Exploring Portfolio Decarbonisation using AI

LIM JERMAINE | A0187866E
AW XIN MIN | A0190383A
CHIA AI FEN | A0188215B
LEE JUN HUI, SEAN | A0189893B
WONG JAO KUEAN | A0200290Y

November 12, 2021

TABLE OF CONTENTS

PROJECT OVERVIEW 01

PROBLEM STATEMENT

Develop an end-to-end data driven solution to help NatWest Markets in efficient and effective retrieval and visualisation of portfolio decarbonisation information from reports across different Financial Institutions (FIs)

DELIVERABLES

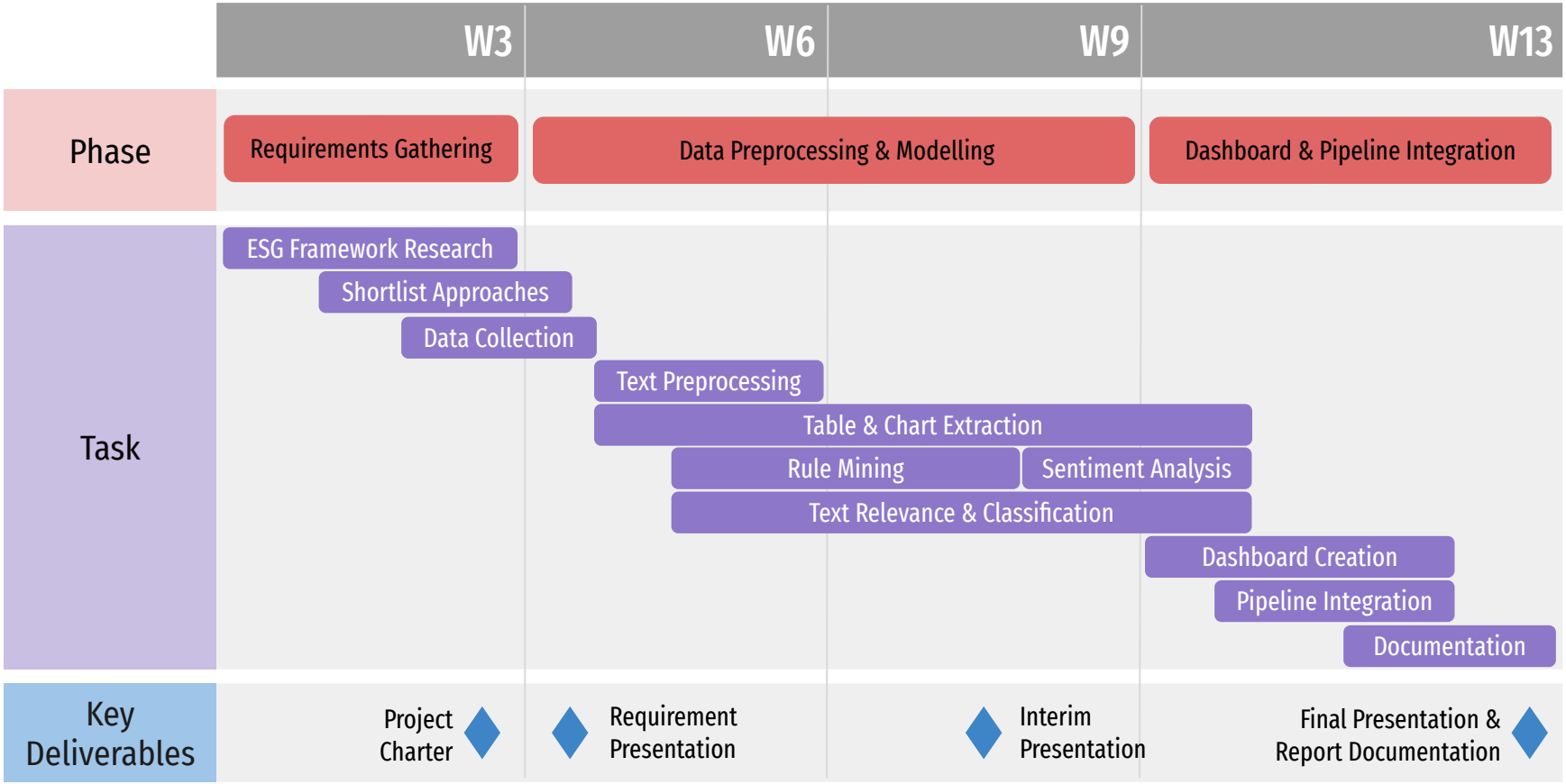
ACHIEVED

1. Build an **end-to-end pipeline** that is able take a PDF report as input and returns extracted information as output.
2. Use **NLP techniques** to perform **text extraction** of decarbonisation-related information from reports.
3. Build a **dashboard** that captures decarbonisation information with the capability to upload and analyse new reports

ADDITIONAL ACHIEVEMENTS

1. Employed **Computer Vision** techniques for **image processing** to extract relevant tables and charts.
2. Performed **Optical Character Recognition** for **tabular data extraction** to convert table images into dataframes.

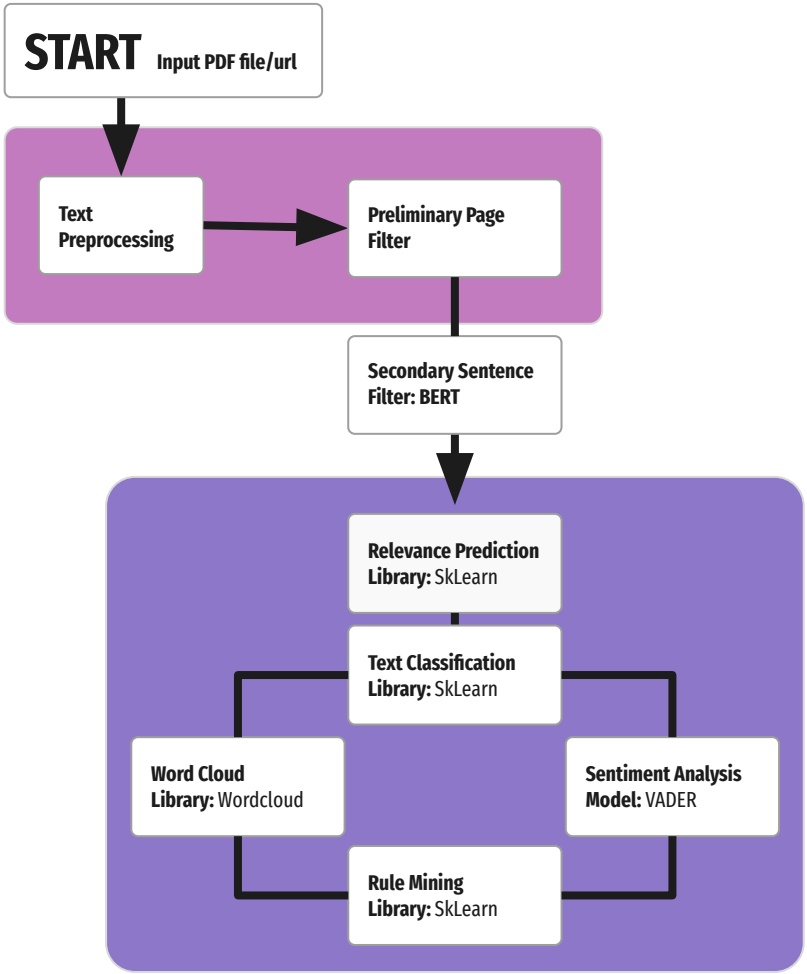
PROJECT TIMELINE



02 DATA PIPELINE

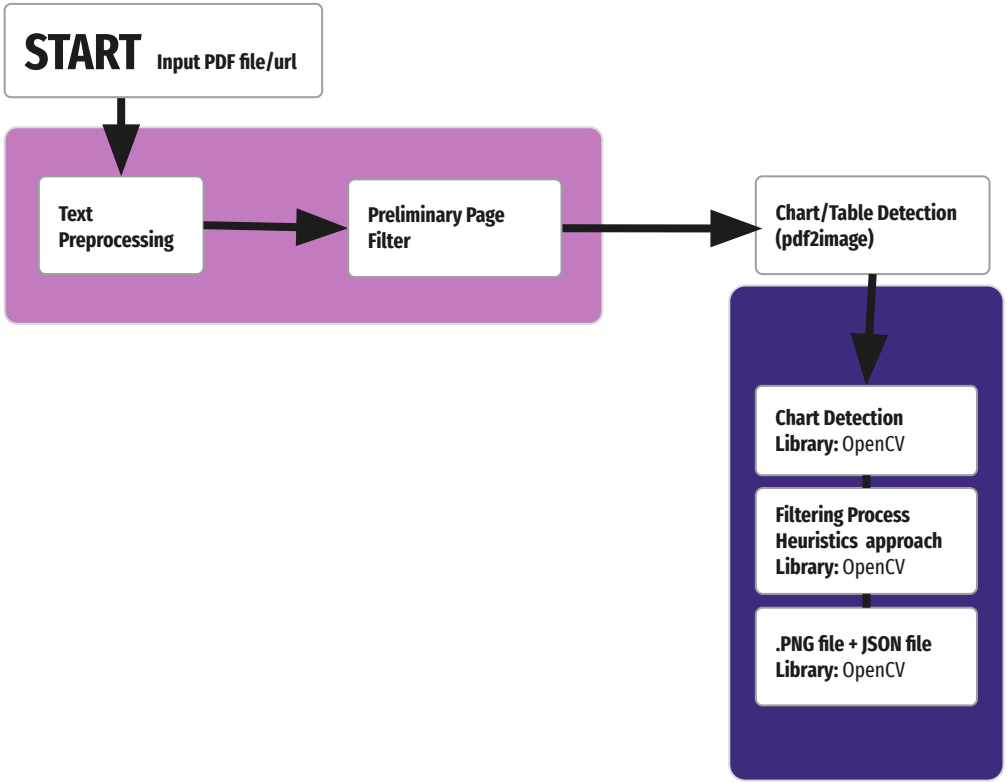
DATABASE SCHEMA


```
1:
    company:                "CMBC Capital"
    year:                   "2017"
  ▶ url:                    "http://en.cmbc.com.cn/up…/2017%20ESG%20REPORT.pdf"
  ▶ text_output:            {…}
  ▶ wordcloud_img_path:     […]
  ▶ sentiment_score:        […]
  ▶ table_keywords:         {…}
  ▶ table_image_keywords:   {…}
  ▶ table_images:           {…}
  ▶ chart_images:           {…}
  ▶ chart_images_keywords:  {…}
```



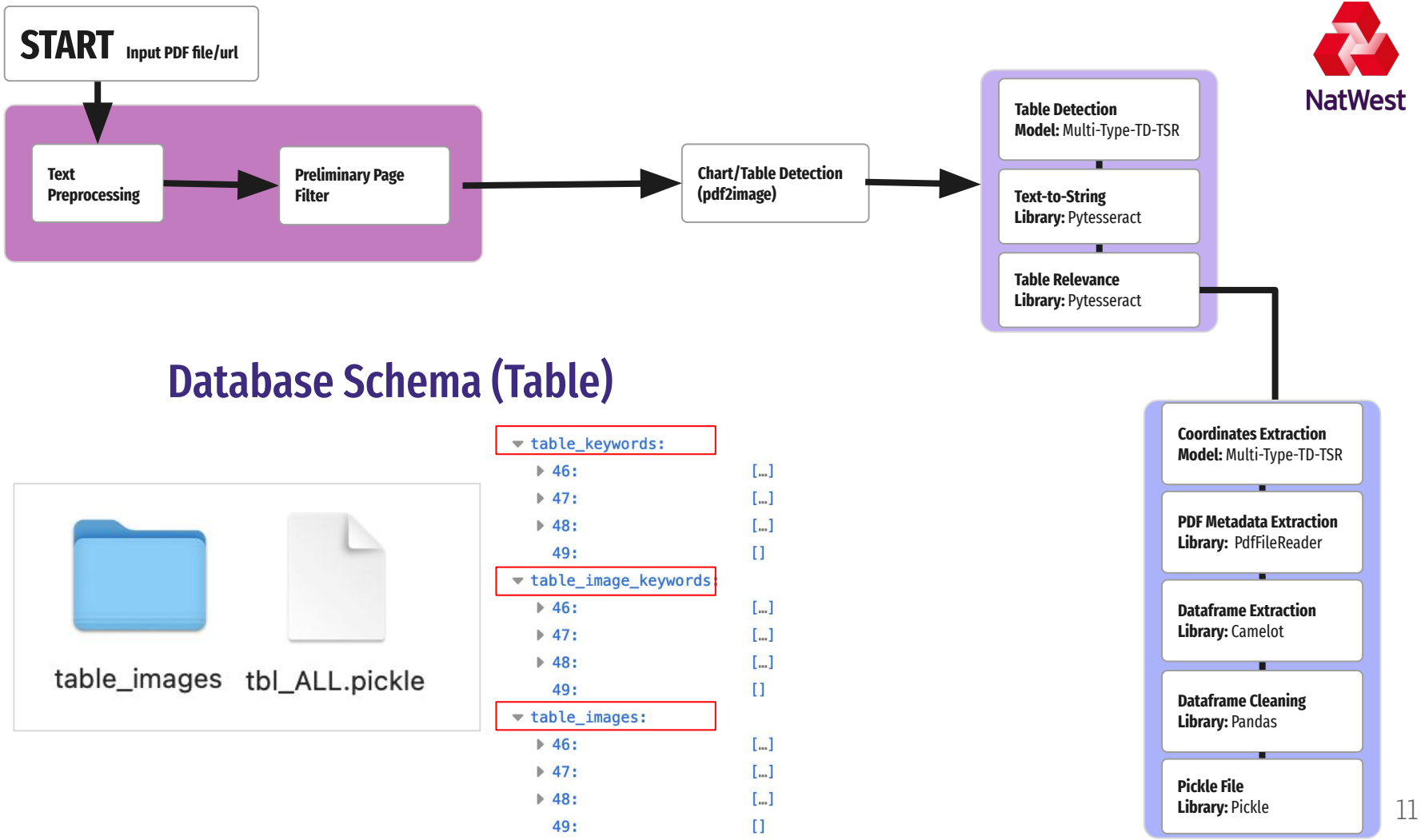
DATABASE SCHEMA (TEXT)

```
▼ text_output:
    ▶ page:                  […]
    ▶ sentence:              […]
    ▶ relevance_prob:        […]
    ▶ carbon_class:          […]
    ▶ mined_text:            […]
  ▶ wordcloud_img_path:      […]
  ▼ sentiment_score:
      0:                     0.2933
      1:                     0.3715
      2:                     null
      3:                     0.8481
      4:                     0.4814
```



Database Schema (Chart)

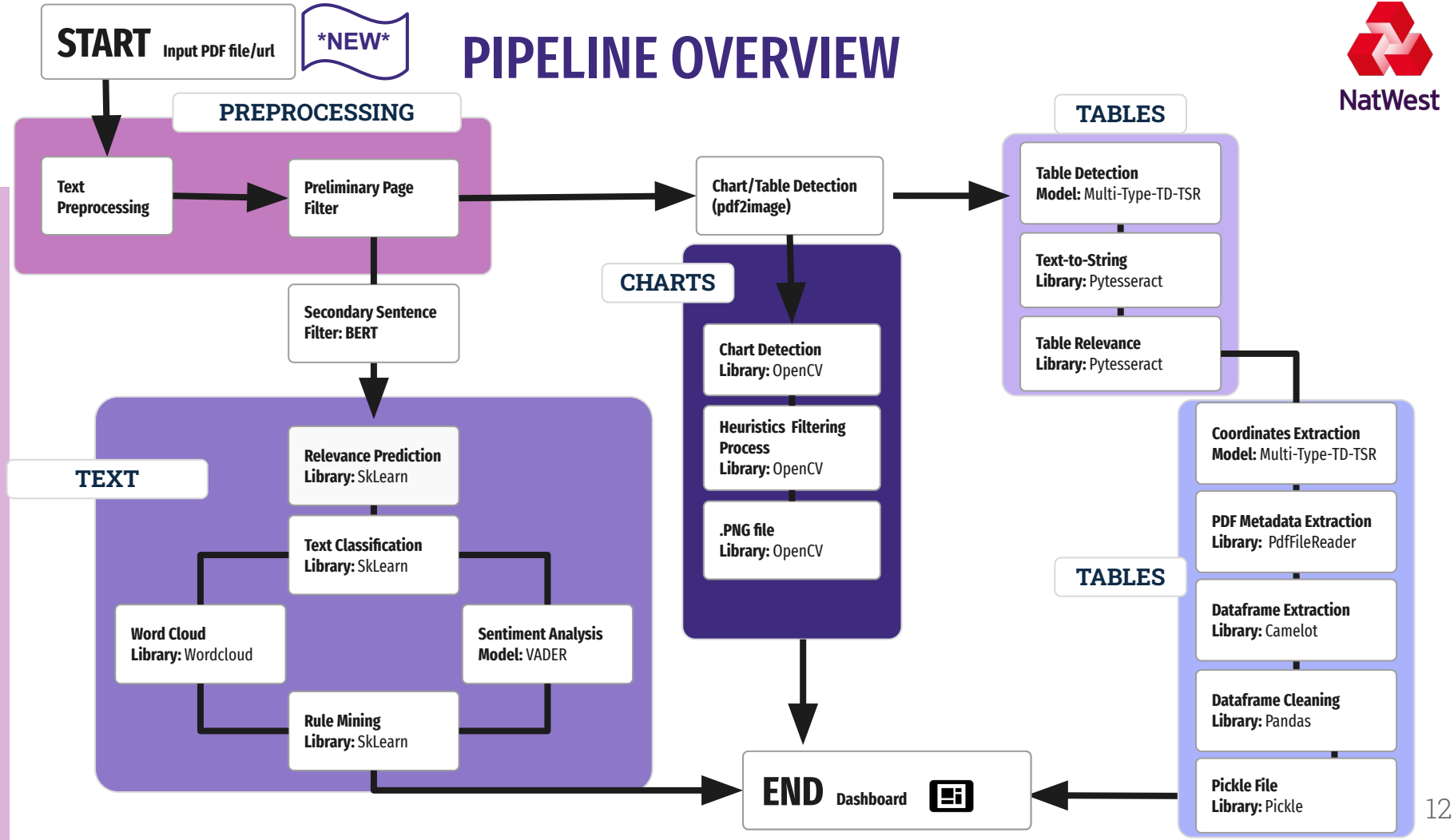

```
chart_images:
  1:
    0: "data/dashboard/ChartExtr…tion_2019/ROI_1_0_4.png"
chart_images_keywords:
  1:
    0:
      0: "emissions"
      1: "emission"
      2: "tonnes"
      3: "carbon"
```



NatWest

Database Schema (Table)

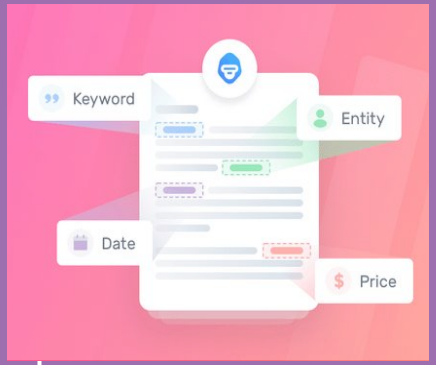
PIPELINE OVERVIEW



03 DATA EXTRACTION

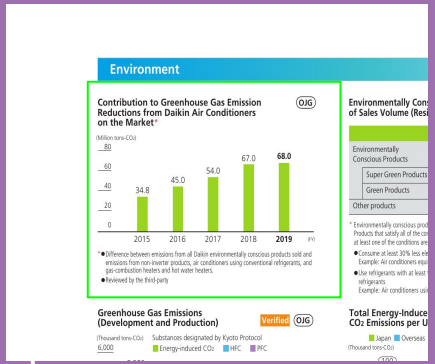
TYPES OF EXTRACTION

TEXT EXTRACTION



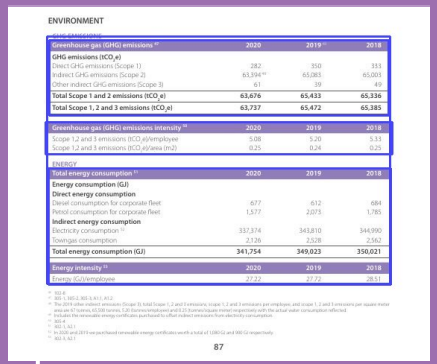
Identify relevant decarbonisation information from text

CHART EXTRACTION



Detect charts in PDF reports that display relevant metrics

TABLE EXTRACTION



ENVIRONMENT

Greenhouse gas (GHG) emissions [1]	2020	2019 [2]	2018
GHG emissions (tCO_2e)			
Direct GHG emissions (Scope 1)	282	350	353
Indirect GHG emissions (Scope 2)	63,394 [3]	65,085	65,003
Other indirect GHG emissions (Scope 3)	61	39	49
Total Scope 1 and 2 emissions (tCO_2e)	63,676	65,433	65,336
Total Scope 1, 2 and 3 emissions (tCO_2e)	63,737	65,472	65,385

Greenhouse gas (GHG) emissions intensity [4]	2020	2019	2018
Scope 1,2 and 3 emissions (tCO_2e)/employee	5.08	5.20	5.33
Scope 1,2 and 3 emissions (tCO_2e)/area ($m2$)	0.25	0.24	0.25

ENERGY

Total energy consumption [1]	2020	2019	2018
Energy consumption (GJ)			
Direct energy consumption			
Diesel consumption for corporate fleet	677	612	684
Petrol consumption for corporate fleet	1,577	2,073	1,785
Indirect energy consumption			
Electricity consumption [1]	337,374	343,810	344,990
Towngas consumption	2,126	2,528	2,562
Total energy consumption (GJ)	341,754	349,023	350,021

Energy intensity [5]	2020	2019	2018
Energy (GJ)/employee	27.22	27.72	28.51

87

Detect tables in PDF reports to display common metrics

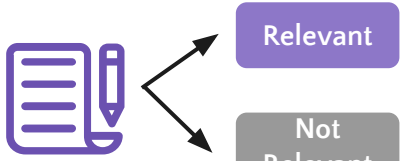
TEXT PREPROCESSING

TEXT EXTRACTION & PROCESSING

NatWest

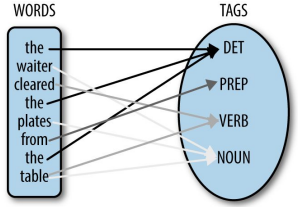

```
{
    "company": "DBS",
    "year": "2020",
    "pdf_url": "https://www.dbs.com/
      'report_sentences_preprocessed':
       [],
       ['introduction introduction 1 ov
     ruption and headwind from covid-19
        'as part of our ongoing effort
     lish the sustainable and transitio
        'we also recently raise our sus
     operational carbon emission across
        'in recognition of the bank env
     corporate social responsibility in
}
```


RELEVANCE PREDICTION



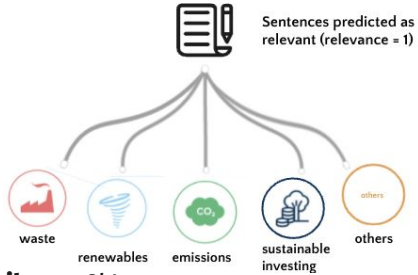
Library: SkLearn

RULE MINING



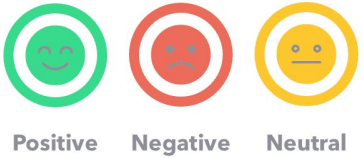
Library: SpaCy

TEXT CLASSIFICATION



Library: SkLearn

SENTIMENT ANALYSIS



Model: VADER

VISUALIZE ON DASHBOARD

RELEVANCE PREDICTION

NatWest

Purpose

Additional classification layer to improve word embedding filtering accuracy

Labelling

Shuffled dataset to break ordering, labelled 5200+ sentences as **relevant/non-relevant** (1,0)

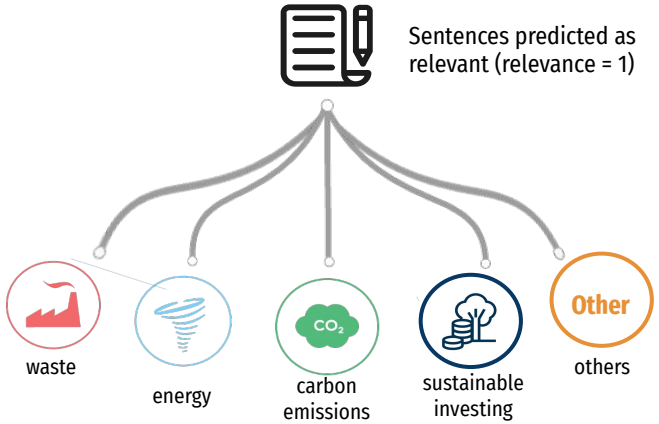
Best Model Results

Model	Best Processing	Weight Avg F1	Accuracy	F1 (Class 1)	F1 (Class 0)
Log Reg	SMOTE, BERT	0.90	0.87	0.48	0.87
SVM	SMOTE, Clean Sentence, Tfidf	0.91	0.91	0.45	0.95
NB	SMOTE, Raw Sentence, BoW	0.90	0.89	0.43	0.94
RF	SMOTE, Clean Sentence, Tfidf	0.92	0.92	0.49	0.96
Voting (Soft)		0.91	0.91	0.48	0.95
Voting (Hard)		0.91	0.91	0.50	0.95
Stacking	SVM Metamodel	0.93	0.92	0.53	0.92

TEXT CLASSIFICATION

Purpose

Allows for **filtering** on dashboard & sentiment analysis on **each category's decarbonisation progress**



Unique Factors

Choose Soft Voting Ensemble over RF due to **context** of small data set; RF has high chance of **overfitting data**

Best Model Results

Model	Vect	Proc	Weight Avg F1	Accuracy
Log Reg	Tfidf	Clean	0.66	0.68
SVM	Tfidf	Raw	0.67	0.67
NB	Tfidf	Clean	0.62	0.64
RF	BoW	Clean	0.70	0.71
CatBoost	Tfidf	Clean	0.67	0.67
Word Heuristic			0.61	0.61
Voting (Soft)			0.70	0.71
Voting (Hard)			0.68	0.70

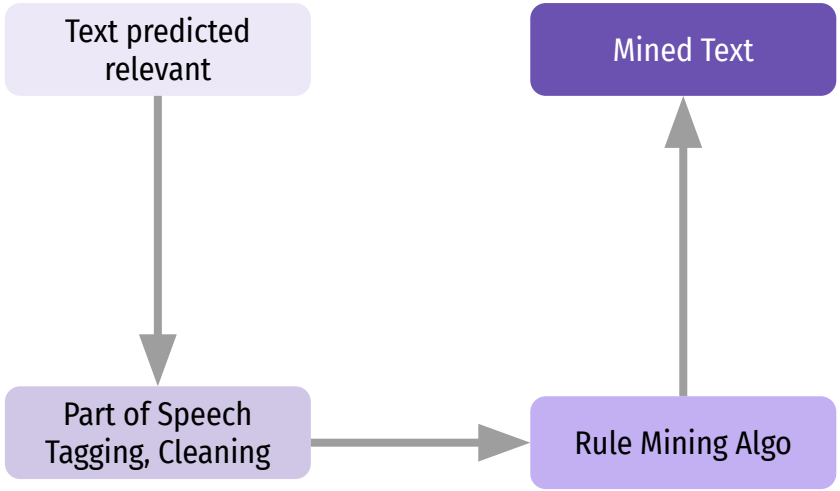
INFORMATION EXTRACTION - RULE MINING

NatWest

Purpose

Bold key portions quantifying decarbonisation progress on dashboard, **reduce time** to read chunky text

Algorithm



Results

Excluded Fine Grain Verb POS	Lemma Match Accuracy	Non-Lemma Match Accuracy
None	72.7%	72.9%
VBP	77.1%	74.3%
VB	65.4%	73.0%
VBP & VB	67.1%	74.5%

Example

"In 2019, Citi **financed 74 million of subordinate lien bonds that were certified green**, given the project's environmental aspects"

SENTIMENT SCORING

NatWest

Purpose

Quantify a FI's progress in each decarbonisation category

Methodology

- Leverage **VADER**, a unsupervised lexicon & rule based approach due to small sample size
- Calculate **polarity compound score** ∈ [-1,1] for each sentence using the pretrained base VADER model

Ensuring Optimal Results - Tuning VADER Model

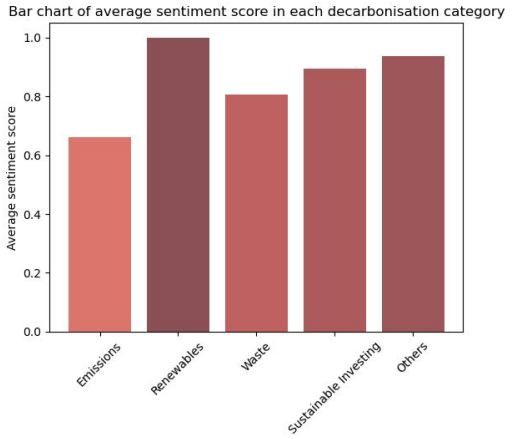
"At the end of 2019, we had reduced our GHG emissions by 71% compared to baseline year 2004."

Before 0.0

Added 20+ **context specific** sentiment mappings to tune base VADER

After 0.25

Visual



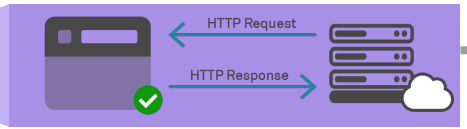
TEXT PREPROCESSING

CHART EXTRACTION

NatWest

INPUT JSON

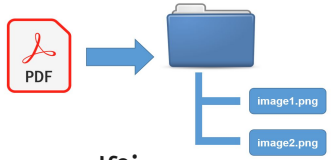

```
{
    "company": "DBS",
    "year": "2020",
    "pdf_url": "https://www.dbs.com/
    "filtered_report_tables_direct": {
        "21": "pillar 1 : responsible banl
        "24": "23 to manage our loan book
        "38": "37 our approach as a finan
        "39": "pillar 2 : responsible bus
        "51": "51 information on environm
        "64": "methodology for tcfd metho
        "65": "65 methodologyonquantifyin
        "69": "69 independent limited assi
    },
}
```

Library: Requests

CONVERT PDF TO IMAGE

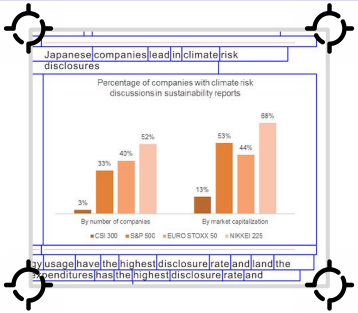
PDF TO IMAGE



Library: pdf2image

CHART DETECTION

COORDINATES EXTRACTED



Library: OpenCV
(Bounding Coordinates)

FILTER NOISY IMAGE

HEURISTIC FILTERING PROCESS

Library: OpenCV
(Image Thresholding)

OUTPUT JSON


```
"chart_images_keywords": {
    "58": [
        [
            "million",
            "energy",
            "emissions",
            "scope",
            "total"
        ]
    ],
    "71": []
},
"chart_images": {
    "58": [
        "data/new_report/ChartExtraction_Output/
    ],
    "71": []
}
```


OUTPUT

CHART EXTRACTION

Filtering away irrelevant images

NatWest

Image Properties

1	Text ratio
2	Unique Keywords
3	Color Dilation

In the wake of the horrific terror attack in Pulwama, the SBI Chairman issued an appeal to all staff members to contribute for the welfare of affected soldiers and their families. This was done through the dedicated Home Ministry, GoI Portal – "Bharat Ke Veer", to help the cause of security forces at a time of national distress.

To show the Bank's solidarity with the forces, ₹30 lakh was paid to each martyr's family under the Personal Accident Insurance scheme within 5 days. SBI also waived loans for 21 of the nation's heroes who had personal loans and accrued interests worth

A Memorandum of Undertaking (MoU) was signed on 12th July 2018 between the Indian Army and SBI on Defence Salary Package. The MoU was signed by the Chief General Manager (PB) and attended by the Managing Director (R&DB), Dy. Managing Director (RB) & Chief General Manager (Delhi Circle)

Cardiac and Critical care Ambulance donated to the Army Camp, 92 Base Hospital, Srinagar

203-2, 413-1

Sustainability Report 2018-19

52

RESPONSIBILITY COLUMN

anges in

the problems of poor households who lacked funds, jobs, technology and collateral.

As at the end of 2020, PSBC Ningxia Branch promoted the "Caichuan mode" to 706 administrative villages across Ningxia, and granted 150,000 loans with a total amount of nearly RMB8 billion, allowing more people to share the achievements of the mode. At the same time, PSBC promoted the mode across the Bank to spread the experience of Caichuan Village across China. In 2020, the "Caichuan mode" was displayed on the official website of the United Nations as a classic case of poverty alleviation in China and

The Bank will continue to conserve environment by participating in Shred2Share project and will report the result of its contribution in environmental conservation to executives and employees once a year.

Table Showing Quantity of Documents Destroyed Using Environmentally-Friendly Methods in Shred2Share Project Within Krung Thai Bank Head Offices

Year	Quantity of document (ton)	CO2 emission reduction (ton)	Coal use reduction (ton)	Water use reduction (m³)
2009	39.08	10.16	4.30	1,954
2010	54.72	14.23	6.02	2,737
2011	55.86	15.34	6.12	2,793
2012	62.13	16.38	6.93	3,150
2013	62.66	16.29	6.83	3,133
2014	69.94	18.18	7.69	3,496
2015	68.44	17.79	7.53	3,422
2016	86.55	22.50	9.52	4,328
2017	57.62	14.98	6.34	2,881
Total	559.72	145.84	61.59	27,986.95

Sustainability Report 2017 | 71

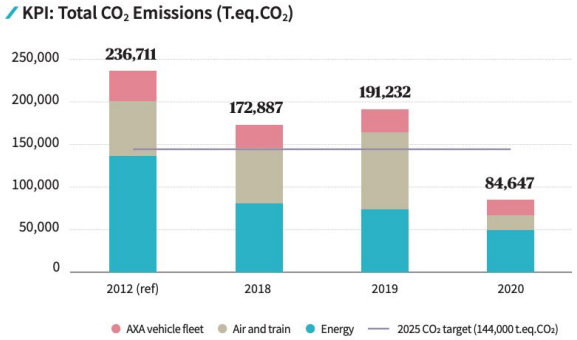
Responsible Banking...Growing to a Sustainable Future

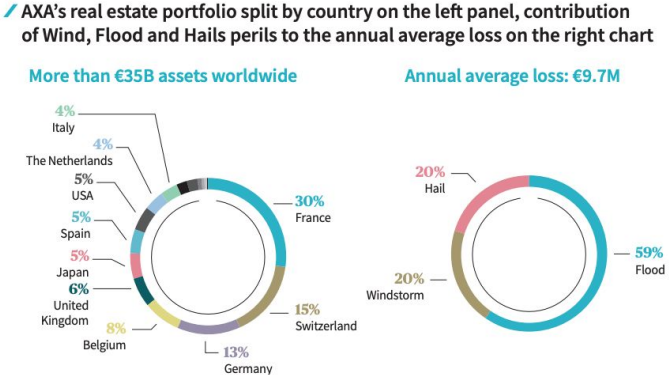
CHART EXTRACTION

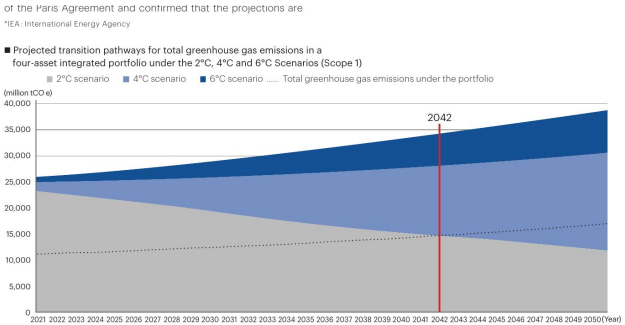
Filtering away irrelevant images

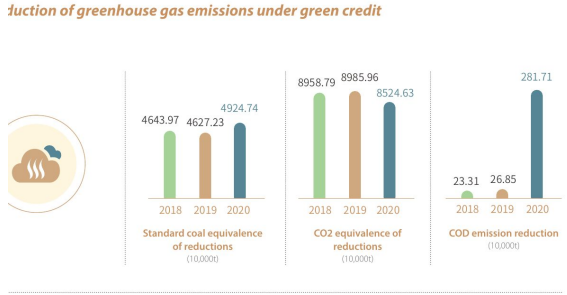
NatWest

Image Properties	
1	Text ratio
2	Unique Keywords
3	Color Dilation









TEST SET EVALUATION

NatWest

Test Set for 9 Reports

Allianz Global Investors	2020
AP Fonden 1	2018
Asahi Life	2020
AXA Insurance	2021
China Merchants Bank	2020
Government Pension Fund	2020
OCBC	2020
UBS Asset Management	2019
UOB Asset Management	2020

*Standardized Across Tables & Charts Extraction

TABLES

Indicator	2018	2019
Total greenhouse gas emission of Head Office (Scope 1 + Scope 2) (tons of CO_2 equivalent)	8,546.83	10,059.62
Paper saved by using e-bills (100 million pieces)	15.48	17.91
Power usage effectiveness (PUE) of data centers	1.72	1.65
Green loan balance (RMB 100 million)	1,660.33	1,767.73

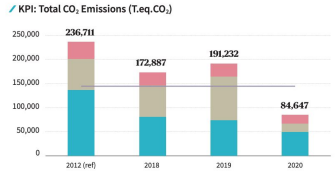
Sector	Segments	Sample population as per cent of total non-bank loans
Cement Manufacturing	Building materials	0.1%
Energy	Utilities, and oil and gas	1.1%
Metals and Mining	Coal, ferrous and non-ferrous mining, and manufacturing of metals	0.5%
Transportation	Land transport, air transport and water transport	1.0%
Agricultural	Agriculture and livestock production, manufacturing of agriculture products, and wholesale and trading of agriculture and livestock products	1.2%
Forestry	Logging, production of wood, and manufacturing of pulp and paper	0.2%
Chemicals	Manufacturing of chemicals	0.2%
Manufacturing of Transport Equipment	Automobile manufacturing	0.2%
Infrastructure	Operations of land, water and infrastructure	0.1%
	Total	5%

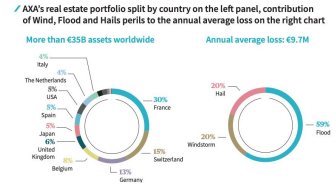
Allianz Global Investors environmental data at a glance

	2019
Total GHG emissions (tons per employee)	3.5
Energy consumption	1.6
Business travel	1.9
Paper consumption	0.03
Share of renewable energy in the mix (%)	
Share of renewable energy	46
Water consumption (cubic metres per employee)	
Water consumption	26

- Single-Row Tables
- Multi-Row Tables with Different Text Alignments
- Fully Bordered Tables
- Partially Bordered Tables
- Borderless Tables
- Tables with Long Headers
- Tables with Merged Cells

CHARTS





- Colourful Charts with Distinct Colour Gradients
- Dull Charts with Similar Colour Gradients
- Different Chart Types (e.g. donut chart, bar chart)
- Charts Near to Text
- Charts Far from Text

SELECTION CRITERIA

CHART EVALUATION METRICS

Goals

- Filter away images that are **irrelevant images** identified from our heuristic rules.
- Maintaining a relatively **low level of noise** in our extraction.

Precision

$$= \frac{\textit{No. of } \textbf{\textit{images}} \textit{ not extracted correctly}}{\textit{No. of } \textbf{\textit{images}} \textit{ extracted } \textbf{\textit{from report}}}$$

Recall

$$= \frac{\textit{No. of } \textbf{\textit{images}} \textit{ not extracted correctly}}{\textbf{\textit{Actual}} \textit{ no. of } \textbf{\textit{irrelevant images}} \textit{ in the report}}$$

TABLE EVALUATION METRICS

NatWest

Goals

- Extract as many **relevant tables** as possible
- Maintaining a relatively **low level of noise** in our extraction.

Precision

$$= \frac{\textit{No. of } \textbf{\textit{tables}} \textit{ extracted correctly}}{\textit{No. of } \textbf{\textit{tables}} \textit{ extracted } \textbf{\textit{from report}}}$$

Recall

$$= \frac{\textit{No. of } \textbf{\textit{tables}} \textit{ extracted correctly}}{\textbf{\textit{Actual}} \textit{ no. of } \textbf{\textit{relevant tables}} \textit{ in the report}}$$

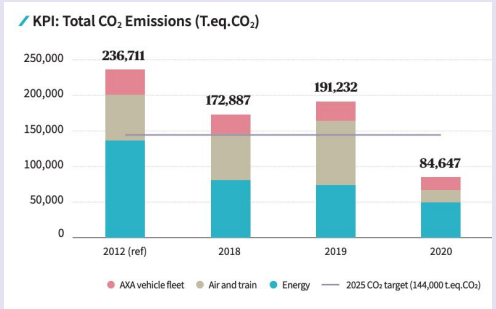
CHART EXTRACTION

PERFORMANCE ON TEST SET

NatWest

False Positive

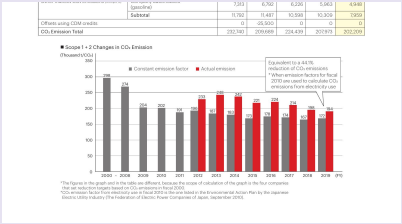
True Positive



PREDICTED

ACTUAL	Irrelevant Charts	Relevant Charts
Not Extracted	248	27
Extracted	11	2

Precision: 0.95
Recall: 0.9
Accuracy: 0.92



True Positive

CHART EXTRACTED

TEXT PREPROCESSING

TABLE DETECTION & EXTRACTION

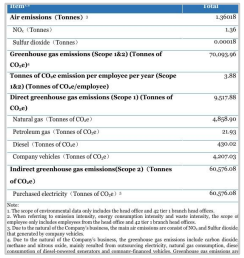
NatWest

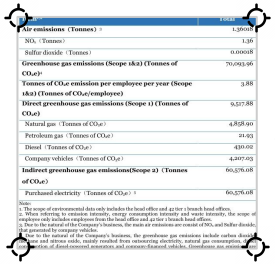
INPUT JSON


```
{
    "company": "CMBC Capital",
    "year": "2017",
    "url": "http://en.cmbc.com.cn/upload/images/2020/11/2017%20ESG%20REPORT.pdf",
    "filtered_report_tables_direct": {
        "46": "environmental datum performance summary emission item1,2 air emissi
        "47": "energy and water consumption item energy consumption(mwh)1 direct e
        "49": "appendixenvironmental , social and governance reporting guide area
    },
    "filtered_report_tables_indirect": {
        "48": "waste management item hazardous wastetonne lead - acid battery ( tc
    }
}
```


TABLE DETECTION

Filter for decarbonisation related tables

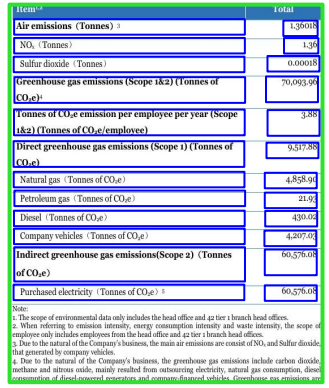




Retrieve table coordinates

Libraries: pdf2image, Multi-Type-TD-TSR, Pytesseract

TABULAR DATA EXTRACTION



Item	Total
Air emissions (Tonnes)	1.36018
NO (Tonnes)	1.36
Sulfur dioxide (Tonnes)	0.00018
Greenhouse gas emissions (Scope 1&2) (Tonnes of CO_2e)	70,093.96
Tonnes of CO_2e emission per employee per year (Scope 1&2) (Tonnes of CO_2e/employee)	3.88
Direct greenhouse gas emissions (Scope 1) (Tonnes of CO_2e)	9,517.88
Natural gas (Tonnes of CO_2e)	4,858.90
Petroleum gas (Tonnes of CO_2e)	21.93
Diesel (Tonnes of CO_2e)	430.02
Company vehicles (Tonnes of CO_2e)	4,207.03
Indirect greenhouse gas emissions(Scope 2) (Tonnes of CO_2e)	60,576.08
Purchased electricity (Tonnes of CO_2e) [3]	60,576.08

Note:
1. The scope of environmental data only includes the head office and 42 tier 1 branch head offices.
2. When referring to emission intensity, energy consumption intensity and waste intensity, the scope of employee only includes employees from the head office and 42 tier 1 branch head offices.
3. Due to the natural of the Company's business, the main air emissions are consist of NO, and Sulfur dioxide, that generated by company vehicles.
4. Due to the natural of the Company's business, the greenhouse gas emissions include carbon dioxide, methane and nitrous oxide, mainly resulted from outsourcing electricity, natural gas consumption, diesel consumption of diesel-powered generators and company-financed vehicles. Greenhouse gas emissions are

TABLE CLEANING & FILTERING

	Item	Total
1	Air emissions (Tonnes)	1.36018
2	NO (Tonnes)	1.36
3	Sulfur dioxide (Tonnes)	0.00018
4	Greenhouse gas emissions (Scope 1&2) (Tonnes of	70093.96
6	Tonnes of COe emission per employee per year (...	3.88
8	Direct greenhouse gas emissions (Scope 1) (Ton...	9517.88
10	Natural gas (Tonnes of COe)	4858.90
11	Petroleum gas (Tonnes of COe)	21.93
12	Diesel (Tonnes of COe)	430.02
13	Company vehicles (Tonnes of COe)	4207.03
14	Indirect greenhouse gas emissions(Scope 2) (Tonnes	60576.08
16	Purchased electricity (Tonnes of CO	60576.08

Library: Pandas

TABLE IMAGE & DATAFRAME STORING

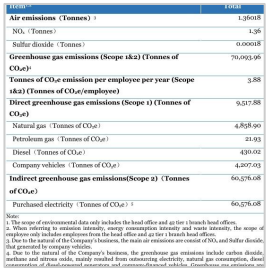


TABLE IMAGE

+

Item	Total
Air emissions (Tonnes)	1.36018
NO (Tonnes)	1.36
Sulfur dioxide (Tonnes)	0.00018
Greenhouse gas emissions (Scope 1&2) (Tonnes of	70093.96
Tonnes of COe emission per employee per year (...	3.88
Direct greenhouse gas emissions (Scope 1) (Ton...	9517.88
Natural gas (Tonnes of COe)	4858.90
Petroleum gas (Tonnes of COe)	21.93
Diesel (Tonnes of COe)	430.02
Company vehicles (Tonnes of COe)	4207.03
Indirect greenhouse gas emissions(Scope 2) (Tonnes	60576.08
Purchased electricity (Tonnes of CO	60576.08

DATAFRAME

OUTPUT JSON + PICKLE

Library: Camelot

TABLE DETECTION RESULTS

NatWest

PREDICTED

ACTUAL	+	-
+	27	7
-	16	30

Precision: 0.63
Recall: 0.79
Accuracy: 0.71

PAGE33_IMAGE0
False Positive

PAGE38_IMAGE1
False Positive

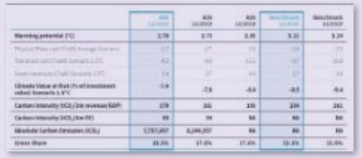
PAGE39_IMAGE2
True Positive

BEFORE TABLE CLEANING

PREDICTED

ACTUAL	+	-
+	27	7
-	7	39

Precision: 0.79
Recall: 0.79
Accuracy: 0.83



PAGE39_IMAGE0
True Positive

AFTER TABLE CLEANING

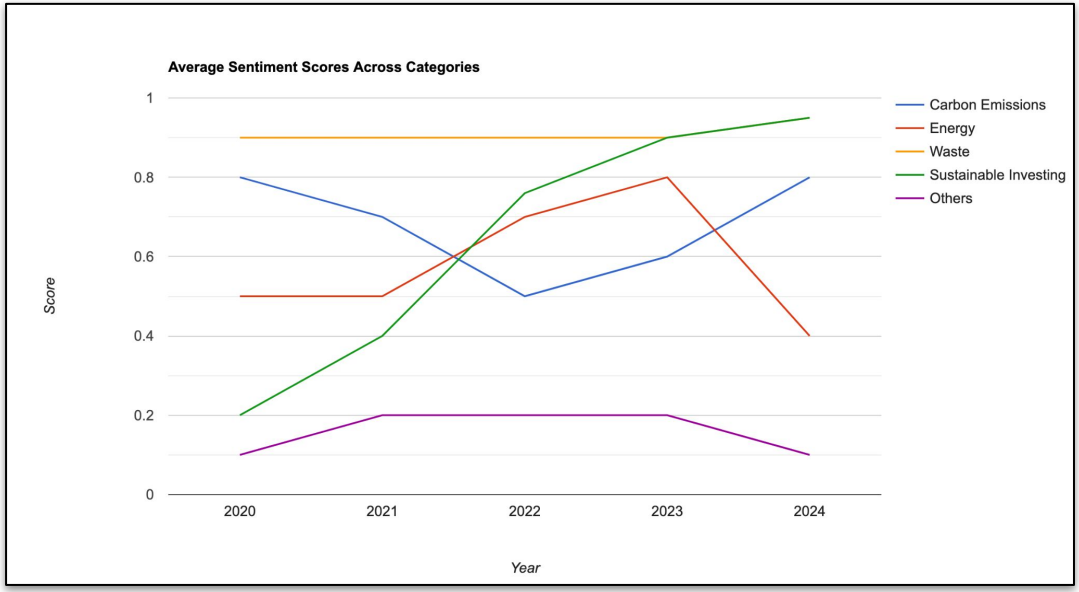
04 DASHBOARD DEMO

FUTURE EXTENSIONS 05

FUTURE EXTENSION(S)

Dataframe Processing	• Generates **temporal visualisations** based on data points collated from dataframes. • Limited utility due to **insufficient data points** in our extracted dataframes.
Sentiments Trendline	• Generates a trendline that displays the **sentiment scores over time**. • Limited utility due to **lack of comprehensive ESG disclosure** over time.

SENTIMENTS TRENDLINE



Benefit

Allows users to view how the ESG performance changes over time

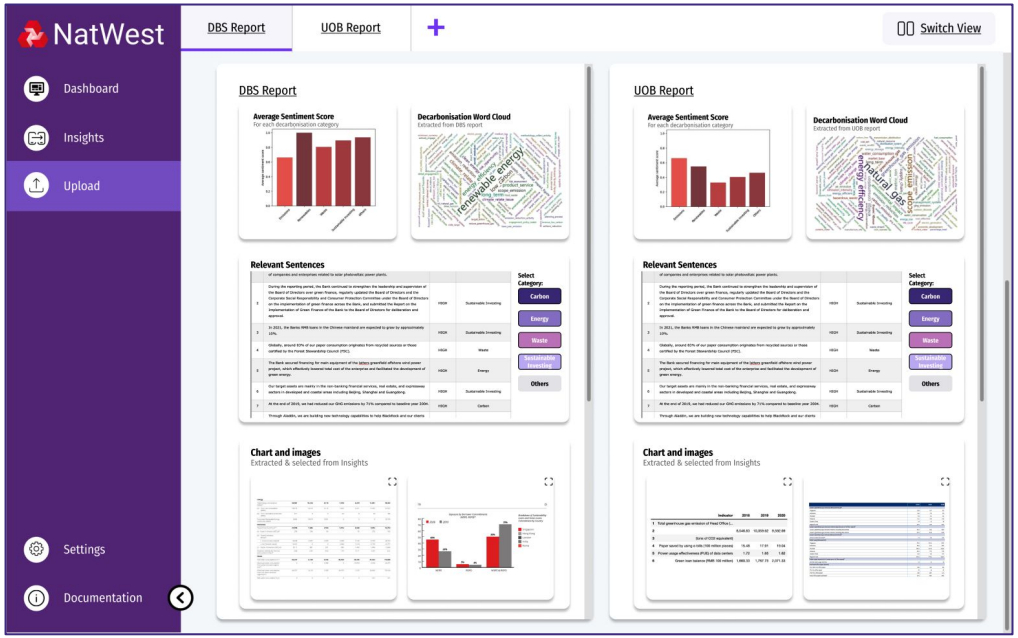
Incomplete ESG Disclosure

Sustainability disclosure being a relatively new phenomenon

FUTURE EXTENSION(S)

Dataframe Processing	• Generates **temporal visualisations** based on data points collated from dataframes. • Limited utility due to **insufficient data points** in our extracted dataframes.
Sentiments Trendline	• Generates a trendline that displays the **sentiment scores over time**. • Limited utility due to **lack of comprehensive ESG disclosure** over time.
Multi-tab Layout	• Allows users to open **multiple Dashboard views** concurrently. • **Lack of standardised metrics** for comparisons across multiple firms.

MULTI-TAB DASHBOARD INTERFACE

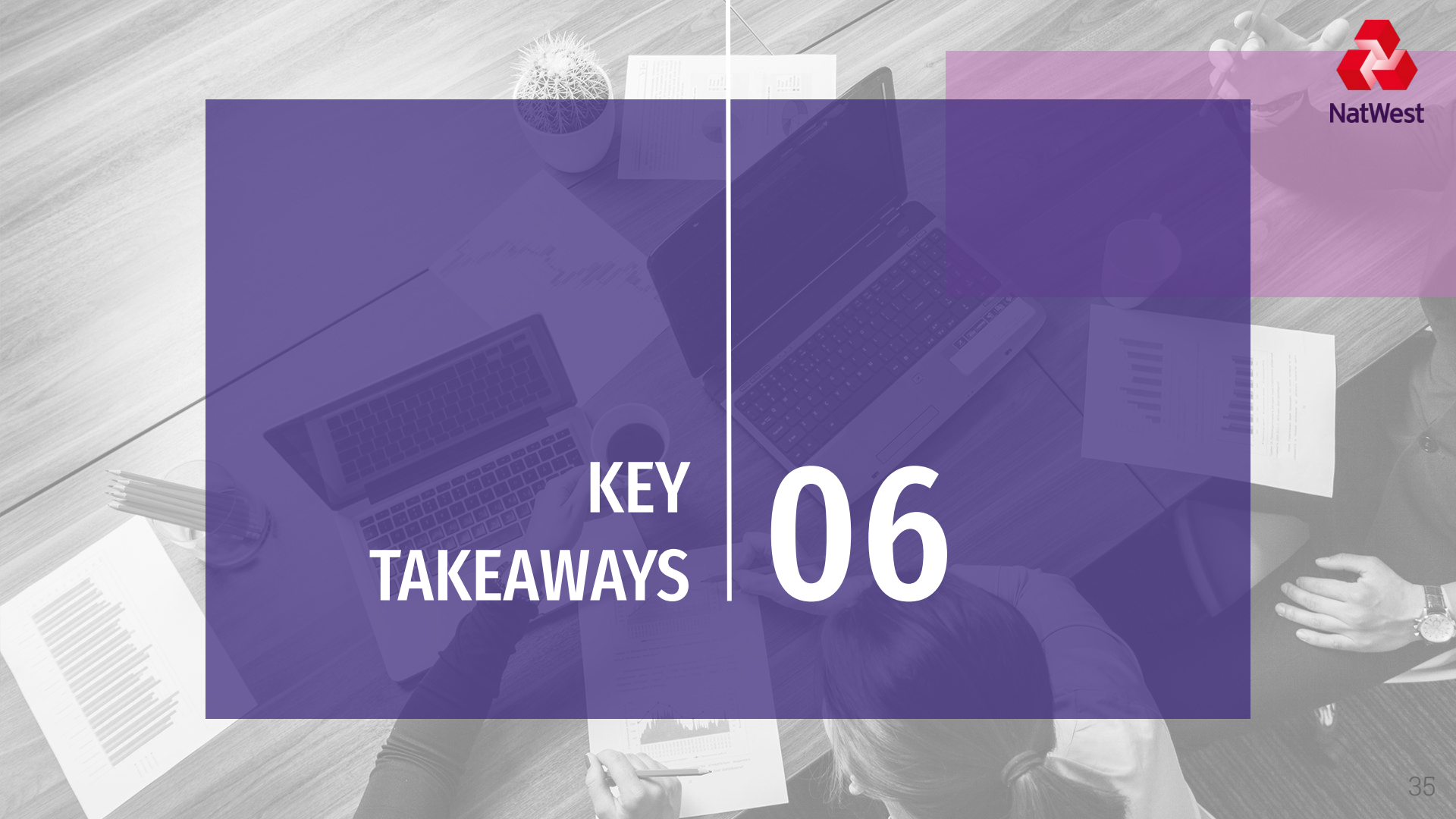


Benefit

Allows users to compare ESG performance across firms easily

Non-uniformity in Disclosure

Challenging to standardize the metrics for comparisons



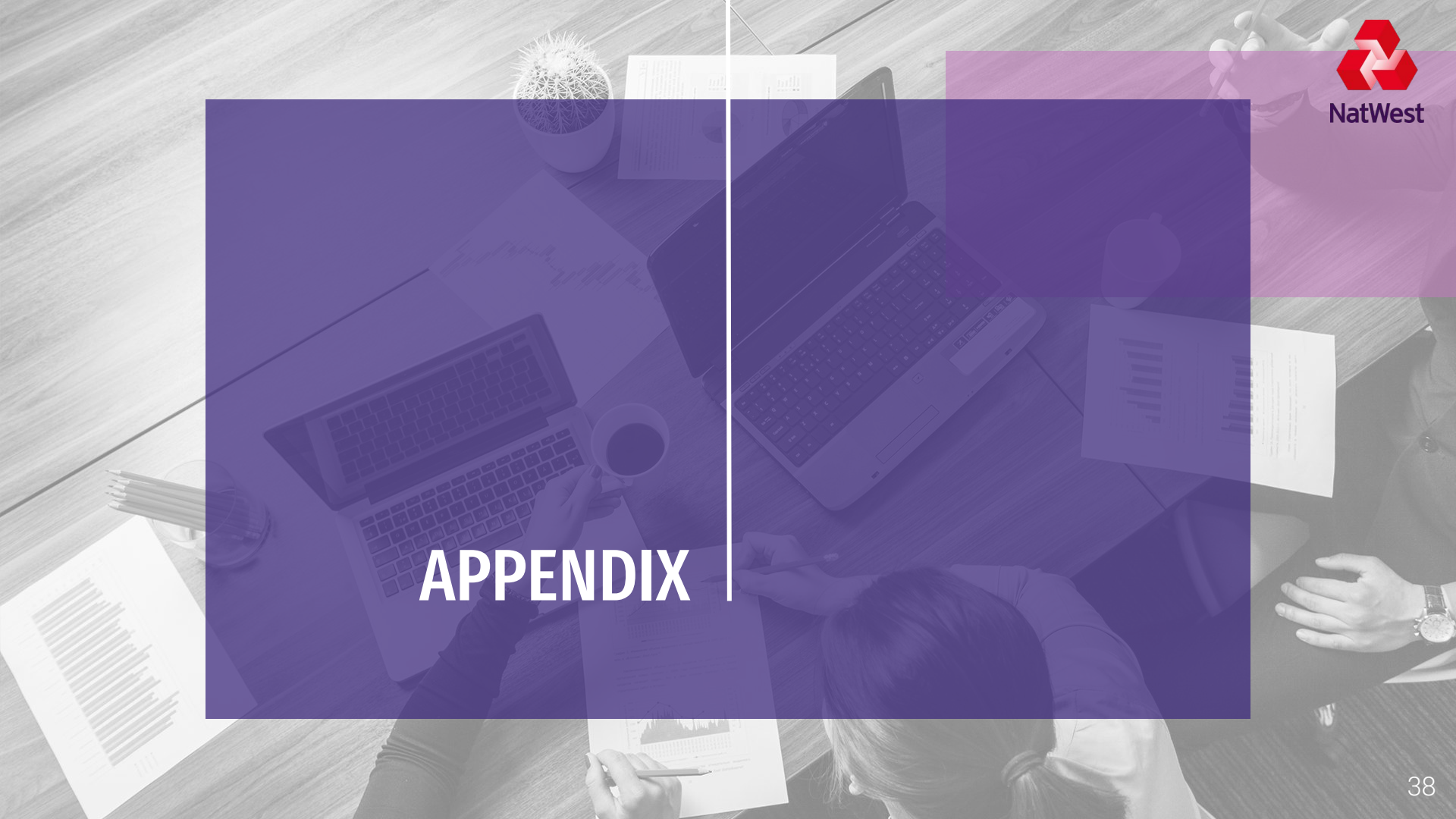
KEY TAKEAWAYS 06

KEY TAKEAWAYS

NatWest

Domain Knowledge	• Understanding of the latest **ESG investing trends** and **importance of ESG reporting**. • First-hand knowledge on the **management of climate-related risks** by a large FI.
Technical Knowledge	• Exposure to **advanced state-of-the-art libraries** like BERT, Detectron2, OpenCV. • **Application of BZA knowledge** for tailoring and selection of advanced algorithms.
Professional Skills	• **Professionalism** in communications to better understand the needs and expectations. • Utilisation of **digital meeting and collaboration tools** to stay connected.

THANK YOU!



APPENDIX

RELEVANCE MODELLING - DATA DISTRIBUTION

NatWest

Training Data

Class	Size	%
0	2867	91.2
1	276	**8.8**

Validation Data

Class	Size	%
0	965	92.1
1	83	**7.9**

Test Data

Class	Size	%
0	943	90.0
1	105	**10.0**

TEXT CLASSIFICATION - DATA DISTRIBUTION

NatWest

Carbon Class Data

Class	Size	%
0	139	30.2
1	80	17.4
2	25	5.4
3	135	29.2
4	82	17.8

RULE MINING-PROCESS

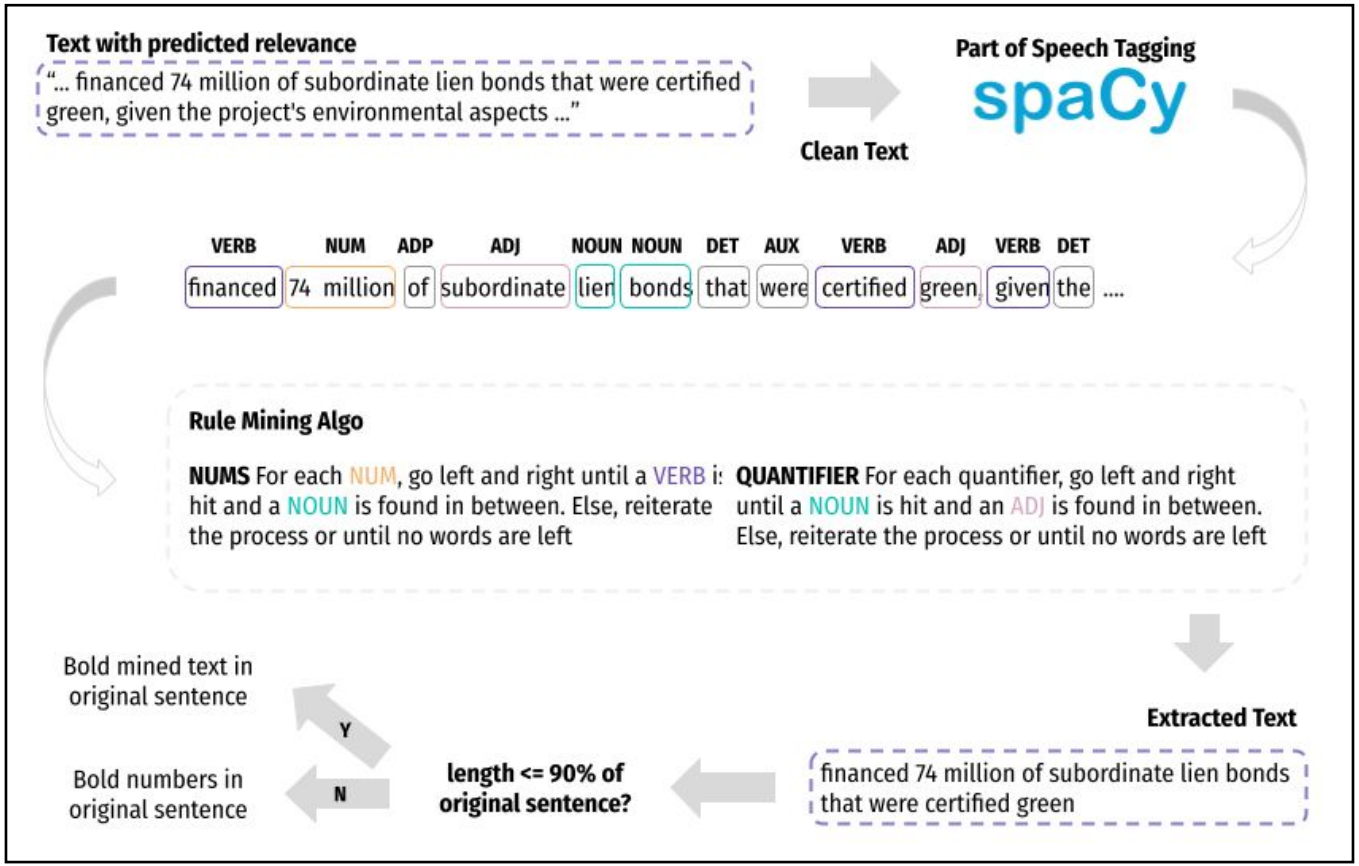


CHART EXTRACTION

Filtering away irrelevant images

NatWest

Image Properties

1	Text ratio
2	Unique Keywords
3	Digit ratio
4	Color Dilation

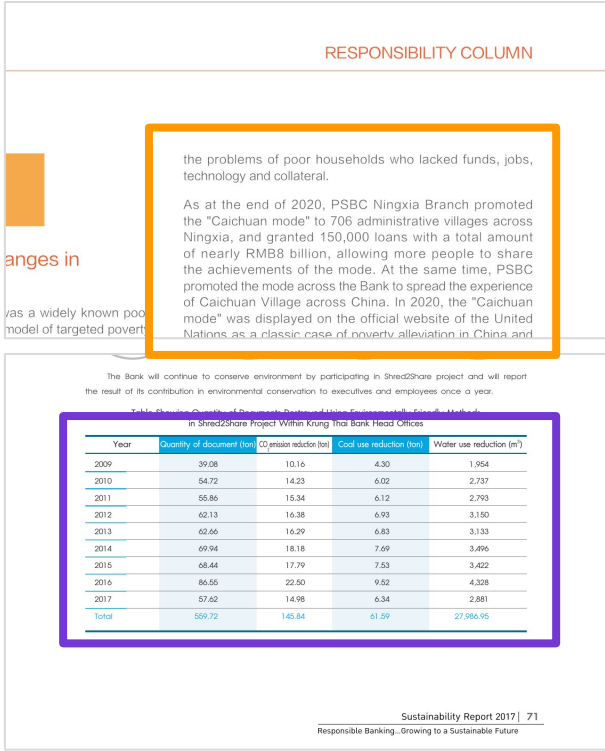


CHART EXTRACTION

Filtering away irrelevant images

Key Metrics			Description	Filter Rationale
Color Property of an Image				
1		Color Dilation (dr)	Total number of colored pixels in an image calculated from using adaptive image thresholding to detect colors and tabulate dilation results. Used adaptive threshold gaussian as optimal threshold compared to about 5 other combinations of image thresholding in OpenCv package.	Filter images with little colors as most graphs usually have a lot of colors.
2		White pix	Total number of white pixels in an image	Filter away images with too much white space indicating that image is a possible white space image.
3		Black pix	Total number of black pixels in an image	
4		Black to White pixels Ratio (BW ratio)	Black pixels compared to white pixels in an image	
Text Content of an Image				
5		Keywords	Total count of unique keywords.	Filter only relevant charts with decarbonisation-related text.
6		Total length of string (total_len)	Total occurrence of string value present in the image.	Filter away images with too much text
7		Total Text count (textonly_len)	Total occurrence of a text present in the image (exclude punctuations)	
8		Text to String Ratio (tt_ratio)	Total occurrence of text divided by total occurrence of string in the image (exclude punctuations)	
8		ta_ratio	Total occurrence of text divided by total area of image (exclude punctuations)	

CHART EXTRACTION

Filtering away irrelevant images

Key Metrics			Description	Filter Rationale
Numerical Content of an Image				
10		Total Number count (numonly_len)	Total occurrence of numbers present in the image.	Filter away images with too little numbers.
11		Number to Area Ratio (na_ratio)	Total occurrence of number divided by total area of image (exclude punctuations)	
12		Number to Text Ratio (nt_ ratio)	Total occurrence of numbers divided by total occurrence of string in the image (exclude punctuations)	
Dimensions of an Image				
13		Height	Height of image	Filter away images with very small height value
14		Width	Width of image	Filter away images with very small width value
15		Area	Area of image	Filter away images with very small area value

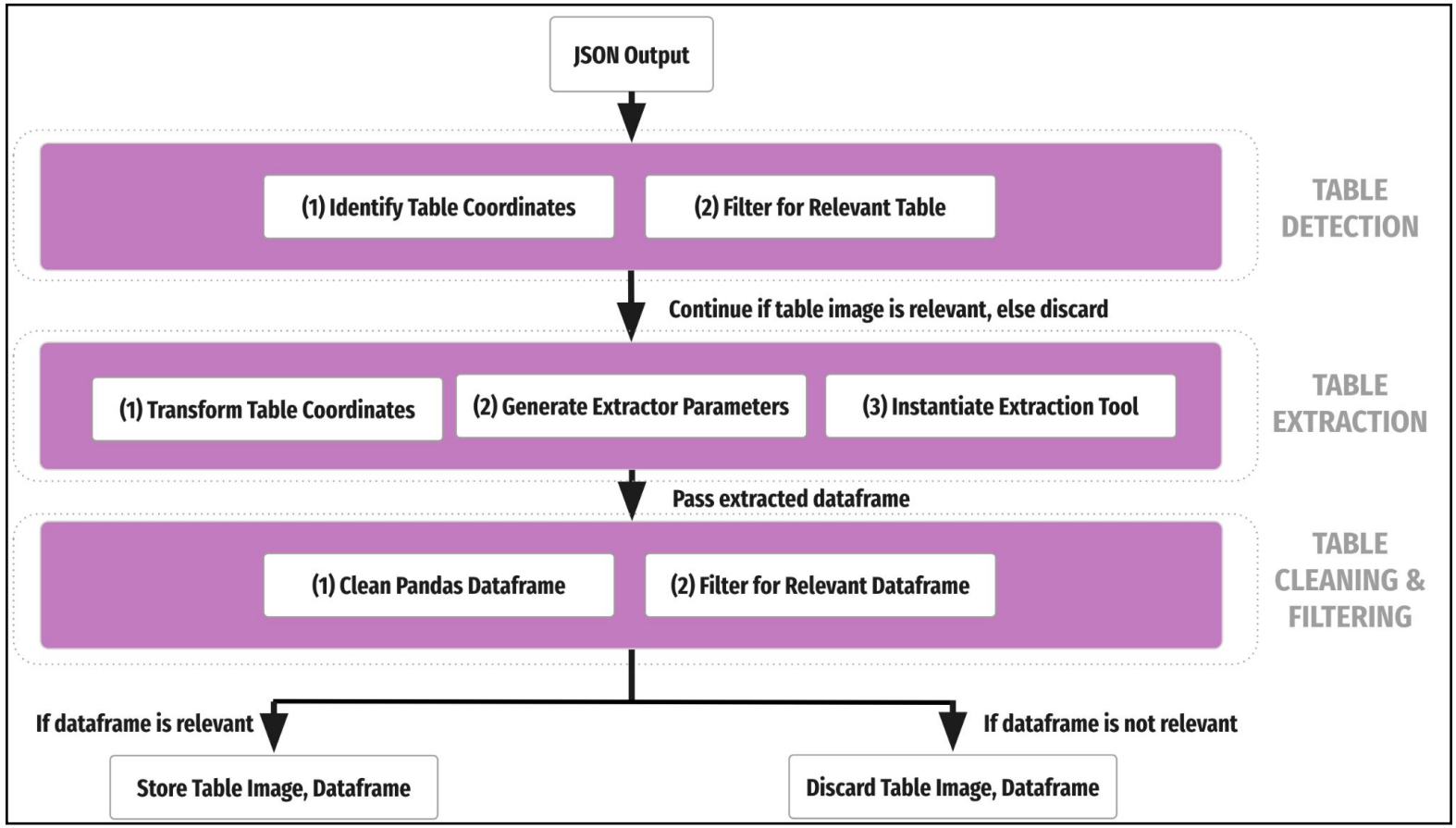
CHART EXTRACTION

Filtering away irrelevant images

NatWest

Column1	source	actual	dr	total_len	keywords	textonly_len	numonly_len	tt_ratio	nt_ratio	height	width	channels	area	ta_ratio	na_ratio	white_pix	black_pix	wb_ratio	bw_ratio
40	ChartExtracti	WRONG	14	4	0	4	0	1	0	467	1537	3	717779	0.56	0	42188	675591	0.06	16.01
156	ChartExtracti	WRONG	15	25	0	25	1	1.04	0.04	467	1534	3	716378	3.63	0.14	43233	673145	0.06	15.57
114	ChartExtracti	WRONG	10	33	0	25	3	0.91	0.15	467	1546	3	721982	4.16	0.69	29847	692135	0.04	23.19
116	ChartExtracti	WRONG	13	13	0	12	2	1	0.15	467	1546	3	721982	1.8	0.28	36923	685059	0.05	18.55
172	ChartExtracti	WRONG	9	25	0	22	0	1	0	468	1407	3	658476	3.8	0	23979	634497	0.04	26.46
138	ChartExtracti	WRONG	24	116	2	82	3	0.97	0.07	667	1654	3	1103218	10.15	0.73	107662	995556	0.11	9.25
179	ChartExtracti	WRONG	12	13	0	13	1	1	0.08	468	1546	3	723528	1.8	0.14	35689	687839	0.05	19.27
149	ChartExtracti	WRONG	21	92	4	39	33	0.58	0.52	759	1500	3	1138500	4.66	4.22	94423	1044077	0.09	11.06
162	ChartExtracti	WRONG	10	24	0	21	2	0.96	0.17	468	1597	3	747396	3.08	0.54	31937	715459	0.04	22.4
63	ChartExtracti	WRONG	14	14	0	12	2	0.86	0.14	471	1537	3	723927	1.66	0.28	40971	682956	0.06	16.67
158	ChartExtracti	WRONG	16	16	0	15	0	1	0	512	1654	3	846848	1.89	0	53499	793349	0.07	14.83
188	ChartExtracti	WRONG	19	10	0	9	1	0.9	0.1	518	1654	3	856772	1.05	0.12	66195	790577	0.08	11.94
67	ChartExtracti	WRONG	11	25	0	21	5	0.88	0.24	521	1654	3	861734	2.55	0.7	38152	823582	0.05	21.59
95	ChartExtracti	WRONG	20	168	7	85	35	0.89	0.27	964	1654	3	1594456	9.34	2.88	129378	1465078	0.09	11.32
164	ChartExtracti	WRONG	9	29	0	23	3	0.9	0.1	547	1342	3	734074	3.54	0.41	28486	705588	0.04	24.77
45	ChartExtracti	WRONG	13	5	0	5	0	1	0	555	1042	3	578310	0.86	0	31208	547102	0.06	17.53
30	ChartExtracti	WRONG	5	4	0	3	1	0.75	0.25	555	1573	3	873015	0.34	0.11	20393	852622	0.02	41.81
33	ChartExtracti	WRONG	5	4	0	4	0	1	0	560	1420	3	795200	0.5	0	18207	776993	0.02	42.68
46	ChartExtracti	WRONG	13	9	0	8	1	0.89	0.11	572	1654	3	946088	0.85	0.11	51263	894825	0.06	17.46
65	ChartExtracti	WRONG	17	20	0	18	1	0.95	0.05	580	959	3	556220	3.42	0.18	37383	518837	0.07	13.88
111	ChartExtracti	WRONG	13	37	0	22	7	0.81	0.24	590	1654	3	975860	3.07	0.92	52027	923833	0.06	17.76
147	ChartExtracti	WRONG	11	65	0	52	1	0.98	0.05	595	1510	3	898450	7.12	0.33	38785	859665	0.05	22.16
135	ChartExtracti	WRONG	10	16	0	13	2	0.81	0.19	595	972	3	578340	2.25	0.52	24887	553453	0.04	22.24
154	ChartExtracti	WRONG	6	5	0	5	0	1	0	596	934	3	556664	0.9	0	13604	543060	0.03	39.92
160	ChartExtracti	WRONG	6	9	0	9	0	1	0	596	1680	3	1001280	0.9	0	25990	975290	0.03	37.53
72	ChartExtracti	WRONG	14	46	0	40	1	1	0.04	598	1147	3	685906	6.71	0.29	38736	647170	0.06	16.71
58	ChartExtracti	WRONG	15	38	0	32	2	0.92	0.08	598	1207	3	721786	4.85	0.42	44499	677287	0.07	15.22
94	ChartExtracti	WRONG	13	28	0	25	1	1	0.04	600	1013	3	607800	4.61	0.16	32883	574917	0.06	17.48
132	ChartExtracti	WRONG	12	27	0	22	3	0.89	0.11	601	1017	3	611217	3.93	0.49	28914	582303	0.05	20.14
66	ChartExtracti	WRONG	12	32	0	24	3	0.88	0.13	601	1871	3	1124471	2.49	0.36	56045	1068426	0.05	19.06
181	ChartExtracti	WRONG	15	51	0	38	7	0.84	0.16	601	1027	3	617227	6.97	1.3	37617	579610	0.06	15.41
157	ChartExtracti	WRONG	4	9	0	9	0	1	0	602	1489	3	896378	1	0	17325	879053	0.02	50.74
192	ChartExtracti	WRONG	20	71	3	51	8	1.04	0.15	678	1500	3	1017000	7.28	1.08	81804	935196	0.09	11.43
55	ChartExtracti	WRONG	4	5	0	5	0	1	0	608	933	3	567264	0.88	0	10755	556509	0.02	51.74
68	ChartExtracti	WRONG	6	9	0	9	0	1	0	608	1681	3	1022048	0.88	0	25644	996404	0.03	38.86

TABULAR DATA PIPELINE - FULL PIPELINE



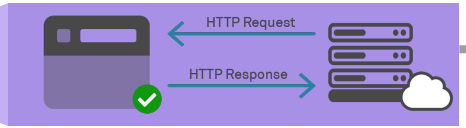
TEXT PREPROCESSING

TABLE DETECTION

NatWest

JSON INPUT

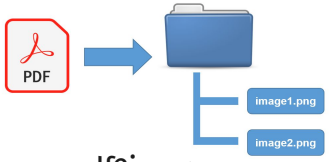

```
{
    "company": "DBS",
    "year": "2020",
    "pdf_url": "https://www.dbs.com/
    "filtered_report_tables_direct": {
        "21": "pillar 1 : responsible banl
        "24": "23 to manage our loan book
        "38": "37 our approach as a finan
        "39": "pillar 2 : responsible bus
        "51": "51 information on environm
        "64": "methodology for tcfd metho
        "65": "65 methodologyonquantifyin
        "69": "69 independent limited ass
    },
}
```

Library: Requests

CONVERT PDF TO IMAGE

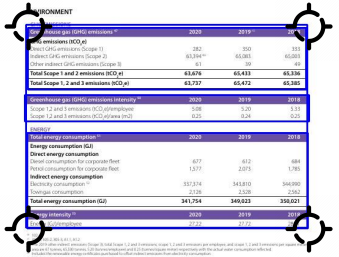
PDF TO IMAGE



Library: pdf2image

TABLE DETECTION

COORDINATES EXTRACTED



Model: Multi-Type-TD-TSR

CONVERT IMAGE TO STRING

TABLE IMAGE TO TEXT

Library: Pytesseract

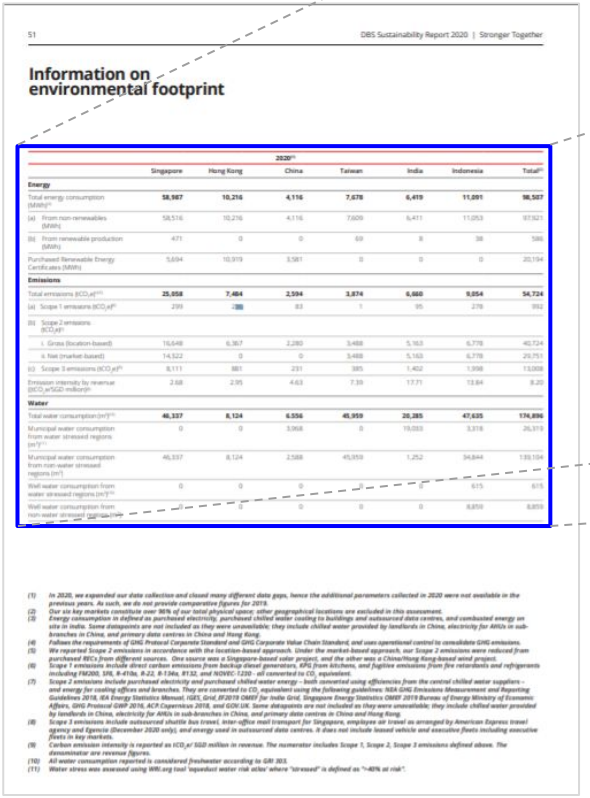
LOOK FOR KEYWORDS IN STRING

RELEVANT IMAGE FOUND

TABULAR DATA EXTRACTION

TABLE DETECTION

using Multi-Type-TD-TSR



51 DBS Sustainability Report 2020 | Stronger Together

Information on environmental footprint

	2020[1]						
	Singapore	Hong Kong	China	Taiwan	India	Indonesia	Total[2]
Energy							
Total energy consumption (MWh)[3]	**58,987**	**10,216**	**4,116**	**7,678**	**6,419**	**11,091**	**98,507**
(a) From non-renewables (MWh)	58,516	10,216	4,116	7,609	6,411	11,053	97,921
(b) From renewable production (MWh)	471	0	0	69	8	38	586
Purchased Renewable Energy Certificates (MWh)	5,694	10,919	3,581	0	0	0	20,194
Emissions							
Total emissions (tCO_2e)[4,5]	**25,058**	**7,484**	**2,594**	**3,874**	**6,660**	**9,054**	**54,724**
(a) Scope 1 emissions (tCO_2e)[6]	299	236	83	1	95	278	992
(b) Scope 2 emissions (tCO_2e)[7]							
i. Gross (location-based)	16,648	6,367	2,280	3,488	5,163	6,778	40,724
ii. Net (market-based)	14,322	0	0	3,488	5,163	6,778	29,751
(c) Scope 3 emissions (tCO_2e)[8]	8,111	881	231	385	1,402	1,998	13,008
Emission intensity by revenue (tCO_2e/SGD million)[9]	2.68	2.95	4.63	7.39	17.71	13.84	8.20
Water							
Total water consumption (m^3)[10]	**46,337**	**8,124**	**6,556**	**45,959**	**20,285**		
Municipal water consumption from water stressed regions (m^3)[11]	0	0	3,968	0	19,033		
Municipal water consumption from non-water stressed regions (m^3)	46,337	8,124	2,588	45,959	1,252		
Well water consumption from water stressed regions (m^3)[11]	0	0	0	0	0	6,	
Well water consumption from non-water stressed regions (m^3)	0	0	0	0	0	8,859	

(1) *In 2020, we expanded our data collection and closed many different data gaps, hence the additional parameters collected in 2020 were not available in the previous years. As such, we do not provide comparative figures for 2019.*
(2) *Our six key markets constitute over 96% of our total physical space; other geographical locations are excluded in this assessment.*
(3) *Energy consumption is defined as purchased electricity, purchased chilled water cooling to buildings and outsourced data centres, and combusted energy on site in India. Some datapoints are not included as they were unavailable; they include chilled water provided by landlords in China, electricity for ATMs in sub-branches in China, and primary data centres in China and Hong Kong.*
(4) *Follows the requirements of GHG Protocol Corporate Standard and GHG Corporate Value Chain Standard, and uses operational control to consolidate GHG emissions.*
(5) *We reported Scope 2 emissions in accordance with the location-based approach. Under the market-based approach, our Scope 2 emissions were reduced from purchased RECs from different sources. One source was a Singapore-based solar project, and the other was a China/Hong Kong-based wind project.*
(6) *Scope 1 emissions include direct carbon emissions from backup diesel generators, KPG from kitchens, and fugitive emissions from fire retardants and refrigerants including FM200, 3M, R-410a, R-22, R-134a, R132, and NOVEC 1230 - all converted to CO_2 equivalent.*
(7) *Scope 2 emissions include purchased electricity and purchased chilled water energy – both converted using efficiencies from the central chilled water suppliers – and energy for cooling offices and branches. They are converted to CO_2 equivalent using the following guidelines: NEA GHG Emissions Measurement and Reporting Guidelines 2018, IEA Energy Statistics Manual, IGES, Grid_EF:2019 CMEF for India Grid, Singapore Energy Statistics CMIEF 2019 Bureau of Energy Ministry of Economic Affairs, GHG Protocol GWP 2016, ACP Copernicus 2018, and GCM.UK. Some datapoints are not included as they were unavailable; they include chilled water provided by landlords in China, electricity for ATMs in sub-branches in China, and primary data centres in China and Hong Kong.*
(8) *Scope 3 emissions include outsourced shuttle bus travel, inter-office mail transport for Singapore, employee air travel as arranged by American Express travel agency and Egencia (December 2020 only), and energy used in outsourced data centres. It does not include leased vehicle and executive fleets including executive fleets in key markets.*
(9) *Carbon emission intensity is reported as tCO_2e/ SGD million in revenue. The numerator includes Scope 1, Scope 2, Scope 3 emissions defined above. The denominator are revenue figures.*
(10) *All water consumption reported is considered freshwater according to GRI 303.*
(11) *Water stress was assessed using WRI.org tool 'aqueduct water risk atlas' where "stressed" is defined as ">40% at risk".*

Source: DBS Sustainability Report 2020

	2020[1]						
	Singapore	Hong Kong	China	Taiwan	India	Indonesia	Total[2]
Energy							
Total energy consumption (MWh)[3]	**58,987**	**10,216**	**4,116**	**7,678**	**6,419**	**11,091**	**98,507**
(a) From non-renewables (MWh)	58,516	10,216	4,116	7,609	6,411	11,053	97,921
(b) From renewable production (MWh)	471	0	0	69	8	38	586
Purchased Renewable Energy Certificates (MWh)	5,694	10,919	3,581	0	0	0	20,194
Emissions							
Total emissions (tCO_2e)[4,5]	**25,058**	**7,484**	**2,594**	**3,874**	**6,660**	**9,054**	**54,724**
(a) Scope 1 emissions (tCO_2e)[6]	299	236	83	1	95	278	992
(b) Scope 2 emissions (tCO_2e)[7]							
i. Gross (location-based)	16,648	6,367	2,280	3,488	5,163	6,778	40,724
ii. Net (market-based)	14,322	0	0	3,488	5,163	6,778	29,751
(c) Scope 3 emissions (tCO_2e)[8]	8,111	881	231	385	1,402	1,998	13,008
Emission intensity by revenue (tCO_2e/SGD million)[9]	2.68	2.95	4.63	7.39	17.71	13.84	
Water							
Total water consumption (m^3)[10]	**46,337**	**8,124**	**6,556**	**45,959**	**20,285**		
Municipal water consumption from water stressed regions (m^3)[11]	0	0	3,968	0	19,033		
Municipal water consumption from non-water stressed regions (m^3)	46,337	8,124	2,588	45,959	1,252		
Well water consumption from water stressed regions (m^3)[11]	0	0	0	0	0	6,	
Well water consumption from non-water stressed regions (m^3)	0	0	0	0	0	8,859	

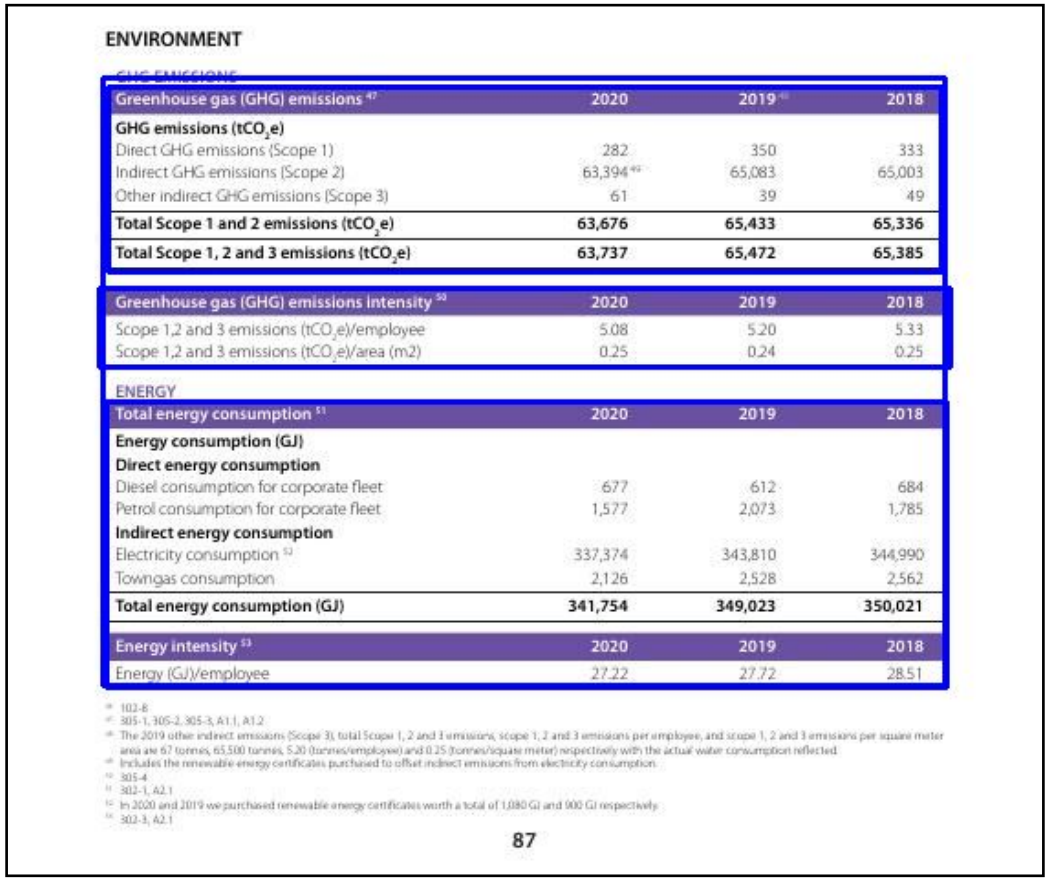
RELEVANT

at least 1 keyword
+
1 unit of measurement found

```
# list of keywords
keywords = ['direct ghg emissions', 'indirect ghg emissions', 'scope 1',
            'scope 2', 'scope 3', 'energy use', 'energy consumption',
            'paper consumption', 'green bonds', 'renewable energy',
            'water consumption', 'water usage', 'carbon intensity',
            'carbon emissions', 'waste management', 'waste output',
            'waste generated', 'electricity consumption',
            'weighted average carbon intensity', 'WACI']

# list of units
units = ['tonnes', 'tons', 'kWh', ' kg ', 'kilogram', 'kilowatt hour',
         'gigajoules', 'GJ', 'litre', 'liter', 'CO2e', 'tCO', 't CO', 'MWh',
         'megawatt hour', '%', 'cubic metres', 'per employee']
```

TABLE DETECTION - OVERLAPPING REGIONS



ENVIRONMENT

GHG EMISSIONS

Greenhouse gas (GHG) emissions [47]	2020	2019 [48]	2018
GHG emissions (tCO_2e)			
Direct GHG emissions (Scope 1)	282	350	333
Indirect GHG emissions (Scope 2)	63,394 [49]	65,083	65,003
Other indirect GHG emissions (Scope 3)	61	39	49
Total Scope 1 and 2 emissions (tCO_2e)	**63,676**	**65,433**	**65,336**
Total Scope 1, 2 and 3 emissions (tCO_2e)	**63,737**	**65,472**	**65,385**

Greenhouse gas (GHG) emissions intensity [50]	2020	2019	2018
Scope 1,2 and 3 emissions (tCO_2e)/employee	5.08	5.20	5.33
Scope 1,2 and 3 emissions (tCO_2e)/area (m2)	0.25	0.24	0.25

ENERGY

Total energy consumption [51]	2020	2019	2018
Energy consumption (GJ)			
Direct energy consumption			
Diesel consumption for corporate fleet	677	612	684
Petrol consumption for corporate fleet	1,577	2,073	1,785
Indirect energy consumption			
Electricity consumption [52]	337,374	343,810	344,990
Towngas consumption	2,126	2,528	2,562
Total energy consumption (GJ)	**341,754**	**349,023**	**350,021**

Energy intensity [53]	2020	2019	2018
Energy (GJ)/employee	27.22	27.72	28.51

[47] 102-8
[48] 305-1, 305-2, 305-3, A1.1, A1.2
[49] The 2019 other indirect emissions (Scope 3), total Scope 1, 2 and 3 emissions, scope 1, 2 and 3 emissions per employee, and scope 1, 2 and 3 emissions per square meter area are 67 tonnes, 65,500 tonnes, 5.20 (tonnes/employee) and 0.25 (tonnes/square meter) respectively with the actual water consumption reflected.
[50] Includes the renewable energy certificates purchased to offset indirect emissions from electricity consumption.
[51] 305-4
[52] 302-1, A2.1
[53] In 2020 and 2019 we purchased renewable energy certificates worth a total of 1,080 GJ and 900 GJ respectively.
[54] 302-3, A2.1

87

IMAGE-BASED TABLE EXTRACTION

TABLE EXTRACTION

NatWest

RELEVANT IMAGE FOUND


forming countries for e-participation in Asia are the Republic of Korea and Japan. Finally, Australia and New Zealand still lead Oceania.

Table ES.2 World and regional e-participation leaders

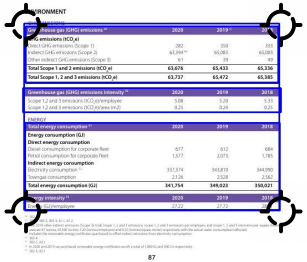
World e-participation leaders	Regional e-participation leaders	
Netherlands	AFRICA	Morocco
Republic of Korea		Kenya
Uruguay	AMERICAS	Uruguay
France		Chile
Japan	ASIA	Republic of Korea
United Kingdom		Japan
Australia	EUROPE	Netherlands
Chile		France
United States of America	OCEANIA	Australia
Singapore		New Zealand

The critical need for whole-of government-approaches and collaborative governance

Although sustainable development challenges have significantly changed over

EXTRACT IMAGE COORDINATES

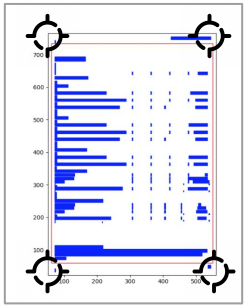
IMAGE COORDINATES


ENVIRONMENT

Model: Multi-Type-TD-TSR

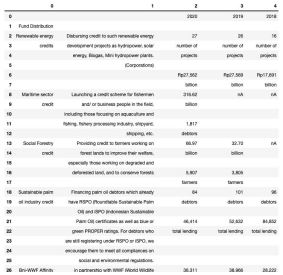
COORDINATES TRANSFORMATION

PDF COORDINATES



PASS PDF COORDINATES, EDGE_TOL, SPLIT_TEXT= True INTO TABLE EXTRACTOR

DATAFRAMES EXTRACTED



	0	1	2	3	4
0			2020	2019	2018
1	Fund Distribution				
2	Renewable energy	Disbursing credit to such renewable energy	27	28	16
3	credits	development projects as hydropower, solar	number of	number of	number of
4		energy, Biogas, Mini hydropower plants.	projects	projects	projects
5			(Corporations)		
6			Rp27,562	Rp27,559	Rp17,691
7			billion	billion	billion
8	Maritime sector	Launching a credit scheme for fishermen	319.62	nA	nA
9	credit	and/ or business people in the field,	billion		
10		including those focusing on aquaculture and			
11		fishing, fishery processing industry, shipyard,	1,817		
12		shipping, etc.	#ebitions		
13	Social Forestry	Providing credit to farmers working on	66.97	32.73	nA
14	credit	forest lands to improve their welfare,	billion	billion	
15		especially those working on degraded and			
16		deforested land, and to conserve forests	5,907	3,805	
17			farmers	farmers	
18	Sustainable palm	Financing palm oil debtors which already	84	101	96
19	oil industry credit	have RSPO (Roundtable Sustainable Palm	debtors	debtors	debtors
20		Oil) and ISPO (Indonesian Sustainable			
21		Palm Oil) certificates as well as blue or	48,414	52,632	84,652
22		green PROPER ratings. For debtors who	total lending	total lending	total lending
23		are still registering under RSPO or ISPO, we			
24		encourage them to meet all compliances on			
25		social and environmental regulations.			
26	Bio-WWF Afforde	In partnership with WWF (World Wildlife	36,311	38,966	28,222

Library: Camelot

CLEAN UP & ASSESS RELEVANCE OF DATAFRAME

RELEVANT TABLE FOUND

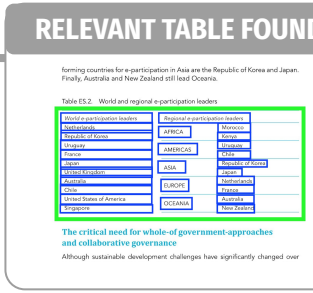

forming countries for e-participation in Asia are the Republic of Korea and Japan. Finally, Australia and New Zealand still lead Oceania.

Table ES.2 World and regional e-participation leaders

World e-participation leaders	Regional e-participation leaders	
Netherlands	AFRICA	Morocco
Republic of Korea		Kenya
Uruguay	AMERICAS	Uruguay
France		Chile
Japan	ASIA	Republic of Korea
United Kingdom		Japan
Australia	EUROPE	Netherlands
Chile		France
United States of America	OCEANIA	Australia
Singapore		New Zealand

The critical need for whole-of government-approaches and collaborative governance

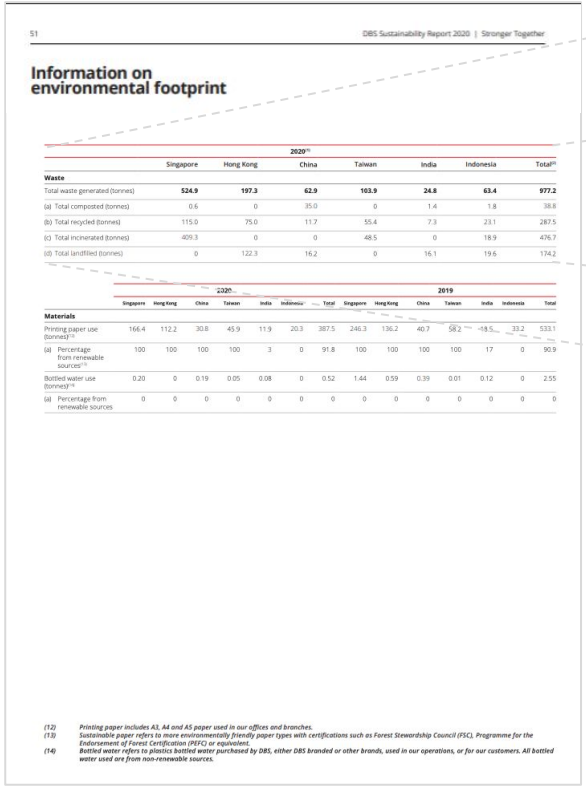
Although sustainable development challenges have significantly changed over

STORE THE RELEVANT DATAFRAME

TABLE EXTRACTION

using Multi-Type-TD-TSR & Camelot

at least 1 numeric metric
+
no "page" keyword found



51

Information on environmental footprint

	Singapore	Hong Kong	China	Taiwan	India	Indonesia	Total[5]
Waste							
Total waste generated (tonnes)	**524.9**	**197.3**	**62.9**	**103.9**	**24.8**	**63.4**	**977.2**
(a) Total composted (tonnes)	0.6	0	35.0	0	1.4	1.8	38.8
(b) Total recycled (tonnes)	115.0	75.0	11.7	55.4	7.3	23.1	287.5
(c) Total incinerated (tonnes)	409.3	0	0	48.5	0	18.9	476.7
(d) Total landfilled (tonnes)	0	122.3	16.2	0	16.1	19.6	174.2

	2020							2019						
	Singapore	Hong Kong	China	Taiwan	India	Indonesia	Total	Singapore	Hong Kong	China	Taiwan	India	Indonesia	Total
Materials														
Printing paper use (tonnes)[12]	166.4	112.2	30.8	45.9	11.9	20.3	387.5	246.3	136.2	40.7	58.2	18.5	33.2	533.1
(a) Percentage from renewable sources[13]	100	100	100	100	3	0	91.8	100	100	100	100	17	0	90.9
Bottled water use (tonnes)[14]	0.20	0	0.19	0.05	0.08	0	0.52	1.44	0.59	0.39	0.01	0.12	0	2.55
(a) Percentage from renewable sources	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(12) Printing paper includes A3, A4 and A5 paper used in our offices and branches.
(13) Sustainable paper refers to more environmentally friendly paper types with certifications such as Forest Stewardship Council (FSC), Programme for the Endorsement of Forest Certification (PEFC) or equivalent.
(14) Bottled water refers to plastics bottled water purchased by DBS, either DBS branded or other brands, used in our operations, or for our customers. All bottled water used are from non-renewable sources.

Source: DBS Sustainability Report 2020

	0	1	2	3	4	5	6	7
0	2020<s>(1)</s>							
1		Singapore	Hong Kong	China	Taiwan	India	Indonesia	Total<s>(2)</s>
2	Waste							
3	Total waste generated (tonnes)	524.9	197.3	62.9	103.9	24.8	63.4	977.2
4	(a) Total composted (tonnes)	0.6	0	35.0	0	1.4	1.8	38.8
5	(b) Total recycled (tonnes)	115.0	75.0	11.7	55.4	7.3	23.1	287.5
6	(c) Total incinerated (tonnes)	409.3	0	0	48.5	0	18.9	476.7
7	(d)\t Total\tlandfilled\t(tonnes)	0	122.3	16.2	0	16.1	19.6	174.2

Regular Expression in Pandas

	0	1	2	3	4	5	6	7
0				2020				
1		Singapore	Hong Kong	China	Taiwan	India	Indonesia	Total
2	Waste							
3	Total waste generated (tonnes)	524.9	197.3	62.9	103.9	24.8	63.4	977.2
4	(a) Total composted (tonnes)	0.6	0	35.0	0	1.4	1.8	38.8
5	(b) Total recycled (tonnes)	115.0	75.0	11.7	55.4	7.3	23.1	287.5
6	(c) Total incinerated (tonnes)	409.3	0	0	48.5	0	18.9	476.7
7	(d)\t Total\tlandfilled\t(tonnes)	0	122.3	16.2	0	16.1	19.6	174.2

TABLE EXTRACTION - UNCLEANED DF

	Environmental
1	A1 Emissions
2	
3	General Disclosure: Information on the po...
4	compliance with relevant laws and regulations ...
5	a significant impact on the issuer relat...
6	greenhouse gas emissions discharges into ...
7	land and generation of hazardous and non-hazar...
8	waste
9	
10	A1.1 The types of emissions and respective emi...
11	
12	A1.2 Direct (Scope 1) and energy indirec...
13	2) greenhouse gas emissions (in tons) an...
14	appropriate intensity (e.g. per unit of produc...
15	per facility)
16	

	Indirect Economi	
1		50
2	G4-DMA	71
3		71
4		71
6	G4-DMA	97
7		97
9	G4-DMA	93
10		93
12	G4-DMA	92
13		92
15	G4-LA12	96
18	G4-DMA	124
19		124
21	G4-DMA	103
22		103
24	G4-DMA	103
25		105
28	G4-DMA	52
31		52
32		50
33	G4-FS14	107
34	G4-FS16	110
35		ity Report 2016

TABLE EXTRACTION - RELEVANT TABLES EG

Indicator	2018	2019	2020
Total greenhouse gas emission of Head Office (Scope 1 + Scope 2) (tons of CO_2 equivalent)	8,546.83	10,059.62	9,592.69
Paper saved by using e-bills (100 million pieces)	15.48	17.91	19.04
Power usage effectiveness (PUE) of data centers	1.72	1.65	1.62
Green loan balance (RMB 100 million)	1,660.33	1,767.73	2,071.33

Sector	Segments	Sample population as per cent of total non-bank loans	Sample population weighted carbon intensity (tCO_2e/$million)
Cement Manufacturing	Building materials	0.1%	7,602
Energy	Utilities, and oil and gas	1.1%	1,813
Metals and Mining	Coal, ferrous and non-ferrous mining, and manufacturing of metals	0.5%	648
Transportation	Land transport, air transport and water transport	1.0%	648
Agricultural	Agriculture and livestock production, manufacturing of agriculture products, and wholesale and trading of agriculture and livestock products	1.2%	433
Forestry	Logging, production of wood, and manufacturing of pulp and paper	0.2%	396
Chemicals	Manufacturing of chemicals	0.2%	355
Manufacturing of Transport Equipment	Automobile manufacturing	0.2%	44
Infrastructure	Operations of land, water and infrastructure	0.1%	41
	Total	5%	–

Energy [1] [2]	2018 [7]	2019 [7]	2020
Electricity consumption (MWh)	94,903	123,042	**114,887**
Electricity Intensity (kWh/square feet)	21.93	21.21	**21.29**
Emissions [1] [2] [3] [4]	**2018** [7]	**2019** [7]	**2020**
Scope 2 emissions (tonnes CO_2e)	47,880	71,361	**65,188**
Scope 2 emissions intensity (tonnes CO_2e/square feet)	0.0111	0.0123	**0.0121**
Water [2] [5]	**2018**	**2019** [7]	**2020**
Total water usage (ML)	501.46	591.94	**507.51**
Water usage intensity (ML/square feet)	0.0001	0.0001	**0.00009**
Paper [6]	**2018**	**2019**	**2020**
Office paper usage (tonnes)	258	239	**204**

Allianz Global Investors environmental data at a glance	2019	2020
Total GHG emissions (tons per employee)	3.5	1.9
Energy consumption	1.6	1.4
Business travel	1.9	0.5
Paper consumption	0.03	0.02
Share of renewable energy in the mix (%)		
Share of renewable energy	46	47
Water consumption (cubic metres per employee)		
Water consumption	26	21
Waste output (kg per employee)		
Waste output	149	90