** A *float* representing the RMSE of the model.
`mae_lr(data, type_of_fi)`	To calculate the Mean Absolute Error (MAE) of the linear regression model used in sentiment analysis. **Input:** *dataframe* that contains the predicted sentiment scores and *string* type_of_fi indicating which financial institution the MAE is applicable to. **Output:** A *float* representing the MAE of the model.
`rmse_dt(data, type_of_fi)`	To calculate the Root Mean Square Error (RMSE) of the decision tree algorithm used in sentiment analysis. **Input:** *dataframe* that contains the predicted sentiment scores and *string* type_of_fi indicating which financial institution the RMSE is applicable to. **Output:** A *float* representing the RMSE of the model.
`mae_dt(data, type_of_fi)`	To calculate the Mean Absolute Error (MAE) of the decision tree algorithm used in sentiment analysis. **Input:** *dataframe* that contains the predicted sentiment scores and *string* type_of_fi indicating which financial institution the MAE is applicable to. **Output:** A *float* representing the MAE of the model.
`decision_tree_evaluation(df)`	To generate the evaluation metrics of the decision tree classifier. **Input:** *dataframe* that is generated by the bag of words model.

	Output: A *float* indicating the accuracy of the model, *plot* displaying the confusion matrix and a classification report which includes the precision, recall and F1-Score of the model.
`logistic_regression_evaluation(df)`	To generate the evaluation metrics of the logistic regression model. **Input:** *dataframe* that is generated by the bag of words model. **Output:** A *float* indicating the accuracy of the model, *plot* displaying the confusion matrix and a classification report which includes the precision, recall and F1-Score of the model.
`naive_bayes_evaluation(df)`	To generate the evaluation metrics of the naive bayes classifier. **Input:** *dataframe* that is generated by the bag of words model. **Output:** A *float* indicating the accuracy of the model, *plot* displaying the confusion matrix and a classification report which includes the precision, recall and F1-Score of the model.

2.3.4 Dashboard & Visualisation

To view the dashboard on local machine, toggle to the directory containing the python script and run *dashboard.py*

Function Name	**Description**
`update_dropdown_tab1(type_of_fi)`	To update the dropdown menu in the first tab that allows users to select a company.
`update_graph_tab1_sentiment(type_of_fi, company)`	To update the sentiment gauge chart in the first tab.
`update_graph_tab1_percentage(type_of_fi, company)`	To update the bar chart displaying the decarbonization

	disclosure (%) in the first tab.
update_graph_tab1_initiativecount(type_of_fi, company)	To update the bar chart displaying the count number of global initiatives and standards in the first tab.
update_graph_tab1_initiativetable(company)	To update the table displaying all the global initiatives and standards in the first tab.
update_graph_tab1_bigram(company)	To update the bar chart displaying the top ten bigrams and respective TF-IDF score in the first tab.
update_dropdown_tab2_fi(type_of_fi)	To update the dropdown menu in the second tab that allows users to select the first company.
update_dropdown_tab2_company(company1, type_of_fi)	To update the dropdown menu in the second tab that allows users to select the second company.
update_graph_tab2_percentage(type_of_fi, company1, company2)	To update the pie charts comparing the decarbonization disclosure (%) in the second tab.
update_graph_tab2_sentiment(type_of_fi, company1, company2)	To update the bar chart comparing the sentiment level in the second tab.
update_graph_tab2_bigram1(company1)	To update the bar chart displaying the top ten bigrams of the first company in the second tab.
update_graph_tab2_bigram2(company2)	To update the bar chart displaying the top ten bigrams of the second company in the second tab.

3 User guide

There are two important python files, the first is ***WIN_BT4103_Decarbonization.ipynb***, which is an ipynb file that can be run on a jupyter notebook loaded from an anaconda environment. *WIN_BT4103_Decarbonization.ipynb* will handle the installation, pre-processing of data, machine learning and data analysis. This is important when new ESG reports are added and needs to be scraped and preprocessed for the first time in order to be visualised and rated on the dashboard.

The second file ***dashboard.py*** will create and update the dashboard using data generated from *WIN_BT4103_Decarbonization.ipynb*. To load the dashboard, run *dashboard.py*, which is a python file that can only be run locally via command prompt (Windows) and terminal (Mac). Navigate to the repository that was cloned on the device where *dashboard.py* is located via the command prompt (cmd), type "python dashboard.py" on the cmd window to run the dashboard on a local server. Copy the local server internet address in the format of (https://127.28…) onto your internet browser to view the dashboard.

An alternative to view the dashboard, use the link https://bt4103-esg-dashboard.herokuapp.com/

3.1 Step 1: Cloning GitHub

Users need to clone data from our github link (https://github.com/cl-xy/bt4103_esg) for one-time only. After the folder has been cloned, users can update files locally without touching git anymore.

3.2 Step 2: Setup the Repository

First, users need to put ***WIN_BT4103_Decarbonization.ipynb*** to the repository where they want to be. After the settings are done, users can import all packages required for *WIN_BT4103_Decarbonization.ipynb*. It can be done by using environment.yml file with the command:

```
conda env create -f environment.yml
```

3.3 Step 3: Updating Dataset (if necessary)

To upload new reports for analysis, users can simply add those reports to specific file folders based on the type of financial institution the company is. Similarly, users can delete unwanted files by accessing the four sub folders. There are 4 different folders namely asian_banks, asset_managers, insurance and pension_funds under the path, each corresponding to a type of financial institution. The path where the reports of the four respective financial institutions will be something like: /Users/user/Documents/GitHub/bt4103_esg/data/type/asian_banks.

Users should follow the current naming convention for any newly added files to ensure that the system is able to run smoothly.

- For companies with only one file in any given year, the naming convention is as follows:

  CompanyName-Year.pdf (eg. DBS-2020.pdf)

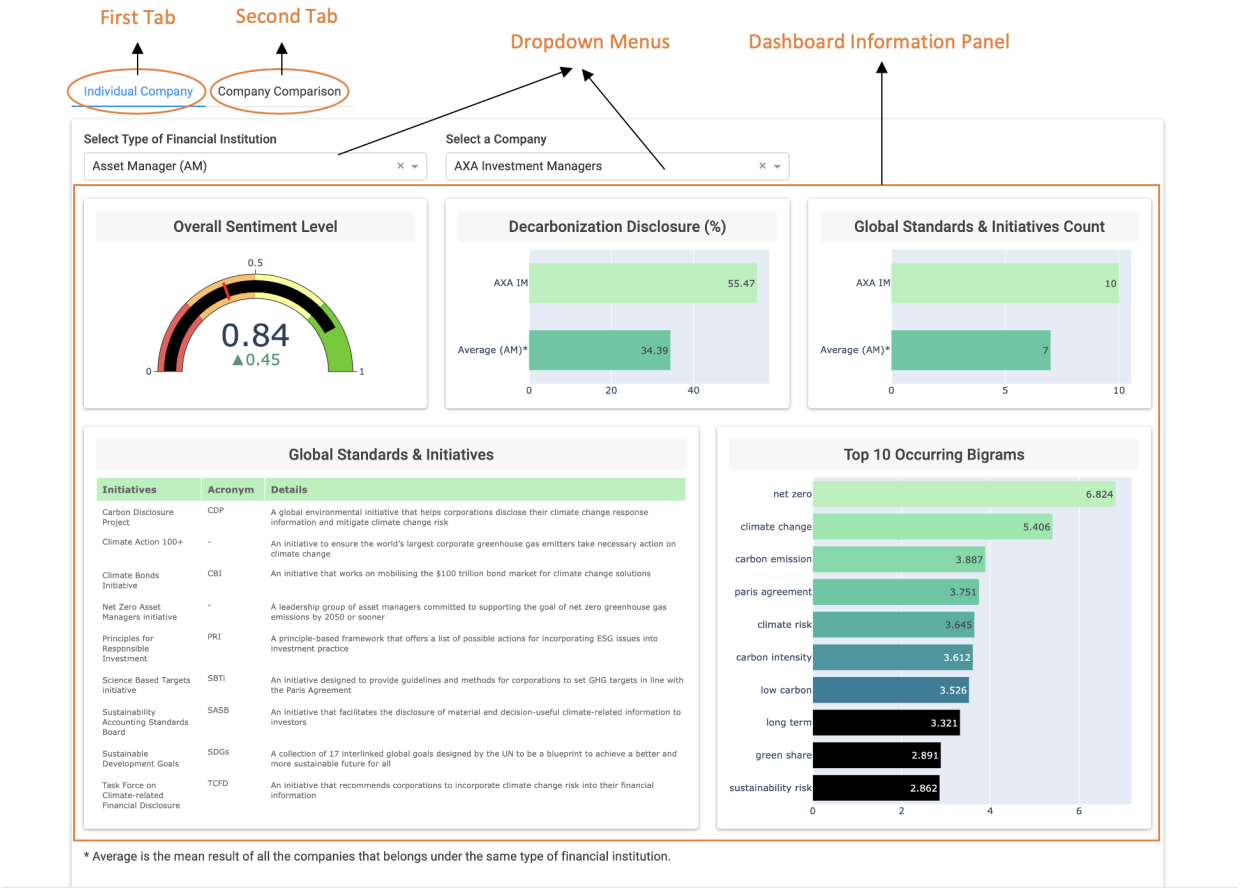
- For companies with multiple files in any given year, the naming convention is as follows:

  CompanyName-Number-Year.pdf (eg. Ashmore-2-2019.pdf, Ashmore-3-2019.pdf)

Whenever a new company is added into the dataset, users are required to update the companylabels.csv which is found under bt4103_esg/data/companylabels.csv. They simply need to enter the company's full name, short form of its name and the type of financial institution it belongs to (**ab** - Asian Banks; **am** - Asset Managers; **ins** - Insurance; **pf** - Pension Funds).

4 Dashboard Documentation

4.1 Dashboard GUI



4.2 Dashboard Features

4.2.1 First Interface

- Dropdown menu to select one out of four types of financial institution
- Dropdown menu with search option for user to select a company
- Gauge chart displaying overall sentiment level
- Bar chart displaying percentage of decarbonization disclosure of the company and the average
- Bar chart displaying the number of global initiatives and standards the company is committed to and the average
- Table displaying all the global standards and initiatives the company is a participant of

- Bar chart displaying top ten occurring bigrams of the company's report

4.2.2 Second Interface

- Dropdown menu to select one out of four types of financial institution
- Dropdown menu with search option for user to select the first company
- Dropdown menu with search option for user to select the second company
- Bar chart comparing the overall sentiment level between both companies and the average
- Pie charts displaying the breakdown of the percentage of decarbonization disclosure between both companies and the average
- Bar charts displaying the top ten occurring bigrams of each of the two selected companies

5 Appendices

5.1 Current List of Dataset

Asian Banks

File Name	Company	Description
AIIB-2020.pdf	Asian Infrastructure Investment Bank	2021 Sustainable Development Bond Framework
BDO-2018.pdf	BDO Unibank	2018 Sustainability Report
BDO-2019.pdf	BDO Unibank	2019 Sustainability Report
BDO-2020.pdf	BDO Unibank	2020 Sustainability Report
BEA-2020.pdf	Bank of East Asia	2020 ESG Report
BNI-2020.pdf	Bank Negara Indonesia	2020 Sustainability Report
BOC-2020.pdf	Bank of China	2020 Annual Report
BRI-2019.pdf	Bank Rakyat Indonesia	2019 Sustainability Report
Bangkok Bank-2019.pdf	Bangkok Bank	2019 Sustainability Report
CCB-2020.pdf	China Construction Bank	2020 CSR Report
CIMB-2017.pdf	CIMB	2017 Sustainability Report
CIMB-2018.pdf	CIMB	2018 Sustainability Report
CIMB-2019.pdf	CIMB	2019 Sustainability Report
CIMB-2020.pdf	CIMB	2020 Sustainability Report
CMB-2020.pdf	China Merchants Bank	2020 Sustainability Report
CNCBI-2020.pdf	China CITIC Bank	2020 Sustainability Report
Chong Hing-2020.pdf	Chong Hing Bank	2020 ESG Report
CitiBank-2021.pdf	Citibank	2021 Sustainability Report
DBS-2020.pdf	DBS Bank	2020 Sustainability Report
Dah Sing-2020.pdf	Dah Sing Bank	2020 ESG Report
HSBC-2020.pdf	HSBC	2020 Annual Report
Hana Bank-2020.pdf	Hana Financial Group	2020 Sustainability Report

Hong Leong-2019.pdf	Hong Leong Bank	2019 Sustainability Report
Hong Leong-2020.pdf	Hong Leong Bank	2020 Sustainability Report
ICBC-2020.pdf	Industrial and Commercial Bank of China	2020 CSR Report
KTB-2017.pdf	Krungthai Bank	2017 Sustainability Report
KTB-2018.pdf	Krungthai Bank	2018 Sustainability Report
KTB-2019.pdf	Krungthai Bank	2019 Sustainability Report
KTB-2020.pdf	Krungthai Bank	2020 Sustainability Report
MUFG-2020.pdf	MUFG Bank	2020 Sustainability Report
MUFG-2021.pdf	MUFG Bank	2021 Integrated Report
Maybank-2018.pdf	Maybank	2018 Sustainability Report
Maybank-2019.pdf	Maybank	2019 Sustainability Report
Maybank-2020.pdf	Maybank	2020 Sustainability Report
Maybank-2021.pdf	Maybank	2021 Sustainability Framework
OCBC-2018.pdf	OCBC Bank	2018 Sustainability Report
OCBC-2019.pdf	OCBC Bank	2019 Sustainability Report
OCBC-2020.pdf	OCBC Bank	2020 Sustainability Report
PSBC-2020.pdf	Postal Savings Bank of China	2020 CSR Report
RHB-2017.pdf	RHB Bank	2017 Sustainability Report
RHB-2018.pdf	RHB Bank	2018 Sustainability Report
RHB-2019.pdf	RHB Bank	2019 Sustainability Report
RHB-2020.pdf	RHB Bank	2020 Sustainability Report
SBI-2019.pdf	State Bank of India	2019 Sustainability Report
SCB-2020.pdf	Siam Commercial Bank	2020 Sustainability Report
Shinhan-2020.pdf	Shinhan Financial Group	2020 ESG Report
UOB-2020.pdf	United Overseas Bank	2020 Sustainability Report

Asset Managers

File Name	Company	Description
AXA IM-2020.pdf	AXA Investment Managers	2020 Active Ownership & Stewardship Report
AXA IM-2021.pdf	AXA Investment Managers	2021 TCFD Report
Aegon-2018.pdf	Aegon Asset Management	2018 Annual Report
Aegon-2019.pdf	Aegon Asset Management	2019 Responsible Investment Report
Aegon-2020.pdf	Aegon Asset Management	2020 Responsible Investment Report
Allianz-2020.pdf	Allianz Global Investors	2020 Sustainability Report
Ameriprise-2020.pdf	Ameriprise Financial	2020 Responsible Business Report
Ameriprise-2021.pdf	Ameriprise Financial	2021 Responsible Business Report
Ashmore-2019.pdf	Ashmore Group	2019 CSR Report
Ashmore-2-2019.pdf	Ashmore Group	2019 Sustainability Report
Ashmore-2020.pdf	Ashmore Group	2020 CSR Report
Aviva-2020.pdf	Aviva Investors	2020 Responsible Investment Report
BNP Paribas-2020.pdf	BNP Paribas	2020 Integrated Report
BNY Mellon-2020.pdf	BNY Mellon	2020 ESG Report
BOCHK-2019.pdf	Bank of China (Hong Kong)	2019 Sustainability Report
BOCHK-2020.pdf	Bank of China (Hong Kong)	2020 Sustainability Report
BOCOM-2020.pdf	BOCOM International	2020 CSR Report
Barings-2020.pdf	Barings	2020 Sustainability Policy
BlackRock-2018.pdf	BlackRock	2018 Sustainability Disclosure
BlackRock-2019.pdf	BlackRock	2019 Sustainability Disclosure
BlackRock-2020.pdf	BlackRock	2020 TCFD Report
BlackRock-2-2020.pdf	BlackRock	2020 Sustainability Disclosure

BofA-2020.pdf	Bank of America	2020 Annual Report
CICC-2019.pdf	China International Capital Corporation	2019 Annual Report
CICC-2020.pdf	China International Capital Corporation	2020 Annual Report
CITIC Capital-2019.pdf	CITIC Capital	2019 ESG Report
CMBC-2018.pdf	China Minsheng Banking Corporation	2018 ESG Report
CMBC-2019.pdf	China Minsheng Banking Corporation	2019 ESG Report
CMBC-2020.pdf	China Minsheng Banking Corporation	2020 Sustainability Report
Capital Group-2020.pdf	Capital Group	2020 ESG Report
Credit Suisse-2020.pdf	Credit Suisse	2020 Sustainability Report
Deutsche-2020.pdf	Deutsche Bank	2020 Sustainable Finance Framework
E Fund-2020.pdf	E Fund Management	2020 Active Ownership Statement
E Fund-2021.pdf	E Fund Management	2021 Responsible investment Statement
Edward Jones-2019.pdf	Edward Jones Investments	2019 Sustainability Report
Edward Jones-2021.pdf	Edward Jones Investments	2021 Purpose, Inclusive & Citizenship Report
Everbright-2017.pdf	China Everbright Limited	2017 ESG Report
Fidelity-2020.pdf	Fidelity International	2020 Sustainable investing Report
Fidelity-2021.pdf	Fidelity International	2021 ESG Report
First Sentier-2019.pdf	First Sentier Investors	2019 Responsible Investment Report
First Sentier-2020.pdf	First Sentier Investors	2020 Responsible Investment Report
Fisch-2020.pdf	Fisch Asset Management	2020 Sustainability Report
Fullerton-2021.pdf	Fullerton Fund Management	2021 Sustainability Policy
GAM-2021.pdf	GAM Investments	2021 Responsible Investment Policy

Goldman Sachs-2020.pdf	Goldman Sachs	2020 Sustainability Report
Hang Seng-2018.pdf	Hang Seng Bank	2018 CSR Report
Hang Seng-2019.pdf	Hang Seng Bank	2019 CSR Report
Hang Seng-2020.pdf	Hang Seng Bank	2020 ESG Report
Insight-2021.pdf	Insight Investment	2021 Responsible Investment Annual Report
Invesco-2019.pdf	Invesco	2019 CSR Report
Invesco-2020.pdf	Invesco	2020 ESG Report
Invesco-2-2020.pdf	Invesco	2020 CSR Report
JPMorgan Chase-2019.pdf	JPMorgan Chase	2019 ESG Report
JPMorgan Chase-2020.pdf	JPMorgan Chase	2020 ESG Report
L&G-2020.pdf	Legal & General	2020 Sustainability Report
Manulife-2020.pdf	Manulife	2020 Sustainability Report
Morgan Stanley-2018.pdf	Morgan Stanley	2018 Sustainability Report
Morgan Stanley-2019.pdf	Morgan Stanley	2019 Sustainability Report
Morgan Stanley-2020.pdf	Morgan Stanley	2020 Sustainability Report
Muzinich-2021.pdf	Muzinich & Co.	2021 Sustainable and Responsible Investment Policy
NB-2020.pdf	Neuberger Berman	2020 ESG Report
Ninety One-2021.pdf	Ninety One	2021 Sustainability Report
PGIM-2020.pdf	PGIM Real Estate	2020 Sustainability Report
PIMCO-2020.pdf	PIMCO	2020 ESG Report
Pictet-2021.pdf	Pictet	2021 Responsible Investment Report
Schroders-2020.pdf	Schroders	2020 Annual Report
Sinopac-2018.pdf	Sinopac Financial Holdings	2018 CSR Report

Sinopac-2019.pdf	Sinopac Financial Holdings	2019 CSR Report
Sinopac-2020.pdf	Sinopac Financial Holdings	2020 CSR Report
Standard Life-2019.pdf	Standard Life Aberdeen	2019 CSR Report
Standard Life-2020.pdf	Standard Life Aberdeen	2020 ESG Report
SuMi TRUST-2020.pdf	Sumitomo Mitsui Trust Holdings	2020 Sustainability Report
T. Row Price-2020.pdf	T. Rowe Price	2020 Sustainability Report
TIAA-2020.pdf	TIAA	2020 Responsible Business Report
UBS-2019.pdf	UBS	2019 Sustainability Report
UBS-2020.pdf	UBS	2020 Sustainability Report
UOB AM-2021.pdf	UOB Asset Management	2021 Sustainable Investment Approach
Value Partners-2019.pdf	Value Partners	2019 Annual Report
Value Partners-2020.pdf	Value Partners	2020 Annual Report
Vanguard-2020.pdf	Vanguard	2020 Annual Report
Wellington-2020.pdf	Wellington Management	2020 Sustainability Report

Insurance

File Name	Company	Description
AIA-2018.pdf	AIA Group	2018 ESG Report
AIA-2019pdf	AIA Group	2019 ESG Report
AIA-2020.pdf	AIA Group	2020 ESG Report
AXA-2018.pdf	AXA Insurance	2018 Annual Report
AXA-2-2018.pdf	AXA Insurance	2018 Integrated Report
AXA-2019.pdf	AXA Insurance	2019 Annual Report
AXA-2-2019.pdf	AXA Insurance	2019 Integrated Report
AXA-2020.pdf	AXA Insurance	2020 Annual Report

AXA-2020.pdf	AXA Insurance	2020 Integrated Report
AXA-2021.pdf	AXA Insurance	2021 Climate Report
Asahi Life-2020.pdf	Asahi Mutual Life Insurance	2020 Annual Report
Cathay Life-2020.pdf	Cathay Life Insurance	2020 ESG Report
DGB-2019.pdf	DGB Financial Group	2019 Sustainability Report
Daiichi Life-2018.pdf	Dai-ichi Life	2018 Sustainability Report
Daiichi Life-2019.pdf	Dai-ichi Life	2019 Sustainability Report
Daiichi Life-2020.pdf	Dai-ichi Life	2020 Sustainability Report
FWD-2020.pdf	FWD Insurance	2020 Code of Conduct
FWD-2021.pdf	FWD Insurance	2021 ESG Report
Fukoku Life-2020.pdf	Fukoku Mutual Life Insurance	2020 Annual Report
JP Insurance-2020.pdf	Japan Post Insurance	2020 CSR Report
KB Financial-2018.pdf	KB Financial Group	2018 Sustainability Report
KB Financial-2019.pdf	KB Financial Group	2019 Sustainability Report
KB Financial-2020.pdf	KB Financial Group	2020 Sustainability Report
Kyobo Life-2019.pdf	Kyobo Life Insurance	2019 Sustainability Report
Kyobo Life-2020.pdf	Kyobo Life Insurance	2020 Sustainability Report
Meiji Yasuda-2020.pdf	Meiji Yasuda Life	2020 Annual Report
Nan Shan Life-2019.pdf	Nan Shan Life Insurance	2019 Sustainability Report
Nippon Life-2018.pdf	Nippon Life	2018 Annual Report
Nippon Life-2020.pdf	Nippon Life	2020 Annual Report
Ping An-2018.pdf	Ping An Insurance	2018 Sustainability Report
Ping An-2019.pdf	Ping An Insurance	2019 Sustainability Report
Ping An-2020.pdf	Ping An Insurance	2020 Sustainability Report
Prudential-2020.pdf	Prudential	2020 ESG Report
Prudential-2-2020.pdf	Prudential	2020 Sustainability Report

Shin Kong-2019.pdf	Shin Kong Insurance	2019 CSR Report
Shin Kong-2020.pdf	Shin Kong Insurance	2020 CSR Report
Sun Life-2018.pdf	Sun Life Financial	2018 Sustainability Report
Sun Life-2019.pdf	Sun Life Financial	2019 Sustainability Report
Sun Life-2020.pdf	Sun Life Financial	2020 Sustainability Report
Sun Life-2021.pdf	Sun Life Financial	2021 Sustainability Report
Taiwan Life-2020.pdf	Taiwan Life Insurance	2020 Sustainability Report
Tokio Marine-2018.pdf	Tokio Marine Holdings	2018 Sustainability Report
Tokio Marine-2019.pdf	Tokio Marine Holdings	2019 Sustainability Report
Tokio Marine-2020.pdf	Tokio Marine Holdings	2020 Sustainability Report

Pension Funds

File Name	Company	Description
ABP-2019.pdf	Stichting Pensioenfonds ABP	2019 Sustainable & Responsible Investment
AP Fonden 1-2018.pdf	AP Fonden 1	2018 Annual Report
AP Fonden 1-2019.pdf	AP Fonden 1	2019 Annual Report
AP Fonden 1-2020.pdf	AP Fonden 1	2020 Annual Report
AP Fonden 2-2018.pdf	AP Fonden 2	2018 Sustainability Report
AP Fonden 2-2019.pdf	AP Fonden 2	2019 Sustainability Report
AP Fonden 2-2020.pdf	AP Fonden 2	2020 Sustainability Report
AP Fonden 3-2018.pdf	AP Fonden 3	2018 Annual Report
AP Fonden 3-2019.pdf	AP Fonden 3	2019 Annual Report
AP Fonden 3-2-2019.pdf	AP Fonden 3	2019 Climate Report
AP Fonden 3-2020.pdf	AP Fonden 3	2020 Annual Report
AP Fonden 4-2018.pdf	AP Fonden 4	2018 Annual Report
AP Fonden 4-2019.pdf	AP Fonden 4	2019 Annual Report
AP Fonden 4-2020.pdf	AP Fonden 4	2020 Annual Report
AP Fonden 7-2018.pdf	AP Fonden 7	2018 Sustainability Report

AP Fonden 7-2019.pdf	AP Fonden 7	2019 Sustainability Report
AP Fonden 7-2020.pdf	AP Fonden 7	2020 Sustainability Report
CPP-2018.pdf	CPP	2018 Report on Sustainable Investing
CPP-2019.pdf	CPP	2019 Report on Sustainable Investing
CPP-2020.pdf	CPP	2020 Report on Sustainable Investing
CalPERS-2019.pdf	California Public Employees' Retirement System	2019 Sustainability Principles
CalSTRS-2020.pdf	California State Teachers' Retirement System	2020 Sustainability Report
FDC-2020.pdf	Fonds de Compensation	2020 Sustainable Investor Report
Future Fund-2018.pdf	Future Fund	2018 Annual Report
Future Fund-2019.pdf	Future Fund	2019 Annual Report
Future Fund-2020.pdf	Future Fund	2020 Annual Report
GEPF-2018.pdf	Government Employees Pension Fund	2018 Annual Report
GEPF-2019.pdf	Government Employees Pension Fund	2019 Annual Report
GEPF-2020.pdf	Government Employees Pension Fund	2020 Annual Report
GPF-2018.pdf	Government Pension Fund Global	2018 Responsible Investment Report
GPF-2019.pdf	Government Pension Fund Global	2019 Responsible Investment Report
GPF-2020.pdf	Government Pension Fund Global	2020 Responsible Investment Report
GPIF-2019.pdf	Government Pension Investment Fund	2019 ESG Report
GPIF-2020.pdf	Government Pension Investment Fund	2020 Climate Report
ISIF-2020.pdf	Ireland Strategic Investment Fund	2020 Sustainability & Responsible Investment Strategy
NPS-2018.pdf	National Pension Service	2018 Annual Report

NPS-2019.pdf	National Pension Service	2019 Annual Report
NPS-2020.pdf	National Pension Service	2020 Annual Report

5.2 Full list of environment.yml

Stages	**Package name**	**Version**
Dashboard deployment (to be installed on virtual environment in dashboard repository)	argcomplete	1.10.3
	beautifulsoup4	4.8.2
	blis	0.7.5
	brotli	1.0.9
	catalogue	2.0.6
	chardet	3.0.4
	charset-normalizer	2.0.7
	click	8.0.3
	compressed-rtf	1.0.6
	cycler	0.11.0
	cymem	2.0.6
	dash	2.0.0
	dash-bootstrap-component	1.0.0
	dash-core-components	2.0.0
	dash-html-components	2.0.0
	dash-table	5.0.0

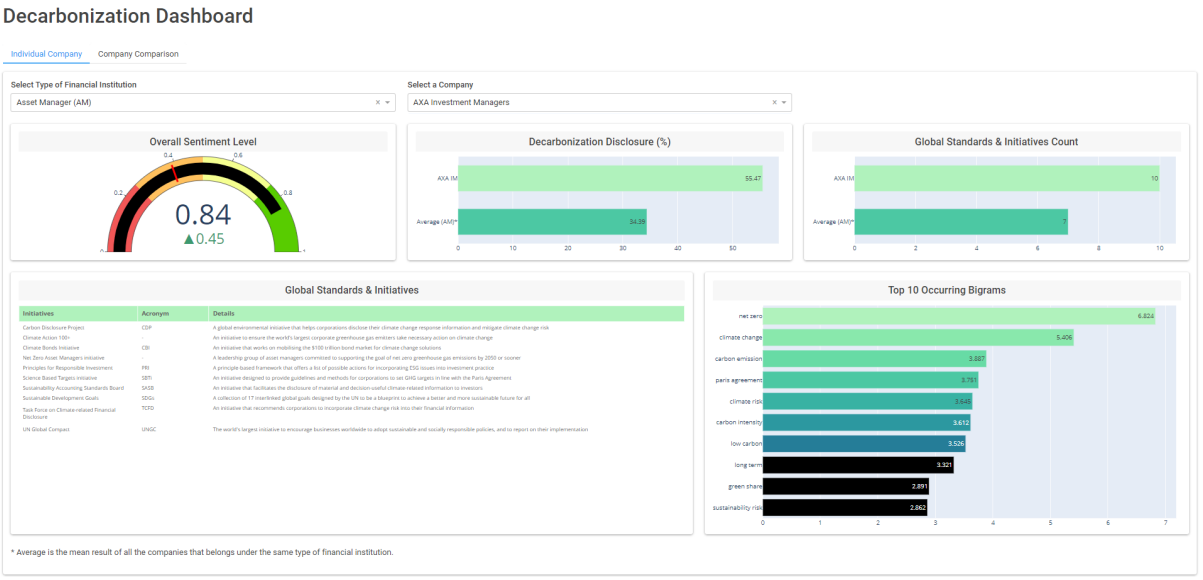
	docx2txt	0.8
	ebcdic	1.1.1
	ebooklib	0.17.1
	en-core-web-sm	3.2.0
	extract-msg	0.28.7
	flask	2.0.2
	flask-compress	1.10.1
	gensim	3.6.0
	idna	3.3
	imapclient	2.1.0
	itsdangerous	2.0.1
	joblib	1.1.0
	kiwisolver	1.3.2
	langcodes	3.3.0
	lxml	4.6.4
	matplotlib	3.4.3
	murmurhash	1.0.6
	nltk	3.6.5
	numpy	1.21.4
	olefile	0.46
	pandas	1.3.4
	pathy	0.6.1

	pdfminer-six	20191110
	pillow	8.4.0
	pip	21.3.1
	plotly	5.4.0
	preshed	3.0.6
	pycryptodome	3.11.0
	pydantic	1.8.2
	pymupdf	1.19.1
	python-pptx	0.6.21
	pytz	2021.3
	pytz-deprecation-shim	0.1.0.post0
	regex	2021.11.10
	requests	2.26.0
	scikit-learn	1.0.1
	scipy	1.7.2
	seaborn	0.11.2
	six	1.12.0
	sklearn	0.0
	smart-open	5.2.1
	sortedcontainers	2.4.0
	soupsieve	2.3.1
	spacy	3.2.0

	spacy-legacy	3.0.8
	spacy-loggers	1.0.1
	speechrecognition	3.8.1
	srsly	2.4.2
	tenacity	8.0.1
	textblob	0.17.1
	textract	1.6.4
	thinc	8.0.13
	threadpoolctl	3.0.0
	tqdm	4.62.3
	typer	0.4.0
	typing-extensions	4.0.0
	tzdata	2021.5
	tzlocal	4.1
	urllib3	1.26.7
	wasabi	0.8.2
	werkzeug	2.0.2
	xlrd	1.2.0
	xlsxwriter	3.0.2

5.3 Screenshots of the Dashboard

5.3.1 First Interface



5.3.2 Second Interface