



Where and how machine learning plays a role in climate finance research

Andres Alonso-Robisco, Javier Bas, Jose Manuel Carbo, Aranzazu de Juan & Jose Manuel Marques

To cite this article: Andres Alonso-Robisco, Javier Bas, Jose Manuel Carbo, Aranzazu de Juan & Jose Manuel Marques (30 Jun 2024): Where and how machine learning plays a role in climate finance research, Journal of Sustainable Finance & Investment, DOI: [10.1080/20430795.2024.2370325](https://doi.org/10.1080/20430795.2024.2370325)

To link to this article: <https://doi.org/10.1080/20430795.2024.2370325>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 30 Jun 2024.



Submit your article to this journal [↗](#)



Article views: 723



View related articles [↗](#)



View Crossmark data [↗](#)

Where and how machine learning plays a role in climate finance research

Andres Alonso-Robisco^{a,b}, Javier Bas^b, Jose Manuel Carbo^a, Aranzazu de Juan^b and Jose Manuel Marques^a

^aBanco de España, Madrid, Spain; ^bUniversidad Autonoma de Madrid, Madrid, Spain

ABSTRACT

The financial sector, by mobilizing capital, is fundamental to adapt and mitigate the impact of climate change in the economy. This has led to the emergence of a new research field, climate finance, where experts are starting to harness Machine Learning (ML) as a tool to solve new problems, due to the need to use big datasets and to model complex non-linear relationships. We propose a review of the academic literature that goes beyond the existing bibliometric studies in the field, with the aim of identifying relevant application domains of this technology to inform ML experts where and how their modeling expertise may add value in climate finance. To achieve this, we first assemble a corpus of texts from three scientific databases and use Latent Dirichlet Allocation (LDA) for topic modeling, to uncover seven research areas which we label as: natural hazards, biodiversity, agricultural risk, carbon markets, energy economics, Environmental, Social and Governance (ESG) factors & investing, and climate data. Second, we perform an analysis of publication trends, which confirms that ML is growing both in breadth and depth in climate finance, in particular topics related to energy economics, ESG factors and climate data. Interestingly, some methods stand out in each area, based on data characteristics and modeling requirements.

ARTICLE HISTORY


Received 12 December 2022
Accepted 11 June 2024

KEYWORDS

Climate finance; machine learning; literature review; LDA

1. Introduction

The financial sector has the potential to become an important ally in alleviating the adverse consequences of climate change. This was recognized by The Paris Agreement, article 2.c. ‘*Making finance flows consistent with a pathway towards low greenhouse gas emissions and climate-resilient developments the financial system*’ (UNEP 2015). In fact, since the signing of the Paris Agreement, the scope of topics and the amount of articles on the intersection between economics, finance and sustainability increased dramatically from 2015 onwards, as reported by Malhotra and Thakur (2020). The recognition of the role of finance in the fight against climate change has led to the emergence of a new field

CONTACT Andres Alonso-Robisco  andres.alonso@bde.es  Calle Alcalá 48, CP: 28014, Madrid, Spain

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

in the literature called climate finance, which is generally referred as ‘*finance for activities aiming to mitigate or adapt to the impacts of climate change*’ by the *United Nations Framework Convention on Climate Change*. However, the definition we follow in this work, is that of Giglio, Kelly, and Stroebl (2021): ‘*the tools of financial economics designed for valuing and managing risk which can help society assess and respond to climate change*’, as we find it to be more comprehensive. Interestingly, the inception of the field is usually linked to the seminal work of Nobel Laureate William Nordhaus, who modeled the interactions between climate change and the economy. From there, more specifically on finance, early work mainly addressed concerns on corporate governance and social investing (Capelle-Blancard and Monjon 2012), but the number of topics covered increased significantly over the years.

In fact, one characteristic of climate finance literature is how fragmented the research is. This is not only a bibliographic concern, as it also makes it difficult to join efforts from different academic fields to develop specific research. In a literature review performed by Cunha, Meira, and Orsato (2021), the authors highlight the difficulty of defining the field and differentiating it from traditional finance, due to the poor theorization of the concept of ‘sustainability’, an opinion shared by several other experts (Capelle-Blancard and Monjon 2012; Talan and Sharma 2019; D. Zhang, Zhang, and Managi 2019; Giglio, Kelly, and Stroebl 2021; Liang and Renneboog 2021). This calls for a precautionary need to define the scope of our review of climate finance. To shed light on this, we propose to only use the term climate finance in this paper, although we acknowledge that three concepts are used indistinctly in the academic literature, such as green finance, climate finance and carbon finance (D. Zhang, Zhang, and Managi 2019). Similarly, we will leave out of our scope any work not touching upon climate change, and exclusively focusing on social topics, like corporate governance, impact investing, social investing and financial inclusion, which would fall under the label of sustainable finance. Nevertheless, a limitation exists as some studies do not disentangle environmental from governance and social factors, for instance, those focusing on the impact of Environmental, Social & Governance (ESG) scores on corporate performance. In this sense, some work from sustainable finance will be included.

Another characteristic of climate finance as a research field is the difficulties experts have to face to perform a solid empirical analysis. To name some of them: the still limited reliability of a growing amount of climate data and the statistical complexity to model the non-linear behavior of climate change. These problems, specially the latter, imply profound mathematical challenges making inference about the real climate (Stephenson et al. 2012) and its relationship with the economy. In fact, Diaz-Rainey, Robertson, and Wilson (2017) conclude that methodological constraints could explain previous lack of climate finance research in top finance and business journals. Additionally, issues like the presence of endogeneity are cornerstone in climate finance, as the impact of climate on the economy is two-folded due to the existence of a feedback loop. This has been widely recognized by policy makers (European Commission, 2020), academics (Gourdel et al. 2022) and financial supervisors (NGFS 2019, NGFS 2021). Overall, this presents an opportunity for researchers and experts to recur to Machine Learning (ML) as this technology is particularly well suited to deal with some of these problems, such as non-linearities, complexity and prediction accuracy. Though, in the field of ML some challenges yet to be addressed are explainability of algorithms (Molnar 2020), and

how to deal with endogeneity, a problem inherent to causal inference (Chernozhukov et al. 2018; Athey and Imbens 2019; Bakhitov 2022).

Most climate finance experts still use traditional statistical tools to analyze the impact of climate change in the economy, existing only a subset of (emerging) studies harnessing ML to solve key topics in this field. Special publications like Musleh Al-Sartawi, Hussainey, and Razzaque (2022) cover the role of this technology in sustainable financial markets, as well as Bril, Kell, and Rasche (2022) which gathers examples of the potential of ML to shape the financial markets and set-up new climate-aligned investing strategies. Therefore, based on the proliferation of articles in climate finance, the fragmentation of the literature, and an increasing use of ML in finance (Goodell et al. 2021) and the financial industry (Buchanan and Wright 2021), in this article we propose a systematic review of studies that rely on ML to solve climate finance problems. To face the challenge of heterogeneity of topics within the field, this review leverages on Natural Language Processing (NLP). In particular, we implement a Latent Dirichlet Allocation (LDA) model to statistically uncover latent topics which we identify as relevant application domains. To the best of our knowledge, this is the first survey that systematically covers ML-based studies in climate finance. Our work complements and extends the work of Ghoddusi, Creamer, and Rafizadeh (2019), Ullah et al. (2021) and de Souza et al. (2019), probably the closest to this study. However, we assemble a significantly larger set of papers, covering the whole field of climate finance, leveraging on searches in different public databases, such as Web of Science, Google Scholar and Dimensions.ai. Moreover, we map which ML methods are mostly used in each climate finance topic, aiming to facilitate a profound understanding of how ML can enable climate finance to scale up. This could be useful for future researchers interested in joining this academic field, as well as policy makers looking for ways to better design climate finance instruments and policies.

The main contribution of our work is to take a step further the more traditional bibliometric reviews that are common in this field identifying research topics where ML experts can add value to climate finance, spotting mature and emerging thematic areas, publication trends, choice of methodologies, as well as research gaps; all by applying state-of-the-art methodologies. Moreover, while there exist several studies reviewing climate finance as a research field (Long et al. 2022; Su, Lucey, and Jha 2023) or the role of ML in general finance topics like risk management (Leo, Sharma, and Maddulety 2019), or banking (Königstorfer and Thalmann 2020), to the best of our knowledge no thorough review has been performed on the use of ML in climate finance. As developed below, our results support the relevance of ML in climate-related issues, an idea that is starting to gain traction within Economic journals (Musleh Al-Sartawi, Hussainey, and Razzaque 2022), but still required empirical evidence, a research gap that we cover now. Notwithstanding this, researchers should be aware of two drawbacks regarding the use of ML & Artificial Intelligence (AI) in climate finance. First, it is an intensive energy consuming technology, therefore any analysis on its potential shall always be complemented with its carbon footprint, a concern by itself that has given birth to a new field of study labeled as ‘Green AI’. Furthermore, we also found empirical evidence that ML cannot solve everything. When only low-quality data is available, ML models cannot deliver better predictive performance as for instance Nguyen et al. (2022) show for Scope 3 carbon emissions. This indicates that further research is needed to assess the potential of technology to increase the quantity and quality of climate-related information (Huntingford et al. 2019; Rolnick et al. 2022).

The remaining part of this paper is organized as follows. Section 2 explores the role of ML in climate finance through a literature review. Section 3 explains the data collection process and the methodology for topic modeling, and Section 4 details the results of this study, first the thematic clustering and then an analysis of publication trends. Section 5 concludes.

2. The role of machine learning in climate finance

2.1. What is machine learning

Machine learning can be understood as the set of tools by which machines can learn from data. The nature of the data and the particular problem at hand give rise to two different approaches: supervised and unsupervised learning. Regarding the former, a dependent variable that can be used to check the accuracy of the estimates exists, whether the problem is a regression (most used general algorithms include Regularized linear regression, Regression Trees and Random Forests, or Neural Networks) or a classification one (Support Vector Machines, Classification Trees and Random Forests, Logistic Regression, Neural Networks). Unsupervised learning, on the other hand, addresses problems of a different nature, in which there is no dependent variable to be estimated, covering tasks such as clustering (k-Means, DBSCAN) or dimensionality reduction (Principal Component Analysis, Variational Autoencoders). ML algorithms present significant differences with respect to traditional econometric models, one of the most important being their non-parametric nature. Instead of making an a-priori assumption about the functional relationship between the outcome and the features (in the supervised learning case) or among features (in the case of unsupervised learning), ML algorithms look for these functional relationships, sometimes highly complex and nonlinear. On the other hand, it is worth mentioning that there are also ML methodologies that address the case of time series. Particularly promising in this area are certain neural networks such as the Long-short Term Memory, Convolutional Neural Networks or the so-called Temporal Fusion Transformers. All of them seem to offer astonishing advances over more traditional techniques such as ARIMA models, at the cost, however, of greater computational efforts.

2.2. Strengths and limitations of machine learning in climate finance

First, the usual large size of the data, primarily characterized by volume, variety (structured and unstructured) and time-varying nature, involved in climate finance may require a scalable technology like ML for efficient storage, manipulation, management and analysis. Datasets in climate finance are not only highly-dimensional but also increasingly complex. They usually contain unstructured data, including information from news articles,¹ voice recordings or satellite images, which fall beyond the grasp of usual econometric analysis (López de Prado 2019). Another source of complexity relates to discrepancies in raw data such as differences in the output of models, from various laboratories around the world, that inform the Intergovernmental Panel on Climate Change (IPCC), with data for over 100 years (Monteleoni et al. 2011). In fact, an advancement in numerical methods has been long awaited by econometricians to

better estimate the impact of climate change factors in the economy and the society (Hsiang 2016).

Second, big datasets may contain flexible relationships between the variables that are not suitable for simple linear models. It has been largely recognized that ML methods, such as Random Forests, Support Vector Machines, Neural Networks and others, may allow for more effective ways to model complex financial and economic relationships (Varian 2014; Athey 2018; Athey and Imbens 2019). The key advantage of many ML methods is that they use data-driven model selection, treating the data-generating process (DGP) as unknown, allowing researchers to deal with large datasets without imposing restrictive assumptions (Breiman 2001). As noted by Huntingford et al. (2019) and Castle and Hendry (2022), shared characteristics of financial and climate time series make ML tools appropriate for studying many aspects of observational climate-change data and its economic impact. For instance, green-house gas emissions are a major cause of climate change as they accumulate in the atmosphere. As these emissions are currently mainly due to economic activity, financial and climate time series have commonalities, including considerable inertia, stochastic trends, possible non-linearities, omitted variables and abrupt distributional shifts. Moreover, both disciplines lack complete knowledge of their respective DGPs, so data-driven model selection allowing for shifting distributions is important. In this context, the appeal of ML is that it manages to uncover generalizable patterns. In fact, the success of ML is largely due to its ability to discover non-trivial relationships that were not specified in advance. Moreover, ML methods can fit flexible functional forms to the data, avoiding overfitting, and performing well out-of-sample (Mullainathan and Spiess 2017).

Motivated by these two factors, extensive datasets and complex non-linearity, researchers can harness ML to explain relationships that have the potential for huge societal impact (Hoepner et al. 2021).² Indeed, the effects of climate change are increasingly visible, usually represented as tail risks, or low-probability and high-impact events with material impact on the economy and well-being of people. Storms, droughts, fires and flooding have become stronger and more frequent (Kruczkiewicz et al. 2022). Global ecosystems are changing, including the natural resources and agriculture on which humanity depends. Yet, year after year, these emissions rise, giving only a pause during Covid-19 lockdown. However, it is intrinsically hard to forecast where, how or when climate change will impact the stock price of a given company, or even the debt of an entire country. Financial short-termism fails to incentivize the prediction of medium or long-term risks, which include most climate change-related exposures such as the physical impact on assets like factories or premises. As we will see, ML can help us to close this ‘inter-temporal’ gap which prevent humans (investors) for perceiving the risks and taking actions. A very illustrative example is given by researchers from the Quebec AI Institute (2021), who warned during the last COP26 that preventing climate-related catastrophic consequences will require changes in both policy-making and individuals’ behavior. However, many cognitive biases (like abstraction and myopic term discount) might prevent us from taking action today. To tackle this market failure, they developed ‘*This Climate does not Exist*’, a research project that harness ML (in particular Generative Adversarial Networks or GANs) to create images of personalized climate impacts which will be especially powerful in overcoming the barriers to action and raising climate change awareness.³

However, it is important to bring to this discussion both sides of the impact of ML on climate change. New technologies do not only bring us opportunities. Kaack et al. (2020) explain ways in which AI and ML can be detrimental to efforts addressing climate change, warning of those uses that might harm our planet. For example, remote sensing algorithms for satellite image analysis can be used to gather information on agricultural productivity, but can also be used to accelerate oil and gas exploration. Self-driving cars can make driving more efficient, but they could also increase the amount people drive. And finally, ML include computationally expensive programming, which is an energy intensive activity. Additionally, AI or AI-driven technologies can become pollutants and net emitters of GHG emissions or produce e-waste with harmful environmental impact, depending on the types of applications and the circumstances of their deployment. This overall concern has minted the term ‘Green AI’ (Schwartz et al. 2020), referring to responsible and low carbon intensive coding and good practices relating the training and deployment of complex algorithms in the academic industry (Strubell, Ganesh, and McCallum 2019; Hershcovich et al. 2022). Building on this, Ligozat et al. (2021) propose to study the possible negative impact of AI systems presenting different methodologies used to assess this impact, in particular life cycle assessment. In 2019, researchers (Strubell, Ganesh, and McCallum 2019) in a pioneer paper estimated the consumption of large NLP models, comparing it in CO₂ equivalents with illustrative general life examples. They conclude that training a big Transformer with neural architecture search can emit up to six times what a car produces (including fuel) in its lifetime. Therefore the authors recommend to grant research equitable access to computation resources and suggest to prioritize computationally efficient hardware and algorithms. In another work, these pioneering researchers (Strubell, Ganesh, and McCallum 2020) extend their work to modern language models like BERT, or GPT-2, an issue also tackled by S. Zhang et al. (2022)

We need accurate reporting of energy and carbon usage. It is essential to understand the potential climate impacts of ML research to incentivize responsible research. To this purpose, Henderson et al. (2020) introduce a framework that makes this easier by providing a simple interface for tracking ML models’ real-time energy consumption and carbon emissions, making carbon accounting easier. Lacoste et al. (2019) present as well a *Machine Learning Emissions Calculator* as a tool for research to better understand the environmental impact of training their models. In a position paper, Schwartz et al. (2020) advocate a practical solution by making efficiency an evaluation criterion for research alongside accuracy and related measures, like Hershcovich et al. (2022) who propose a climate performance model scorecard to increase awareness about the environmental impact of NLP research.⁴

2.3. Solving a policy relevant question

It shall be stressed that the importance of bringing together AI & ML experts with climate finance researchers is also supported from outside the discipline of economics. For instance, computer science communities, like *Climate Change AI*, today includes a category of climate finance in its Climate Change AI Directory; or the Association for the Advancement of Artificial Intelligence (AAAI), currently dedicates several workshops on understating the potential of this technology for Environmental, Social and Governance (ESG) analysis. But most remarkably, this has become an internationally policy

relevant question, as shown back in the Conference of the Parts (COP26), where it has been explicitly stated that AI & ML can play a key role in important climate-related topics like prediction, mitigation and adaptation, in ways we cannot afford to ignore (Clutton-Brock et al. 2021). Building on this, Gailhofer et al. (2021) specifically discuss about the role of AI in the European Green Deal, and even the United National Climate Change's platform has created an initiative on 'AI for Climate Action' (#AI4ClimateAction). Similarly, in a position paper Kaack et al. (2020) hope that recent breakthroughs in ML can help us get closer to achieving the United Nations' Sustainable Development Goals, and Kumar et al. (2022) think that new-age technologies applied to sustainability can make significant contributions to the green transition. On the same vein, both Al-Sartawi et al. (2021) and Avgouleas (2021) suggest that cutting-edge financial technology encompassing AI, ML and blockchain can be critical in terms of boosting sustainable finance, as well as Inampudi and Macpherson (2020) who thinks that there is a great potential for AI to contribute towards global economic activity, especially ESG investing. Indeed, rooted on this rationale, the community of international Central Banks decided back in 2019 that Green Finance would be a thematic focus area to experiment in innovative projects under the umbrella of the Bank of International Settlements Innovation Hub (BISIH). A good example is how the combination of AI & ML and blockchain technology is been used to track, monitor and validate climate-related information, giving trust and reliability for instance to new financial instruments like green bonds in [Project Genesis](#), or [Project Gaia](#), that leverages different technologies like large language models (LLMs), to help analysts search corporate climate-related disclosures quickly and efficiently, enabling climate-risk analysis.⁵

Interestingly, other literature reviews on sustainable finance, like Rolnick et al. (2022), show partially how ML can contribute to climate finance, for instance applying deep learning both for tilting portfolio selection towards low carbon emitting corporates, and investment timing, or Spyridou, Polyzos, and Samitas (n.d.) who examines the environmental impact of green assets using machine learning and impulse responses by local projections. Also, in a more general fashion, Akomea-Frimpong et al. (2022), Long et al. (2022), Su, Lucey, and Jha (2023), Gasparini and Tufano (2023) and Carlin et al. (2023) all suggest future research directions in sustainable finance, including the role of technology for climate change, however, we note that no thorough review of the ML-based literature in climate finance has been done before, to the best of our knowledge.

Overall, our work aims to complement and extend more specifically previous efforts to assess the value of ML in climate finance-related topics, like Ghoddusi, Creamer, and Rafizadeh (2019), Ullah et al. (2021) and de Souza et al. (2019), probably the closest studies to this one. However, notably, we assemble a significantly larger set of papers, completing the analysis over all climate finance, by implementing a systematic review of publications using topic modeling to tackle the fragmentation of thematic research areas.

3. Methodology

We adopt and implement the Scientific Procedures and Rationales for Systematic Literature Reviews (SPAR-4-SLR) protocol, developed by J. Paul et al. (2021), who have pooled

their expertise and experience of authoring, editing and reviewing literature reviews to develop a rigorous review protocol which consists of three major stages: assembling, arranging and assessing of articles (see Table 1 for further details).

Our final collection of documents adds up to 217 research articles, from which we extract the abstracts, which will comprise the sample of texts (corpus) of our study. Our goal will be to discover the hidden or latent (unobservable) topics in the corpus of documents (observable) using NLP. Topic modeling assumes a person approaches writing a document with a collection of topics in mind and the words chosen will represent this topic mixture. For instance, a climate finance researcher applying ML to solve a problem will, for example, write a paper with a topic mixture of 50% climate change, 30% finance and 20% ML modeling. The key task for the topic modeling researcher is therefore to reverse engineer the latent topics from the observed words. This will help us understand documents analyzing the presence of words. Often the term ‘topic’ is used in a technical, statistical sense, but ultimately the last phase of any topic modeling approach involves expert analysis to uncover through inspection a more usual theme that aligns with each topic, allowing to name each of them with a more economic meaningful name. In addition, we aim to rank the topics according to

Table 1. SPAR-4-SLR protocol. Assembling, arranging and assessing.

Assembling	
Search Keywords:	
ALL = ('climate change' OR 'ESG' OR 'sustainable finance' OR 'green finance' OR 'climate finance') AND AB = (finance OR 'financial market*' OR bond* OR investment*OR corporate* OR funding OR financing) AND ALL = ('lasso' OR 'random forest*' OR 'extreme gradient' OR 'xgboost' OR CART OR 'deep learning' OR 'neural network' OR 'machine learning')	
Search Databases:	
1. Web of Science (WoS)	
2. Google Scholar (GS)	
3. Dimensions.ai (D.AI)	
Search Result:	
1. Web of Science (WoS): 125 documents	
2. Google Scholar (GS): 18,300 documents -- 45 search pages screened, approx. 450 results.	
3. Dimensions.ai (D.AI): 127 documents	
Arranging	
Organizing Filters:	
Filtered Year for Inclusion:	1999–2022
Filtered Area for Inclusion:	Environmental Science, Computer Science, Economics Finance
Filtered Document Type for Inclusion:	Article
Filtered Publication Stage for Inclusion:	Final
Filtered Source Type for Inclusion:	Journal Article, Working Paper, Conference Proceedings, Book chapter.
Filtered Language for Inclusion:	English
Find duplicates:	Using Endnote bibliographic manager
Ex-post external validation:	Based on field expertise (Human-in-the-Loop).
Filtered Search Result:	217 documents
Assessing	
Analysis Methods:	
Performance analysis:	Publication trend, Evolution of model choice by topic, Breakdown of Journal and Publication type
Topic modelling	Latent Dirichlet Allocation (LDA), using Python.
Agenda Proposal Method:	Reading of articles and reflection of text extracts including mention on machine learning models.
Reporting Convention:	Figures, tables and words.
Limitations:	Accuracy of search results, specially in GS.
	Completeness of references in Environmental Science with a focus on finance.
Support:	No funding received

their prevalence (Sievert and Shirley 2014), a recognized visualization tool for the exploration and presentation of the topics.⁶ Among the different techniques suitable for topic modeling, we rely for this task on the so-called Latent Dirichlet Allocation model (Blei, Ng, and Jordan 2003), given its advantages over the rest. While other methods base their classification on the frequency of concurring words, like Latent Semantic Analysis (Landauer et al. 2013), or on correlating the probability of words appearing in the text with the probability of it belonging to a certain topic, like Naive Bayes (Mosteller and Wallace 1963), or Support Vector Machines (Shawe-Taylor and Cristianini 2004), the approach of LDA is richer, as fully described below.

3.1. Data collection

To assemble the corpus of articles on ML-based climate finance, we identified relevant keywords relating to climate finance from a preliminary assessment of literature reviews on both sustainable (carbon, or green) finance, energy economics and ML in finance (Ghodusi, Creamer, and Rafizadeh 2019; Aziz et al. 2022; Kumar et al. 2022). After determining a reasonable combination of words we experimented with some other variations of terms for both ML and climate change, finding no meaningful number of articles variation, suggesting we got a good convergence on a suitable corpus of texts. Following the identification of these words in climate finance and ML (this led to a combination of 20 keywords⁷), we conducted a search for articles using an advanced search string in the category ALL ('article title, abstract, and keywords'), and AB ('abstract only') on Google Scholar, Web of Science, and Dimensions.ai,⁸ as shown in **Expression 1**. The start date was selected to be 1st January 1999, being the last update as of 22 April 2022.

Expression 1

ALL = ('climate change' OR 'ESG' OR 'sustainable finance' OR 'green finance' OR 'climate finance') AND AB = (finance OR 'financial market' OR bond* OR investment* OR corporate* OR funding OR financing) AND ALL = ('lasso' OR 'random forest*' OR 'extreme gradient' OR 'xgboost' OR CART OR 'deep learning' OR 'neural network' OR 'machine learning')*

The data collection included a systematic 'Human-In-the-Loop' (HIL) approach. It consists of proceeding to an automated data collection with an ex-post validation based on human field expertise.⁹ For instance, a total of 45 search pages (showing 10 items each) were screened in Google Scholar by an expert, while the process of checking potential duplicates between different databases was performed automatically using the references manager EndNote.¹⁰ Contrary to other literature reviews, we aim to focus on a narrow definition of ML in climate finance. This means our results should be familiar to climate finance experts and not relying too heavily on environmental or engineering science with no connection of the research question or conclusion to an economic (or finance) theme or discourse.¹¹ It is important to highlight that our approach, incorporating a screening phase in Google Scholar, allows a richer understanding of a research field that is growing so fast, and therefore relevant research is still in working paper status, waiting to be published by peer-reviewed journals, and consequently does not yet appear in the results retrieved from more standardized databases like Web of Science or Dimensions.ai.

3.2. Topic modeling

As pointed out above, we rely for the topic modeling task on the LDA algorithm given its advantages over other simpler methods. The key practical advantage of LDA is that it allows to treat documents like a mixture of different topics, while topics are presented as a mixture of words. Furthermore, no label of the documents is required. This makes it highly flexible and applicable to a wide range of domains and datasets, which fits the reality observed in climate finance studies, since different topics can partially overlap within a document. Interestingly, LDA is based on a generative probabilistic model, learning the topic-word distributions and the document-topic proportions from the data. Last but not least, LDA is easily scalable, as it handles large-scale datasets efficiently, which makes it valuable to fulfill our task at hand.

The procedure for extracting the topics consists of a variety of steps required for training, tuning and applying the resulting LDA model to the corpus, as an unsupervised learning technique. We will briefly describe the most important ones.¹²

A necessary first step in topic modeling is processing the corpus of documents by tokenizing each document into a collection of their individual words where order is unimportant (i.e. each document is treated as a ‘bag of words’). Then, stop-words that have no topic context (such as ‘and’, ‘of’, ‘the’), are removed, as well as common terms that are highly repeated in the corpus, which we identify because they appear in more than half of the documents, or rare terms for which we set a threshold of being in less than two documents. We deem that both categories of terms contain little meaning to contribute to a relevant topic.¹³ Remaining words in a document are stemmed to generate the words’ root and accurately capture unique terms usage.¹⁴ This means suffixes are removed to create common stem terms, e.g. finance, financial and finances might be reduced to the common ‘financ’ root. In theory, a token can have any number of words (usually monograms are used, but we could have bi- and trigrams). For simplicity, we keep our analysis to single word tokens as we find that it allows us to easily label the topics at the final stage.

Once the corpus is preprocessed, we count with D documents that together contain N unique tokens that we can represent by an $N \times D$ matrix W with entries $w_{n,d}$ that are the number of occurrences of token n in document d . Thus the total number of tokens in document d is $N_d = \sum_{n=0}^N w_{n,d}$. The LDA model consists of two matrices, $\beta_{N \times K}$ and $\theta_{K \times D}$, where K is the total number of topics. For topic k , the vector β_k contains the N token weights, which act as the probabilities $P(n|k)$ that the token n contributes to a document’s bag of words, conditional on the topic k contributing to the document. That is, $P(n|k) = \beta_k$, i.e. the weight of token n in topic k . Therefore, $\sum_{n=1}^N \beta_{n,k} = 1$. For document d , the vector θ_d contains the K topic weights – which act as the probabilities $P(k|d)$ that the topic k appear in the document. Thus $P(k|d) = \theta_{k,d}$, i.e. the weight of topic k in document d . Similarly, $\sum_{k=1}^K \theta_{k,d} = 1$. When these probabilities are significant, we may say that a topic k is relevant in document d . Finally, this setting allows us to decompose in the next equation the probability of a token n in a document d as Hofmann (2001):

$$P(n|d) = \sum_{k=1}^K P(n|k) \cdot P(k|d) = \sum_{k=1}^K \beta_{n,k} \cdot \theta_{k,d} \quad (1)$$

Topic modeling involves reducing the dimensions of these matrices to end up with the same number of rows (documents) but a restricted number of columns which represent the topics. To this purpose, LDA assumes a particular Dirichlet distribution that can be used to produce probability vectors β_k and θ_d that allow an assumption to be made about how topics are distributed across tokens and documents. Using two external inputs, α and β as Dirichlet priors, we can determine the generative process in the LDA (Blei 2012; Blei, Ng, and Jordan 2003) α is a parameter that determines θ_d or per-document topic distribution, and β is a parameter that determines β_k or per-topic token distribution. The LDA posteriors are a result of the trade-off between two inherently conflicting goals. First, that only a relatively small number of topics are expected in a well-written document, and second that only high probability should be assigned to a small number of tokens that belong to highly informative topics. The trade-off exists because if we assign, for instance, a single topic to a single document, thus succeeding at the first goal, the second goal becomes difficult to achieve because all tokens in the document must have a relatively high probability of belonging to that topic. The estimation of the LDA model requires a Bayesian updating from its initial semi-random allocation of topics to tokens and documents, to converge to a probabilistic distribution of topics across documents. Technically, the process will be completed when we find matrices $\beta_{N \times K}$ and $\theta_{K \times D}$ that most likely have produced the observed data W . In our case, the Gensim implementation in Python, based on a Bayesian approach, finds the best configuration of the model automatically as well as several settings related to numerical efficiency (Hofmann 2001). In order not to stop at a local optimum we use a high enough number of iterations, in particular we needed 40,000 passes to reach a stable solution.

4. Results

4.1. Thematic clustering

There are two main challenges when it comes to clustering topics in a corpus of texts. First, there is no easy way to find the optimal number of topics. To this purpose, in the literature several scores are suggested, like Perplexity (Blei, Ng, and Jordan 2003)¹⁵ or Coherence (Röder, Both, and Hinneburg 2015).¹⁶ Increasing the number of topics usually improves these statistical measures during topic modeling, however, we must at the same time account for a higher computational cost of training the model as the number of topics increase, and more importantly, the complexity for a human to discern the economic meaning of more topics will also increase. In our case, we decide to estimate our LDA model with 10 topics, informed by the Rate of Perplexity Change (Zhao et al. 2015), as shown in Figure A1 in the Appendix.¹⁷

A second challenge with topic modeling is that topics that make ML-sense do not necessarily make human sense. Therefore, to ensure a correct labeling of the resulting topics we do a qualitative check with human expert judgment to ensure that the words determined for each topic make sense within the existing climate finance literature. When the LDA model is estimated, we inspect the topics in three ways: first, we look at the tokens with the highest probability per topic β_k . The initial set of results are shown in Table 2. At this stage, we are able to label a total of 7 a-priori reasonable topics, having to discard 3 of them.¹⁸ This initial set of labels is: (i) natural hazards,

(ii) biodiversity, (iii) carbon markets, (iv) agricultural risk, (v) Environmental, Social & Governance (ESG) factors & investing, (vi) energy economics and (vii) climate data.

After this initial inspection, we sample $d = 20$ documents and check whether the highest probability $\theta_{k \times D}$ of each document d belonging to a topic k matches the thematic area identified by a human expert in advance (who read the abstract)¹⁹; and finally we look at the tokens ranked according to topic relevance Sievert and Shirley (2014). The relevance r of token n to topic k , having a tuning parameter λ is given in by

$$r(n, k|\lambda) = \log(\beta_{N \times K}) + (1 - \lambda) \cdot \log\left(\frac{\beta_{N \times K}}{\sum_{k=1}^K \beta_{N \times K}}\right) \quad (2)$$

where the term $\log\left(\frac{\beta_{N \times K}}{\sum_{k=1}^K \beta_{N \times K}}\right)$ is called token's lift. The higher the marginal probability of token n over the corpus, the higher is its lift and the more exclusive a token is for a topic. With $\lambda = 1$, tokens of top relevance equal the top words, even if these do not show up exclusively in that particular topic. With $\lambda = 0$, tokens of top relevance are the ones exclusive to the given topic. By varying $\lambda \in (0, 1)$ and studying the different resulting ranking of tokens, we get a good understanding of the words that contribute to a topic. Following the recommendation of Sievert and Shirley (2014), we fix $\lambda = 0.66$.

By completing this process, we successfully arrive to a meaningful understanding of the concepts covered by each one. For instance, using as example Figure 1 for Topic 9, in the right-hand side panel, we find highly ranked (nearly) exclusive terms like 'energi', 'emiss', 'carbon', 'ghg' or 'greenhous', as well as overlapping terms like 'predict', 'carbon' and 'build'. Varying the values of λ , we can easily label this topic as Energy economics, understanding this as a cluster of research papers dealing with ML to solve problems, for instance, related to GHG emissions, air pollution, carbon price, energy forecasting, energy consumption or buildings efficiency. The remaining topics also show characteristic unique terms as well as shared ones, but overall, we are able to confirm that the labeling makes economic sense after inspection of the respective relevance rankings, allowing us to fine-tune the final name of each topic in detail.²⁰

As a first conclusion, we observe that currently ML is applied for a majority of topics related to climate change in finance. For instance, we identify relevant studies covering five out of the seven topics listed in Kumar et al. (2022),²¹ and four out of six topics identified in Debrah, Darko, and Chan (2023),²² which could serve as a benchmark survey describing the field of sustainable finance as a whole.

4.2. Publication trends and analysis

From a total of 217 unique documents, out of the 7 identified latent topics, we can group them in three broad-scale areas, well known in climate finance literature (Kumar et al. 2022): Physical risks, Transition risks and Corporate & Social Responsibility (CSR), noting that they capture a similar share of total publications. Table 3 shows a summary of the descriptive statistics for each one of the topics.

We observe that the three areas, Physical risks, Transition risks and CSR capture a similar share of total publications, around 31–35%. From the former Table 3, we can conclude that Physical risk is an homogeneously mature research area as the majority of

Table 2. Probabilities of tokens, per topic.

LDA Topic	1		2		3		4		5	
	Tokens	Probability of token per topic (β_k)	Tokens	Probability of token per topic (β_k)	Tokens	Probability of token per topic (β_k)	Tokens	Probability of token per topic (β_k)	Tokens	Probability of token per topic (β_k)
Econ. Label	activ	0.026	risk	0.026	sustain	0.018	biodivers	0.012	carbon	0.028
	csr	0.017	flood	0.023	chang	0.011	financi	0.012	soil	0.024
	valu	0.012	algorithm	0.012	studi	0.01	green	0.011	invest	0.016
	flood	0.01	predict	0.011	method	0.009	base	0.011	predict	0.011
	storg	0.01	price	0.01	financ	0.009	develop	0.01	power	0.011
	correl	0.01	term	0.009	polici	0.009	invest	0.01	polici	0.011
	corpor	0.01	differ	0.007	research	0.008	resourc	0.01	emiss	0.01
	signific	0.01	impact	0.007	topic	0.008	cost	0.009	studi	0.01
	avail	0.009	provid	0.007	inform	0.008	conserv	0.008	result	0.009
	base	0.009	studi	0.007	train	0.008	research	0.008	forecast	0.009
		<i>*discarded*</i>				<i>*discarded*</i>		Biodiversity		<i>*discarded*</i>
Natural hazards			Natural hazards		Natural hazards		Natural hazards		Natural hazards	
LDA Topic	6		7		8		9		10	
	Tokens	Probability of token per topic (β_k)	Tokens	Probability of token per topic (β_k)	Tokens	Probability of token per topic (β_k)	Tokens	Probability of token per topic (β_k)	Tokens	Probability of token per topic (β_k)
Econ. Label	carbon	0.026	chang	0.027	esg	0.07	energi	0.03	compani	0.02
	price	0.023	crop	0.024	invest	0.024	predict	0.019	corpor	0.019
	market	0.02	yield	0.019	rat	0.022	emiss	0.016	report	0.018
	emiss	0.018	futur	0.014	social	0.021	carbon	0.015	financi	0.018
	firm	0.016	agricultur	0.013	portfolio	0.021	forest	0.012	disclosur	0.017
	green	0.015	adapt	0.011	compani	0.013	result	0.01	csr	0.016
	financ	0.013	product	0.011	perform	0.013	chang	0.009	perform	0.014
	paper	0.012	hybrid	0.011	stock	0.013	use	0.008	risk	0.013
	stock	0.01	project	0.01	risk	0.012	random	0.008	relat	0.012
	sector	0.01	suitabl	0.01	score	0.012	impact	0.008	environment	0.011
								Energy economics		Climate data
Carbon markets			Agricultural risk		ESG factors investing		Energy economics		Climate data	

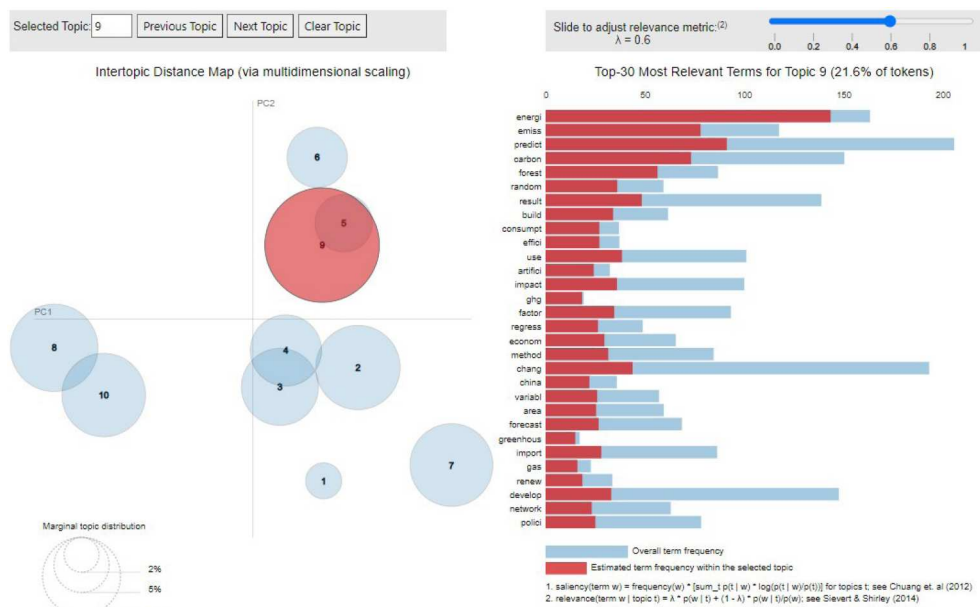


Figure 1. Visualization of topic 9 (energy economics).

publications in each topic of this thematic area are released in peer-reviewed journals. However, for instance in Transition risks, while Energy economics presents high level of peer-reviewed publications, Carbon markets is clearly a younger, emerging research area, relying still more on working paper format. This is specially visible in Climate data, where more than half of the articles gathered are still not published in a journal, as well as ESG factors & investing, where close to half of the documents are not peer-reviewed.

Interestingly, as shown in Figure 2, we see a sharp increase of Energy economics (yellow line) in the last 3 or 4 years, followed by Climate data (darkest blue line) and ESG factors & Investing (green line). However, other research areas have experienced more steady growth rates, like Carbon markets (dark blue line), and Agricultural risks (gray line).

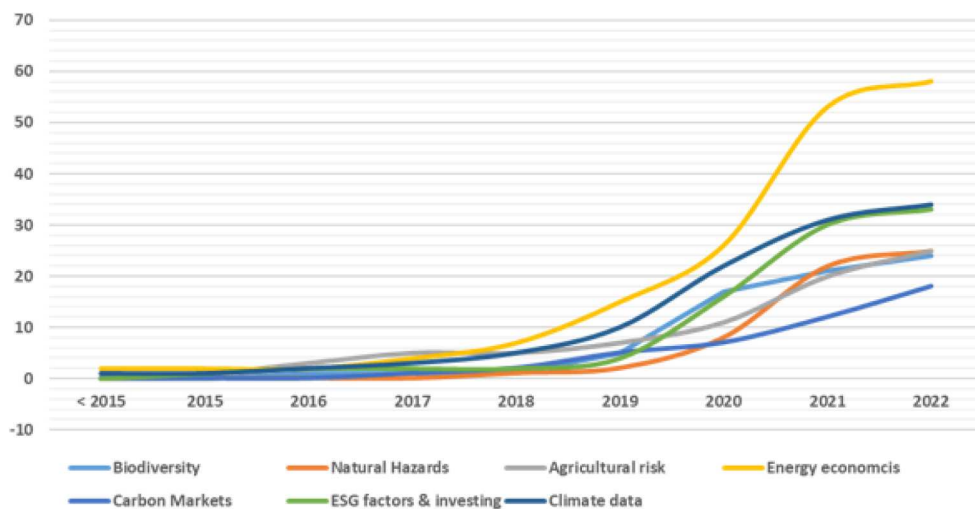
By delving deeper into the research in journal format (totaling 139), it is possible to identify publications in heterogeneous knowledge domains, such as Environmental Sciences, Computer Sciences, or Economic journals. In Figure 3, we plot this breakdown, showing evidence that Economic journals still pay more attention to topics related to CSR and Climate data, lagging behind other scientific journals that publish more work on Physical risk and its socio-economic impact. Yet, due to a lack of granular bibliometric information of the publications in our corpus, we are not able to assess the scientific relevance of articles based on quality rankings or metrics (e.g. H-index), leaving this for further research.

We are also interested in investigating which are the most used ML methods per topic, to provide valuable insights to ML experts willing to look into new fields where these tools are useful. In Figure 4, we show the result by the overall publications, and in Appendix in Table A2 we show the exact figures that support this plot.²³ Also, in Appendix

Table 3. Descriptive statistics of corpus.

	Peer-reviewed	Non-peer reviewed				Total
	Journal	Working Paper	Conf. Proceeding	PhD Dissertation	Book Chapter	
Biodiversity	15	6	3	0	0	24
Natural Hazards	19	2	3	1	0	25
Agricultural risk	17	4	3	0	1	25
Physical Risks: Subtotal	51	12	9	1	1	94
Energy economics	44	10	2	1	1	58
Carbon Markets	12	2	1	2	1	18
Transition Risks: Subtotal	56	12	3	3	2	76
ESG factors investing	17	14	2	0	0	33
Climate data	12	13	6	3	0	34
CSR: Subtotal	29	27	8	3	0	67
Total	136	51	20	7	3	217

Figures A9, A10 and A11 we show the breakdown for Physical risks, Transition risks and CSR respectively.²⁴ While overall Artificial Neural Networks lead as the main ML method used in our corpus of papers, we appreciate further insights analyzing the methods used per research area. For instance, a strong usage of image recognition tools, usually associated with the need to handle newly available (unstructured) data from remote sensing, text, and satellites is appreciated in Physical risks. Therefore, Random forests and Convolutional Neural Networks are widely used in this field. However, in Transition risks, while Artificial Neural Networks still dominate, we can highlight that a relevant share of studies in this domain still use more traditional techniques like Lasso and other penalized-type regressions, and Decision trees, usually associated with an increased need from researchers to present interpretable results due to policy requirements. Finally, the case of CSR is particularly different, as Climate data is a genuinely new topic generated thanks to the use of NLP.

**Figure 2.** Number of publication (cumulative), per year and topic.

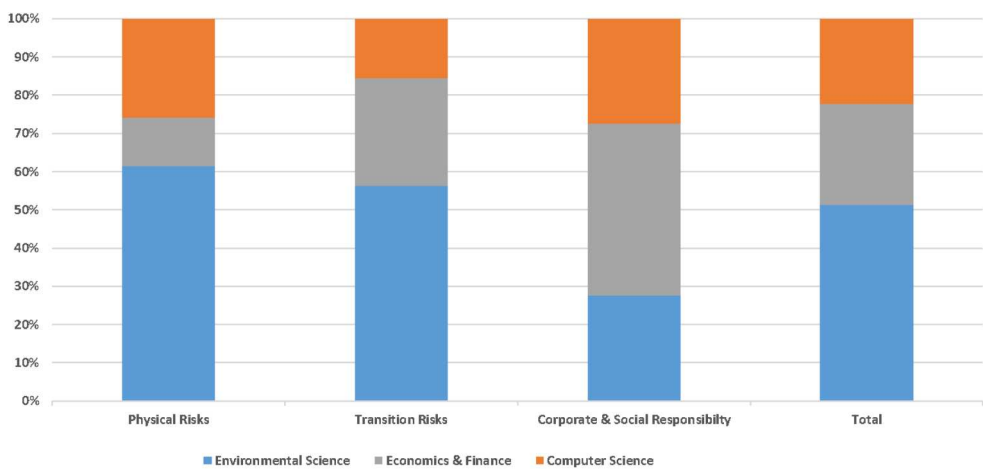


Figure 3. Total number of publication, by type of publication science.

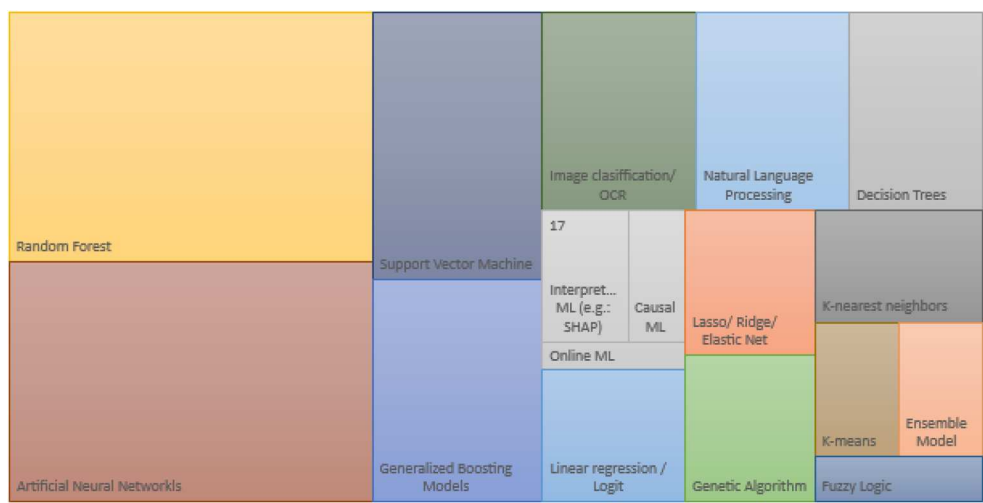


Figure 4. ML methods used, overall.

5. Conclusion

One of the main characteristics of climate finance literature is how fragmented the research is. Other issues facing climate finance experts are the limited reliability of an increasing amount of climate data, and the statistical complexity of modeling the non-linear behavior of climate change. These types of problems create profound mathematical challenges for making inference about the real climate and its relationship with the economy (Stephenson et al. 2012). Aware of these factors, researchers are starting to harness ML to explain these interactions, with highly successful results. This trend has started to gain traction in the academic industry only incipiently in thematic journals (Musleh Al-Sartawi, Hussainey, and Razzaque 2022). However, financial authorities

have been aware of the importance of this technology to scale up climate finance since for some years, as shown by the G20-BIS Techsprint 2021, a race horse between private sector players leveraging technologies to solve a series of pre-identified problem statements.²⁵ With a longer term vision, the Bank of International Settlements has created a series of Innovation Hubs (BISIH) worldwide to experiment with new technologies which might have a big impact for Central Banks' activity, and a Network (BISIN) that is monitoring new developments in technology, being climate finance innovation one of their main focus of interest.²⁶ Also, the success of ML applied to climate finance issues is corroborated by a new wave of projects and market-driven solutions which are flourishing in the private sector, giving birth to a new market segment currently labeled as 'green fintechs' (Macchiavello and Siri 2022).

With the aim to improve the knowledge of ML experts willing to work in climate finance, we perform a review of the literature using ML methods in this research field. Using topic modeling, a new methodology in this field, we uncover up to seven research topics that are coherent with current sustainable finance literature reviews and illustrate the areas where ML is adding more value, which represents a novel contribution never explored before climate finance. To this purpose, we assemble a corpus of relevant articles and we estimate an LDA model to uncover latent topics in the literature, finding three broad-level areas and seven granular application domains which we are able to label with economic meaning that significantly describe where ML is being used within climate finance. To the best of our knowledge, this is the first study that relies on NLP to review this fragmented research field, offering researchers, market experts and policy makers a mean to assess emerging topics, and well as laggards. We hope this will enable a better knowledge of this innovative field, aiding climate finance to scale up to become mainstream in the near future.

Importantly, we observe that currently ML is applied in most of the topics been investigated in climate finance. For instance, we identify relevant studies covering five out of the seven topics listed in Kumar et al. (2022), and four out of six topics identified in Debrah, Darko, and Chan (2023), which could serve as benchmarks describing the field of sustainable finance as a whole.

We also identify a research area like Physical risks that remains mainly covered by Environmental journals, while Economic journals seem to prioritize research on ESG and Carbon markets. This finding supports that the relevance of climate finance is still a work in progress in the top academic arena in Economics. In fact, this is a concern shared by financial authorities like the European Central Banks (ECB), as stated by Tuominen (2022), from the Supervisory Board, referring to its recent report (March, 2022) on banks' progress towards transparent disclosure of their climate-related and environmental risk profiles noted that *'although both physical and transition risks are becoming increasingly material, banks continue to focus their strategies more on transition risks than on physical risks'*.²⁷

Interestingly, we observe that although ML has been initially applied to solve Physical risks problems, like weather and natural hazards forecasting, and issues related to Energy economics, it is possible to observe a recent growth of publications in relevant areas like responsible investing, ESG factors and measuring corporate's compliance with climate data regulatory disclosures. This is also supported by higher ratios of peer-reviewed publications versus working papers format in topics like Agricultural risk,

Natural hazards, Biodiversity and Energy economics, showing a mature research field status, in comparison with topics like Climate data and ESG factors & investing have higher rates of publications in working paper format, being categorized as emerging, younger topics.

Finally, we discover that some ML methods standout within each field of interest. Overall, Random forests and Artificial Neural Networks are the mostly used ones, but for instance, in Physical risk we appreciate a strong usage of image recognition tools, usually associated with the need to handle data collected from AI-driven technologies like remote sensing and satellites, relying therefore heavily on Convolutional Neural Networks and Random forests. However, in Transition risks, Artificial Neural Networks dominate within our subset of documents, usually benefiting from access to big datasets to study energy-related topics. Finally, in CSR, interestingly the access to bigger amounts of data is still challenging, and the requirements on the specifications of the models and the interpretability of results push towards more linear techniques like Ridge and/or Elastic net regularization in multiple types of regressions, and Decision trees, presenting also a notable share of studies introducing techniques from explainable AI (xAI), like Shapley values (Lundberg and Lee 2017).

Interestingly, inspecting [Tables A3, A4 and A5](#), it is observed that supervised ML, both for classification and regression problems, is predominant, which is consistent with the goal of improving accuracy in prediction problems.²⁸ Only a minority of the work reviewed applies unsupervised ML for clustering (such as K-means), aiming to discover hidden patterns in the data without labeled input variables. In a similar fashion, most studies use ML to solve research questions using cross-sectional datasets, while time series problems, requiring particular model architectures like Long-Short Term Memory (LSTM) neural networks, are not as frequent.

To conclude, in this work we studied the potential of ML in climate finance, but we also noted that there are limits to the use of ML in climate finance. For instance, technology cannot improve badly reported carbon data, as shown by Nguyen et al. (2022), who found low predictive capabilities of ML methods to estimate Scope 3 carbon emissions of corporates due to low-quality reported data. However, there is evidence that AI-driven technologies offer great potential to capture and validate climate-related information, improving notably its quality, a lesson which should be taken by policy makers and regulators (Huntingford et al. 2019; Rolnick et al. 2022). Here, a good example is the project ‘Climate Trace’, a source for independent GHG tracking, using satellite data.

Also, ML is an intensive energy consuming technology and therefore, its usage should be promoted in an environmental responsible way. To this purpose, several studies are promoting frameworks to account for the environmental impact of programming complex ML algorithms like Henderson et al. (2020); Strubell, Ganesh, and McCallum (2020). This issue has given birth to a research field by itself, known as ‘Green AI’ (Kaack et al. 2020; OECD 2022). Further research could be carried out to better investigate and communicate about these two limitations of ML. In this vein of exploring new research related to our study, it would be interesting to count with bibliometric information of a greater granularity that would make it possible to assess, based on based on quality metrics, the relevance of the articles that conform the corpus. Also, on a more methodological vein, to check the robustness of results when using bi-grams and tri-grams, in contrast to their exclusion.

Last but not least, it is worth noting that this work is not intended to supporting ML to the detriment of other statistical modeling techniques, such as econometrics. Finance is a field where notions like causality are of greater importance, not only predictive accuracy. Therefore, we understand ML as a tool to add value, which might assist research achieving some particular objectives in climate finance. A great example of this cooperation between both statistical modeling approaches is given by DeepAg (Gurrapu et al. 2021).²⁹ In this same line, we acknowledge that between the ML methods reviewed herein, researchers are increasingly recurring to new techniques like post-hoc interpretability techniques brought from Explainable AI, and causal ML to balance forecasting accuracy with transparency and explainability of the results. All in all, our research shows how climate finance is a multidisciplinary research field which surely benefits from cooperation and collaborations from different skilled experts and professionals.

Notes

1. This is a general trend in economics, where the use of text as data is growing significantly (Gentzkow, Kelly, and Taddy 2019).
2. As we will see in Section 4, most of the research focus on classification and regression problems in cross-sectional set-ups, using supervised ML with labeled data, while only a smaller sample of the studies reviewed applies either unsupervised ML for clustering problems, or ML model architectures to deal with time-series datasets.
3. See link: [here](#).
4. We note that an additional big challenge remains open regarding new methods being currently developed to achieve a trustworthy and scalable ML. For instance, ML model's interpretability require computationally expensive ad-hoc techniques like SHAP (Lundberg and Lee 2017), or the cost of differential privacy which requires extensive experiments (Tornede et al. 2021). Similarly, this happens with Automated ML (Naidu et al. 2021).
5. Notably, the digitization of climate finance has also led to the birth of a fintech private sector that comprises technology-backed innovative business models for finance, something that has been given the name of 'green fintech' (see Macchiavello and Siri (2022), Kwong, Kwok, and Wong or GDFA (2023) for a taxonomy of green fintech solutions).
6. Also used other literature reviews like Kvamsdal et al. (2021).
7. The symbol * is used to capture singular and plural forms of the words.
8. As a robustness check we verified that all the studies tagged as 'climate finance and economics' in the expert network hosted in <https://www.climatechange.ai/> were included.
9. We strove not to alter the distribution of articles included in the corpus, inspecting all the search pages equally irrespective of their order of appearance.
10. For the sake of clarification, no duplication of works concurred, in the sense of counting with a non-peer reviewed as well as with its corresponding peer-reviewed publication.
11. This was actually a drawback appreciated in other literature reviews like Warin and Stojkov (2021), or Kumar et al. (2022), where on the other hand, the size of the corpus analyzed was one order of magnitude bigger.
12. We apply the Gensim implementation of LDA in Python (Rehurek and Sojka 2010).
13. We decide not to include bi-grams or tri-grams in this process as we deem that common candidates like 'climate change' or 'green bonds' would fall under the definition of common terms when split into two. Therefore, we do not expect to change our results. Though, further research could be carried out to perform this robustness check.

14. Stemming consists on the removal of suffixes without considering the context or the actual meaning of the word, which can sometimes lead to the generation of non-interpretable words. Therefore, for the sake of clarification we provide Table A1 in the Appendix, showing some examples of the root words which have originated in our corpus of documents the main tokens included in our topics.
15. It measures how surprised the model is to new data it has not seen before.
16. It measures the degree of semantic similarity between highly relevant tokens in a topic.
17. We include in Appendix Figure A2 the plot using the coherence score, from which we extract similar conclusions. Both metrics are plotted together with the computer cost (latency, in seconds) to estimate the LDA model, informing us that in the range of topics that we work around, it is not a significant feature to consider at the time of stopping the model.
18. We find that their composition is either mainly comprised of methodological terms (e.g. in topics 1 and 3 we encounter tokens like 'activ', 'correl', 'signific', 'algorithm', 'term', 'price', 'differ', etc.) or repetitive with other topics (e.g. in topic 5 we find concepts related to carbon markets like 'emiss', 'carbon' and 'soil', but commingled with low relevant tokens like 'studi', 'result' and 'forecast').
19. All results present herein pass this test, with a threshold of at least 50% success rate.
20. We invite the reader to consult the Appendix for the details of the ranking of each topic (see Figure A7 for Topic 2, Figure A6 for Topic 4, Figure A3 for Topic 6, Figure A8 for Topic 7, Figure A4 for Topic 8, and Figure A5 for Topic 10).
21. Seven clusters were identified in this study, namely: Socially responsible investing, Climate financing, Green financing, Impact investing, Carbon financing, Energy financing, Governance of sustainable financing and investing. Inspecting their uncovered tokens per topic, we find coincidence of terms in all of the clusters but Green financing, and Governance of sustainable financing and investing.
22. Six clusters were identified in this study, namely: Green bond market and greenium, Green credit, Carbon investment and market, Green banking, Market stress, and Climate finance policies. Inspecting their uncovered tokens per topic, we find coincidence of terms in all of the clusters but Green banking, and Market stress.
23. Several papers implement more than one ML method, therefore, the overall count to 380.
24. We only plot the breakdown by overall research areas because the embedded topics in each area usually share same characteristics regarding modeling complexity, and therefore, the choice of models usually are alike. For detailed list of ML methods per research area, please see the Appendix Figures A3, A4, A5.
25. In particular, climate data collection, analysis of climate-related financial risks, and better connecting projects with investors.
26. Additional examples at international level would be the global Fintech Hackcelerator for a greener financial system sponsored in 2021 by the Monetary Authority of Singapore, or the 2021 Green Fintech Challenge, hosted by the Federal Conduct Authority in UK.
27. See the full report [here](#) 'Supervisory assessment of institutions' climate-related and environmental risks disclosures. ECB report on banks' progress towards transparent disclosure of their climate-related and environmental risk profiles'.
28. We pursued the conservative approach of using the same nomenclature as the authors themselves did, as we do not know the details of each model used.
29. A framework that employs econometrics to determine the relationship between financial indices and production of agricultural commodities and then uses Artificial Neural Networks to identify and measure the effect of outlier events on the global economy, based on interdependent relationships.

Acknowledgements

The authors appreciate all comments received from participants in the following seminars: Banco de España, Green Finance Research Advances (2022), Sustainable and Socially Responsible

Finance Conference - University of Bologna (2022), Annual Symposium of the Yale Initiative on Sustainable Finance (2022), and AAAI 2022 Fall Symposium: The Role of AI in Responding to Climate Challenges.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors acknowledge financial help provided by grant PID2022-139614NB-C21 funded by MCIN/AEI /10.13039/501100011033/FEDER, EU.

References

- Abdullah, Akibu Mahmoud, Raja Sher Afgun Usmani, Thulasyammal Ramiah Pillai, Mohsen Marjani, and Ibrahim Abaker Targio Hashem. **2021**. An optimized artificial neural network model using genetic algorithm for prediction of traffic emission concentrations. *International Journal of Advanced Computer Science and Applications* 12 no. 6: 794–803.
- Acheampong, Alex O., and Emmanuel B. Boateng. **2019**. Modelling carbon emission intensity: Application of artificial neural network. *Journal of Cleaner Production* 225: 833–856.
- Akomea-Frimpong, Isaac, David Adeabah, Deborah Ofosu, and Emmanuel Junior Tenakwah. **2022**. A review of studies on green finance of banks, research gaps and future directions. *Journal of Sustainable Finance & Investment* 12 (4): 1241–1264.
- Al-Sartawi, Abdalmuttaleb, Manaf Al-Okaily, Azzam Hannon, and Azam Abdelhakeem Khalid. **2021**. Financial technology: Literature review paper. In *The international conference on global economic revolutions*, 194–200. Cham: Springer.
- Amel-Zadeh, Amir, Mike Chen, Mussalli George, and Michael Oliver Weinberg. **2021**. NLP for SDGs: Measuring corporate alignment with the sustainable development goals. *Columbia Business School Research Paper*.
- Anders, Erik. **2021**. Classification of corporate social performance.
- Antonic, Madelyn. **2020**. Uncovering hidden signals for sustainable investing using Big Data: Artificial intelligence, machine learning and natural language processing. *Journal of Risk Management in Financial Institutions* 13 no. 2: 106–113.
- Athey, Susan. **2018**. The impact of machine learning on economics. *The economics of artificial intelligence: An agenda*, 507–547.
- Athey, Susan, and Guido W. Imbens. **2019**. Machine learning methods that economists should know about. *Annual Review of Economics* 11: 685–725.
- Avand, Mohammadtaghi, Ali Nasiri Khiavi, Majid Khazaei, and John P. Tiefenbacher. **2021**. Determination of flood probability and prioritization of sub-watersheds: A comparison of game theory to machine learning. *Journal of Environmental Management* 295: 113040.
- Avgouleas, Emiliós. **2021**. Resolving the sustainable finance conundrum: Activist policies and financial technology. *Law and Contemporary Problems* 84: 55.
- Aziz, Saqib, Michael Dowling, Helmi Hammami, and Anke Piepenbrink. **2022**. Machine learning in finance: A topic modeling approach. *European Financial Management* 28 no. 3: 744–770.
- Bakhitov, Edvard. **2022**. *Automatic debiased machine learning in presence of endogeneity*. Technical Report. Working Paper. https://edbakhitov.com/assets/pdf/jmp_edbakhitov.pdf.
- Bala, Greg, Hendrik Bartel, James P. Hawley, and Yung-Jae Lee. **2015**. Tracking “real-time” corporate sustainability signals using cognitive computing. *Journal of Applied Corporate Finance* 27 no. 2: 95–102.
- Bastien-Olvera, Bernardo A., and Frances C. Moore. **2021**. Use and non-use value of nature and the social cost of carbon. *Nature Sustainability* 4 no. 2: 101–108.

- Bayle, Federico, Nabil Kawas, Alejandra Mortarini, Carlos Rufin, Alfredo Stein, Lidia Torres, and Daniel Tsai. 2020. Identification of climate change-related hazards in informal communities through the application of machine learning to satellite images. In *2020 World bank conference on land and poverty*.
- Belhadi, Amine, Sachin S. Kamble, Venkatesh Mani, Imane Benkhathi, and Fatima Ezahra Touriki. 2021. An ensemble machine learning approach for forecasting credit risk of agricultural SMEs' investments in agriculture 4.0 through supply chain finance. *Annals of Operations Research* 1–29.
- Ben Ayed, Rayda, and Mohsen Hanana. 2021. Artificial intelligence to improve the food and agriculture sector. *Journal of Food Quality* 2021: 5584754.
- Benites-Lazaro, Lira Luz, Leandro Giatti, and Angelica Giarolla. 2018. Sustainability and governance of sugarcane ethanol companies in Brazil: Topic modeling analysis of CSR reporting. *Journal of Cleaner Production* 197: 583–591.
- Berg, Florian, Julian F. Kölbel, Anna Pavlova, and Roberto Rigobon. 2021. ESG confusion and stock returns: Tackling the problem of noise. *Available at SSRN* 3941514.
- Best, Kelsea B., Jonathan M. Gilligan, Hiba Baroud, Amanda R. Carrico, Katharine M. Donato, Brooke A. Ackerly, and Bishawjit Mallick. 2021. Random forest analysis of two household surveys can identify important predictors of migration in Bangladesh. *Journal of Computational Social Science* 4 no. 1: 77–100.
- Biesbroek, Robbert, Shashi Badloe, and Ioannis N. Athanasiadis. 2020. Machine learning for research on climate change adaptation policy integration: An exploratory UK case study. *Regional Environmental Change* 20 no. 3: 1–13.
- Biffis, Enrico, and Erik Chavez. 2017. Satellite data and machine learning for weather risk management and food security. *Risk Analysis* 37 no. 8: 1508–1521.
- Bingler, Julia Anna, Mathias Kraus, Markus Leippold, and Nicolas Webersinke. 2022. Cheap talk and cherry-picking: What climatebert has to say on corporate climate risk disclosures. *Finance Research Letters* 47: 102776.
- Bjånes, Alexandra, Rodrigo De La Fuente, and Pablo Mena. 2021. A deep learning ensemble model for wildfire susceptibility mapping. *Ecological Informatics* 65: 101397.
- Blei, David M. 2012. Probabilistic topic models. *Communications of the ACM* 55 no. 4: 77–84.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3 no. Jan: 993–1022.
- Bouyé, Eric, and Diane Menville. 2020. The convergence of sovereign environmental, social and governance ratings. *Social and Governance Ratings (December 21, 2020)*.
- Breiman, Leo. 2001. Statistical modeling: The two cultures. *Statistical Science* 16 no. 3: 199–215.
- Bril, Herman, Georg Kell, and Andreas Rasche, eds. 2022. *Sustainability, technology, and finance: Rethinking how markets integrate ESG*. Taylor & Francis.
- Bua, Giovanna, Daniel Kapp, Federico Ramella, and Lavinia Rognone. 2022. Transition versus physical climate risk pricing in European financial markets: A text-based approach.
- Buchanan, G. Bonnie, and Danika Wright. 2021. The impact of machine learning on UK financial services. *Oxford Review of Economic Policy* 37 (3): 537–563.
- Caldecott, Ben, Lucas Krutwagen, Matthew McCarten, Xiaoyan Zhou, David Lunsford, Oliver Marchand, Phanos Hadjikyriakou, et al. 2018. Climate risk analysis from space: Remote sensing, machine learning, and the future of measuring climate-related risk.
- Calvo-Pardo, Hector F, Tullio Mancini, and Jose Olmo. 2022. Machine learning the carbon footprint of bitcoin mining. *Journal of Risk and Financial Management* 15 no. 2: 71.
- Capelle-Blancard, Gunther, and Stéphanie Monjon. 2012. Trends in the literature on socially responsible investment: Looking for the keys under the lamppost. *Business Ethics: A European Review* 21 no. 3: 239–250.
- Carlin, David, Josefine Falk, Drew Johnson, Wenmin Li, and Lea Lorkowski. 2023. The 2023 climate risk landscape.
- Castle, Jennifer L., and David F. Hendry. 2022. Econometrics for modelling climate change. In *Oxford research encyclopedia of economics and finance*.
- Cepni, Oguzhan, Riza Demirer, and Lavinia Rognone. 2022. Hedging climate risks with green assets. *Economics Letters* 212: 110312.

- Cesarini, Luigi, Rui Figueiredo, Beatrice Monteleone, and Mario L.V. Martina. 2021. The potential of machine learning for weather index insurance. *Natural Hazards and Earth System Sciences* 21 no. 8: 2379–2405.
- Chang, Ran, Liya Chu, T.U. Jun, Bohui Zhang, and Guofu Zhou. 2021. ESG and the market return.
- Chen, Junfei, Qian Li, Huimin Wang, and Menghua Deng. 2020. A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: A case study of the Yangtze River Delta, China. *International Journal of Environmental Research and Public Health* 17 no. 1: 49.
- Chen, Qian, and Xiao-Yang Liu. 2020. Quantifying ESG alpha using scholar big data: An automated machine learning approach. In *Proceedings of the first ACM international conference on AI in finance*, 1–8.
- Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. Double/debiased machine learning for treatment and structural parameters.
- Citterio, Alberto. n.d. The role of ESG in predicting bank financial distress: Cross-country evidence.
- Clarkson, Peter M., Jordan Ponn, Gordon D. Richardson, Frank Rudzicz, Albert Tsang, and Jingjing Wang. 2020. A textual analysis of US corporate social responsibility reports. *Abacus* 56 no. 1: 3–34.
- Clutton-Brock, Peter, David Rolnick, Priya L. Donti, and Lynn Kaack. 2021. *Climate change and AI. Recommendations for government action*. Technical Report. GPAI, Climate Change AI, Centre for AI & Climate.
- Coca-Castro, Alejandro, Aaron Golden, and Louis Reymondin. n.d. A multi-source, end-to-end solution for tracking climate change adaptation in agriculture.
- Cojoianu, Theodor, Andreas G.F. Hoepner, Georgiana Ifrim, and Yanan Lin. 2020. Greenwatching: Using AI to detect greenwashing. *AccountancyPlus-CPA Ireland*.
- Coqueret, Guillaume, Sascha Stiernegrip, Christian Morgenstern, James Kelly, Johannes Frey-Skött, and Björn Österberg. 2021. Boosting ESG-based optimization with asset pricing characteristics. Available at SSRN 3877242.
- Cortés, Andrés J., and Felipe López-Hernández. 2021. Harnessing crop wild diversity for climate change adaptation. *Genes* 12 no. 5: 783.
- Cunha, Felipe Arias Fogliano de Souza, Erick Meira, and Renato J. Orsato. 2021. Sustainable finance and investment: Review and research agenda. *Business Strategy and the Environment* 30 no. 8: 3821–3838.
- D’Amato, Valeria, Rita D’Ecclesia, and Susanna Levantesi. 2022. ESG score prediction through random forest algorithm. *Computational Management Science* 19 no. 2: 347–373.
- Dao, David, Catherine Cang, Clement Fung, Ming Zhang, Nick Pawlowski, Reuven Gonzales, Nick Beglinger, and Ce Zhang. 2019. GainForest: Scaling climate finance for forest conservation using interpretable machine learning on satellite imagery. In *ICML climate change AI workshop*. Vol. 2019.
- Dao, David, Johannes Rausch, Iveta Rott, and Ce Zhang. n.d. Xingu: Explaining critical geospatial predictions in weak supervision for climate finance.
- da Silveira, Camila Brasil Louro, Gil Marcelo Reuss Strenzel, Mauro Maida, Ana Lúcia Bertoldi Gaspar, and Beatrice Padovani Ferreira. 2021. Coral reef mapping with remote sensing and machine learning: A nurture and nature analysis in marine protected areas. *Remote Sensing* 13 no. 15: 2907.
- Debnath, Ramit, and Ronita Bardhan. 2020. India nudges to contain COVID-19 pandemic: A reactive public policy analysis using machine-learning based topic modelling. *PLoS One* 15 no. 9: e0238972.
- Debone, Daniela, Vinicius Pazini Leite, and Simone Georges El Khouri Miraglia. 2021. Modelling approach for carbon emissions, energy consumption and economic growth: A systematic review. *Urban Climate* 37: 100849.

- Debrah, Caleb, Amos Darko, and Albert Ping Chuen Chan. 2023. A bibliometric-qualitative literature review of green finance gap and future research directions. *Climate and Development* 15 (5): 432–455.
- De Lucia, Caterina, Pasquale Pazienza, and Mark Bartlett. 2020. Does good ESG lead to better financial performances by firms? Machine learning and logistic regression models of public enterprises in Europe. *Sustainability* 12 no. 13: 5317.
- de Souza, Jovani Taveira, Antonio Carlos de Francisco, Cassiano Moro Piekarski, Guilherme Francisco do Prado, and Leandro Gasparello de Oliveira. 2019. Data mining and machine learning in the context of sustainable evaluation: A literature review. *IEEE Latin America Transactions* 17 no. 3: 372–382.
- Dhokley, Waheeda, Umair Shaikh, Samad Ansari, and Nehal Ansari. 2018. Machine learning approach to predict farmer's loan/credit repayability using weather prediction and credit history. In *IC-CSOD-2018 conference proceedings*, 283.
- Diaz-Rainey, Ivan, Becky Robertson, and Charlie Wilson. 2017. Stranded research? Leading finance journals are silent on climate change. *Climatic Change* 143 no. 1–2: 243–260.
- Diggelmann, Thomas, Jordan Boyd-Graber, Jannis Bulian, Massimiliano Ciaramita, and Markus Leippold. 2020. Climate-fever: A dataset for verification of real-world climate claims. arXiv preprint [arXiv: 2012.00614](https://arxiv.org/abs/2012.00614).
- Diniz, Écio Souza, Alexandre Simões Lorenzon, Nero Lemos Martins de Castro, Gustavo Eduardo Marcatti, Osmarino Pires dos Santos, José Carlos de Deus Júnior, Rosane Barbosa Lopes Cavalcante, Elpídio Inácio Fernandes-Filho, and Cibele Hummeldo Amaral. 2021. Forecasting frost risk in forest plantations by the combination of spatial data and machine learning algorithms. *Agricultural and Forest Meteorology* 306: 108450.
- Donner, Simon D., Milind Kandlikar, and Sophie Webber. 2016. Measuring and tracking the flow of climate change adaptation aid to the developing world. *Environmental Research Letters* 11 no. 5: 054006.
- Drei, Angelo, Théo Le Guenedal, Frédéric Lepetit, Vincent Mortier, Thierry Roncalli, and Takaya Sekine. 2019. ESG investing in recent years: New insights from old challenges. *Available at SSRN* 3683469.
- Dudás, Fanni, and Helena Naffa. 2020. The predictive role of country-level ESG indicators. *Economy and Finance: English-Language Edition of Gazdaság És Pénzügy* 7 no. 4: 441–453.
- Ehrhardt, Adrien, and Minh Tuan Nguyen. 2021. Automated ESG report analysis by joint entity and relation extraction. In *Joint European conference on machine learning and knowledge discovery in databases*, 325–340. Springer.
- Engle, Robert F., Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebe. 2020. Hedging climate change news. *The Review of Financial Studies* 33 no. 3: 1184–1216.
- Erhardt, Dr. 2020. The search for ESG alpha by means of machine learning—A methodological approach. *Available at SSRN* 3514573.
- Evans, Jeffrey S., Melanie A. Murphy, Zachary A. Holden, and Samuel A. Cushman. 2010. Modeling species distribution and change using random forest. In *Predictive species and habitat modeling in landscape ecology: Concepts and applications*, 139–159. New York: Springer.
- Fang, Zheng, Jianying Xie, Ruiming Peng, and Sheng Wang. 2021. Climate finance: Mapping air pollution and finance market in time series. *Econometrics* 9 no. 4: 43.
- Feng, Xi, Huanping Shi, Jian Wang, and Shaoguang Wang. 2021. Green intelligent financial system construction paradigm based on deep learning and concurrency models. *Concurrency and Computation: Practice and Experience* 33 no. 12: e5784.
- Feng, Puyu, Bin Wang, De Li Liu, Cathy Waters, and Qiang Yu. 2019. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in South-Eastern Australia. *Agricultural and Forest Meteorology* 275: 100–113.
- Floreano, Isabela Xavier, and Luzia Alice Ferreira de Moraes. 2021. Land use/land cover (LULC) analysis (2009–2019) with Google Earth Engine and 2030 prediction using Markov-CA in the Rondônia State, Brazil. *Environmental Monitoring and Assessment* 193 no. 4: 1–17.

- Friederich, David, Lynn H. Kaack, Alexandra Luccioni, and Bjarne Steffen. 2021. Automated identification of climate risk disclosures in annual corporate reports. arXiv preprint [arXiv: 2108.01415](https://arxiv.org/abs/2108.01415).
- Gailhofer, Peter, Anke Herold, Jan Peter Schemmel, Cara-Sophie Scherf, Cristina Urrutia de Stebelski, Andreas R. Köhler, and Sibylle Braungardt. 2021. The role of artificial intelligence in the European green deal.
- Gasparini, Matteo, and Peter Tufano. 2023. The evolving academic field of climate finance. Available at SSRN 4354507.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. Text as data. *Journal of Economic Literature* 57 no. 3: 535–574.
- Ghaffarian, Saman, Mariska van der Voort, João Valente, Bedir Tekinerdogan, and Yann de Mey. 2022. Machine learning-based farm risk management: A systematic mapping review. *Computers and Electronics in Agriculture* 192: 106631.
- Ghoddusi, Hamed, Germán G. Creamer, and Nima Rafizadeh. 2019. Machine learning in energy economics and finance: A review. *Energy Economics* 81: 709–727.
- Giglio, Stefano, Bryan Kelly, and Johannes Stroebe. 2021. Climate finance. *Annual Review of Financial Economics* 13: 15–36.
- Goodell, John W., Satish Kumar, Weng Marc Lim, and Debidutta Pattnaik. 2021. Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance* 32: 100577.
- Gourdel, Régis, Irene Monasterolo, Nepomuk Dunz, Andrea Mazzocchi, and Laura Parisi. 2022. The double materiality of climate physical and transition risks in the euro area.
- Gümüşçü, Abdülkadir, Mehmet Emin Tenekeci, and Ali Volkan Bilgili. 2020. Estimation of wheat planting date using machine learning algorithms based on available climate data. *Sustainable Computing: Informatics and Systems* 28: 100308.
- Guo, Tian, Nicolas Jamet, Valentin Betrix, Louis-Alexandre Piquet, Emmanuel Hauptmann, and R.A. Investments. n.d. A deep learning framework for climate responsible investment.
- Gupta, Akshat, Utkarsh Sharma, and Sandeep Kumar Gupta. 2021. The role of ESG in sustainable development: An analysis through the lens of machine learning. In *2021 IEEE international humanitarian technology conference (IHTC)*, 1–5. IEEE.
- Gurrapu, Sai, Feras A. Batarseh, Pei Wang, Md Nazmul Kabir Sikder, Nitish Gorentala, and Munisamy Gopinath. 2021. DeepAg: Deep learning approach for measuring the effects of outlier events on agricultural production and policy. In *2021 IEEE symposium series on computational intelligence (SSCI)*, 1–8. IEEE.
- Han, You, Achintya Gopal, Liwen Ouyang, and Aaron Key. 2021. Estimation of corporate greenhouse gas emissions via machine learning. arXiv preprint [arXiv: 2109.04318](https://arxiv.org/abs/2109.04318).
- Haro, Abi, Alma Mendoza-Ponce, Óscar Calderón-Bustamante, Julián A. Velasco, and Francisco Estrada. 2021. Evaluating risk and possible adaptations to climate change under a socio-ecological system approach. *Frontiers in Climate* 3: 674693.
- Henderson, Peter, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. 2020. Towards the systematic reporting of the energy and carbon footprints of machine learning. *Journal of Machine Learning Research* 21 no. 248: 1–43.
- Hershcovich, Daniel, Nicolas Webersinke, Mathias Kraus, Julia Anna Bingler, and Markus Leippold. 2022. Towards climate awareness in NLP research. arXiv preprint [arXiv: 2205.05071](https://arxiv.org/abs/2205.05071).
- Hilario-Caballero, Adolfo, Ana Garcia-Bernabeu, Jose Vicente Salcedo, and Marisa Vercher. 2020. Tri-criterion model for constructing low-carbon mutual fund portfolios: A preference-based multi-objective genetic algorithm approach. *International Journal of Environmental Research and Public Health* 17 no. 17: 6324.
- Hisano, Ryohei, Didier Sornette, and Takayuki Mizuno. 2020. Prediction of ESG compliance using a heterogeneous information network. *Journal of Big Data* 7 no. 1: 1–19.
- Hoang, D.T., Pr L. Yang, L.D.P Cuong, P.D. Trung, N.H. Tu, L.V. Truong, T.T. Hien, and V.T. Nha. 2020. Weather prediction based on LSTM model implemented AWS machine learning platform. *International Journal for Research in Applied Science & Engineering Technology* 8 no. 5: 283–290.

- Hoepner, Andreas G.F., David McMillan, Andrew Vivian, and Chardin Wese Simen. 2021. Significance, relevance and explainability in the machine learning age: An econometrics and financial data science perspective. *The European Journal of Finance* 27 no. 1-2: 1–7.
- Hofmann, Thomas. 2001. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning* 42 no. 1/2: 177–196.
- Hong, Xiangjun, Xian Lin, Laitan Fang, Yuchen Gao, and Ruipeng Li. 2022. Application of machine learning models for predictions on cross-border merger and acquisition decisions with ESG characteristics from an ecosystem and sustainable development perspective. *Sustainability* 14 no. 5: 2838.
- Hou, Deyi, Nanthi S Bolan, Daniel C.W. Tsang, Mary B Kirkham, and David O'Connor. 2020. Sustainable soil use and management: An interdisciplinary and systematic approach. *Science of the Total Environment* 729: 138961.
- Hsiang, Solomon. 2016. Climate econometrics. *Annual Review of Resource Economics* 8: 43–75.
- Huntingford, Chris, Elizabeth S. Jeffers, Michael B. Bonsall, Hannah M. Christensen, Thomas Lees, and Hui Yang. 2019. Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters* 14 no. 12: 124007.
- Inampudi, Kalyani, and Martina Macpherson. 2020. The impact of AI on environmental, social and governance (ESG) investing: Implications for the investment value chain. In *The AI book: The artificial intelligence handbook for investors, entrepreneurs and FinTech visionaries*, 129–131.
- Inyang, Udoinyang Godwin, Emem Etok Akpan, and Oluwole Charles Akinyokun. 2020. A hybrid machine learning approach for flood risk assessment and classification. *International Journal of Computational Intelligence and Applications* 19 no. 02: 2050012.
- Jaycocks, Amber. 2019. Climate finance and green bond evolution. PhD diss., Pardee Rand Graduate School.
- Jha, Manish. 2021. Essays in corporate finance and machine learning. PhD diss., Washington University in St. Louis.
- Joshi, Himanshu, and Rajneesh Chauhan. 2024. Determinants of price multiples for technology firms in developed and emerging markets: Variable selection using shrinkage algorithm. *Vision* 28 (1): 55–66.
- Kaack, Lynn, Priya Danti, Emma Strubell, and David Rolnick. 2020. Artificial intelligence and climate change: Opportunities, considerations, and policy levers to align AI with climate change goals.
- Keys, Patrick W., Elizabeth A. Barnes, and Neil H. Carter. 2021. A machine-learning approach to human footprint index estimation with applications to sustainable development. *Environmental Research Letters* 16 no. 4: 044061.
- Khan, Mohd Jawad Ur Rehman, and Anjali Awasthi. 2019. Machine learning model development for predicting road transport GHG emissions in Canada. *WSB Journal of Business and Finance* 53 no. 2: 55–72.
- Kheradmand, Elham, Didier Serre, Manuel Morales, and Cedric B. Robert. n.d. A NLP-based analysis of alignment of organizations' climate-related risk disclosures with material risks and metrics.
- Klusak, Patrycja, Matthew Agarwala, Matt Burke, Moritz Kraemer, and Kamiar Mohaddes. 2021. Rising temperatures, falling ratings: The effect of climate change on sovereign creditworthiness.
- Kluza, Krzysztof, Magdalena Ziolo, and Anna Spoz. 2021. Innovation and environmental, social, and governance factors influencing sustainable business models-meta-analysis. *Journal of Cleaner Production* 303: 127015.
- Königstorfer, Florian, and Stefan Thalmann. 2020. Applications of artificial intelligence in commercial banks—A research agenda for behavioral finance. *Journal of Behavioral and Experimental Finance* 27: 100352.
- Krappel, Tim, Alex Bogun, and Damian Borth. 2021. Heterogeneous ensemble for ESG ratings prediction. arXiv preprint [arXiv: 2109.10085](https://arxiv.org/abs/2109.10085).
- Kruczkiewicz, Andrew, Fabio Cian, Irene Monasterolo, Giuliano Di Baldassarre, Astrid Caldas, Moriah Royz, Margaret Glasscoe, Nicola Ranger, and Maarten van Aalst. 2022. Multiform

- flood risk in a rapidly changing world: What we do not do, what we should and why it matters. *Environmental Research Letters* 17 no. 8: 081001.
- Kulkarni, Shruti. 2021. Using machine learning to analyze climate change technology transfer (CCTT).
- Kumar, Satish, Dipasha Sharma, Sandeep Rao, Weng Marc Lim, and Sachin Kumar Mangla. 2022. Past, present, and future of sustainable finance: Insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research* 1–44.
- Kvamsdal, Sturla F, Ivan Belik, Arnt Ove Hopland, and Yuanhao Li. 2021. A machine learning analysis of the recent environmental and resource economics literature. *Environmental and Resource Economics* 79 no. 1: 93–115.
- Kwong, Raymond, Man Lung Jonathan Kwok, and Helen Wong. 2023. Green fintech as a future research direction: A bibliometric analysis on green finance and fintech. *Sustainability*.
- Lacoste, Alexandre, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. 2019. Quantifying the carbon emissions of machine learning. arXiv preprint [arXiv: 1910.09700](https://arxiv.org/abs/1910.09700).
- Landauer, Thomas K., Danielle S. McNamara, Simon Dennis, and Walter Kintsch, eds. 2013. *Handbook of latent semantic analysis*. Psychology Press.
- Lanza, Ariel, Enrico Bernardini, and Ivan Faiella. 2020. Mind the gap! machine learning, ESG metrics and sustainable investment. *Machine learning, ESG metrics and sustainable investment (June 26, 2020)*. Bank of Italy Occasional Paper (561).
- Leo, Martin, Suneel Sharma, and Koilakuntla Maddulety. 2019. Machine learning in banking risk management: A literature review. *Risks* 7 no. 1: 29.
- Levi, Sebastian. 2021. Why hate carbon taxes? Machine learning evidence on the roles of personal responsibility, trust, revenue recycling, and other factors across 23 European countries. *Energy Research & Social Science* 73: 101883.
- Li, Mingyu, Dongxiao Niu, Zhengsen Ji, Xiwen Cui, and Lijie Sun. 2021. Forecast research on multidimensional influencing factors of global offshore wind power investment based on random forest and elastic net. *Sustainability* 13 no. 21: 12262.
- Li, Qin, Shengqun Xia, Suwei Liu, and Jingyi Yao. 2020. Driving factors of green climate fund leverage. *World Scientific Research Journal* 6 no. 5: 20–32.
- Li, Kai, and Tingyu Yu. 2022. A machine learning based anatomy of firm-level climate risk exposure. Available at SSRN 4025598.
- Liang, Hao, and Luc Renneboog. 2021. Corporate social responsibility and sustainable finance. In *Oxford research encyclopedia of economics and finance*.
- Ligozat, Anne-Laure, Julien Lefèvre, Aurélie Bugeau, and Jacques Combaz. 2021. Unraveling the hidden environmental impacts of AI solutions for environment. arXiv preprint [arXiv: 2110.11822](https://arxiv.org/abs/2110.11822).
- Lima, Valdeir Pereira, Renato Augusto Ferreira de Lima, Fernando Joner, Ilyas Siddique, Niels Raes, and Hans Ter Steege. 2022. Climate change threatens native potential agroforestry plant species in Brazil. *Scientific Reports* 12 no. 1: 1–14.
- Lin, Boqiang, and Rui Bai. 2022. Machine learning approaches for explaining determinants of the debt financing in heavy-polluting enterprises. *Finance Research Letters* 44: 102094.
- Liu, Lixia, and Xueli Zhan. 2019. Analysis of financing efficiency of Chinese agricultural listed companies based on machine learning. *Complexity* 2019: 9190273.
- Long, Suwan, Brian Lucey, Satish Kumar, Dayong Zhang, and Zhiwei Zhang. 2022. Climate finance: What we know and what we should know? *Journal of Climate Finance* 1: 100005.
- López de Prado, Marcos. 2019. Beyond econometrics: A roadmap towards financial machine learning. Available at SSRN 3365282.
- Luccioni, Alexandra, Emily Baylor, and Nicolas Duchene. 2020. Analyzing sustainability reports using natural language processing. arXiv preprint [arXiv: 2011.08073](https://arxiv.org/abs/2011.08073).
- Luccioni, Alexandra, and Hector Palacios. 2019. Using natural language processing to analyze financial climate disclosures. In *Proceedings of the 36th international conference on machine learning, Long Beach, California*.
- Lundberg, Scott M., and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 30.

- Lyubchich, Vyacheslav, Nathaniel K. Newlands, Azar Ghahari, Tahir Mahdi, and Yulia R. Gel. 2019. Insurance risk assessment in the face of climate change: Integrating data science and statistics. *Wiley Interdisciplinary Reviews: Computational Statistics* 11 no. 4: e1462.
- Ma, Zongming. 2019. Three essays on asset pricing in regime and ESG environments.
- Ma, Ning, Wai Yan Shum, Tingting Han, and Fujun Lai. 2021. Can machine learning be applied to carbon emissions analysis: An application to the CO₂ emissions analysis using Gaussian process regression. *Frontiers in Energy Research* 9: 756311.
- Macadam, Alex, Cameron J. Nowell, and Kate Quigley. 2021. Machine learning for the fast and accurate assessment of fitness in coral early life history. *Remote Sensing* 13 no. 16: 3173.
- Macchiavello, Eugenia, and Michele Siri. 2022. Sustainable finance and fintech: Can technology contribute to achieving environmental goals? A preliminary assessment of 'green fintech' and 'sustainable digital finance'. *European Company and Financial Law Review* 19 no. 1: 128–174.
- Malhotra, Gaurika, and K.S. Thakur. 2020. Evolution of green finance: A bibliometric approach. *Gedrag & Organisatie Review* 33 no. 2: 583–594.
- Manandhar, Achut, Alex Fischer, David J. Bradley, Mashfiqu Salehin, M. Sirajul Islam, Rob Hope, and David A. Clifton. 2020. Machine learning to evaluate impacts of flood protection in Bangladesh, 1983–2014. *Water* 12 no. 2: 483.
- Mansouri, Sasan, and Paul P. Momtaz. 2021. Financing sustainable entrepreneurship: ESG measurement, valuation, and performance in token offerings. *Valuation, and performance in token offerings (May 12, 2021)*.
- Margot, Vincent, Christophe Geissler, Carmine de Franco, Bruno Monnier, France Advestis, and France Ossiam. 2021. ESG investments: Filtering versus machine learning approaches. *Applied Economics and Finance* 8 no. 2: 1–16.
- Michalski, Lachlan, and Rand Kwong Yew Low. 2021. Corporate credit rating feature importance: Does ESG matter? Available at SSRN 3788037.
- Miglionico, Andrea. 2022. The use of technology in corporate management and reporting of climate-related risks. *European Business Organization Law Review* 23 no. 1: 125–141.
- Mitsuzuka, Kanau, Feng Ling, and Hayato Ohwada. 2017. Analysis of CSR activities affecting corporate value using machine learning. In *Proceedings of the 9th international conference on machine learning and computing*, 11–14.
- Molnar, Christoph. 2020. *Interpretable machine learning*. Lulu. com.
- Monteleoni, Claire, Gavin A. Schmidt, Shailesh Saroha, and Eva Asplund. 2011. Tracking climate models. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 4 no. 4: 372–392.
- Moreno, Angel Ivan, and Teresa Caminero. 2022a. Analysis of ESG disclosures in pillar 3 reports. A text mining approach. *A text mining approach (March 8, 2022)*. Banco de Espana Occasional Paper (2204).
- Moreno, Angel Ivan, and Teresa Caminero. 2022b. Application of text mining to the analysis of climate-related disclosures. *International Review of Financial Analysis* 83: 102307.
- Morkner, Paige, Jennifer Bauer, C. Gabriel Creason, Michael Sabbatino, Patrick Wingo, Randall Greenburg, Samuel Walker, David Yeates, and Kelly Rose. 2022. Distilling data to drive carbon storage insights. *Computers & Geosciences* 158: 104945.
- Mosteller, Frederick, and David L. Wallace. 1963. Inference in an authorship problem: A comparative study of discrimination methods applied to the authorship of the disputed Federalist Papers. *Journal of the American Statistical Association* 58 (302): 275–309.
- Mullainathan, Sendhil, and Jann Spiess. 2017. Machine learning: An applied econometric approach. *Journal of Economic Perspectives* 31 no. 2: 87–106.
- Müller, Daniel, Anne Jungandreas, Friedrich Koch, and Florian Schierhorn. 2016. *Impact of climate change on wheat production in Ukraine*. Vol. 41. Kyiv: Institute for Economic Research and Policy Consulting.
- Musleh Al-Sartawi, Abdalmuttaleb M.A., Khaled Hussainey, and Anjum Razzaque. 2022. The role of artificial intelligence in sustainable finance.
- Naidu, Rakshit, Harshita Diddee, Ajinkya Mulay, Aleti Vardhan, Krithika Ramesh, and Ahmed Zamzam. 2021. Towards quantifying the carbon emissions of differentially private machine learning. arXiv preprint [arXiv: 2107.06946](https://arxiv.org/abs/2107.06946).

- Natsume, Shohei, and Ling Feng. 2019. Relationship between CSR and corporate value from the ISO26000 and the GRI guidelines perspective. *Asian Journal of Management Science and Applications* 4 no. 2: 141–162.
- Nay, John. 2016. Predicting and understanding law-making with machine learning. *PLoS ONE* 12 no. 5: e0176999.
- Nay, JJ. 2017. Predicting and understanding law-making with word vectors and an ensemble model. *PLoS ONE* 12 no. 5: e0176999.
- Nguyen, Quyen, Ivan Diaz-Rainey, Adam Kitto, Ben McNeil, Nicholas Pittman, and Renzhu Zhang. 2022. Scope 3 emissions: Data quality and machine learning prediction accuracy.
- Nguyen, Quyen, Ivan Diaz-Rainey, and Duminda Kuruppuarachchi. 2021. Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach. *Energy Economics* 95: 105129.
- Nti, Isaac Kofi, Owusu Nyarko-Boateng, Samuel Boateng, Faiza Umar Bawah, Promise Ricardo Agbedanu, Nicodemus Songose Awarayi, Peter Nimbe, et al. 2021. Enhancing flood prediction using ensemble and deep learning techniques. In *2021 22nd international Arab conference on information technology (ACIT)*, 1–9. IEEE.
- Nugent, Tim, Nicole Stelea, and Jochen L. Leidner. 2020. Detecting ESG topics using domain-specific language models and data augmentation approaches. arXiv preprint [arXiv: 2010.08319](https://arxiv.org/abs/2010.08319).
- Nunnari, Giuseppe, Stephen Dorling, Uwe Schlink, Gavin Cawley, Rob Foxall, and Tim Chatterton. 2004. Modelling SO₂ concentration at a point with statistical approaches. *Environmental Modelling & Software* 19 no. 10: 887–905.
- OECD. 2022. Measuring the environmental impacts of artificial intelligence compute and applications (341). <https://www.oecd-ilibrary.org/content/paper/7babf571-en>.
- Owusu, Abena Fosua. 2020. *Three essays on the application of machine learning for risk governance in financial institutions*. Rensselaer Polytechnic Institute.
- Paul, S.S., N.C. Coops, M.S. Johnson, M. Krzic, A. Chandna, and S.M. Smukler. 2020. Mapping soil organic carbon and clay using remote sensing to predict soil workability for enhanced climate change adaptation. *Geoderma* 363: 114177.
- Paul, Justin, Weng Marc Lim, Aron O’Cass, Andy Wei Hao, and Stefano Bresciani. 2021. Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies* 45 no. 4: 1–16.
- Pearson, Katelin D., Gil Nelson, Myla F.J. Aronson, Pierre Bonnet, Laura Brenskelle, Charles C. Davis, Ellen G. Denny, et al. 2020. Machine learning using digitized herbarium specimens to advance phenological research. *BioScience* 70 no. 7: 610–620.
- Pincet, Arnaud, Shu Okabe, and Martin Pawelczyk. 2019. Linking aid to the sustainable development goals—A machine learning approach.
- Plakandaras, Vasiliios, Periklis Gogas, and Theophilos Papadimitriou. 2018. The effects of geopolitical uncertainty in forecasting financial markets: A machine learning approach. *Algorithms* 12 no. 1: 1.
- Porfirio, Luciana L., David Newth, Ian N. Harman, John J. Finnigan, and Yiyong Cai. 2017. Patterns of crop cover under future climates. *Ambio* 46 no. 3: 265–276.
- Qi, Yawei, Wenxiang Peng, Ran Yan, and Guangping Rao. 2021. Use of BP neural networks to determine China’s regional CO₂ emission quota. *Complexity* 2021 (1): 6659302.
- Raghupathi, Viju, Jie Ren, and Wullianallur Raghupathi. 2020. Identifying corporate sustainability issues by analyzing shareholder resolutions: A machine-learning text analytics approach. *Sustainability* 12 no. 11: 4753.
- Rahman, Mohd Nayyer, Muganda M. Manini, and Zeenat Fatima. 2021. Economic indicators and climate change for BRICS economies in the post COVID-19 world: Neural network approach. *Management & Economics Research Journal* 7 no. 4: 31450.
- Rakova, Bogdana, and Alexander Winter. 2020. Leveraging traditional ecological knowledge in ecosystem restoration projects utilizing machine learning. arXiv preprint [arXiv: 2006.12387](https://arxiv.org/abs/2006.12387).
- Raman, Natraj, Grace Bang, and Armineh Nourbakhsh. 2020. Mapping ESG trends by distant supervision of neural language models. *Machine Learning and Knowledge Extraction* 2 no. 4: 453–468.

- Reed, Michael, Patrick O'Reilly, and Joshua Hall. 2019. The economics and politics of carbon taxes and regulations: Evidence from voting on Washington State's Initiative 732. *Sustainability* 11 no. 13: 3667.
- Rehurek, Radim, and Petr Sojka. 2010. Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks*.
- Reiersen, Gyri, David Dao, Björn Lütjens, Konstantin Klemmer, Xiaoxiang Zhu, and Ce Zhang. 2021. Tackling the overestimation of forest carbon with deep learning and aerial imagery. arXiv preprint [arXiv: 2107.11320](https://arxiv.org/abs/2107.11320).
- Riad, Oumaima, Sahar Saoud, Lamia Boukaya, and Khalid Azami. 2019. Neural networks for measurement of the eco-responsible decision's impact on the governance system: The case of Moroccan companies. In *Proceedings of the 4th international conference on smart city applications*, 1–7.
- Röder, Michael, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on web search and data mining*, 399–408.
- Rohayani, Hetty, Harco Leslie Hendric Spits Warnars, Tuga Mauritsius, and Edi Abdurrachman. 2021. Wind speed forecasting in big data and machine learning: From presents, opportunities and future trends. *Communications in Mathematical Biology and Neuroscience* 2021.
- Rolnick, David, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, et al. 2022. Tackling climate change with machine learning. *ACM Computing Surveys (CSUR)* 55 no. 2: 1–96.
- Sabu, Kiran M., and T.K. Manoj Kumar. 2020. Predictive analytics in agriculture: Forecasting prices of Arecanuts in Kerala. *Procedia Computer Science* 171: 699–708.
- Santamaria, Simona, David Dao, Björn Lütjens, and Ce Zhang. 2020. TrueBranch: Metric learning-based verification of forest conservation projects. arXiv preprint [arXiv: 2004.09725](https://arxiv.org/abs/2004.09725).
- Sautner, Zacharias, Laurence van Lent, Grigory Vilkov, and Ruishen Zhang. 2020. Firm-level climate change exposure. *European Corporate Governance Institute–Finance Working Paper* (686).
- Schmidt, Victor, Alexandra Sasha Luccioni, Mélisande Teng, Tianyu Zhang, Alexia Reynaud, Sunand Raghupathi, Gautier Cosne, et al. 2021. ClimateGAN: Raising climate change awareness by generating images of floods. arXiv preprint [arXiv: 2110.02871](https://arxiv.org/abs/2110.02871).
- Schwabe, Henning, Sumeet Sandhu, and Sergiy Grebenshchikov. n.d. Accelerated data discovery for scalable climate action.
- Schwartz, Roy, Jesse Dodge, Noah A. Smith, and Oren Etzioni. 2020. Green AI. *Communications of the ACM* 63 no. 12: 54–63.
- Seidl, Andrew, Kelvin Mulungu, Marco Arlaud, Onno van den Heuvel, and Massimiliano Riva. 2020. Finance for nature: A global estimate of public biodiversity investments. *Ecosystem Services* 46: 101216.
- Sharma, Utkarsh, Akshat Gupta, and Sandeep Kumar Gupta. 2024. The pertinence of incorporating ESG ratings to make investment decisions: A quantitative analysis using machine learning. *Journal of Sustainable Finance & Investment* 14 (1): 184–198.
- Shawe-Taylor, John, and Nello Cristianini. 2004. *Kernel methods for pattern analysis*. Cambridge University Press.
- Shi, Yan, Yuhuan Shen, and Hao Wu. 2021. Construction and research of regional green finance statistical model based on CVM-MLP neural network. In *2021 2nd international conference on intelligent computing and human-computer interaction (ICHCI)*, 188–193. IEEE.
- Shi, Xunpeng, Keying Wang, Tsun Se Cheong, and Hongwu Zhang. 2020. Prioritizing driving factors of household carbon emissions: An application of the LASSO model with survey data. *Energy Economics* 92: 104942.
- Shu, Xiaolong, Yufeng Ren, Zhe Duan, Xing Liu, Xiaojun Hua, and Huike Lei. 2022. Flood risk assessment in Ya'an, Sichuan, China based on the emergy theory. *Journal of Water and Climate Change* 13 no. 1: 247–259.
- Sievert, Carson, and Kenneth Shirley. 2014. LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, 63–70.

- Škapa, Stanislav, Nina Bočková, Karel Doubravský, and Mirko Dohnal. 2023. Fuzzy confrontations of models of ESG investing versus non-ESG investing based on artificial intelligence algorithms. *Journal of Sustainable Finance & Investment* 13 (1): 763–775.
- Sokolov, Alik, Kyle Caverly, Jonathan Mostovoy, Talal Fahoum, and Luis Seco. 2021. Weak supervision and Black–Litterman for automated ESG portfolio construction. *The Journal of Financial Data Science* 3 no. 3: 129–138.
- Sokolov, Alik, Jonathan Mostovoy, Jack Ding, and Luis Seco. 2020. Building machine learning systems to automate ESG index construction.
- Sokolov, Alik, Jonathan Mostovoy, Jack Ding, and Luis Seco. 2021. Building machine learning systems for automated ESG scoring. *The Journal of Impact and ESG Investing* 1 no. 3: 39–50.
- Spyridou, Anastasia, Efsthios Polyzos, and Aristeidis Samitas. n.d. Green assets are not so green: Assessing. Available at SSRN 4448447.
- Stephenson, David B., Matthew Collins, Jonathan C. Rougier, and Richard E. Chandler. 2012. Statistical problems in the probabilistic prediction of climate change. *Environmetrics* 23 no. 5: 364–372.
- Strubell, Emma, Ananya Ganesh, and Andrew McCallum. 2019. Energy and policy considerations for deep learning in NLP. arXiv preprint [arXiv: 1906.02243](https://arxiv.org/abs/1906.02243).
- Strubell, Emma, Ananya Ganesh, and Andrew McCallum. 2020. Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI conference on artificial intelligence*. Vol. 34, 13693–13696.
- Su, Yang, Brian M. Lucey, and Ashish Jha. 2023. Finance research and the UN sustainable development goals—An analysis and forward look. Available at SSRN 4696722.
- Sun, Chenghao. 2022. The correlation between green finance and carbon emissions based on improved neural network. *Neural Computing and Applications* 34 no. 15: 12399–12413.
- Sun, Qiancheng, Andrea Zela-Koort, Ava Stokes, and Salahaldin Alshatshati. 2021. How COVID-19 impacted CO₂ emissions based on electricity usage: A machine learning approach. *Journal of Energy & Technology (JET)* 1 no. 2: 24–29.
- Svanberg, Jan, Tohid Ardeshiri, Isak Samsten, Peter Öhman, Tarek Rana, and Mats Danielson. 2022. Prediction of environmental controversies and development of a corporate environmental performance rating methodology. *Journal of Cleaner Production* 344: 130979.
- Talan, Gaurav, and Gagan Deep Sharma. 2019. Doing well by doing good: A systematic review and research agenda for sustainable investment. *Sustainability* 11 no. 2: 353.
- Taleb, Walid, Théo Le Guenedal, Frédéric Lepetit, Vincent Mortier, Takaya Sekine, and Lauren Stagnol. 2020. Corporate ESG news and the stock market. Available at SSRN 3723799.
- Talukdar, Swapan, Mohd Waseem Naikoo, Javed Mallick, Bushra Praveen, Pritee Sharma, Abu Reza Md Towfiqul Islam, Swades Pal, and Atiqur Rahman. 2022. Coupling geographic information system integrated fuzzy logic-analytical hierarchy process with global and machine learning based sensitivity analysis for agricultural suitability mapping. *Agricultural Systems* 196: 103343.
- Teoh, T.-T., Q.K. Heng, J.J. Chia, J.M. Shie, S.W. Liaw, M. Yang, and Y.-Y. Nguwi. 2019. Machine learning-based corporate social responsibility prediction. In *2019 IEEE international conference on cybernetics and intelligent systems (CIS) and IEEE conference on robotics, automation and mechatronics (RAM)*, 501–505. IEEE.
- Tidake, Ankita, Dr Sharma, Yogesh Kumar, and Dr Deshpande. 2020. Implementation of farmer informative system for maximizing the farming yield using machine learning approach. In *2nd international conference on communication & information processing (ICCIP)*.
- Tiwari, Aviral Kumar, Emmanuel Joel Aikins Abakah, David Gabauer, and Richard Adjei Dwumfour. 2022. Dynamic spillover effects among green bond, renewable energy stocks and carbon markets during COVID-19 pandemic: Implications for hedging and investments strategies. *Global Finance Journal* 51: 100692.
- Tornede, Tanja, Alexander Tornede, Jonas Hanselle, Marcel Wever, Felix Mohr, and Eyke Hüllermeier. 2021. Towards green automated machine learning: Status quo and future directions. arXiv preprint [arXiv: 2111.05850](https://arxiv.org/abs/2111.05850).
- Tuominen, Anneli. 2022. The way towards banks' good climate change risk management. 09. Keynote speech by Anneli Tuominen, Member of the Supervisory Board of the ECB, at the

- 9th Conference on the Banking Union, https://www.bankingsupervision.europa.eu/press/speeches/date/2022/html/ssm.sp220922_bb043aa0bd.en.html.
- Ullah, A.K.M., Samiha Sultana, Fahim Faisal, Md. Rahi, Muzahidul Islam, Md. Alam, Md. Alam, and Golam Rabiul. 2021. A brief review of responsible AI and socially responsible investment in financial and stock trading.
- UNEP. 2015. Paris agreement. <https://wedocs.unep.org/20.500.11822/20830>.
- Varian, Hal R. 2014. Big data: New tricks for econometrics. *Journal of Economic Perspectives* 28 no. 2: 3–28.
- Vishwakarma, Sandeep. 2019. Predictive impact assessment of climate change on crop yield using semi parametric neural network. *Advance and Innovative Research* 352.
- Vo, Nhi N.Y., Xuezhong He, Shaowu Liu, and Guandong Xu. 2019. Deep learning for decision making and the optimization of socially responsible investments and portfolio. *Decision Support Systems* 124: 113097.
- Wang, Wenhui, Xiangqing Ma, Syed Moazzam Nizami, Chao Tian, and Futao Guo. 2018. Anthropogenic and biophysical factors associated with vegetation restoration in Changting, China. *Forests* 9 no. 6: 306.
- Warin, Thierry, and Aleksandar Stojkov. 2021. Machine learning in finance: A metadata-based systematic review of the literature. *Journal of Risk and Financial Management* 14 no. 7: 302.
- Wei, Sun, Wang Yuwei, and Zhang Chongchong. 2018. Forecasting CO₂ emissions in Hebei, China, through moth-flame optimization based on the random forest and extreme learning machine. *Environmental Science and Pollution Research* 25 no. 29: 28985–28997.
- Wen, Hui. 2018. An empirical study of global corporate social responsibility reporting regulation and practice over 2000–2015 period. PhD diss., University of Illinois at Urbana-Champaign.
- Yan, Jiabao, and Biyun Meng. 2021. Stock market reactions to firms' first publication of sustainability reports: Evidence from multiple industries in Nordic countries, the United States, and China.
- Yang, Peng, Jun Xia, Xiangang Luo, Lingsheng Meng, Shengqing Zhang, Wei Cai, and Wenyu Wang. 2021. Impacts of climate change-related flood events in the Yangtze River Basin based on multi-source data. *Atmospheric Research* 263: 105819.
- Yao, Jiaxiong, and Yunhui Zhao. 2022. Structural breaks in carbon emissions: A machine learning analysis.
- Yu, Baojun, Changming Li, Nawazish Mirza, and Muhammad Umar. 2022. Forecasting credit ratings of decarbonized firms: Comparative assessment of machine learning models. *Technological Forecasting and Social Change* 174: 121255.
- Yu, Guangliang, Yukun Liu, William Cheng, and Chun-Te Lee. 2022. Data analysis of ESG stocks in the Chinese stock market based on machine learning. In *2022 2nd international conference on consumer electronics and computer engineering (ICCECE)*, 486–493. IEEE.
- Zhang, Jun, and Xuedong Chen. 2021. Socially responsible investment portfolio construction with a double-screening mechanism considering machine learning prediction. *Discrete Dynamics in Nature and Society* 2021 (1): 7390887.
- Zhang, Susan, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, et al. 2022. Opt: Open pre-trained transformer language models. arXiv preprint [arXiv: 2205.01068](https://arxiv.org/abs/2205.01068).
- Zhang, Dayong, Zhiwei Zhang, and Shunsuke Managi. 2019. A bibliometric analysis on green finance: Current status, development, and future directions. *Finance Research Letters* 29: 425–430.
- Zhao, Weizhong, James J. Chen, Roger Perkins, Zhichao Liu, Weigong Ge, Yijun Ding, and Wen Zou. 2015. A heuristic approach to determine an appropriate number of topics in topic modeling. In *BMC bioinformatics*, Vol. 16, 1–10. BioMed Central.
- Zhou, Jianguo, Xuechao Yu, and Xiaolei Yuan. 2018. Predicting the carbon price sequence in the Shenzhen emissions exchange using a multiscale ensemble forecasting model based on ensemble empirical mode decomposition. *Energies* 11 no. 7: 1907.
- Zhu, Bangzhu, and Julien Chevallier. 2017. Carbon price forecasting with a hybrid Arima and least squares support vector machines methodology. *Pricing and forecasting carbon markets: Models and empirical analyses* 87–107.

Appendix

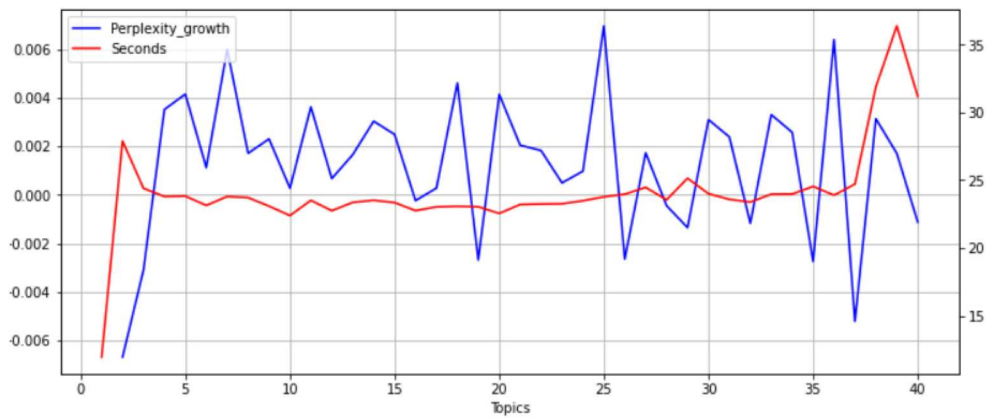


Figure A1. Rate of perplexity change and latency of training.

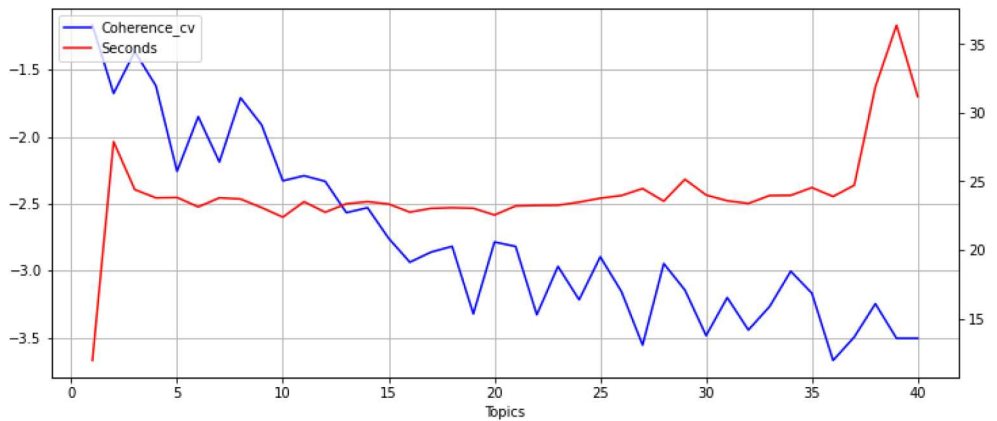


Figure A2. Coherence score and latency of training.

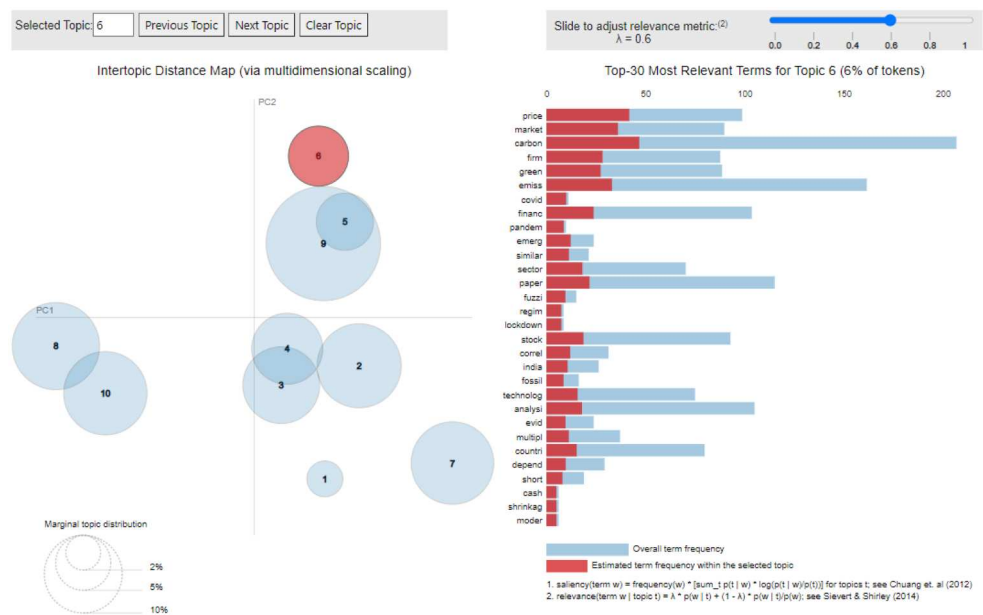


Figure A3. Visualization of topic 6 (carbon markets).

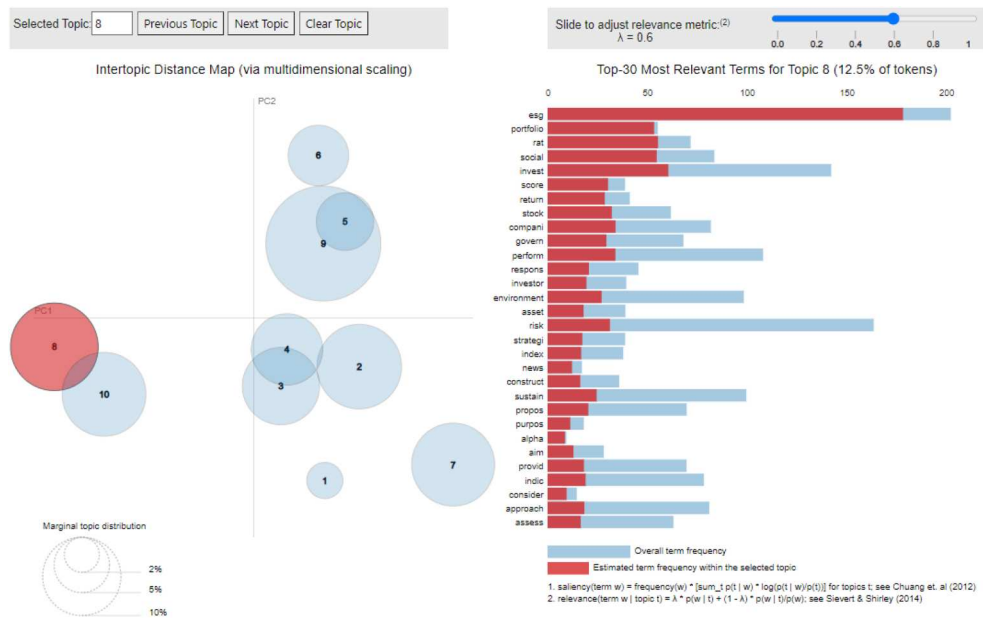


Figure A4. Visualization of topic 8 (ESG factors & investing).

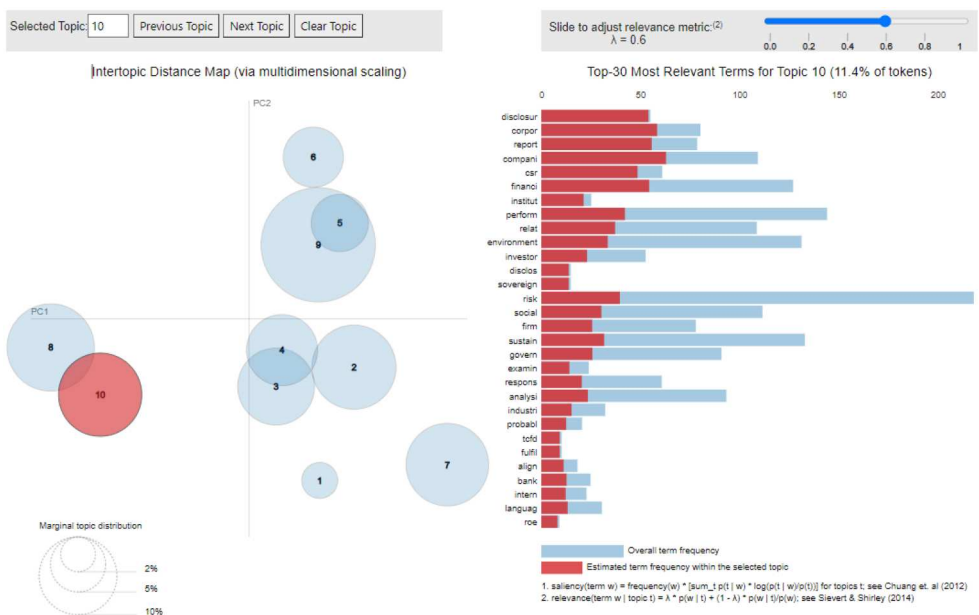


Figure A5. Visualization of topic 10 (climate data).

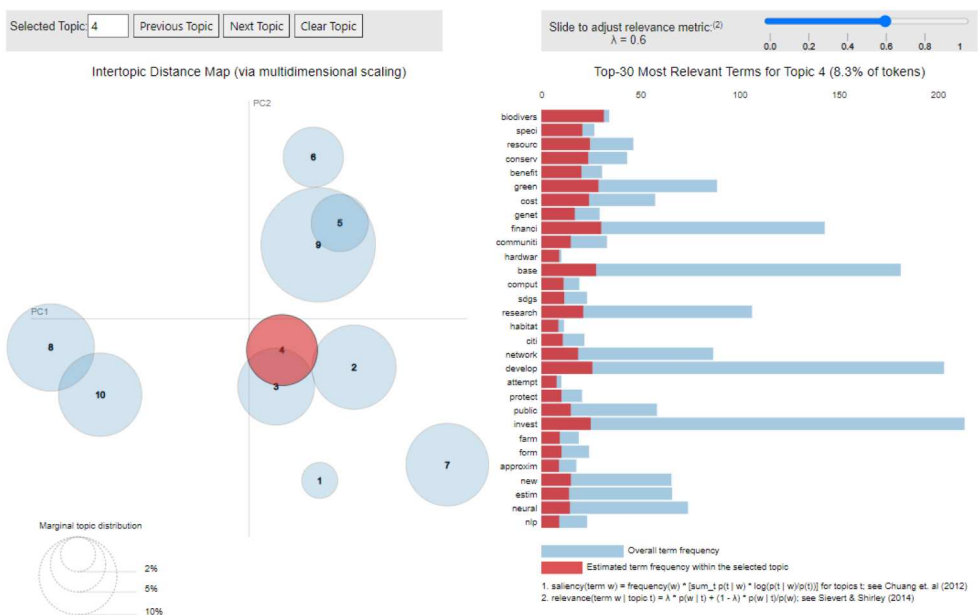


Figure A6. Visualization of topic 4 (biodiversity).

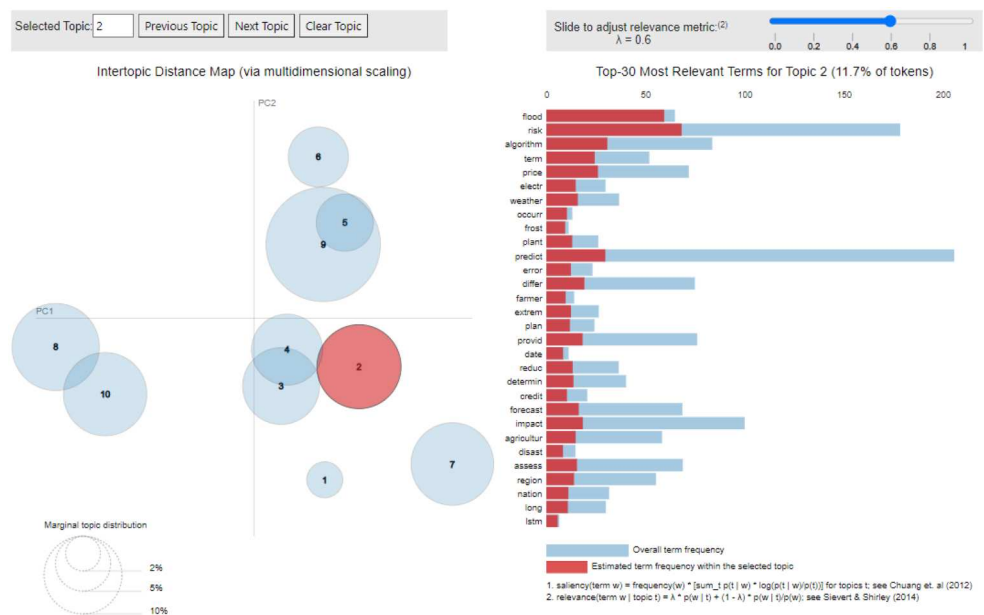


Figure A7. Visualization of topic 2 (natural hazards).

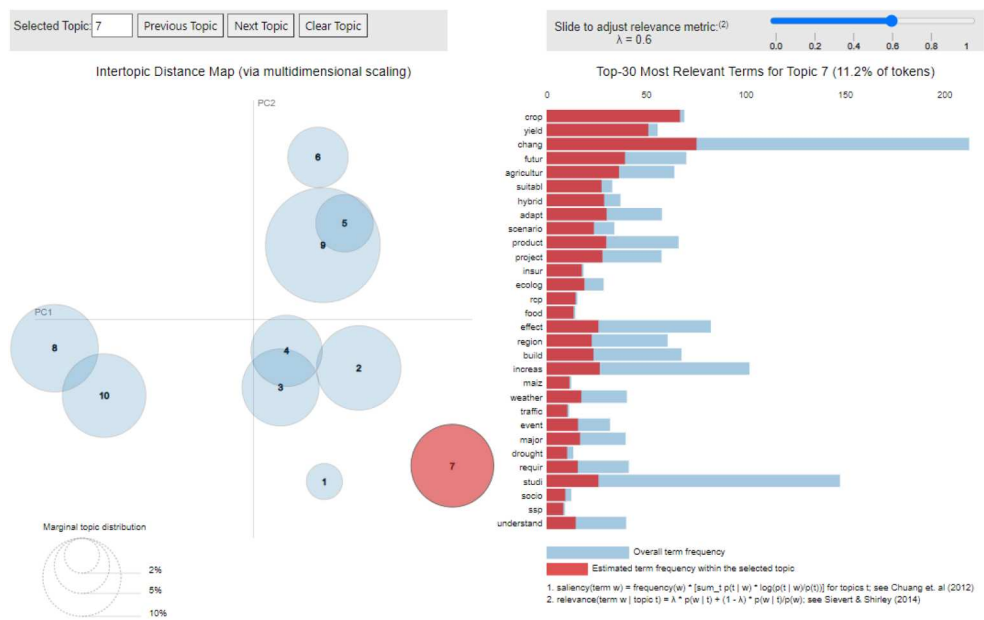


Figure A8. Visualization of topic 7 (agricultural risk).

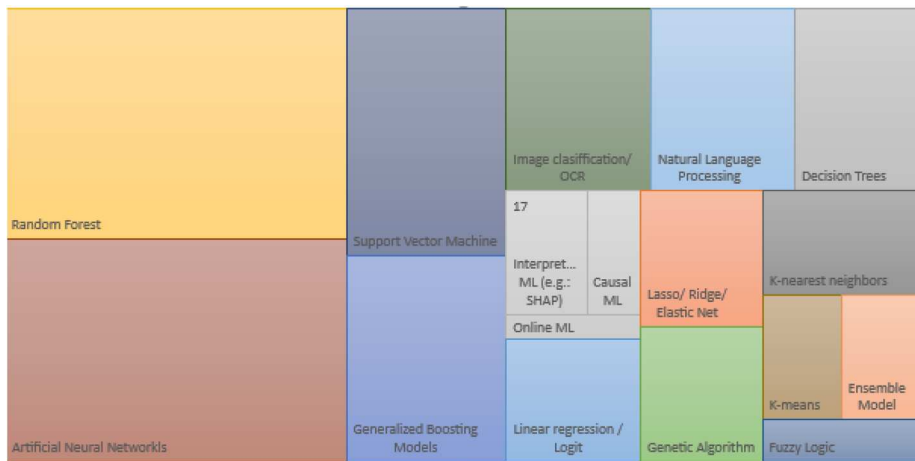


Figure A9. ML methods used in physical risks.

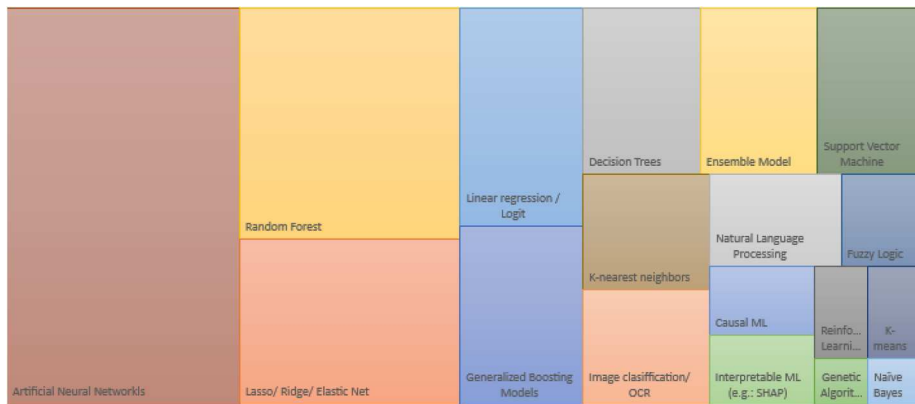


Figure A10. ML methods used in transition risks.

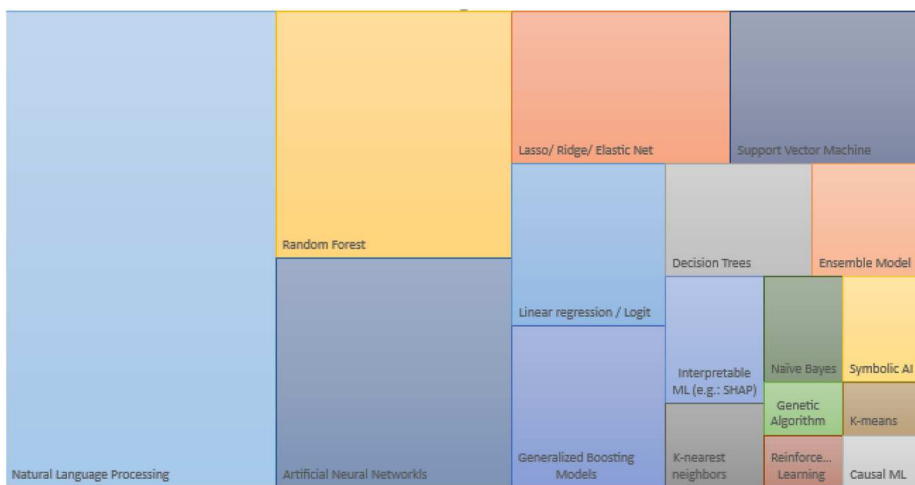


Figure A11. ML methods used in CSR.

Table A1. Original root words from stem.

Stem	Original words
activ	activity, active
adapt	adapt, adaptation
agricultur	agriculture, agricultural
avail	availability, available
base	based, baseline
biodivers	biodiversity, biodiverse
chang	change, changing, changer
compani	company, companies
conserv	conservation
corpor	corporate, corporation, corporatism
correl	correlation, correlate, correlated
cost	cost, costly, costing
develop	development, developing
differ	difference, different, differently
disclosur	disclosure, disclosures, disclosed
emiss	emission, emissions, emitting
energi	energy, energies
financ	finance, finances, financial, financially
financi	financial, financially
firm	firm, firms
flood	flooding, floods, flood
futur	future, futures
hybrid	hybrid hybridization
inform	inform, information, informative
invest	investment, investor, investing
method	method, methodology, methodological
polic	policy, policies, policymaker
predict	prediction, predictability, predictable
product	product, production, productive
project	project, projects, projecting
provid	provide, provider, providing
rat	rating, rated
relat	relate, related, relationship
resource	resource, resources, resourceful
risk	risk, risky, riskiness
signific	significance, significant, significantly
stock	stock, stocks
storag	storage, store, stored
studi	study, studied, studying
suitabl	suitable, suitability, suitably
sustain	sustainability, sustainable, sustainably
term	term, terminology
train	training
valu	value, valuation, valuable

Table A3. Corpus of documents. ML methods, by topic (Physical Risk).

Application domain		List of papers	List of ML models
Physical Risks	Natural Hazards	Bayle et al. (2020), Manandhar et al. (2020), Biffis and Chavez (2017), Chen et al. (2020), Cesarini et al. (2021), Lyubchich et al. (2019), Hoang et al. (2020), Inyang, Akpan, and Akinyokun (2020), Bjånes, De La Fuente, and Mena (2021), Nti et al. (2021), Rohayani et al. (2021), Avand et al. (2021), Shu et al. (2022), Yang et al. (2021), Diniz et al. (2021), Best et al. (2021)	Markov-CA (deep learning), Image classification, Random forest, Genetic algorithms, K-means, ANN, SVM, XGBoost, LSTM (Recurrent Neural Network), Extra trees, Regression model, CART (Decision Trees), Multi-layer Perceptron (deep learning), Adaptive Neuro Fuzzy Inference System, Back-propagation Neural Network, Ensemble model
	Biodiversity	Floreano and de Moraes (2021), Wang et al. (2018), Dao et al. (n.d.), Lima et al. (2022), da Silveira et al. (2021), Keys, Barnes, and Carter (2021), Macadam, Nowell, and Quigley (2021), Pearson et al. (2020), Dao et al. (2019), Santamaria et al. (2020), Reiersen et al. (2021), Rakova and Winter (2020), Hou et al. (2020), Seidl et al. (2020), Evans et al. (2010), Bastien-Olvera and Moore (2021)	Linear Regression, Decision Tree, Naive Bayes, Support Vector Machine, Random forest, Artificial Neural Network (ANN), K-Nearest Neighbours, Boosting Ensemble meta-algorithm, Reinforcement learning, Deep multi-agent reinforcement learning, Kernel Extreme Learning Machine, Stacked denoising autoencoders, Wavelet Neural Network, Genetic Algorithm, Particle Swarm Optimization, Bagging, Causal Direction from Dependency (D2C) algorithm, LightGBM (Gradient boosted decision trees), CatBoost, XGBoost, SHAP, Optical Character Recognition (OCR), Natural Language Processing (NLP), Interpretable trees, K-means, LSTM (Recurrent Neural Network), Double debiased ML, Radial Artificial Neural Network, Lasso, Causal forest, Causal boosting, Local Interpretable Model-Agnostic Explanation (LIME), Passive Aggressive Regressor, Linear Regression, Box-Cox, K-NN, Multilayer perceptron, Ridge, Elastic Net, RidgeCV, Least Angle Regression, Extra Trees, AdaBoost, Gradient Boosting, Failing rule (decision tree) Stacking, SHAP
	Agricultural Risk	P. Feng et al. (2019), Porfirio et al. (2017), Dhokley et al. (2018), Tidake et al. (2020), Ben Ayed and Hanana (2021), Talukdar et al. (2022), Ghaffarian et al. (2022), Liu and Zhan (2019), Coca-Castro, Golden, and Reymondin (n.d.), Gümüşçü, Tenekeci, and Bilgili (2020), Vishwakarma (2019), Belhadi et al. (2021), Sabu and Kumar (2020), S.S. Paul et al. (2020), Cortés and López-Hernández (2021), Müller et al. (2016), Haro et al. (2021)	Random forest, SVM, C4.5 classifier, Decision Trees, Gradient Boosting, Random forest, Multi-layer Perceptron (deep learning), SVM, Logit Boosting, Rotation Forest, Genetic algorithm, Multiple linear regression, Bayesian network, Convolutional Neural Network, Least-squares SVM, Extreme machine learning (feed-forward neural network), Ensemble model, LSTM (Recurrent Neural Networks), K-NN, ANN, Fuzzy logic, K-means, Generalized Boosted Regression, AdaBoost, Gradient Boosting Machines, Radial Basis Function Neural Network, Bagging, Boosting

Table A4. Corpus of documents. ML methods, by topic (Transition Risk).

Application domain		List of papers	List of ML models
Transition Risks	Carbon Markets	Zhu and Chevallier (2017), Zhou, Yu, and Yuan (2018), Levi (2021), Reiersen et al. (2021), Qi et al. (2021), Morkner et al. (2022), Reed, O'Reilly, and Hall (2019), C. Sun (2022), Y. Shi, Shen, and Wu (2021), Biesbroek, Badloe, and Athanasiadis (2020), Kulkarni (2021), Debnath and Bardhan (2020), Donner, Kandlikar, and Webber (2016), Nay (2016), Nay (2017), Pincet, Okabe, and Pawelczyk (2019), X. Feng et al. (2021), Q. Li et al. (2020), Jaycocks (2019), Schmidt et al. (2021), Abdullah et al. (2021), Caldecott et al. (2018), Nguyen, Diaz-Rainey, and Kuruppuarachchi (2021), Han et al. (2021), Yao and Zhao (2022), Khan and Awasthi (2019), M. Li et al. (2021), Fang et al. (2021), Debone, Leite, and Miraglia (2021), Nunnari et al. (2004), Acheampong and Boateng (2019), N. Ma et al. (2021), Q. Sun et al. (2021), Calvo-Pardo, Mancini, and Olmo (2022), Wei, Yuwei, and Chongchong (2018), X. Shi et al. (2020), Rahman, Manini, and Fatima (2021)	Least Squares Support Vector Machines, Extreme learning machine (Deep learning), SVM, Natural language processing (NLP), Back-propagation Neural Network, OLS, Lasso, Genetic Algorithm, ANN, Random Forest, Decision Tree, Convolutional Neural Networks, Multiple Linear regression, OLS, Elastic Net, K-NN, Random forest, Extreme Gradient Boosting Decision Tree, Fuzzy logic, Multilayer perceptron, Multinomial logistic regression, Ensemble model, Convolutional-Long Short Term Memory, Artificial neural network with backpropagation, Gaussian Process Regression, Feed-forward neural network, Extreme machine learning, Lasso, Natural Language Processing (NLP), Latent Dirichlet Allocation (LDA), Machine-coding (Symbolic AI), Fuzzy comprehensive evaluation model, GAN
	Climate data	Diggelmann et al. (2020), Schwabe, Sandhu, and Grebenshchikov (n.d.), Nugent, Stelea, and Leidner (2020), Owusu (2020), Sautner et al. (2020), K. Li and Yu (2022), Antoncic (2020), Kheradmand et al. (n.d.), Luccioni and Palacios (2019), Moreno and Caminero (2022b), Friederich et al. (2021), Luccioni, Baylor, and Duchene (2020), Cojoianu et al. (2020), Miglionico (2022), Raghupathi, Ren, and Raghupathi (2020), Bingler et al. (2022), Benites-Lazaro, Giatti, and Giarolla (2018), Raman, Bang, and Nourbakhsh (2020), Bala et al. (2015), Moreno and Caminero (2022a), Amel-Zadeh et al. (2021), Clarkson et al. (2020), Ehrhardt and Nguyen (2021), Wen (2018), Mansouri and Momtaz (2021)	Natural language Processing (NLP), Natural language understanding (NLU), Context-based algorithms, Keyword discovery algorithm, LDA, Word2vec, Doc2Vec (word embeddings), Text mining, Automated language systems, Text analytics, ClimateBert, Neural language modeling, SVM, Fully-connected neural network, Computer-based textual analysis, Logistic classifier, Lasso, Joint entity, Relation extraction, ANN

Table A5. Corpus of documents. ML methods, by topic (CSR).

Application domain		List of papers	List of ML models
Corporate & Social Responsibility	ESG factors & Investing	Engle et al. (2020), Hilario-Caballero et al. (2020), B. Yu et al. (2022), Lanza, Bernardini, and Faiella (2020), Jha (2021), Margot et al. (2021), Klusak et al. (2021), Vo et al. (2019), Guo et al. (n.d.), Q. Chen and Liu (2020), Erhardt (2020), J. Zhang and Chen (2021), Sokolov et al. (2021), G. Yu et al. (2022), Bua et al. (2022), Cepni, Demirer, and Rognone (2022), Plakandaras, Gogas, and Papadimitriou (2018), Taleb et al. (2020), Tiwari et al. (2022), Hisano, Sornette, and Mizuno (2020), Drei et al. (2019), Chang et al. (2021), Coqueret et al. (2021), Škapa et al. (2023), De Lucia, Paziienza, and Bartlett (2020), Teoh et al. (2019), Sokolov et al. (2020), Mitsuzuka, Ling, and Ohwada (2017), Gupta, Sharma, and Gupta (2021), Sokolov et al. (2021), Krappel, Bogun, and Borth (2021), D'Amato, D'Ecclesia, and Levantesi (2022), Svanberg et al. (2022), Lin and Bai (2022), Bouyé and Menville (2020), Berg et al. (2021), Citterio (n.d.), Kluza, Ziolo, and Spoz (2021), Natsume and Feng (2019), Z. Ma (2019), Anders (2021), Yan and Meng (2021), Joshi and Chauhan (2024), Michalski and Low (2021), Dudás and Naffa (2020), Riad et al. (2019), Hong et al. (2022), Sharma, Gupta, and Gupta (2024)	Textual analysis, Genetic algorithm, Multiobjective evolutionary algorithms, Classification and Regression Trees, Random forest, ANN, SVM, Decision Trees, Support Vector Regression (SVR), Deterministic ML (Symbolic AI), Multivariate Bidirectional Long Short-Term Memory neural network, Deep reinforcement learning, Deep learning, Ensemble model, XGBoost, Fuzzy reasoning, K-NN, AdaBoost, OLS, Lasso, Elastic Net, PLS, Ordered Logistic regression, Ridge, K-NN, SVM, Naive Bayesian, Multilayer Perceptron (MLP) Neural Networks, Long Short-Term Memory (LSTM) Neural Networks, Natural language processing (NLP), Extremely randomized trees, Linear regression, Feed-forward neural network, AdaBoost, CatBoost, XGBoost, Ensemble model, Kohonen neural network, Naïve Bayes, Gradient boosting, Logistic regression, Radial basis function (RBF), SVM, SHAP, Classification tree, Lasso, SHAP
	Climate data	Diggelmann et al. (2020), Schwabe, Sandhu, and Grebenshchikov (n.d.), Nugent, Stelea, and Leidner (2020), Owusu (2020), Sautner et al. (2020), K. Li and Yu (2022), Antoncic (2020), Kheradmand et al. (n.d.), Luccioni and Palacios (2019), Moreno and Caminero (2022b), Friederich et al. (2021), Luccioni, Baylor, and Duchene (2020), Cojoianu et al. (2020), Miglionico (2022), Raghupathi, Ren, and Raghupathi (2020), Bingler et al. (2022), Benites-Lazaro, Giatti, and Giarolla (2018), Raman, Bang, and Nourbakhsh (2020), Bala et al. (2015), Moreno and Caminero (2022a), Amel-Zadeh et al. (2021), Clarkson et al. (2020), Ehrhardt and Nguyen (2021), Wen (2018), Mansouri and Momtaz (2021)	Natural language Processing (NLP), Natural language understanding (NLU), Context-based algorithms, Keyword discovery algorithm, LDA, Word2vec, Doc2Vec (word embeddings), Text mining, Automated language systems, Text analytics, ClimateBert, Neural language modeling, SVM, Fully-connected neural network, Computer-based textual analysis, Logistic classifier, Lasso, Joint entity, Relation extraction, ANN