Machine Learning Predictive Modeling for assessing Climate Risk in Finance

MASSIMILIANO FERRARA[1], TIZIANA CIANO[2], ALESSIO CAPRIOTTI[3], SILVIA MUZZIOLI[3]

[1] Department of Law, Economics and Human Sciences
University "Mediterranea'" of Reggio Calabria
Via dell'Università, 25 - 89125 Reggio Calabria (RC)
ITALY

[2] Department of Economics and Political Sciences
University of Aosta Valley
Strada Cappuccini, 2 - 11100 Aosta (AO)
ITALY

[3] Marco Biagi Department of Economics
University of Modena and Reggio Emilia
Via Jacopo Berengario, 51, 41121 Modena (MO)
ITALY

Abstract: - We investigate how the application of advanced predictive models could help investors to assess and manage climate risk in their portfolios, contributing to the development of more sustainable and resilient investment practices. We highlight the possible applications of predictive analytics as a key tool in climate finance. It emerges how emerging technologies (blockchain and Artificial Intelligence) can improve transparency, efficiency, and climate risk analysis in sustainable investments. Further lines of research are highlighted, focusing on how investors and portfolio managers can develop strategies to manage the risks associated with climate events and the integration of climate risks into the management of Supply Chain Finance to ensure greater resilience and sustainability. Some generalized models are analyzed focusing the most important aspects and features by which modeling Climate risks and related issues in financial frameworks.

Key-Words: - Climate Risk, Machine Learning, Supply Chain Finance, Blockchain, Predictive Models, Generalized Models.

Received: June 23, 2024. Revised: November 7, 2024. Accepted: December 8, 2024. Published: December 16, 2024.

1 Introduction

Climate change significantly impacts the worldwide economy and financial markets, pointing out climate risk and uncertainty as central themes in asset pricing and relating modeling. Potential financial damages and financial crashes connected to severe weather events and slow changes like sea level rise are examples of climate hazards. These hazards may have an impact on supply (value) chains, infrastructure stability and business profitability and the related dynamics. The most important sources of the uncertainty around climate change are the complexity of natural systems and the unpredictability of human actions. Despite an increasing activities by which producing improvements in scientific investigation, there are still unknown aspects, such as the rate or severity of climate change. All aspects directly or indirectly involved, asset price and investment decisions are significantly influenced by climate risk and the uncertainty viewed as primary consequence. Climate risk abatement and environmental sustainability are critical to the long-term health of global economies and financial systems. Predictive and forecasting analytics are effective key tools in climate finance that provides investors with critical information about the risks and possibilities related to climate change. Emerging challenges with data quality, model uncertainty, regulatory complexity, and the inclusion of climate-related factors, predictive analytics has the capacity to modify enhancing the resilience and sustainability of investment portfolios. This study's findings could significantly implement the operational procedures and stages, policy decisions of regulatory agencies and Financial Institutions. This paper analyses the centrality of robust regulatory frameworks that promotes the integration of climate-related financial risks into policymakers' financial decisions. Businesses could be oriented to provide comprehensive climate change disclosures, establish standardized risk assessment methods, and promote climate-resilient investment strategies. The study highlights the strategical importance of incorporating machine learning models into risk management strategies for financial institutions; the involved process could arrange conducting climate stress tests, assessing portfolio climate risk using prediction models, and identifying investment opportunities in sustainable industries and related policies. The findings can suggest regulators enhancing oversight and enforcement by mandating Financial Institutions to use advanced climate risk models, conduct frequent assessments, and incorporate climate-related scenarios in stress tests. Furthermore, promoting collaboration among researchers, policymakers, and financial institutions is crucial to ensure the effective implementation of these findings, leading to more sustainable and resilient financial systems (Figure 1).

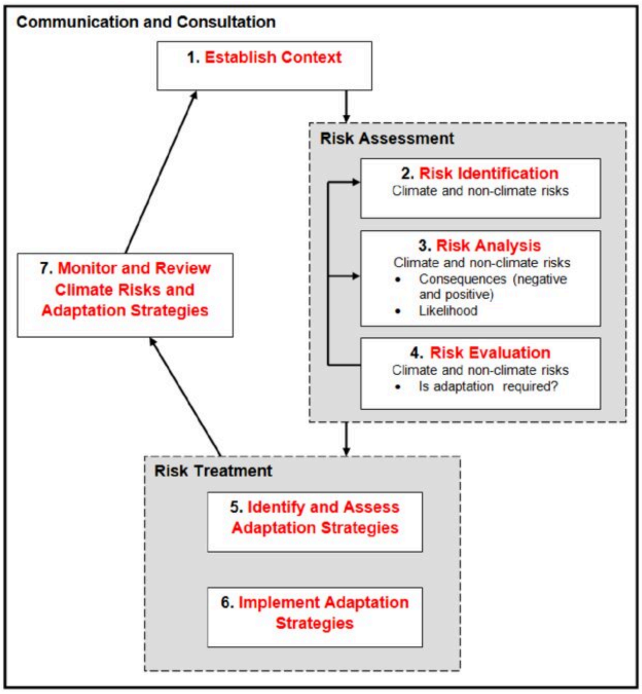

Figure 1: Climate Risks Modeling.

The paper proceeds as follows: section 2 reviews recent studies based on machine learning techniques that measure financial exposure to climate risks, which are understood as physical risks (extreme weather events such as floods and hurricanes) and transition risks (legislative changes and shifts in investment preferences). Section 3 suggests new directions for future research, based on machine learning methods. Section 4 is devoted to supply chain finance. Section 5 presents a list of generalized models can be built following the ML approach which arises from the existing literature. The last section concludes.

2 Machine learning approaches to climate risk, uncertainty and financial dynamics of related assets

Climate finance is a growing field in the literature, using machine learning (ML) to analyze climate-related information and model complex relationships.

In [1] authors point out that the traditional approach in econometrics involves specifying a target, or estimand, which is a functional of a joint distribution of the data. This target is usually a parameter of a statistical model that describes the distribution of a set of variables, typically conditional on some other variables, using a finite or infinite set of parameters. Let us consider a random sample from the population of interest, the parameter of interest and the nuisance parameters are estimated by identifying the parameter values that best fit the full sample, utilizing an objective function such as the sum of squared errors or the likelihood function. The emphasis is on the quality of the estimators of the target, traditionally assessed through large sample efficiency, and there is often interest in constructing confidence intervals. Researchers generally report point estimates and standard errors. On the other hand, machine learning (ML) literature primarily focuses on developing algorithms. The primary goal of these algorithms is typically to make predictions about certain variables given others or to classify units based on limited information, such as classifying handwritten digits based on pixel values.

In [2] authors review the academic literature to assess how machine learning is enabling the growth of climate finance. They identify seven granular areas where machine learning plays a significant role in the climate finance literature: natural hazards, biodiversity, agricultural risk, carbon markets, energy economics, ESG factors, and climate investing and data. The paper examines publishing trends and the application of machine learning methods in different research areas.

Predictive analytics is a vital tool in climate finance, offering investors valuable insights into climate-related risks and opportunities. This strategy forecasts and formalizes risks like extreme weather events and sea level rise, while identifying emerging opportunities in sustainable sectors like renewable energy and clean technologies. Challenges as data quality, (model) uncertainty, regulatory complexities, and climate-related factors need interdisciplinary collaboration, robust risk assessment structures, and continuous innovation. Predictive analytics and Forecasting modeling, despite its challenges, holds pivotal potential in enhancing the resilience and sustainability of investment portfolios management, contributing to a low-carbon economy. Assessing financial exposure to climate change is complex due to physical resource damage, low-carbon economies, changing social attitudes, and technological developments, making business disclosure texts valuable resources [3].

In [4] the authors employ context-sensitive machine learning techniques (LDA, word2vec, and FinBERT) to classify climate-related sentences, enhancing the objectivity and interpretability of enterprise-wide climate change exposure measures. The study highlights risks, mitigation strategies, and physical exposures, revealing unique economic implications and reducing human bias.

In [5], the authors employ advanced machine learning techniques to forecast oil prices during the Covid-19 pandemic, using a large dataset and SHApely Additive ExPlanations values for model analysis. The results show that high GER and ESG values lead to lower crude oil prices, promoting climate change mitigation and economic prosperity through green energy resources.

In [6] one examine the impact of the investment horizon on the optimal equity-bond-liquid portfolio in a dynamic model with climate change uncertainty. The equity risk premium is assumed to be an affine function of the global mean temperature and an unobserved factor estimated by Bayesian learning. The optimal investment strategy was found to be sensitive to climate uncertainty, with potentially significant welfare losses.

The impacts of climate change and water availability on corporate operational performance pose significant risks to investors, shareholders, and capital markets. The Task Force on Climate-Related Disclosures (TCFD) obliges companies to disclose their exposures to financial risks, with the aim of incentivizing investments in climate resilience through the management of natural resources. In [7] develops quantitative approaches to understand the financial relevance of water risk for corporate accounting and market performance. Indicators such as water intensity relative to turnover, operating profit and net fixed assets were assessed for companies representing nine industry sectors. It has shown that low water intensity results in better returns than the benchmark, return on equity and long-term valuation. An imputation methodology was developed that combines econometric models and machine learning techniques to predict water intensity parameters for companies that do not disclose water use risks. The results show that markets are rewarding water-intensive companies with higher yields, although the effect is attenuated after the implementation of TCFD.

In [8] authors present a new method for estimating crude oil prices based on socio-political and economic factors in the context of green finance. They use the LASSO (Least Absolute Shrinkage and Selection Operator) model, evaluating six factors and green finance. The model outperforms other models and provides insights into the temporal association between socio-political factors, green finance, and crude oil. The study finds that global steel production, the Kilian index, the green finance index, the value of the dollar, and terrorist attacks are important drivers of demand. In [9] one study ML models for predicting the credit rating of eco-friendly companies. They use 355 Eurozone companies ranked by climate change score from 2010 to 2019. The results show that classification and regression trees have the highest accuracy for credit rating predictions, even in investment grade or predefined categories. A set of random forests can also be used to predict predefined assessments. These findings are crucial for assessing the credit risk of pro-green firms.

The financial sector is incorporating sustainability into risk assessment models, using machine learning concepts to predict the likelihood of default. In [10] authors use regression and classification models, including Linear Regression, Decision Tree Regression, Support Vector Regression, Logistic Regression, Random Forest, Decision Tree Classifier, XGBoost Classifier, and Bagging Classifier. Label encoding converts categorical characteristics to numeric ones. The XGBoost classifier performs best in identifying defaulting companies, while the decision tree regressor predicts defaults. The authors explore the impact of machine learning on investment decisions, the potential benefits, limitations, and role in identifying sustainability-related risks and opportunities for sustainable growth. Machine learning literature focus also on analysing text data. Text analysis (TA) is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights [11] and [12]. The main advantage of this approach is the possibility to evaluate the existence of climate risk premium for physical and transition risks, jointly. The first paper in this stream of literature is [13]. Focusing on the United States, they create an index based on climate news from The Wall Street Journal (WSJ) and another from Crimson Hexagon (CH), which gathers a vast corpus of news articles and social media posts, filtering for climate change relevance. They use E scores (from both MSCI and Sustainalytics) to measure each firm's climate risk exposure and construct a hedge portfolio for innovations in climate news.

3 Machine Learning and Climate Finance: Towards a New Era of Risk Analysis and Forecasting

We see that collaboration across disciplines, such as economics, environmental science, and data science, improves understanding of climate risks and sustainable investment opportunities. Therefore, starting from the vast body of literature concerning climate risk, uncertainty, and the financial dynamics of related assets, we propose possible lines of theoretical research to be developed in the future. We note, at least to the best of our knowledge, that based on the above information, a possible specific and not widely explored area of research could be the integration of machine learning models for assessing climate risk exposure into real-time investment portfolios. Therefore, this line of research can focus on collecting real-time data and using up-to-date, real-time data from various sources, including climate, financial, and business data, to gain a more accurate view of climate risk exposure in investment portfolios. The application of machine learning techniques (neural networks, decision trees, or ensemble models) to develop advanced predictive models helps assess the climate risk exposure of investment portfolios. Investors may benefit greatly from this line of inquiry, which would enable them to more accurately evaluate and control the exposure of their portfolios to climate risk in real time. Moreover, it could potentially lead to the development and the related promotion of more robust and sustainable investment strategies and their decisions making. The present paper explores the correlation between financial resilience and climate change analyzing the influence of uncertainty and extreme weather events on investment portfolios and financial resilience of Institutions and Agencies. The work focuses on the development of risk management methodologies by investors portfolio managers and practioners and the role of portfolio diversification in reducing these kind of risks. The use of machine learning and deep learning techniques to predict climate change risks and their impact on businesses' financial performance was deeply studied. I. particular the use of sustainable practices or clean technology significantly improve risk-reduction tactics. Climate change impacts business models, particularly in carbon-intensive industries, through company-level climate risk assessment, examining how businesses can modify operational plans in reducing risk and enhancing sustainability. In this optic, the present work could explore the impact of emerging technologies like Blockchain, Fintech, and Artificial Intelligence Algorithms on climate finance; a special focus is fundamental for that concerns Blockchain which provides transparency and traceability in financial transactions, making it beneficial for monitoring the allocation of funds for green projects. The main goal is to permit that funds raised through green bonds are utilized for the declared purposes. AI algorithms and Fintech technologies can implement and favorite the analysis of climate risks and ESG data, aiding investors in portfolio management and informed decision-making and and its operational reinforcement. The main objective is to examine the utilization of these technologies in climate finance, focusing on their enhancement of efficiency and transparency in sustainable investment management by analyzing the potential risks associated with blockchain technology, such as potential risks in sustainable investments and carbon trade traceability, can be mitigated by examining the application of artificial intelligence.

4 The field of Supply Chain Finance is a new area of investigation.

Supply Chain Management and related approach, when combined with climate finance and advanced technology like blockchain and AI, can significantly promote ethical and sustainable supply chain practices. The Supply Chain Finance (from now SCF) is a set of financial products elaborated in studying the efficient movement of funds from the supplier to the final consumer throughout the supply chain: the concrete opportunity is provided to utilize advanced technologies in incorporating sustainable practices throughout the supply chain. This approach could lead to improved risk management, efficiency, and transparency, by which facilitating the transition to a low-carbon economy. The implementation of blockchain technology and the related approach in SCF enhances supply chain transaction transparency and traceability allowing businesses to monitor carbon emissions, verifying the origin of goods, and ensure suppliers in adhering to social and environmental regulations. Artificial intelligence can effectively optimize SCF operations such as payment management, demand forecasting, and data analysis to identify areas for its improvement. This could lead to reduced waste and increased efficiency. The SCF can encourage suppliers to adopt sustainable practices by offering more favorable financing terms in exchange for environmental and social obligations. Suppliers who utilize recycled materials or reduce carbon emissions may be offered lower credit rates. Along the supply chain, emerging technologies can assist in identifying and reducing risks, such as supplier-related social or environmental concerns. Blockchain-based supply chain (SCF) can enhance business continuity and reputation by facilitating collaboration among manufacturers, suppliers, and financiers. This could lead to more information sharing and a better understanding of common goals and needs. One possible analysis, in our opinion, is to include climate risks in the management of the SCF. This means looking at how climate risks could be included in SCF management models to ensure greater resilience and sustainability. To answer these questions, we believe it is useful to analyze how the integration of climate risks into SCF management models affects environmental sustainability and the valuation of financial assets, and to propose solutions to optimize the resilience and sustainability of supply chains. We also ask how the integration of climate risks into SCF management models improves the resilience and environmental sustainability of supply chains? What is the impact of ESG adoption on the valuation of financial assets along the supply chain? What innovative strategies can be adopted to optimize the financial management of supply chains under conditions of climate uncertainty? Therefore, it is crucial to understand the role of climate risks in SCF and their impact on environmental sustainability and the valuation of financial assets, but at the same time, it is crucial to identify innovative strategies and solutions to optimize the financial management of supply chains in conditions of climate uncertainty. Providing recommendations for companies interested in improving the resilience and sustainability of their supply chains through effective climate risk management is another aspect to focus on. Therefore, this line of research could help provide new perspectives on the financial management of supply chains in relation to climate risks, helping companies become more resilient and sustainable.

5 Machine Learning Modeling and Climate Risk Forecasting: a technical excursus of the main models in the existing literature and further developments

By analyzing the existing literature in this framework, some generalized mathematical representations of common concepts used in a great number of articles were selected in this survey. These are conceptual frameworks, not the actual equation-type used in the selected works. Our goal is to offer to the readers a scientific platform by which driving the conceptualization of the forecasting modeling approach following in the last decade.

1. *Predictive Modeling of Climate Risk*:
   A simplified representation of predicting financial losses due to climate risk might be:

   $$L = f(C, E, M)$$

   where:
   L: Financial losses.
   C: Climate variables (temperature, precipitation, sea level).
   E: Economic factors (market conditions, asset values).
   M: Mitigation measures (investments in resilience).
   f: A machine learning model (for example, a neural network, support vector machine).

2. *Risk Assessment by using Machine Learning*:
   We can represent the prediction of a risk score (R) as a function of various inputs (X):

   $$R = g(X)$$

   where:
   R: Risk score (a probability, a severity level).
   X: Input features (for example, climate data, financial data, geographic location).
   g: A machine learning model (for example, a regression model, classification model).

3. *Portfolio Optimization under Climate Uncertainty*:
   Portfolio optimization aims to maximize returns while minimizing risk. Incorporating climate uncertainty might look like:

   $$\begin{gathered}\text{Maximize } E[Rp] \\ \text{Subject to } Var[Rp] \leq \sigma^2, \\ \text{and} \\ C(p) \leq C_{max}\end{gathered}$$

   where:
   $E[Rp]$: Expected return of the portfolio
   $Var[Rp]$: Variance (risk) of the portfolio.
   σ^2: Acceptable risk threshold.
   $C(p)$: Carbon footprint of the portfolio.
   C_{max}: Maximum allowable carbon footprint
   Let us delve deeper into predictive modeling of climate risk, providing more nuanced examples. Recall that these are still generalized frameworks. The specific mathematical details would depend on the machine learning algorithm chosen and the data used.

Example 1: *Regression Model for Predicting Property Damage from Floods*
We could use a regression model to predict the monetary cost of flood damage (D) to properties:

$$D = \beta_0 + \beta_1 A + \beta_2 H + \beta_3 P + \beta_4 M + \varepsilon$$

where:
D: Monetary cost of flood damage (dependent

variable).
A: Area of the property (in square meters, for instance).
H: Height of the property above sea level (or flood plain elevation).
P: Precipitation during the flood event (in millimeters).
M: Flood mitigation measures implemented (for example, flood barriers. This might be a categorical variable that needs encoding).
$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$: Regression coefficients (estimated by the model).
ϵ: Error term (accounts for uncaptured factors).

This is a linear regression model. Other regression techniques (for example, polynomial regression, generalized linear models) could also be used, depending on the nature of the relationship between the variables.
The main aim of this kind of model is to predict the monetary cost of flood damage (D) to properties using a regression model. This is a supervised learning task; we need labeled data where we know the damage cost and other relevant features for each property. Data Requirements:

- Dependent Variable (D): Monetary cost of flood damage (in currency units, for example, dollars, euros). This should ideally represent the total cost of repairs, replacement, and any loss of value. Data sources could include insurance claims, government records, or post-flood assessments.
- Independent Variables (Predictors): These are the features that might influence the amount of flood damage.

The original example mentioned:

- A: Area of the property (square meters). Larger properties generally suffer greater damage.
- H: Height of the property above sea level (meters). Properties at lower elevations are more vulnerable.
- P: Precipitation during the flood event (millimeters). Higher rainfall leads to more severe flooding.
- M: Flood mitigation measures (categorical variable).

This could represent the presence or absence of flood barriers, elevation improvements, or other protective measures, which would require encoding (for example, 0 for no measures, 1 for measures implemented).

- Additional Relevant Variables: To improve the model's accuracy, consider including these:
- Distance to river/water body: Proximity to a water source strongly influences flood risk.
- Soil type: Different soil types have varying water absorption capacities.
- Building material and construction quality: The building's material and construction influence its resilience to flood damage.
- Floodplain zone: Properties located in designated floodplain zones face higher risks.
- Flood depth: The depth of floodwaters significantly impacts the damage level.
- Flood duration: Longer durations of flooding cause more extensive damage.

Model Selection and Training: While the original example suggests linear regression, other regression models are equally suitable, depending on the data and the relationships between variables:

- Linear Regression: Simple and interpretable, but assumes a linear relationship between predictors and the dependent variable.
- Polynomial Regression: Accounts for non-linear relationships.
- Ridge or Lasso Regression: Useful for handling multicollinearity (high correlation between predictor variables), which might occur with numerous variables related to flood characteristics.
- Random Forest Regression: A robust, ensemble method that often produces high accuracy but can be less interpretable than linear regression.

The selected model will be trained on a historical dataset using an appropriate algorithm. This algorithm will learn the relationships between the predictor variables and flood damage costs, thereby creating a model to predict damage based on new input features. After training, the model's performance must be assessed using various metrics, including:

- The Mean Squared Error (MSE), which measures the discrepancy between the predicted and actual damage costs, expressed as the average squared difference.
- R-squared, which shows the proportion of variance in the damage costs that the model can explain.

- The splitting of data into training and validation sets, which helps assess the model's ability to generalize to new data, a technique known as cross-validation.

- Data Availability: It may be difficult or impossible to obtain accurate and comprehensive data on flood damage and related factors, especially for past events.

- Model Complexity: Models that are too complex might overfit the data, reducing their ability to make predictions on new, unseen situations.

- Unforeseen Events: The model might struggle to handle extreme weather events that exceed historical trends. The accuracy of the model depends directly on the quality of the data used for training.

This approach highlights that developing an effective model for predicting flood damage requires careful data collection, thoughtful algorithm selection, rigorous performance evaluation, and an understanding of inherent limitations. Examination of Mitigation Measures (M): Mitigation measures (M) should be treated as categorical variables, as they represent different types of preventive actions rather than a continuous quantity. Examples of mitigation measures that could apply to a property include:

- No mitigation measures

- Installation of flood barriers

- Elevated foundation

- Improved drainage systems

- Combination of measures

Encoding Categorical Variables: Category variables can be converted into numerical representations that can be used in regression models using a variety of encoding schemes:

- One-Hot Encoding: For every category, a new binary variable is created. For example, if we have three categories of mitigation measures (None, Barriers, Elevated), we would create three new binary variables:
  M_None: 1 if no measures, 0 otherwise.
  M_Barriers: 1 if barriers installed, 0 otherwise.
  M_Elevated: 1 if elevated foundation, 0 otherwise.

- Label Encoding: Assigns a unique integer to each category (for example, 0 for None, 1 for Barriers, 2 for Elevated). Although this is more straightforward, it may suggest an ordinal relationship between categories that does not actually exist.

- Binary Encoding: In situations where there are numerous categories, binary encoding is helpful. It uses a binary code to represent each category.

The model and number of categories determine the encoding option. Regression models typically benefit from one-hot encoding in order to prevent imposing erroneous ordinal associations. One-hot encoding is a method of converting qualitative data (categories) into quantitative data (numbers) since regression models operate with numbers rather than categories. Let's break down, each line:

- M_None: 0 otherwise, 1 if no measures. This creates a new binary variable called M_None.

For each property in your dataset:

- If the property has no mitigation measures implemented, the value of M_None will be set to 1.

- If the property does have any mitigation measures (barriers, elevated foundation.), then the value of M_None will be set to 0.

- M_Barriers: 1 if barriers installed, 0 otherwise.

This creates another binary variable, M_Barriers. For each property:

- If flood barriers are installed, M_Barriers is 1.

- If flood barriers are not installed, M_Barriers is 0.

- M_Elevated: 1 if elevated foundation, 0 otherwise.

Similarly, this creates a binary variable M_Elevated:

- 1 if the property has an elevated foundation.

- 0 if the property does not have an elevated foundation.

By using separate binary variables for each category avoids imposing an artificial order or numerical relationship between the different mitigation strategies. If we simply assigned numbers (0, 1, 2) to (None, Barriers, Elevated), the model might incorrectly interpret "2" (Elevated) as being twice as effective as "1" (Barriers), which isn't necessarily true (see table 1). One-hot encoding ensures that

Table 1. Imagine a dataset with three properties

Property	Mitigation Measures
A	None
B	Barriers
C	Elevated Foundation

each category is treated independently by the model. Imagine a dataset with three properties:

After one-hot encoding, the data would look like this:

Table 2. Imagine a dataset with three properties

Property	M_None	M_Barriers	M_Elevated
A	1	0	0
B	0	1	0
C	0	0	1

This numerical representation can now be easily used as input features in the regression model to predict flood damage (see table 2).

Linearity Assumption and Non-Linear Models: In our discussion by which presenting the problem for a vast audience in understanding the methodology followed in modeling, we can start considering that the linear assumption in a standard linear regression model might oversimplify the relationships between predictors and flood damage. Anyway, flood damage is likely to have non-linear relationships with many predictors. For example:

- The relationship between flood depth and damage cost is not linear; a small increase in depth may lead to minor damage, but a large increase can lead to catastrophic damage.
- The effectiveness of mitigation measures isn't necessarily linear; having some measures may offer partial protection, while more comprehensive measures provide greater reduction in damage.

Therefore, non-linear models should be considered:

- Polynomial Regression: Incorporates polynomial terms of the predictors to model curvature.
- Regression Trees (Decision Trees, Random Forests, Gradient Boosting Machines): These models are non-parametric and can capture complex non-linear relationships effectively. Trees are particularly good for handling the interaction effects that may exist between the different mitigation measures and other explanatory variables.
- Neural Networks: Can model highly non-linear relationships but can be more complex to train and interpret.

Using non-linear models generally improves the accuracy and robustness of the prediction model for flood damage. The choice of the appropriate model would depend on the characteristics of the data and the complexity of the relationships. It's often necessary to compare the performance of several models using cross-validation to select the best one.

Example 2: *Classification Model for Assessing Risk of Coastal Erosion*
To classify the risk level of coastal erosion (R), we could use a classification model:

$$R = g(S, C, P)$$

where:
R: Risk level (for example, low, medium, high - a categorical variable).
S: Slope of the coastline.
C: Coastal sediment composition (for example, percentage of sand, silt, and clay).
P: Predicted sea level rise in the next 20 years.
g: A classification model (for example, logistic regression, support vector machine, random forest).

The classification model learns the relationship between the input variables (slope, sediment composition, sea level rise) and the risk level. The output would be a probability or a direct assignment to a risk category.

Example 3: *Time Series Model for Predicting Changes in Crop Yields*

To model the impact of climate change on crop yields (Y) over time, we could use a time series model:

$$Y_t = f(Y_{t-1}, Y_{t-2}, \ldots, T_t, P_t, R_t)$$

where:
Y_t: Crop yield in year t.
Y_{t-1}, Y_{t-2}, ...: Crop yields in previous years (autoregressive component).
T_t: Temperature in year t.
P_t: Precipitation in year t.
R_t: Other relevant factors (for example, irrigation, fertilizer use).
f: A time series model (for example, ARIMA, LSTM recurrent neural network).

This model accounts for the temporal dependence of crop yields and the influence of climate variables.

Starting from the technical survey of the modeling approach have been emerged in the related literature, some useful considerations should be pointed out. In particular in our opinion some aspects and /or conceptual phases must be explored in finding new research paths in the next future.

Feature Engineering and Selection:

- Relevance and Redundancy. The selection of input features (variables) is critical. Including irrelevant features can lead to overfitting (the model performs well on training data but poorly on new data), while redundant features can increase computational cost and reduce model interpretability. Feature engineering techniques (creating new features from existing ones) can significantly improve model performance.
- Handling Categorical Variables. Many climate-related variables are categorical (for example, soil type, land use). These need to be appropriately encoded (for example, one-hot encoding, label encoding) before being used in most machine learning algorithms.
- Data Scaling and Normalization. Different variables might have different scales (for example, temperature in Celsius, precipitation in millimeters). Scaling or normalizing the data is often necessary to prevent features with larger values from dominating the model.

Model Interpretability and Explainability:

- Some machine learning models, such as deep neural networks, are considered ”black boxes” because their decision-making processes are not transparent. In contrast, ”white box” models (such as linear regression or decision trees) are more easily interpretable. In climate risk assessment, understanding why a model makes a specific prediction is essential for ensuring its reliability and practical use. Model interpretability can be enhanced using methods like SHAP (SHapley Additive exPlanations).
- It is crucial to assess the uncertainty in model predictions. This can involve creating uncertainty maps, probability distributions, or confidence intervals. Responsible risk management requires informing decision-makers about the levels of uncertainty associated with predictions.

Generalization and Extrapolation:

- Models must perform well when applied to new, unseen data. Poor performance on new data is often a result of overfitting, where the model learns the training data, including noise, too well. On the other hand, underfitting occurs when the model is too simple to capture the underlying patterns accurately. Techniques like regularization and cross-validation are commonly used to address these issues and improve the model’s ability to generalize.
- Climate models often need to forecast outcomes outside the scope of the training data, such as predicting extreme events that have not been observed before. This can be challenging, and the limitations of the model must be carefully considered when making predictions in new scenarios.

Data Availability and Limitations:

- The accuracy of these models is significantly influenced by the quantity and quality of the input data. Biases, measurement errors, and missing data can all have a substantial impact on the model’s predictions.
- The choice of machine learning technique should be guided by the nature of the problem and the characteristics of the data. Depending on the need, techniques such as regression, classification, or time series analysis might be chosen.
- Rigorous assessment techniques are essential for evaluating a model’s performance. These approaches include metrics like accuracy, precision, recall, and similar measures.
- Climate data can sometimes have inconsistent spatial and temporal resolutions, which should be carefully considered as they may affect the model’s ability to predict outcomes accurately.
- Bias in climate data can arise from historical factors, measurement errors, or sampling techniques. It is crucial to address these biases to prevent distorted predictions.
- In some areas, especially in developing countries, there may be a lack of high-quality climate data. This can result in less accurate risk estimates in those regions.

Computational Resources and Costs:

- Training advanced machine learning models, such as deep learning models, can require significant amounts of time and computational power.

- The model requires ongoing effort to ensure its proper mechanism design in real-world applications, including monitoring its performance and retraining it as new data becomes available. These extended considerations point out the challenges involved in building and implementing effective predictive models for climate risk assessment. Addressing these issues is crucial for generating reliable and actionable results that can inform effective decision-making.

6 Conclusions

By this study we have selected some future research's lines including the involvement of machine learning models to assess climate risk exposure in investment portfolios in real-time and a sketch of some generalized models which summarize the different approach have been captured in the related literature. Our main aim was to point out how the use of real-time data and machine learning and deep learning techniques can sunstain the development of advanced predictive models to manage climate risk exposure and improving investment strategies at the same time. Emerging technologies such as blockchain and artificial intelligence are candidate as concrete and robust promising for improving transparency, efficiency, and climate risk analysis in sustainable investing. It explores the importance of understanding the role of these technologies in climate finance and investigating the possible associated risks. In fact, we have highlighted how the role of Supply Chain Finance (SCF) is crucial in promoting more sustainable and responsible practices along the supply chain, in relation to climate finance and emerging technologies such as blockchain and artificial intelligence. This improves transparency, efficiency, and risk management, contributing to the transition to a low-carbon economy. Therefore, we have highlighted how the integration of climate risks into the management of the SCF improves the resilience and sustainability of supply chains. Emerging technologies can help monitor and mitigate climate risks along the supply chain, facilitating collaboration between different parties and contributing to environmental sustainability and the valuation of financial assets. Future research should prioritize the development of real-time climate risk assessment models. This necessitates the integration of diverse, high-frequency data sources (climate data, financial market information, and real-time business intelligence) using sophisticated machine learning techniques. Specifically, exploring advanced ensemble methods is crucial. Research should focus on tacking and blending by investigating on the optimal combination of different base learners (for istance, regression trees, neural networks, support vector machines) to improve predictive accuracy and robustness. To enhance overall performance, it is recommended to utilize each model's proportional advantages and disadvantages. We have to observe on ”Stacking and blending” by examining the best way to combine various base learners (such as support vector machines, neural networks, and regression trees) to increase the resilience and accuracy of predictions. What challenges for the next future? We are going to focus on Ensemble model explainability: by using of techniques such as SHAP values to improve complex ensemble models' interpretability and to propose new interpretation of ”Coalition” not formed by agents but to the contrary from ”Features” of some informations involving for predictive analysis. This kind of approach aids to gain trust driving decision-making require an understanding of the reasons behind a model's prediction of a particular risk level. We have to point out on data scarcity and quality: establishing methods for dealing with missing and unequally dispersed data, which is particularly important in such areas with a scarse quantity of historical financial and climatic records. This could entail generating synthetic data, by using data imputation techniques, or integrating other data sources. In our opinion we need to study much more the relationships between supply chain dynamics, financial resilience, and climate change in detail is also essential. This should involve developing increasingly complex portfolio optimization models that particularly account for climate-related risks and uncertainties, as well as evaluating the effectiveness of different SCF techniques under different climate scenarios. Beyond the purely academic contributions, there are significant practical implications. The results may have an impact on regulatory policy by providing a scientific basis for developing standards for climate risk disclosure, developing stress tests for financial institutions, and rewarding climate resilient investments. Better risk assessment models should be integrated into financial institutions' operations to enhance portfolio management, evaluate the climate exposure of various assets, and create more sustainable investment strategies. Developing guidelines and resources for the real-world implementation of these models in financial institutions is a crucial next step, with a focus on model transparency, explainability, and responsible data use. In conclusion, more deep research needs to transform these promising findings into practical policies and sound financial judgments, even if machine learning offers a lot of potential for controlling financial risks related to climate change. This involves a special emphasis on explainability, data constraints, model development, and applications so useful for practioners.

Acknowledgments:
This work was funded by European Union under the NextGeneration EU Programme within the Plan "PNRR - Missione 4 "Istruzione e Ricerca" - Componente C2 Investimento 1.1 "Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN)" by the Italian Ministry of University and Research (MUR), Project title: "Climate risk and uncertainty: environmental sustainability and asset pricing". Project code "P20225MJW8" (CUP: E53D23016470001), MUR D.D. financing decree n. 1409 of 14/09/2022.

References:

[1] Athey, Susan, and Guido W. Imbens. (2019). "Machine learning methods that economists should know about". Annual Review of Economics, 11, pp. 685-725.

[2] Alonso, A., Carbó, J. M., and Marqués, J. M. (2023). Machine Learning methods in climate finance: a systematic review.

[3] Ofodile, O. C., Oyewole, A. T., Ugochukwu, C. E., Addy, W. A., Adeoye, O. B., and Okoye, C. C. (2024). Predictive analytics in climate finance: Assessing risks and opportunities for investors. GSC Advanced Research and Reviews, 18(2), 423-433.

[4] Hu, W., Li, K., and Yu, T. (2022). A Machine Learning-based Anatomy of Firm-level Climate Risk Exposure. Available at SSRN.

[5] Jabeur, S. B., Khalfaoui, R., and Arfi, W. B. (2021). The effect of green energy, global environmental indexes, and stock markets in predicting oil price crashes: Evidence from explainable machine learning. Journal of Environmental Management, 298, 113511.

[6] Rubtsov, A., & Shen, S. (2024). Dynamic portfolio decisions with climate risk and model uncertainty. Journal of Sustainable Finance & Investment, 14(2), 344-365.

[7] Tian, M. (2023). Impact of Climate Water Risk on Corporate Operational and Capital Markets Performance: A Machine Learning Approach (Doctoral dissertation).

[8] Mohsin, M., and Jamaani, F. (2023). Green finance and the socio-politico-economic factors' impact on the future oil prices: Evidence from machine learning. Resources Policy, 85, 103780.

[9] Yu, B., Li, C., Mirza, N., and Umar, M. (2022). Forecasting credit ratings of decarbonized firms: Comparative assessment of machine learning models. Technological Forecasting and Social Change, 174, 121255.

[10] Sufian, M. A., and Levesley, J. (2023, October). Machine Learning and Sustainability Metrics: Optimising Risk Assessment and Default Prediction. In Proceedings of the Future Technologies Conference (pp. 377-414). Cham: Springer Nature Switzerland.

[11] Blei, D. M., Ng, Y., Andrew, and Jordan, M. I. (2003). Latent dirichlet allocation. J. Mach. Learn. Re, 993–1022.

[12] Gentzkow, M., Kelly, B., and Taddy, M. (2019). Matt Text as Data. Journal of Economic Literature, 57(3), 535-74.

[13] Engle, R. F., Giglio, S., Kelly, B., Lee, H., and Stroebel, J. (2020). Hedging Climate Change News. The Review of Financial Studies, 33(3).

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)
The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
This work was supported and funded by European Union under the NextGeneration EU Programme within the Plan "PNRR - Missione 4 "Istruzione e Ricerca" - Componente C2 Investimento 1.1 "Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN)" by the Italian Ministry of University and Research (MUR).

Conflicts of Interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Statement During the preparation of this work the authors used GPT4-o in order to support:

1. the selection of the most important papers and items to cite in the literature review (see paragraph 1 and 2) 2. To arrange the final version of the Conclusions.

After using this tool, the authors reviewed and edited the content adding fundamental improvements as needed and takes full responsibility for the content of the publication.