Automated Corporate Climate Risk Disclosure Extraction and TCFD Readiness Assessment

NASDAQ ESG Advisory

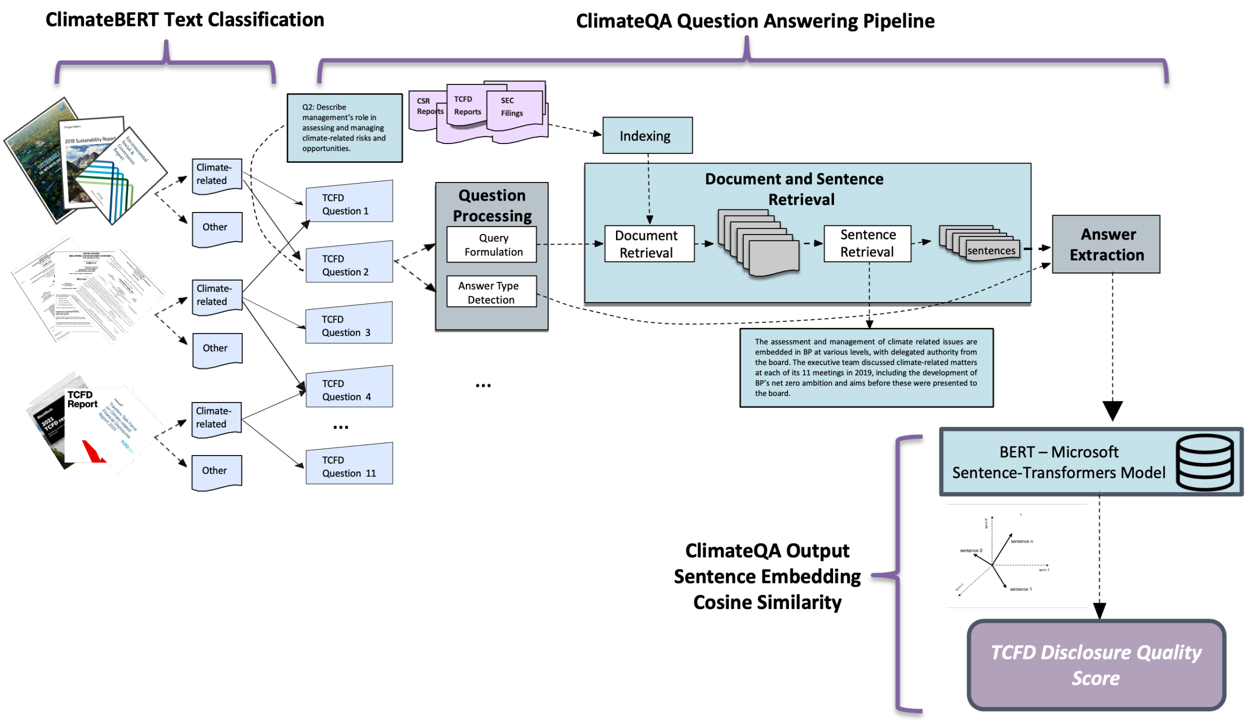
Abstract

Climate-related financial risk disclosure by companies helps investors understand companies' readiness for climate change. Disclosures based on the recommendations of the Task Force for Climate-related Financial Disclosures (TCFD) are designed to provide consistent and transparent information to investors about companies' carbon-related assets and their exposures to climate-related risks and opportunities [1]. The abundance of largely qualitative data constrains analysts to go through hundreds of pages of company filings to identify and extract TCFD-relevant company disclosures for readiness assessment. Given that TCFD recommendations on climate-related financial disclosures are widely adopted across sectors and the diversity of climate topics they cover, we propose an end-to-end pipeline of ML models trained on 11 general TCFD recommendations to identify climate-related disclosures in company filings, respond to TCFD disclosure recommendations, and evaluate companies' climate disclosure readiness.

In particular, neural language models have been designed to:

- Identify climate-related texts in company reports by further training ClimateBERT model [2]
- Automatically extract TCFD recommended disclosures based on a question-answering approach by fine-tuning ClimateQA model [3]
- Conduct company TCFD readiness assessment and classify across weak, good, and excellent categories
- Generate automated company climate benchmarking reports

Below, we present the proposed climate disclosure text classification, TCFD question answering, and readiness assessment pipeline. The process involves gathering all available company-published reports including newsroom articles and SEC filings, applying our fine-tuned ClimateBERT model for climate-relevant text identification, providing extracted climate texts to our fine-tuned ClimateQA model to retrieve company disclosures across each TCFD topic, quantifying a company's readiness across each TCFD disclosure topic, and reporting company and peer benchmark results in excel and PDF.



A Proposed Climate Disclosure Text Identification and TCFD Question Answering Model Pipeline

---

1. Methodology & Approach

Natural language models have significantly advanced the state of the art for NLP tasks [4]. In particular, BERT is the first fine-tuning based representation model that achieves state-of-the-art performance on a large suite of task-specific architectures from text classification to question answering and natural language inference (NLI) [4]. In recent years, different variations of original BERT architectures have also been used for domain-specific applications, from financial or biomedical text mining to sentiment analysis, and have already been trained on financial documents and climate change-related documents [2]. BERT requires a fine-tuning process in great detail with domain-specific datasets to train the algorithm for specific downstream tasks. In 2019, Facebook AI research team found that BERT was significantly undertrained and suggested an improved recipe for its training, called RoBERTa [5]. We adopted this Transformer-based approach, namely — ClimateBERT and ClimateQA — tools for extracting climate-relevant passages and apply question-answering in company reports based on TCFD recommendations. Both of our models are based on RoBERTa architecture.

2. RoBERTa-based Climate-related Text Classification

We first leverage ClimateBERT model [2] developed by a group of researchers from Germany and Switzerland for our climate-related text classification task. We frame it as a binary classification problem (climate-related vs non-climate-related).

According to Webersinke et al. (2021) climate-related textual analysis often involved using predefined set of presumably relevant words and then simply searched for these words in reports. For example, search the word "target" or "committment" in TCFD Metrics & Targets c) — describe the targets used by the organization to manage climate-related risks and opportunities and performance against targets.

However, keyword-based methods do not account for context. This lack of context may cause a considerable confusion given the ambiguity of many climate-related words such as "environment," "target", "sustainable," or even "climate". For example, a separate passage from Microsoft's Sustainability Report [6] below will not be identified as an "answer" to TCFD Governance Disclosure Recommendation by a simple automated key word search because non of "board", "management", or "climate" appears in the text below. We do not know the context on what "this team" is being referred to (i, e., Board vs Management, etc.), which is in this case referred to their "Carbon Program Manager, part of the Environmental Sustainability (ES) team that leads Microsoft's carbon mitigation efforts":

Table 1: Example Confusion of Automated Rule-based Disclosure Identification

TCFD Disclosure	Company Dislcosure against TCFD recommendations
Describe management's role in assessing and managing climate-related risks and opportunities.	"For guidance on globally changing dynamics, ***this team*** engages with experts around the world, including internal finance, regulatory/policy, technology and environmental professionals, as well as external subject matter experts."

However, our transformers-based models are capable of accounting for the context of words and as we will see in the following sections have outperformed traditional approaches by large margins. ClimateBERT is a transformer-based language model that, further pretrained on over 1.6 million paragraphs of climate-related texts, gathered from various sources such as common news, research articles, and climate reporting of companies [2]. According to Webersinke et al. (2021) ClimateBERT leads to a 46% improvement on a masked language model objective which, in turn, leads to lowering error rates by 3.57% to 35.71% for various climate-related downstream tasks like text classification, sentiment analysis, and fact-checking.

The table below illustrates the results of climate text classification task using ClimateBERt that led to improvements of 32.64% in terms of cross-entropy loss and a reduction in the error rate of the F1 score by 35.71%.

Hence, we leverage state-of-the-art ClimateBERT to achieve our first goal. We took the pretrained ClimateBERT model as a base model and further fine-tuned it for our downstream binary text classification task. For the text classification experiment, authors of Cli-

Table 2: Results on climate text classification task. Reported are the average cross-entropy loss and the average weighted F1 score from ClimateBERT paper [2]

	Text classification	
Model	Loss	F1
DistilRoBERTa	0.242	0.986
ClimateBERT	0.163	0.991

mateBERT used a dataset consisting of hand-selected paragraphs from companies' annual reports or sustainability reports. All paragraphs were annotated as yes (climate-related) or no (not climate-related) by at least four experts from the field using the software prodigy [2]. We further hand-labeled equally-weighted 1000 paragraphs of climate vs non-climate texts from companies' sustainability reports and fine-tuned ClimateBERT for our use-case. After accurately identifying and extracting climate-related paragraphs from company-published reports, we move forward to our core task — TCFD Question-Answering Model.

3. RoBERTa-based TCFD Question-Answering Model

The next step in the pipeline has been to build a "closed domain" question-answering model that allows for automated climate disclosure extraction to respond to TCFD disclosures. We leverage previous work done by Luccioni et al. (2020) [3] who trained ClimateQA model based on 14 TCFD recommended disclosures. The team hand-labeled climate disclosures for 14 TCFD recommendations and trained RoBERTa model for automated answer extraction. More specifically, the team trained ClimateQA model on 15,000 negative examples and 1,500 positive examples, with the development set comprised of 7,500 negative examples and 750 positive examples, and the test set having 1,200 negative and 400 positive examples. The model achieved a significant success in terms of its accuracy and consistency. However, the model has only been trained to respond "yes or no" to whether each of these 14 TCFD recommendations existed in the reports. We further worked with the ClimateQA and its data output to design a pipeline to iterate across each climate-related paragraph and respond to TCFD questions to actually retrieve qualitative answers.

Another adjustment we were required to make to the ClimateQA model has been to go from 14 TCFD recommendations to 11 recommendations. This adjustment has further improved the accuracy by ~3% on average across each question. We also sourced training data from sustainability and TCFD reports and trained our model on 200 new positive question-answer examples, in areas where the model needed improvement (i.e., Q3, Q11, etc.). Yet another improvement we made was to create a binary classification model for the TCFD Q11 — describe the targets used by the organization to manage climate-related risks — to filter out non-relevant targets identified by the companies without strictly compromising the quality of answer extraction in other areas. In other words, it was identified as operationally more pragmatic not to add biased training data in areas in which QA model needed improvement but rather build a separate classification "yes/no" model.

Table 3: Examples of Question-Answer pairs from our training corpus

TCFD Strategy a)	Risk	Description	Answer Passage
Describe the climate-related risks and opportunities the organization has identified over the short, medium, and long term.	Transition Risk	Market	Reduced demand for goods and services due to shift in consumer preferences or changes in purchasing power
		Technology	Increased costs related to data center resiliency
		Policy and Legal	Increased compliance costs and potential disruption related to new mandates and regulations on existing products
	Physical Risks	Acute	Increased costs from repairing or restoring damaged locations
		Chronic	Increased cost related to relocation due to sea level rise

Fine-tuned ClimateQA results by each TCFD question is reported in Table 4. We achieved a significant improvement in terms of the accuracy and efficiency of ClimateQA model by only providing the model climate-related paragraphs, further training it for our use-case on custom training data, and adapting to the new TCFD framework recommendations.

4. Siamese BERT-Network-based TCFD Readiness Assessment Score

The final layer in our pipeline is the design of a TCFD disclosure quality score to measure companies' climate disclosure readiness for climate change across each TCFD disclosure by engaging sentence similarity neural language model — Sentence BERT (SBERT) [7].

All of the studies conducted on climate disclosure assessment have been focused on just identification or classification of climate-related disclosures. However, the materiality and quality of such disclosures remain a major challenge due to their complexity and diversity. To the best of our knowledge, this is the first study attempting to not only identify and extract but also score companies' TCFD disclosure readiness.

4.1. Semantic Textual Similarity

A good majority of NLP model architectures rely on similarity in highly-dimensional spaces. Generally, an NLP classification, regression, or question answering solution will take a text and process it to create a big array/vectors representing that, and then perform certain transformations. The difference between using cross-encoder architecture for sentence similarity with RoBERTa and SBERT is that SBERT drops the final classification layer, and attempts to process each sentence at a time. Further to this, unlike other BERT architecture, SBERT is fine-tuned on sentence pairs using a Siamese architecture (two identical BERTs in parallel).

The general logic behind applying Siamese architecture-based text similarity models in our context would be to:

- Take a TCFD recommendation text, convert it into a vector representation (sentence embedding)
- Take company TCFD disclosure assessment (i.e., ClimateQA output) and convert them into vectors
- Score disclosure from the smallest to largest distance (Euclidean) or smallest angle (cosine similarity) between source and each extracted disclosure

4.2. ***BERT Sentence Embeddings***

A brief introduction into BERT's ability to embed the meaning of each word into densely packed vectors will be useful. BERT packed vectors are called dense vectors because every value within the vector has a value and has a reason for being that value [4].

BERT is designed to create these dense vectors, and each of the below encoder layer (there are 12) outputs are a set of dense vectors [4].

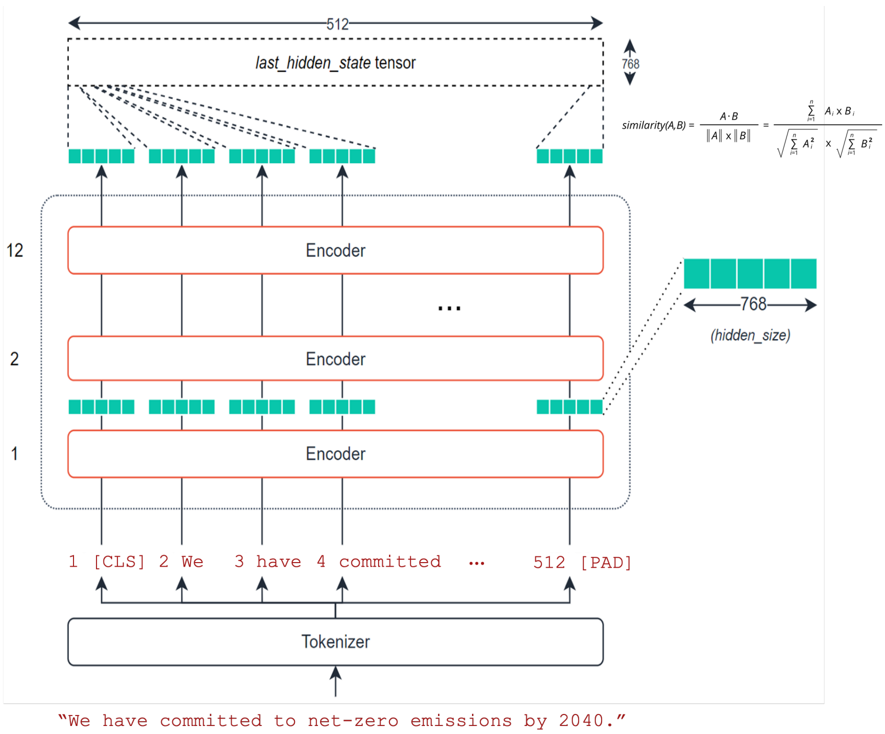


Figure 1: BERT Sentence Embeddings

In case of BERT base, we will have a vector containing 768 values. Those 768 values contain our numerical representation of a single token — which we can use as contextual word embeddings. Given that there is one vector representing each token (output by each encoder), we are looking at an "array" of 768 by 512 — the number of tokens. We take these "array"-s — and create semantic representations of the input sequence from them. After that, we take our similarity metric (i.e., cosine similarity) and calculate the respective similarity between different sequences [8].

Each of our sentences will have 512 tokens with respective 768 values. We then use a mean pooling operation to take the mean of all token embeddings and compress them into a single 768 vector space — creating a "sentence vector". We perform this operation for each piece of text and then take those embeddings and find the cosine similarity between each. The original pre-trained model applied for our task has first been introduced by Microsoft's team under "MINILM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers" [9].

4.3. ***Example Use-Case of BERT Sentence Embedding Representations: Score Generation***

Table 4 shows an example of output in which we define the source text as the TCFD recommendation adjusted based on the NASDAQ TCFD Gap Analysis Tool. For company assessment example, we chose to retrieve examples on TCFD Governance Disclosure a). from Teck and Autodesk reports. The end goal of the readiness assessment is to classify each TCFD disclosure topic across weak/good/excellent categories. In this instance, Teck assessment will fall into the "Excellent" category and the Autodesk assessment will fall into "Weak" category. While both companies disclose information on TCFD Governance part a, we were able to clearly distinguish between excellent and weak disclosure by representing each disclosure as a sentence embedding in a vector space and calculating cosine angle between the reference/source disclosure text and each assessment.

5. Conclusion and Next Steps

In this paper, we developed a pipeline of language models to automatically identify TCFD disclosures in corporate reports and conduct TCFD readiness assessment. We designed an internal TCFD Readiness Assessment Tool to help us evaluate companies' preparedness to the emerging regulatory and investor pressure to report on TCFD recommendations, identify gaps, and benchmark against peers. We automatically evaluate a company's climate related risks and opportunities in line with TCFD recommendations in each sector and give a readiness score to each of the 11 recommendations. Based on this assessment, we can then provide companies with a detailed gap analysis to help improve their compliance with TCFD recommendations and help them understand where their strengths and weaknesses lie relative to their peers. Such internal tool will also serve as an integrated platform to select companies and upload and download their corporate annual TCFD or Sustainability reports.

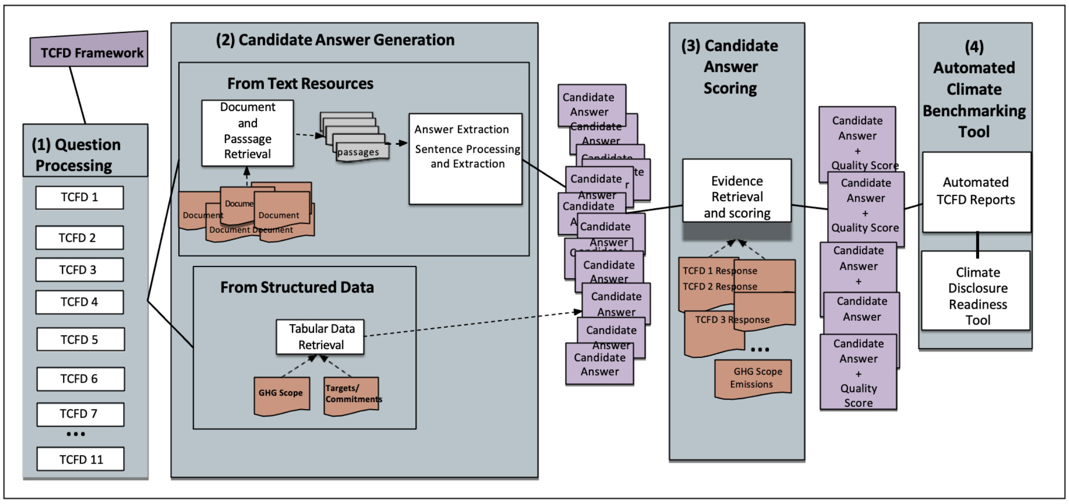


Figure 2: Final Question Answering and Readiness Assessment Pipeline

While TCFD Readiness Assessment Tool will be highly useful in helping companies understand, report, and benchmark their climate risks and opportunities, we also go one step further to design automated climate reports with peer benchmarking by analyzing publicly available company information. The intention is to turn unstructured climate-relevant data into standardized climate benchmark reports to help companies understand their gaps across TCFD recommendations and benchmark against peers. This will help companies understand areas in which they are outperforming and underperforming their peer group, find gap/improvement areas, and communicate their competitive advantages to stakeholders.

Finally, the next steps for refining our pipeline will include 1. gathering more training data to retrain Q3, Q7, and Q11 for an even better performance, 2. building an internal tool to allow users dynamically interact and visualize the passages identified by the model within given document and create custom TCFD reports, and 3. generating fully automated peer benchmark reports across TCFD recommendations on a more granular, indicator-by-indicator basis.

[1–5] [6] [7] [8] [9]

References

[1] Task Force on Climate-related Financial Disclosures 2021 Status Report (2021). URL https://assets.bbhub.io/company/sites/60/2021/07/2021-TCFD-Status_Report.pdf

[2] N. Webersinke, et al., CLIMATEBERT: A Pretrained Language Model for Climate-Related Text (2021). URL https://arxiv.org/pdf/2110.12010.pdf

[3] A. S. Luccioni, et al., Analyzing Sustainability Reports Using Natural Language Processing (2020). URL https://arxiv.org/pdf/2011.08073.pdf

[4] J. Devlin, et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (2019). URL https://arxiv.org/pdf/1810.04805.pdf

Table 4: TCFD Quality Disclosure Score Example (Autodesk vs Teck)

General Example TCFD Source Text	Teck Assessment Score: 71.27 (0 to 100 scale)	Autodesk Assessment Score: 44.67 (0 to 100 scale)
Describe the board's oversight of climate-related risks and opportunities, processes and frequency by which the board and/or board committees (e.g. audit, risk, or other committees) are informed about climate-related issues, whether the board and/or board committees consider climate-related issues when reviewing and guiding strategy, major plans of action, risk management policies, ... annual budgets, and business plans as well as setting the organisation's performance objectives, monitoring implementation and performance, and overseeing major capital expenditures, acquisitions, and divestures, andhow the board monitors and oversees progress against goals and targets for addressing climate-related issues.	Our Board of Directors and senior management are involved in assessing climate-related risks and opportunities to enable Teck to plan for these business and market forces, and to maintain resilience. We recognize that timely and transparent disclosures related to our response to climate change are of importance to Teck and our communitiesof interest. At Teck, we understand that investors, lenders and other users of climate-related ... financial disclosures are interested in understanding the role that our Board plays in overseeing climate-related risks and issues, as well as management's rolein assessing and managing those risks and issues. Climate- related risks and issues receive Board and management attention. We consider climate-related issues and risks in strategic planning across our business units....	With oversight from our CEO, the Sustainability & Foundation team has direct responsibility for setting and implementing our corporate sustainability strategy, including our climate change strategy.

[5] Y. Liu, et al., RoBERTa: A Robustly Optimized BERT Pretraining Approach (2019).
URL https://arxiv.org/pdf/1907.11692.pdf

[6] Microsoft, 2020 Environmental Sustainability Report p. 85 (2020).
URL https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RWyG1q

[7] N. Reimers, I. Gurevych, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks (2019).
URL https://arxiv.org/pdf/1908.10084.pdf

Table 5: ClimateQA results by question

TCFD Question	Validation F1 Score	Testing F1 Score	Label
1) Describe the board's oversight of climate-related risks and opportunities.	97.78%	90.83%	Goveranance a)
2) Describe management's role in assessing and managing climate-related risks and opportunities.	96.60%	89.71%	Goveranance b)
3) Describe the climate-related risks and opportunities the organization has identified over the short, medium, and long term.	91.61%	89.5%	Strategy a)
4) Describe the impact of climate-related risks and opportunities on the organization's business, strategy, and financial planning.	90.91%	87.43%	Strategy b)
5) Describe the resilience of the organization's strategy, taking into consideration different climate-related scenarios, including a 2°C or lower scenario	100.00%	100.00%	Strategy c)
6) Describe the organization's processes for identifying and assessing climate-related risks.	89.87%	85.54%	Risk Management a)
7) Describe the organization's processes for managing climate-related risks.	93.54%	84.23%	Risk Management b)
8) Describe how processes for identifying, assessing, and managing climate-related risks are integrated into the organization's overall risk management.	88.35%	74.44%	Risk Management c)
9) Disclose the metrics used by the organization to assess climate-related risks and opportunities in line with its strategy and risk management processes.	88.48%	90.67%	Metrics & Targets a)
10) Disclose Scope 1, Scope 2, and, if appropriate, Scope 3 greenhouse gas (GHG) emissions, and the related risks.	97.50%	94.74%	Metrics & Targets b)
11) Describe the targets used by the organization to manage climate-related risks and opportunities and performance against targets.	90.20%	98.31%	Metrics & Targets c)
Average	93.17%	89.58%	

[8] J. Briggs, Natural Language Processing (NLP) for Semantic Search (2021).
URL https://www.pinecone.io/learn/nlp

[9] W. W. Microsoft Research, MINILM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers.
URL https://arxiv.org/pdf/2002.10957.pdf