UNIVERSITY OF EDINBURGH
Business School

INVESTIGATING THE IMPACT OF CLIMATE CHANGE ON MORTGAGE DEFAULTS IN THE US EAST COAST : a machine learning approach

B213801

Traditional Research Dissertation

Dissertation presented for the Degree of

MSc Banking and Risk

2022/2023

Abstract

Machine learning models, including random forest and neural networks, were developed in this study to explore the influence of extreme weather events, specifically hurricanes and cyclones, on the likelihood of a mortgage defaulting. This study represents a thorough investigation into the intricate relationship between time-varying covariates, weather events, and their joint effect on loan defaults. These models were meticulously compared to traditional logistic regression models to ascertain their predictive prowess. The study is centered around a portfolio of over 74,000 mortgages spanning from 2015 and originating from five US east coast states – Florida, North Carolina, South Carolina, New York, and New Jersey. Crucially, the study unearthed statistically significant and non-linear relationships between weather events and mortgage defaults. Hurricanes of category 4 and 5 emerged as the most influential factors, with their impact showing the highest statistical significance. Notably, category 1, 2, and 3 hurricanes exhibited a notable impact, albeit to a lesser extent. Even cyclones, while presenting the least significance, demonstrated an undeniable influence on mortgage default probabilities. The impact of hurricanes more than doubles when transitioning from a hurricane of category greater than 3 to a lower category. These findings resonate with existing literature, reinforcing the consensus regarding the correlation between extreme weather events and mortgage repayment dynamics. This observation highlights the need for financial institutions to adopt a comprehensive approach that factors in climate change when assessing models and pricing loans.

Table of Contents

CHAPTER 1: INTRODUCTION……………………………………………………………………......4

CHAPTER 2: LITERATURE REVIEW ………………………………………………………………......8

2.1. CLIMATE RISKS……………………………………………………………………………………………….........8

2.2. FINANCIAL RISKS……………………………………………………………………………………………........10

2.3. CREDIT SCORING………………………………………………………………………………………………......14

2.4. DELINQUENCY/ DEFAULT………………………………………………………………………………………..15

2.5. CREDIT SCORING MODELS…………………………………………………………………………………….18

2.6. CREDIT SCORING MODELS' LIMITATIONS………………………………………………………………19

2.7. THE CASE OF THE UNITED STATES……………………………………………………………………………20

2.8. GAPS …….24

CHAPTER 3: METHODOLOGY…………………………………………………………………………….26

3.1. DEVELOPMENT OF HYPOTHESIS…………………………………………………………………………….26

3.1.1. MAIN HYPOTHESIS…………………………………………………………………………………………….26

3.1.2. SUB HYPOTHESIS……………………………………………………………………………………………….26

3.2. RESEARCH DESIGN………………………………………………………………………………………………..27

3.2.1. VARIABLES AND CONFOUNDING FACTORS…………………………………………………………...27

3.2.2. MODELS – MAIN HYPOTHESIS …………………………………………………………………………….28

3.2.3. MODELS – SUB HYPOTHESIS ……………………………………………………………………………….31

CHAPTER 4: DATA AND PRESENTATION …………………………………………………………….35

4.1. FANNIE MAE SINGLE-FAMILY HISTORICAL LOAN PERFORMANCE DATASET……………………….35

4.2. WEATHER EVENTS…………………………………………………………………………………………………39

4.3. CLIMATE DATA……………………………………………………………………………………………………...41

CHAPTER 5: RESULTS…………………………………………………………………………………….43

5.1. DESCRIPTIVE STATISTICS……………………………………………………………………………………43

5.2. CORRELATION ANALYSIS ………………………………………………………………………………………47

5.3. EMPIRICAL RESULTS…………………………………………………………………………………………….49

5.3.1 MAIN HYPOTHESIS…………………………………………………………………………………………….49

5.3.2 SUB HYPOTHESIS……………………………………………………………………………………………….56

5.4. MODEL PERFORMANCE………………………………………………………………………………………62

CHAPTER 6: CONCLUSION…………………………………………………………………………….65

6.1. SUMMARY………………………………………………………………………………………………………..65

6.2. IMPLICATIONS……………………………………………………………………………………………………67

6.3. LIMITATIONS……………………………………………………………………………………………………..69

6.4. FUTURE RESEARCH……………………………………………………………………………………………..71

CHAPTER 7: APPENDIX……………………………………………………………………………….73

CHAPTER 8: REFERENCES ……………………………………………………………………………74

1) Introduction

Climate change, a pervasive and escalating phenomenon, has emerged as an existential threat, with far-reaching implications for our planet and its inhabitants (Intergovernmental Panel on Climate Change, 2014). The compelling scientific consensus says that anthropogenic activities are greatly increasing the accumulation of greenhouse gases in the atmosphere, resulting in a steady warming trend. Evidence suggests that global temperatures are expected to rise by at least 2.7°F by 2100 under most scenarios positing the most aggressive mitigation of greenhouse gas emissions (U.S. Global Change Research Program, 2018). Projections also indicate that ground-level temperatures will increase in a quicker manner over land than over oceans, while some regions are expected to witness a greater increase in temperature when compared to the global average (Harris, Jones, Osborn, & Lister, 2014).

Climate change is a multifaceted problem, inducing shifts in long-term weather patterns and increasing the frequency and intensity of extreme weather events (Melillo et al., 2014). The resultant environmental impacts, from rising sea levels to more severe storms, heatwaves, and floods, pose unprecedented challenges for economic stability and social wellbeing (Hsiang et al., 2017).

In particular, the US residential mortgage market, one of the largest debt markets in the U.S, is valued at over $11 trillion, and is vulnerable to climate-induced risks (Mian & Sufi, 2014). Adverse weather events such as hurricanes, wildfires, and floods can lead to mortgage default and prepayment, disrupting the stability of the residential housing market and the broader financial system (Kousky, Palim, & Pan, 2020). Despite the immense

stakes, the exposure of vulnerable residences to climate change remains uncertain and underexplored (Painter, 2020).
Focusing on the U.S, and specifically the five eastern states of New York, New Jersey, Florida, North Carolina, and South Carolina, this study aims to analyse the intertwined dynamics of climate change impacts and mortgage defaults. The selection of these states is predicated on several salient factors. Firstly, they are constantly subjected to hurricanes and cyclones, and other intense weather phenomena with far-reaching socio-economic consequences (National Oceanic and Atmospheric Administration [NOAA], 2021).
Secondly, these states are more populous compared to others, due to a high-density of urban centres and expansive suburban areas, which amplifies the financial ramifications of these climatic events (U.S. Census Bureau, 2020). The ubiquity of residential structures increases the number of mortgage loans exposed to climate risk, and thereby the potential for widespread default in the event of a severe storm.

Indeed, the US has been incurring financial losses amounting to billions of dollars annually due to the ravages of hurricanes (National Centers for Environmental Information [NCEI], 2021). This adverse fiscal fallout has been on an upward trajectory, mirroring the escalating frequency and intensity of these disasters, the trend is likely fuelled by climate change (Knutson et al., 2019). Recent examples such as Hurricane Harvey in 2017, and the successive blows of Hurricanes Florence and Michael in 2018, inflicted substantial damage, underscoring the urgency and importance of this inquiry (Congressional Budget Office [CBO], 2019).

To navigate the intricate interplay between climate change and mortgage defaults, this study employs a methodological approach that employs both machine learning, specifically the Random Forest algorithm, and logistic regression. The choice of the Random Forest algorithm in this study stems from its unique strengths in dealing with the complexities of large datasets, the kind typically found in studies analysing climatic variables and mortgage default data. This decision tree-based model is adept at identifying non-linear relationships, an attribute that is particularly useful when exploring interactions between multifaceted climatic and financial variables. The model constructs multiple decision trees on different sub-samples of the dataset and then aggregates their predictions. This process mitigates the risk of overfitting, a common problem in machine learning where models perform exceptionally well on training data but poorly on unseen data (Breiman, 2001). The Random Forest algorithm's capacity for pattern recognition and generalization makes it a potent tool for unravelling the intricate correlations between climate change factors and mortgage defaults.

Simultaneously, the study incorporates logistic regression, a more conventional statistical method. The inclusion of logistic regression stems from its attribute of providing clear interpretability of relationships between variables. While machine learning models like Random Forests can capture complex patterns, their 'black box' nature can make it challenging to understand the precise nature of the relationships they identify (Rudin, 2019). In contrast, logistic regression offers an intuitive and straightforward representation of the impact of independent variables (in this case, climate-related factors) on the dependent variable (mortgage default). The odds ratios resulting from logistic regression provide an easily interpretable measure of the effect size of each variable (Menard, 2002).

The integration of Random Forests and logistic regression in this study offers a dual advantage. The Random Forest algorithm enables the capturing of complex patterns and interactions, while logistic regression provides an interpretative understanding of these relationships.

2) Literature Review

2.1) Climate Risk

The Basel Committee's report in 2022 underscores the significance of climate-related risks, which are progressively escalating and can manifest in two forms: Physical risks and Transition risks. Physical risks involve the direct and indirect impacts of climate change, such as extreme weather events and rising sea levels. Transition risks are tied to the shift towards a low-carbon economy (Basel Committee on Banking Supervision, 2022). The amalgamation of these risks yields a substantial influence on our environment and infrastructure. Particularly, coastal areas are the most susceptible to increased exposure and risk, largely due to sea-level rise (Calabrese et al., 2022). This might lead to heightened housing demand in vulnerable regions like coastal areas prone to flooding (Hertin et al., 2003).

Furthermore, predictions suggest a significant surge in floods due to prolonged temperature rise, with the past three decades exhibiting the highest flood frequency in the last 500 years (Fatica et al., 2022). (World Bank, 2015) emphasizes that flooding is a primary cause of natural disasters in developing nations, causing displacement and considerable financial losses. Corroborated by (Regelink et al., 2017) and the Bank of England (BOE), flooding ranks among the major physical risks. Notably, Thailand incurred around US$45 billion in flood damages, with minimal insurance coverage (Scott, Van Huizen, & Jung, 2017). Intensified flood risks, combined with other extreme weather events, amplify the overall impact.

However, our existing measurements and comprehension of flood risk and other physical risks remain restricted, implying that the quantified risks represent only the visible part of the iceberg. Moreover, secondary effects could lead to property value declines once flood damages are quantified, causing investors to avoid high-risk areas (Bikakis, 2020). This situation would also inflate construction expenses, as builders adapt to changing climate conditions, potentially increasing mortgage payments for homeowners (Hertin et al., 2003). The disruptive influence of water damage, including floods and sea-level rise, on firm performance is also well-documented (Fatica et al., 2022).

(Feyen et al., 2020) present a model projecting the long-term consequences of climate change. In a scenario where global temperature increase surpasses 3°C by 2100, Europe could witness half a million flood-exposed individuals annually, triple the current number. River flood losses could surge sixfold, reaching EUR 50 billion/year. Coastal flooding might double, leading to EUR 250 billion/year in losses by 2100, affecting 2.2 million people yearly (Fatica et al., 2022). Notably, the United Kingdom, Spain, and Romania witness such frequent floods that they are more commonplace than exceptional, with the interval between events decreasing (Fatica et al., 2022).

Evidence points to climate change impacting interest rates on new loans. Instances of climate-related disasters influencing firms' funding costs and debt-raising capability have been observed (Pörtner et al., 2022). Both the immediate and long-term physical effects of climate change are expected to have disruptive consequences on firm operations, with implications for climate adaptation policies and access to finance for smaller firms. Notably, the occurrence of a flood within a year before loan origination correlates with an average 0.134% increase in loan interest rates (Barbaglia et al., 2022).

2.2) Financial Risks

In recent years, a surge of attention has been directed toward the financial implications of climate change, a topic that central banks, international organizations, and researchers are now delving into. This surge has seen entities like the G20, the United Nations Environment Programme, the Financial Stability Board, and central banks from Australia, Britain, France, Italy, and the Netherlands issue alarms about the potentially destabilizing consequences (Regelink et al., 2017).

In 2015, Mark Carney, Chairman of the Bank of England, and the Financial Stability Board (FSB), drew attention to the peril climate change poses to global financial stability in his speech on "breaking the tragedy of the horizon" (Carney, 2015). Significant strides have been made by major institutions in grappling with this challenge, although the literature exploring the financial impacts related to credit risk remains in its early stages (Calabrese et al., 2021). The Task Force on Climate-related Financial Disclosures (TCFD), initiated by the FSB, has aimed to provide recommendations for disclosures that enhance the understanding of climate risks for financial market participants (FSB-TCFD, 2017). Additionally, the Network of Central Banks and Supervisors for Greening the Financial System (NGFS) has placed a core objective on the development of environment and climate risk management in the financial sector (NGFS, 2017). A recent report by the US Federal Reserve highlighted the potential escalation of financial shocks and instability due to climate change (Board of Governors of the Federal Reserve System, 2020). The Basel Committee on Banking Supervision has also recently unveiled 18 principles for effective climate-related financial risk management and supervision, underlining the importance of accounting for unique risk characteristics (principle 8) and the need for robust data and metrics (principle 3) (Calabrese et al., 2021).

(Dietz et al., 2016) underscores that financial asset losses can stem from an array of sources. Climate change can amplify the occurrence of natural disasters, leading to property and infrastructure damage. Furthermore, it can disrupt supply chains, creating shortages of goods and services. The risks posed by climate change can be categorized into two primary types. Firstly, direct asset damage can lead to the devaluation of collateral and substantial reevaluation of loans and securities for institutions operating in high-risk regions such as coastal or flood-prone areas. Secondly, these physical risks can trigger disruptions in supply chains and result in asset loss or damage, with a concentration of economic activities intensifying the impact of local events. This could constrain credit provision in high-risk areas, particularly affecting lower-income borrowers and less-capitalized banks (Cortés and Strahan, 2017; Faiella and Natoli, 2018).

Catastrophe risks can have a negative impact on a country's public finances and debt sustainability. This is because disasters can lead to higher fiscal costs, such as increased social assistance expenditures and relief payments, as well as lower tax revenues. Additionally, disasters can require investment in adaptation and risk mitigation, and can result in direct losses on government assets. All these factors can affect a country's credit quality and debt financing rates (Zenios, 2022).

Many investigations into the economic and financial implications of climate change commence with underlying assumptions about forthcoming emissions, the scope of warming, and other dimensions of climate risk, such as shifts in sea levels and alterations in precipitation patterns. In this context, (Tol, 2009) asserts that "Cheap energy cannot coexist without carbon dioxide emissions." As of the year 2000, the global emission of 24 billion metric tons of carbon dioxide (tCO_2) raised an intriguing prospect. If these emissions were assigned a price equivalent to the €15/tCO_2 rate that prevailed in the European Union's Emissions Trading System in January 2009, the cumulative value of carbon dioxide would amount to approximately 1.5 percent of the world's income. Yet, this calculation navigates through a sea of uncertainties intrinsic to climate change, rendering traditional tools for decision-making and learning potentially inapplicable, as pointed out by (Tol, 2009).

A more recent study by (Gagliardi et al., 2022) delves into the consequential relationship between climate change and a country's borrowing capacity. By simulating fiscal disruptions precipitated by natural calamities across 13 EU nations, Gagliardi's analysis unveils an intriguing insight. Projecting the debt-to-GDP ratios in the context of two distinct global warming scenarios—1.5°C and 2°C—reveals a noteworthy trend. The average debt-to-GDP ratio is predicted to climb by approximately 2.3 and 2.7 percentage points by the year 2032 for these respective warming scenarios. Such findings underscore the intricate interplay between climate-driven challenges and a country's fiscal resilience.

In the broader discourse on climate change and its multifaceted repercussions, these perspectives shed light on the complex dynamics that intertwine with economic and financial facets. As these insights emerge, they beckon further exploration into the adaptive strategies that nations must embrace in the face of climate-driven transformations. These impacts and a combination of other risks posed by climate change are likely to lead to lower economic growth, likely to increase poverty and are also likely to lead to more frequent and severe natural disasters (Hertin et al., 2003) finds that the cost of building a new home in the UK could increase by up to 10% by 2050 due to climate change. This is because the need to use more expensive materials and construction methods to make homes more resilient to extreme weather events, such as storms and floods, will increase. The risk of damage to homes from these events could also increase by up to 50% by 2050.

Other studies have found that the economic impacts of climate change could be significant. The Stern Review estimated that the economic impacts of climate change could range from $2 trillion to $28 trillion per year by the end of the 21st century. The costs of inaction on climate change could be even higher, with some estimates suggesting that they could reach $200 trillion per year by the end of the 21st century.

The study by (Tol, 2009) also found that the costs of adaptation to climate change could be significant, ranging from $100 billion to $1 trillion per year by the end of the 21st century. The study finds that the potential losses to financial assets are not evenly distributed. Some sectors, such as the insurance sector, are more vulnerable to the impacts of climate change than others. The study also found that the impacts of climate change are likely to be unevenly distributed, with developing countries being more vulnerable than developed countries.

These studies suggest that climate change is likely to have a significant impact on the UK economy. The cost of building new homes could increase, and the risk of damage to homes from extreme weather events could also increase. The costs of inaction on climate change could also be high, and developing countries are likely to be more vulnerable than developed countries.

2.3) Credit Scoring

Credit scoring is a method that is used to predict the likelihood of a loan applicant or existing borrower defaulting on their loan (Abdou & Pointon, 2011). It is used by lenders to assess the risk of lending money to someone. The approach, originating in the 1950s, has gained significant traction in consumer lending, particularly for credit cards. Furthermore, its application is progressively expanding in the realm of mortgage lending. While its adoption in business lending has historically been limited, the landscape is evolving, witnessing a notable shift (Mester, 1997).

The scoring method involves analysing historical data on the performance of previously made loans to determine which borrower characteristics are most predictive of default (Hand & Jacka, 1998; Thomas et al., 2002). These characteristics can include things like income, debt, employment history, and credit history. The information is then used to create a scorecard, which assigns a numerical value to each borrower. This score is used to rank borrowers in terms of risk, with higher scores indicating lower risk. Lenders typically set a cut-off score, below which they usually do not approve a loan.

Credit scoring is a valuable tool for lenders, as it helps them to make more informed lending decisions. It can also help borrowers to understand their own creditworthiness and to improve their chances of being approved for a loan (Abdou & Pointon, 2011). (Al Amari, 2002) has argued that while a lot of credit scoring models have been used in the field, the following key questions have not yet been answered conclusively: What is the optimal method to evaluate customers? What variables should a credit analyst include to assess their applications? What kind of information is needed to improve and facilitate the decision-making process? What is the best measure to predict the loan quality (whether a customer will default or not)? To what extent can a customer be classified as good or bad?

2.4) Delinquency/Default

“Taking a loan 12 months after a flood correlates with an increase of 0.134% interest on the loan” (Barbaglia et al., 2022).

Delinquency and default are two terms that are often used interchangeably, but they have different meanings. Delinquency arises when a borrower fails to fulfil a scheduled loan payment. As loan payments are typically required monthly, the lending industry conventionally categorizes delinquent loans based on the time they have remained unpaid, usually as 30, 60, 90, or 120 days overdue, depending on the duration of the oldest outstanding payment.

Default, on the other hand, arises when a borrower's missed payments reach a certain threshold, prompting the lender to initiate debt collection actions. It's noteworthy that the count of borrowers falling into delinquency surpasses the number of actual defaults. This is due to the transient nature of delinquency, which can sometimes be rectified as borrowers improve their financial situation.

Research by (Avery et al., 1996) discovered that the likelihood of default is higher among borrowers with lower credit scores and those who have recently encountered adverse credit events. This study employed data from the Home Mortgage Disclosure Act (HMDA) to observe mortgage performance. It also revealed that borrowers who have recently experienced negative credit events are approximately three times more likely to default compared to those without such experiences. These findings underscore the importance of lenders carefully evaluating borrowers' credit scores and history in lending decisions. Individuals with low credit scores or recent negative credit events are more prone to default, necessitating lenders to account for this possibility. (Avery et al., 1996) Furthermore, higher credit scores typically correlate with borrowers who possess stronger financial literacy (e.g., Duca and Kumar, 2014; Bajo and Barbi, 2018).

Default represents a substantial financial burden for lenders, entailing future and probable interest losses, along with foreclosure expenses. As a countermeasure, lenders often incorporate risk premiums or limit credit access to only the most creditworthy applicants. The ability to distinguish between borrowers likely to perform well on their loans and those with a higher likelihood of default enables lenders to offer mortgages at prices that accurately reflect underlying risks. However, default has far-reaching consequences for borrowers and society. It leads to diminished credit ratings and reduced credit access for borrowers, asset losses, and the costs associated with relocating. When concentrated in specific areas, defaults can trigger significant societal effects by reducing property values, discouraging investments in affected neighbourhoods, increasing lending risks, and decreasing credit availability (Avery et al., 1996).

A recent report from the Board of Governors of the Federal Reserve System (2020) highlights the insufficient geographical granularity and time horizons in the climate risk pricing models employed by many banks. The uncertainty surrounding the timeline of climate risk realization also poses a challenge in this context (Barnett, Brock, and Hansen, 2020).

(Cortés and Strahan, 2017) delve into the response of US banks to natural disaster-induced shocks in local mortgage credit demand. Their research reveals that small banks increase credit supply in affected areas but reduce it in other markets where they have lending exposure. Conversely, large banks do not adjust lending in connected markets, possibly due to lower external finance costs. (Koetter et al., 2020) identify interest rates and local economic conditions as the key determinants of loan default. (Barbaglia et al., 2022) contribute to this discussion.

A recent investigation conducted by (Bernstein, Gustafson, and Lewis, 2019) reveals that lending institutions impose higher interest rate spreads on mortgage applications for properties that face greater risk from rising sea levels. This research, utilizing data from the Home Mortgage Disclosure Act (HMDA), uncovers an approximately 7.5 basis points (bps) elevated interest rate spread for mortgages in zip codes characterized by heightened vulnerability to sea level rise. This uptick translates to nearly $9,000 in added financing expenses for the typical borrower. Importantly, this premium related to sea level rise risk isn't solely connected to immediate flooding concerns; it is also applicable to counties that never encounter significant river or coastal floods. The findings suggest that financial establishments might not be fully integrating sea level rise risk into their mortgage pricing strategies, and the exposure to such risk significantly influences how banks factor in this atypical risk when determining pricing (Nguyen et al., 2022).

Despite these difficulties, numerous banks persist in relying on traditional backward-looking models that are built on past losses and exposures. However, these models might not encompass the ever-changing nature of risks posed by climate change. Furthermore, due to the array of risks that banks currently grapple with (such as cybersecurity and geopolitical risks) and the long-term nature of climate change, the consideration of climate risk may not consistently rank as a top priority (Nyberg and Wright, 2015).

2.5) Credit Scoring Models

Credit scoring models are constructed through various methodologies, encompassing linear probability models, logit models, probit models, and discriminant analysis models. These approaches differ in their mechanisms for estimating the probability of default (Mester, 1997).

There is no single "best" statistical technique for credit scoring, as the choice of technique depends on the specific application. However, the most used statistical technique for credit scoring is logistic regression. This technique is effective in predicting loan default, and it is often used in commercial loans (Abdou & Pointon, 2011). An emerging statistical technique in credit scoring is the utilization of neural networks. These artificial intelligence algorithms possess the ability to learn from experience, enabling them to uncover the intricate connections between borrower characteristics and the likelihood of default. Moreover, neural networks facilitate the identification of the most pivotal borrower attributes in predicting default probabilities (Mester, 1997).

The evaluation criteria that are used to assess the performance of credit scoring models include accuracy, discrimination, and calibration. Accuracy is the most important evaluation criterion, followed by discrimination and calibration (Abdou & Pointon, 2011). Credit scoring can be effective in predicting loan default. For example, one study found that a logistic regression model was able to predict loan default with an accuracy of 80%. Another study found that a neural network model was able to predict loan default with an accuracy of 90% (Abdou & Pointon, 2011).

Utilizing credit scoring has the potential to decrease the expenses associated with the credit approval process and the anticipated risk linked with granting a risky loan. Additionally, it can improve the credit evaluation process, while also saving time and effort (Lee et al., 2002; Ong et al., 2005). When constructing a credit scoring model, it is essential to account for the correlations between the variables. It's plausible that a few of the variables initially taken into account might not be included in the final model, as they might not contribute significantly. According to Fair, Isaac and Company, Inc., a prominent developer of scoring models, around 50 to 60 variables could be contemplated during the creation of a typical model, yet only eight to twelve might ultimately be integrated into the final scorecard due to their ability to yield the most predictive amalgamation (Mester, 1997).

2.6) Credit Scoring Models' Limitations

Credit scoring models are not always accurate. In a survey of banks that used credit scoring in their credit card operations, 56% of the banks reported that their models failed to accurately predict loan-quality problems. This was partly due to a new willingness by consumers to declare bankruptcy (Avery et al., 1996).

Credit scores are based on historical data, so they may not be as accurate for borrowers who have recently experienced a change in their financial circumstances. For example, a borrower who recently lost their job may have a lower credit score than they would have if they were still employed. Credit scores can also be affected by factors that are beyond the borrower's control, such as job loss or natural disasters. A borrower who lost their home in a hurricane may have a lower credit score than they would have if they had not experienced the hurricane. Credit scores are found to be more predictive of default for certain types of borrowers, such as those with low credit scores or those who have recently experienced a negative credit event. (Mester, 1997)

2.7) The case of the United States

By the end of the century, the United States is projected to experience a significant rise in average temperatures, with estimates ranging from about 3°F to 12°F, contingent on emissions scenarios and climate models. The annual average temperature in the U.S. has already seen an uptick of around 0.6°C, accompanied by a nationwide increase in precipitation by approximately 5-10%. Notably, research indicates that warming trends in the 21st century are poised to surpass those of the 20th century. Future scenarios underscore the possibility of a temperature surge of 3-5°C within the next century, surpassing the projected global average. The most substantial temperature escalation is predicted for the Southern U.S., signifying a faster temperature increase over the next century than witnessed in the previous 10,000 years ("Future of Climate Change | Climate Change Science | US EPA," n.d.).

The mounting global temperatures are poised to usher in more frequent and intense heatwaves. The prevalence of days with temperatures exceeding 30°C is projected to rise significantly across the United States by the close of this century. Climate models forecast that temperatures ranking among the hottest 5% in the period of 1950-1980 will encompass at least 70% of the time during 2035-2064 ("Future of Climate Change | Climate Change Science | US EPA," n.d.).

Furthermore, average sea levels are anticipated to experience a rise of 10-12 inches over the upcoming three decades. With the average U.S. temperature in 2017 already surpassing historical records by 2.6 degrees Fahrenheit, the effects of climate change are evident. Hurricanes are projected to impose an additional $7.3 billion annual cost on the economy (Climate Change Predictions, 2017). These alarming trends underscore the pressing need for comprehensive climate action and adaptation strategies to mitigate potential economic and social impacts.

Rainfall

Heavy precipitation events will likely be more frequent, even in areas where total precipitation is projected to decrease. Presently, heavy downpours that manifest approximately once every two decades are projected to become more frequent, occurring between two and five times more often by the year 2100, contingent on specific geographical locations.

An observable shift is expected in the proportion of precipitation that falls as rain rather than snow, except in the far northern areas. The trajectory of cold-season storm tracks is anticipated to continue moving northward, a trend that aligns with the projected intensification of the most robust cold-season storms. Over the past half-century, precipitation has shown an average increase of around 5 percent. Climate models underline the likelihood of wetter conditions in northern regions ("Future of Climate Change | Climate Change Science | US EPA," n.d.).

According to climate projections, heavy downpours that currently have a 1-in-20-year recurrence rate are foreseen to take place approximately every 4 to 15 years by the close of this century, contingent on geographic location. Moreover, the intensity of these 1-in-20-year heavy downpours is predicted to increase by 10 to 25 percent by the century's end, a finding echoed by the report "Climate Change Impacts on the United States - Overview Report" (Cambridge University Press, 2000). These changes in precipitation patterns underscore the imperative of adopting robust climate adaptation strategies and policies to mitigate potential risks and impacts on various sectors.

Hurricanes

The future holds the potential for more frequent heavy precipitation events, even in regions where overall precipitation levels are anticipated to decrease. The warming of the oceans is expected to contribute to the heightened intensity of Atlantic hurricanes. Climate models are indicative of a surge in the occurrence of the most potent hurricanes, namely those falling under Categories 4 and 5. Additionally, these models predict an escalation in rainfall rates associated with hurricanes ("Climate Change Predictions," 2017).

Historical data spanning from 1881 to 2008 reveals alternating periods of above-average hurricane activity, notably during the 1800s, mid-1900s, and more recently, since 1995. The strength and frequency of Atlantic hurricanes have exhibited a significant increase in recent decades. Although the total number of hurricanes making landfall has seen relatively little change, this phenomenon can be attributed to a complex interplay of factors influencing landfall, including overarching steering winds, atmospheric stability, and wind shear, among others. This insight is gleaned from the comprehensive report "Climate Change Impacts on the United States - Overview Report" (Cambridge University Press, 2000), underscoring the need for proactive and adaptive measures in the face of changing hurricane patterns.

Hurricanes - Sea level rise

The escalation of sea surface temperatures exerts a direct influence on the intensity of tropical storms, enhancing their wind speeds and augmenting their potential for inflicting significant damage upon landfall. Analyses spanning a 39-year period from 1979 to 2017 have shown an increase in major hurricanes alongside a decrease in smaller ones (Kossin et al., 2020). According to projections based on sophisticated modelling, the National Oceanic and Atmospheric Administration ("Climate Change Predictions," 2017) anticipates a rise in Category 4 and 5 hurricanes, coupled with heightened wind speeds. Elevated sea temperatures also contribute to the amplification of hurricane precipitation, with a projected increase of 10-15 percent in storm-related rainfall. Recent instances, such as Hurricane Harvey in 2017, Hurricane Florence in 2018, and Hurricane Imelda in 2019, underscore the potential for extreme rainfall-induced flooding.

The phenomenon of sea level rise further exacerbates the impact of coastal storms, intensifying their destructiveness. A global rise in average sea level by over half a foot since 1900 and a projected increase of 1 to 2.5 feet within the current century (IPCC, 2021) is a testament to this trend. Coastal regions are particularly susceptible to these effects. Research has demonstrated that higher sea levels led to flood elevations during Hurricane Katrina that were 15-60 percent higher than conditions in 1900 (Irish et al., 2013). Similar findings were observed with Hurricane Sandy, where elevated sea levels significantly augmented flooding likelihood, with additional sea level rise projected to intensify this risk even further (Lin et al., 2016). Furthermore, changes in the atmospheric landscape, including the warming of the Arctic, potentially contribute to additional trends observed in hurricane patterns. A noteworthy alteration is the reduction in the speed of hurricanes compared to historical patterns. Although the precise mechanism behind this slowdown remains subject to debate, the consequences are evident as storms tend to "stall," leading to extended periods of high winds, storm surges, and prolonged heavy rainfall in coastal regions. This phenomenon has been a significant factor in escalating the devastation caused by recent storms across the United States (Center for Climate and Energy Solutions, 2018). The interplay of these factors underscores the imperative of understanding and addressing the intricate dynamics of climate change-induced shifts in hurricane behaviour.

2.8) Gaps

The financial impact of climate change is a complex and rapidly evolving issue. There are a number of reasons why the literature on the topic is limited. First, climate change is a relatively new phenomenon, and it has only been in recent years that we have begun to understand the full extent of its financial risks. Second, the financial impact of climate change is highly complex, and it is difficult to isolate the impact of climate change from other factors that can affect the mortgage market.

Only some research that has begun to explore the financial impact of climate change on mortgages. For example, a study by (Bernstein, Gustafson, and Lewis, 2019) found that lenders charge higher interest rates on mortgages for properties that are exposed to greater sea level rise risk. Another study by (Baldauf, Garlappi, and Yannelis, 2020) found that the value of homes that are exposed to climate change risk is declining. These studies suggest that climate change is likely to have a significant impact on the mortgage market. (Dietz, 2016)

There is a growing body of literature on the topic, but there are still significant gaps. One of the key areas where further research is needed is the impact of climate change on mortgage default rates. More research is needed to understand how climate change will affect the likelihood that borrowers will default on their mortgages. This dissertation aims to fill this very gap by exclusively focusing on how climate change effects mortgage repayment.

3) Methodology

3.1) Development of Hypotheses

In examining the potential relationship between significant weather events and mortgage loan repayment, the study's hypotheses are rooted in empirical concerns over climate change and its potential implications on mortgage repayments. These concerns postulate that severe climatic events can disrupt financial ecosystems, including an individual capacity to fulfil financial obligations.

3.1.1) Main Hypothesis

Null Hypothesis (H_0): *Weather events, specifically hurricanes and cyclones, significantly affect the mortgage loan repayment status.*

Alternate Hypothesis (H_1): *Weather events, specifically hurricanes and cyclones, have no significant impact on the mortgage loan repayment status.*

3.1.2) Sub-hypothesis

Upon confirming the main hypothesis, a secondary line of inquiry comes to the forefront. This sub-hypothesis is instrumental in refining the research focus.

Impact Analysis of Specific Weather Events (Hs): *Which specific weather event exerts the most pronounced influence on mortgage loan defaults?*

3.2) Research Design

The current research adopts a correlational design, aiming to discern the relationship between severe weather events and mortgage loan repayment behaviours. This design type is chosen because it allows for the establishment of relationships between variables without manipulating any of the research variables.

3.2.1) Variables and confounding factors

A crucial aspect of this research design is the selection and understanding of the variables involved. The dataset comprises several variables, with 'Weather_Events' as the primary independent variable of interest. It's segmented into events like hurricanes (categorized above and below category 3) and cyclones, all geocoded by zip codes.

Financial factors encapsulated in variables like 'ORIG_RATE', 'ORIG_UPB', 'OLTV', and 'DTI' are fundamental as they shed light on the borrower's financial landscape, influencing the probability of loan repayment. The credit scores, 'CSCORE_B' and 'CSCORE_C', obtained from the credit rating bureau offer a perspective on the borrower's creditworthiness.

While 'Weather_Events' is the study's focal point, it's paramount to control for other variables that could potentially sway mortgage repayment behaviours. 'ORIG_RATE' or 'CSCORE_B', for instance, might independently influence loan repayment. Their inclusion ensures that the discerned relationship between weather events and loan repayment isn't confounded by these variables.

3.2.2) Models – Main Hypothesis

The modelling phase forms the backbone of the analysis. Four models are constructed, but there are only two primary models, each model has a Random forest version and a corresponding one for logistic regression. The first model incorporates 'Weather_Events', while the other does not. The idea is to discern the importance of weather events in predicting loan defaults.

Model 1A – Logistic Regression:

ln(1−P(Y=1)P(Y=1))= β 0+ β1 ORIG_RATE+ β2 ORIG_UPB+ β3 OLTV+ β4 DTI+ β5CSCORE_B+ β6 CSCORE_C+ β7 FIRST_FLAG+ β8 WEATHER_EVENTS+ ϵ

Dependent Variable DLQ_STATUS [1−P(Y=1) P(Y=1)] - represents the probability of mortgage loan repayment status being 1.

β0, β1, β2, β3, …, β_8 - correspond to the coefficients of the independent variables ORIG_RATE+ORIG_UPB+… +XWEATHER_EVENTS.

ϵ - Error term accounting for unobserved factors that contribute to the variation in the dependent variable.

The application of logistic regression within financial risk assessment and default prediction is firmly substantiated by contemporary literature. Notable studies, such as (Rajagopalan and Ray, 2018) and (Chen and Kou, 2016), underscore its relevance in credit risk prediction and financial distress evaluation, validating its employment for investigating mortgage loan default dynamics. This methodological choice leverages logistic regression's robustness in uncovering intricate relationships within complex datasets, enabling nuanced insights into the convergence of financial, credit-related, and meteorological factors affecting loan defaults.

The chosen approach uses logistic regression due to its efficacy in estimating probabilities in binary classification scenarios. In this context, it serves as the ideal tool to predict binary outcomes, loan defaults. The method hinges on the principle of modelling the log odds of an event occurring, which involves the selected predictor variables.

The dependent variable 'DLQ_STATUS' is regressed against the predictor variables. To ensure the model's reliability, the dataset is divided into distinct training (75%) and testing subsets (25%). This division facilitates rigorous assessment of the model's performance and predictive accuracy. Importantly, the logistic regression framework accommodates nonlinear relationships and furnishes interpretable coefficients, which is pivotal in deciphering the intricate interplay of various determinants that contribute to loan defaults.

Model 1B – Random Forest:

The Random Forest model serves as the primary model in this research for exploring relationships between predictors, including 'WEATHER_EVENTS,' and loan default. This ensemble learning technique creates multiple decision trees during training, and the mode or mean of the individual trees' predictions determines the output class or prediction for classification or regression tasks, respectively (Breiman, 2001). The model’s flexibility captures complex interactions and non-linear relationships in data, giving it an advantage over Logistic Regression.

The Random Forest model's performance is enhanced through the utilization of the Synthetic Minority Over-sampling Technique (SMOTE) and Randomized Grid Search with cross-validation. SMOTE addresses the imbalance of non-default and default loans by creating synthetic instances of the minority class (Chawla et al., 2002). Randomized Grid Search, along with cross-validation, fine-tunes the Random Forest model's hyperparameters, leading to improved performance (Pedregosa et al., 2011). Cross-validation partitions the dataset, allowing more accurate performance estimates (Agresti, 2018).

Model 2A – Logistic Regression

In(1−P(Y=1)P(Y=1))= β 0 + β1 ORIG_RATE + β2 ORIG_UPB + β3 OLTV + β4 DTI + β5 CSCORE_B + β6 CSCORE_C + β7 FIRST_FLAG + ϵ

The second logistic regression model uses a similar approach used in the first one. The only difference here is the model does not incorporate the control variable “WEATHER_EVENTS”

Model 2B – Random Forest

The second random forest model uses a similar approach used in the primary model. The only difference again is the model does not incorporate the control variable "WEATHER_EVENTS"

After analysing the results (shown later), the null hypothesis holds, which paves the path for the Sub Hypothesis elaborated below.

3.2.3) Models - Sub Hypothesis

Upon the main hypotheses' evaluation, a sub-hypothesis emerges: discerning which specific weather event has the highest impact on loan defaults. To study the specific effects of Hurricanes and Cyclones, the "WEATHER_EVENTS" column was split into three distinct columns. These columns indicate if a particular zip code was impacted by a particular event or not (one each for hurricanes greater than category 3, 4 and 5, hurricanes categories 1 and 2, and cyclones), if the zip code corresponding to the event has been impacted it will be indicated by 1 and 0 if not.

Model 3A – Logistic Regression

$$In(1-P(Y=1)P(Y=1)) = \beta\ 0 + \beta 1 \text{ ORIG\_RATE} + \beta 2 \text{ ORIG\_UPB} + \beta 3 \text{ OLTV} + \beta 4 \text{ DTI} + \beta 5 \text{ CSCORE\_B} + \beta 6 \text{ CSCORE\_C} + \beta 7 \text{ FIRST\_FLAG} + \beta 8 \text{ HURRICANES} < 3 + \beta 9 \text{ HURRICANES} > 3 + \beta 10 \text{ CYCLONES} + \epsilon$$

This model was built for the sole purpose of recognising the significance of the 3 weather events mentioned. The model follows a similar approach taken in building the Main Hypothesis model. The only exception here is that Weather events has been added in three separate columns, namely, Hurricane>3, Hurricane<3 and Cyclones.

Model 3B – Random Forest

-

This model again follows a similar approach taken in building the primary model. The exception again here is that Weather events has been added in three separate columns, namely, Hurricane>3, Hurricane<3 and Cyclones. The model's intent was to ascertain the singular impact these weather events have on loan default.

Model 3C – Neural network Model (Including SHAP)

SHAP is used to highlight the main part of the sub hypothesis, the sole purpose of this model is to highlight the importance and significance of different weather events in predicting loan defaults.

The adoption of a neural network (NN) classification model in mortgage default prediction is substantiated by its capability to capture intricate patterns within complex datasets. As illuminated by (Rajagopalan and Ray, 2018) and (Chen and Kou, 2016), neural networks have demonstrated efficacy in credit risk assessment and financial distress prediction, thus underpinning their relevance in evaluating mortgage loan default occurrences. This methodological choice harnesses the neural network's ability to decipher nonlinear relationships, potentially revealing latent interactions among financial, credit, and meteorological determinants that influence loan defaults.

The proposed neural network architecture, implemented using TensorFlow, presents a sophisticated framework for modelling loan default behaviour. Through an iterative process, the model constructs layers of interconnected nodes, with each layer contributing to the extraction of hierarchical features. The initial layer, comprising 64 nodes, employs the ReLU activation function to facilitate nonlinearity and enhance information extraction from input features, which encompass loan attributes, borrower characteristics, and macroeconomic indicators. The subsequent output layer employs the sigmoid activation function, conducive to binary classification tasks, to yield a probability score indicating the likelihood of loan default.

For model training, the binary cross-entropy loss function is employed to optimize the model's parameters using the Adam optimizer. Training is enhanced by early stopping, which curbs overfitting by halting the training process once validation loss ceases to improve. The neural network is iteratively trained over 100 epochs, and its performance is evaluated against a validation dataset, ensuring its generalization capabilities.

Moreover, to unravel the neural network's intricate decision-making process, SHAP (SHapley Additive exPlanations) values are employed. SHAP values provide a comprehensive perspective on feature importance by attributing each feature's contribution to individual predictions. The KernelExplainer from the SHAP library accommodates neural networks and is utilized here to compute SHAP values. These values are then visualized using a summary plot to present a consolidated depiction of feature impact on predictions.

Incorporating a neural network with SHAP analysis enhances mortgage default prediction by amalgamating advanced learning capabilities with transparent interpretability. This multifaceted approach substantiates the model's predictions while providing insights into feature importance, a crucial aspect in guiding informed decision-making for stakeholders within the financial domain.

4) Data & Presentation

This study draws upon a comprehensive dataset to explore the intricate relationship between climate change and mortgage defaults. The primary data sources include Fannie Mae's Single-Family Historical Loan Performance Dataset, renowned for its extensive residential mortgage information, and the USGS' Flood Event data as well as the second-generation North Atlantic hurricane database from the National Hurricane Centre, which provide crucial climate data. Encompassing records spanning from 1851 to 2019, this database delivers a comprehensive repository of pertinent information, including event names, dates, hourly data, event types, geographical coordinates, and maximum wind speeds (measured in knots).

These rich sources of data collectively underpin the investigation into the financial implications of climate change, particularly in the context of mortgage defaults along the U.S. East Coast. The integration of mortgage performance data with climate event records constitutes the bedrock of this study's effort to dissect the interplay between climate events and mortgage default occurrences, contributing valuable insights to the evolving discourse on climate-related financial risks.

4.1) Fannie Mae Single-Family Historical Loan Performance Dataset

The Fannie Mae Mortgage Dataset serves as a crucial resource for understanding the credit performance of loans. This dataset is a subset of Fannie Mae's portfolio, comprising only 30-year and less, fully amortizing, full documentation, single-family, conventional fixed-rate mortgages.

The dataset chosen was the 2020 Q1 data which has all the loan level information leading up to 2020. The dataset contains the details of more than 20 million mortgages, however this total is brought down to roughly 1.3 million loans when filtering down to the 5 East coast states. The dataset also contains monthly updates too to track each loans performance as well as containing information regarding the original loan lenders.

We start with the set of mortgages that originated in 2015 up till Q1 of 2020. This avoids capturing the loans that defaulted due to the COVID pandemic, whilst keeping the research restricted to recent data. The dataset has a huge disparity in the number of loans that have defaulted and the number of loans that have not, the ratio of defaults to no defaults is less than 0.5%. To balance this disparity, the number of loans that have not defaulted is randomly sampled to include only 50,000 loans, this brings the new ratio of defaults to no defaults to approximately 3:4. This is further filtered by dropping “NAN” rows, and then excluding loans that are less than 2 months old as we aim to capture loans that have defaulted only more than 3 months (90+ days delinquent) as defaults. Then the 1% outliers are windsorized according to (Fox & Weisberg, S2019). This brings the unique loan count, down to approximately 74 thousand loans.

The original dataset had 107 columns, 14 of those have been retained and some of these columns have been binned to make the analysis uniform, the variables have been outlined in Table 1 below.

Variable	Definition	Binning Information	Reference Examples
LOAN_ID	An identifier for the mortgage loan.	-	
SELLER	The entity that delivered the loan to Fannie Mae.	-	-
ORIG_RATE	The original interest rate on the mortgage loan.	5 equal bins from 1-5 (Interest ranging from 2.88- 5.25)	(Calabrese et al., 2022)
ORIG_UPB	The dollar amount of the loan sanctioned.	6 equal bins from 1-6 (Sanctioned amount ranging from $65,000 - $650,000)	(Calabrese et al., 2022)
FIRST_PAYMENT	This column was created using two original columns – “ORIG_DATE” and “LOAN_AGE”. The “LOAN_AGE” column was subtracted from the “ORIG_DATE” column to determine when the first repayment was made/supposed to be made on the loan.	-	-
OLTV	Obtained by dividing the amount of the loan at origination by the value of the property.	9 bins from 1-9, values ranging from 9- 97.	(Calabrese et al., 2022)
DTI	Obtained by dividing the total monthly debt expense by the total monthly income of the borrower.	5 bins from 1-5, values ranging from 13 – 50.	(Agarwal et al., 2013)

CSCORE_B	A numerical value used by the financial services industry to evaluate the quality of borrower's credit. It is referring to the "Classic" FICO score developed by Fair Isaac Corporation.	9 equal bins from 1-9, values ranging from 631 - 817	(Calabrese et al., 2022)
CSCORE_C	A numerical value used by the financial services industry to evaluate the quality of Co-borrower's credit. It is referring to the "Classic" FICO score developed by Fair Isaac Corporation.	9 equal bins from 1-9, values ranging from 633 - 817	(Calabrese et al., 2022)
FIRST_FLAG	First time home buyer or not	2 bins. 0 – Not a first time buyer 1 – First time buyer	(Thomas et al., 2002)
STATE	A two-letter abbreviation indicating the state within which the property securing the loan is located.	FL, SC, NC, NY and NJ.	-
ZIP	Limited to the first three digits of the code designated by the U.S. Postal Service where the subject property is located.	-	-
DLQ_STATUS	The number of months the obligor is delinquent.	2 bins. 0 – did not default 1 – has defaulted.	(Calabrese et al., 2022)

Table 1: Variable Description and Critical Information

Variable description is directly obtained from Fannie Mae's Website

4.2) Weather Events

For this study, data was first researched on the USGS Flood event viewer (FEV) and the National Oceanic and Atmospheric Administration's National Hurricane Centre – Tropical Cyclone Reports. Five weather events which have impacted these 5 states (separately), spanning across 2015 to 2019 were chosen for this analysis. These events range from Hurricanes of Category 1 to 5 and also one Nor'easter as well to understand the varied impact these events cause. They are outlined in Table 2 below:

Table 2: Weather events across the time series distribution

No	Name	Category	Year	States	Description
1	**Dorian**	Hurricane : (Category 1 /Category 3)	Aug-Sept 19	FL, NC, SC	Dorian made landfall in the US as a category 1 hurricane, but was captured as a category 3 hurricane only in parts of South Carolina
2	**Michael**	Hurricane : Category 5	Oct-18	FL	Michael made landfall in Florida as a category 5 hurricane, having directly cased damage worth $25 billion.
3	**Florence**	Hurricane : Category 1	Sept-Oct 18	NC, SC	Florence made landfall in the Carolinas as an upper end category 1 hurricane causing significant flooding.
5	**Irma**	Hurricane : (Category 3/ Category 4)	Sep-17	FL, SC	Irma made landfall in Florida as a category 3 hurricane but was also recorded as a category 4 hurricane at a certain point in Florida. Irma had little impact in South Carolina, but caused massive losses in the areas affected.
6	**Nor'easter**	Cyclone/Nor'easter	Jan-16	NY, NJ	This blizzard affected nearly a 100 million US residents dumping 2- 3 foot snow across regions and caused significant coastal flooding.

These weather events have impacted other states too, but only information regarding the 5 states has been captured and shown in this table. The events are not categorised by state but by ZIP code, this table is only for representative purposes.

4.3) Climate Data

To incorporate the impacts of weather events, particularly the influence of climate change, on the U.S. East Coast, data from the National Oceanic and Atmospheric Administration's National Hurricane Centre is used. This centre serves as a vital resource for tracking and forecasting hurricanes and other severe weather phenomena, offering essential information to help communities prepare and respond effectively.

The platform provides comprehensive information about hurricanes, including real-time updates on their locations, trajectories, intensities, and wind speeds. The data gathered and disseminated by the National Hurricane Centre is multidimensional and encompasses a wide range of factors, including atmospheric pressure, sea surface temperature, moisture content, and wind patterns. By integrating these data points with additional spatial datasets, the data paints a comprehensive picture of particular weather events, detailing the spatial extent, magnitude, and temporal attributes of these events in the five digit zip code level. However, at the five-digit ZIP code level, it does not provide information on wind speed. The estimation of wind speed is based on the work of (Willoughby et al., 2006) and is operationalised through the storm wind model R package. This approach allows for the calculation of maximum sustained winds (measured in knots) for individual five-digit ZIP codes, using the trajectory of the tropical cyclone

This data is geocoded in the three digit zip code level as the mortgage data comes with the three digit zip code. The zip codes are obtained from the FEV are aligned with mortgage zip codes. This data is further categorised into a separate column known as weather event where: no events (in the absence of tropical cyclones or wind speeds lower than 35 knots), tropical storm (35–63 knots), hurricane category 1 (64–82 knots), hurricane category 2 (83–95 knots), hurricane category 3+ (greater than 95 knots) (Calabrese et al., 2022).

The occurrence of weather events and their potential impact on loan repayment behaviors is a key focus of this study. Approximately 32% of the loans in the dataset were associated with at least one weather event, providing a significant subset of observations for studying the potential impact of such events on loan repayment behaviors. These weather events are further categorized based on their severity, with about 20% of loans experiencing hurricanes of category 3 or below, and 12% experiencing hurricanes above category 3.

5) Results

5.1) Descriptive Statistics

This study meticulously analyzes a dataset comprised of 74,165 mortgage loan observations originating from five pivotal states along the U.S. East Coast: Florida, South Carolina, North Carolina, New York, and New Jersey. These states were selected based on a strategic framework that emphasized their distinct climatic characteristics and varied exposure to extreme weather events. While each state possesses its own unique climatic profile—shaped by factors such as geographical positioning and prevailing weather patterns—they also experience different intensities and frequencies of these events. For instance, while Florida's peninsular structure makes it a frequent target of hurricanes, New York's northern disposition exposes it to a contrasting set of climatic challenges. Historical data has indicated a direct correlation between regions with pronounced weather event exposure and increased mortgage default rates. By concentrating on these five states, representing a broad spectrum of exposure levels, this study seeks not only to corroborate previous findings but to delve deeper into the intricate interplay between regional environmental conditions and mortgage loan performance. The objective is to illuminate the multifaceted relationship between climatic dynamics and financial stability, providing valuable insights into the broader ramifications of climate change on economic structures.

These loans, issued by 21 different lenders, span a period from May 2015 to January 2020. This specific period was intentionally chosen for its significance and relevance. Firstly, it predates the onset of the COVID-19 pandemic—a crucial distinction, as the subsequent months witnessed an unprecedented surge in loan defaults, largely attributable to the pandemic's economic repercussions, rather than environmental factors. By focusing on the years prior to 2020, this analysis ensures that the loan performance is predominantly influenced by climatic conditions and not overshadowed by the economic turmoil triggered by COVID-19. Moreover, the chosen half-decade period offers a comprehensive snapshot of recent loan behaviors, providing a sufficiently broad yet current dataset for an in-depth evaluation. The confluence of loans from various geographical locales, combined with the diverse lending institutions, further furnishes a rich tapestry for studying the interplay between environmental dynamics and financial practices.

Each observation in the dataset includes multiple variables that capture relevant information about the loan and the borrower. The average original loan balance (ORIG_UPB) across all observations is approximately $275,345.86, with a standard deviation of $117,833.31. Furthermore, the dataset includes variables related to the borrowers' creditworthiness, such as debt-to-income ratio (DTI) and credit scores. The average DTI across all observations is 35.63, with a standard deviation of 9.36, indicating variability in borrowers' indebtedness relative to their income. Credit scores, represented by CSCORE_B and CSCORE_C, show average values of 748.65 and 751.85 respectively. These scores suggest that, on average, the borrowers in this sample are reasonably creditworthy, as these values are comfortably above the 700 threshold, often used to indicate good credit (Sullivan, 2020).

The DLQ_STATUS variable delineates the delinquency status of each loan in the dataset. With an average value registering at 0.67, the data implies that the bulk of the loans within this sample are predominantly current on their payments. It's imperative to note that a loan is classified as delinquent if there's a payment delay surpassing 90 days (Mortgage Bankers Association, 2018). Central to this investigation is the assessment of weather hurricanes and cyclones—major representatives of climatic adversities under the WEATHER_EVENTS variable—bear a correlation with an augmented risk of loan default.

Such an exploration is not merely academic but resonates profoundly with prevailing concerns in the financial realm. Institutions are becoming increasingly cognizant of the looming threats that climate change, especially through catastrophic events like hurricanes and cyclones, poses to the established financial equilibrium. The Bank for International Settlements has notably accentuated the pertinence of evaluating the impacts of these climatic perturbations on financial stability in its recent discourses (Bolton et al., 2020). Hence, this study's endeavor to bridge the gap between climate events and financial risk comes at a pivotal juncture.

Table 3: Descriptive Statistics

Variable	Obs.	Mean	Std. Dev	Min	Max
LOAN_ID	74165	9.92	1.97	98801882	99170000
SELLER	74165	.	.	.	.
ORIG_RATE	74165	3.81	0.42	2.88	5.25
ORIG_UPB	74165	275345.86	117833.31	65000	650000
FIRST_PAYMENT	74165	.	.	2015-05	2020-01
OLTV	74165	73.61	17.05	9	97
FIRST_FLAG	74165	2.02	0.16	2	4
DTI	74165	35.63	9.36	13	50
STATE	74165	.	.	.	.
ZIP	74165	238.19	104.47	70	349
DLQ_STATUS	74165	0.67	2.52	0	32
CSCORE_B	74165	748.65	44.70	631	817
CSCORE_C	74165	751.85	43.80	633	817
HURRICANE <3	74165	0.20	0.28	0	1
HURRICANE >3	74165	0.12	0.15	0	1
CYCLONES	74165	0.12	0.33	0	1
WEATHER_EVENTS	74165	0.2	0.4	0	1

5.2) Correlation Analysis

The correlation matrix presented in the (Appendix) offers a comprehensive snapshot into the intricate web of relationships between the dataset's various components. It also sheds light on our primary research areas, specifically, the impacts of extreme weather events, such as hurricanes and cyclones, on loan default status (highlighted in columns 9 and 10).

From the matrix, a distinct yet moderate association between weather events and the new_DLQ_STATUS variable emerges, showcased by a correlation coefficient of 0.058. This correlation underscores the growing consensus in the literature about the potential link between climatic disturbances, especially hurricanes and cyclones, and spikes in loan default rates. Research by (Burke et al., 2015) establishes that severe climatic deviations can dramatically impact regional economic activities. These disturbances can disrupt local economies, putting downward pressure on individual income levels and, consequently, straining borrowers' ability to service their loans.

Moreover, the relatively minor negative correlations of weather events with original loan balance (ORIG_UPB, -0.028) and the number of borrowers (NUM_BO, -0.157) affirm that these specific financial metrics do not influence the occurrence of weather events. (Auffhammer, 2018) echoes this, suggesting that while climatic variables influence economic behaviors, they remain independent of several financial dynamics.

Delving deeper into the new_DLQ_STATUS variable, it reveals revealing patterns. Its positive correlations with primary financial predictors—like original loan-to-value ratio (OLTV, 0.164), original loan rate (ORIG_RATE, 0.159), and debt-to-income ratio (DTI, 0.260)—highlight the financial challenges many borrowers grapple with. In a globalized economic landscape (Gertler, Rose, and Svensson, 2019) have indicated that such financial pressures are intensifying, leading to increased default rates.

Interestingly, the data presents an intriguing narrative around credit scores. Despite their central role in measuring financial trustworthiness, they don't encapsulate the borrower's entire financial journey, especially when external factors come into play, a sentiment captured by (Keys et al., 2010).

Furthermore, the strong correlation between CSCORE_B and CSCORE_C is pivotal. As (Jappelli and Pagano, 2002) have posited, even though scores from different bureaus might exhibit similarities, their nuanced differences can be revealing. Such distinctions offer richer insights into a borrower's financial behavior and creditworthiness, emphasizing the necessity of multi-bureau credit analysis.

It's crucial to acknowledge the possibility of multicollinearity due to this high correlation. However, a variance inflation factor (VIF) less than 10 for all variables suggests that multicollinearity is not a concern in this study (O'Brien, 2007). A maximum VIF of less than 10 is generally considered acceptable and is indicative of a reasonable degree of multicollinearity (Kutner, Nachtsheim, & Neter, 2004). This ensures that the model's estimates will be unbiased and efficient, and that the statistical inference about the relationship between predictors and the response will be reliable.

5.3) Empirical Results

5.3.1) Main Hypothesis

Null Hypothesis (H_0): *Weather events, specifically hurricanes and cyclones, have no significant impact on the mortgage loan repayment status.*

Alternate Hypothesis (H_1): *Weather events, specifically hurricanes and cyclones, significantly affect the mortgage loan repayment status.*

Model 1A & 2A – Logistic Regression:

Table 4: Logistic Regression Analysis – Main Hypothesis

Variable	1A (Base Model + Weather)		2A (Base Model)	
	Co Efficients	Std Error	Co Efficients	Std Error
Constant	-1.142	0.069	-1.270	0.065
ORIG_RATE	0.137***	0.011	0.139***	0.011
ORIG_UPB	0.140***	0.006	0.139***	0.006
FIRST_FLAG	0.039*	0.042	0.039*	0.043
OLTV	0.116***	0.016	0.115***	0.017
DTI	0.136***	0.006	0.136***	0.006
CSCORE_B	-0.261***	0.012	-0.261***	0.012
CSCORE -C	-0.125***	0.012	-0.125***	0.012
Weather Events	0.227***	0.035	-	-

*Estimates for loan default probability, including weather events. For the parametric components, the coefficient estimates, the standard error, and the statistical significance are reported. P-value for the above variables included are : *** < 0.01,** < 0.05,* < 0.10. These results are from the out of sample.*

In the realm of financial risk assessment and default prediction, Logistic Regression stands as a formidable tool for unravelling intricate relationships between loan attributes and external factors. Logistic Regression's relevance echoes recent research by (Rajagopalan and Ray, 2018) and (Chen and Kou, 2016), underlining its importance in credit risk prognosis and financial distress evaluation.

Upon estimation of the model that includes weather events, the coefficients extracted from the logistic regression equation provide a lens through which the influence of each predictor on the likelihood of loan default becomes apparent. Interestingly, nearly all variables exhibit high levels of significance, marked by p-values < 0.01, signifying their substantial impact on loan default dynamics. The only exception is the credit score variables have a negative correlation to the independent variable.

Among the predictor variables, weather events emerge as a particularly noteworthy factor. With a statistically significant coefficient of 0.227 ($p < 0.01$), weather events substantiate their relevance in the context of loan default prediction. The positive sign of this coefficient indicates that the occurrence of adverse weather events amplifies the odds of loan default, emphasizing their potential influence on borrowers' repayment capabilities.

However, it's intriguing to note that 'FIRST_FLAG' exhibits a p-value of approximately 0.10, indicating a relatively weaker statistical significance. This observation does not diminish its potential relevance but rather encourages a more in-depth exploration of its impact on loan default dynamics. While not every variable reaches the same degree of statistical significance, their collective contribution shapes a comprehensive understanding of credit risk factors.

The application of the AUC-ROC metric substantiates the model's predictive efficacy. With an AUC-ROC value of 0.778, the model effectively distinguishes between loan default outcomes, underscoring its discriminatory power. In tandem, the Kolmogorov-Smirnov (KS) statistic of 0.436 adds further credence to the model's ability to differentiate between predicted default probabilities.

Now, we analyse the model which excludes the "WEATHER_EVENTS" variable, allowing us to deeply investigate other predictors' roles in loan defaults. The same methodology as before is at play, but the omission of the environmental factor uncovers the unique impact of other determinants.

The coefficient estimates echo our earlier narrative. Predictors like "ORIG_RATE," "ORIG_UPB," "FIRST_FLAG," "OLTV," "DTI," "CSCORE_B," and "CSCORE_C" maintain their significance in influencing loan default probabilities. Yet, there's an interesting pattern again. The coefficient for "FIRST_FLAG" doesn't hold strong statistical significance, aligning with previous observations. This reinforces the notion of its minimal impact on predicting loan defaults.

Model 2A's performance yields an AUC of 0.776 and a KS statistic of 0.401. While it effectively differentiates between default and non-default cases, this model experiences a slight reduction in predictive metrics when compared to its weather-inclusive counterpart. This discrepancy highlights "WEATHER_EVENTS'" vital role in bolstering predictive accuracy and cementing the importance of climate-related factors in credit risk assessment.

In summary, these Logistic Regression models with and without "WEATHER_EVENTS" reminds us of the pivotal role climate plays in credit risk assessment. The inclusion of weather events boosts predictive accuracy and highlights their significance in understanding loan default probabilities.

Model 1B & 2B – Random Forest:

In the context of this study, where the convergence of credit risk and climate risk takes center stage, the application of machine learning stands as the primary approach. This approach aims to dissect and comprehend the intricate relationships that underlie mortgage loan defaults. As the cornerstone of this effort, the Random Forest model emerges as the principal tool, adept at navigating the multidimensional landscape of credit risk while also considering the influence of environmental factors.

The model with weather events' interpretability is exemplified by the feature importance ranking, which is found using the numpy and standardscaler packages on python. This illuminates variables like DTI, ORIG_UPB, and CSCORE_B as influential predictors. Of particular significance is the inclusion of the 'Weather_Events' variable, which subtly reshapes the landscape of importance, highlighting the interplay between climatic influences and credit risk.

The model's predictive efficacy is quantified through metrics such as AUC-ROC and accuracy scores. With an AUC-ROC value of 0.855 and a KS statistic of 0.531, the Random Forest model excels in distinguishing between default and non-default outcomes (James et al., 2013). The minimal gap between training (0.893) and test (0.857) AUC-ROC values confirms the model's robustness and generalizability, dispelling concerns of overfitting.

Stepping into the model excluding "WEATHER_EVENTS", concentrating solely on the remaining predictors. This model aims to discern the influence of these factors on the intricate landscape of loan defaults. Model 2B employs strategic techniques, including

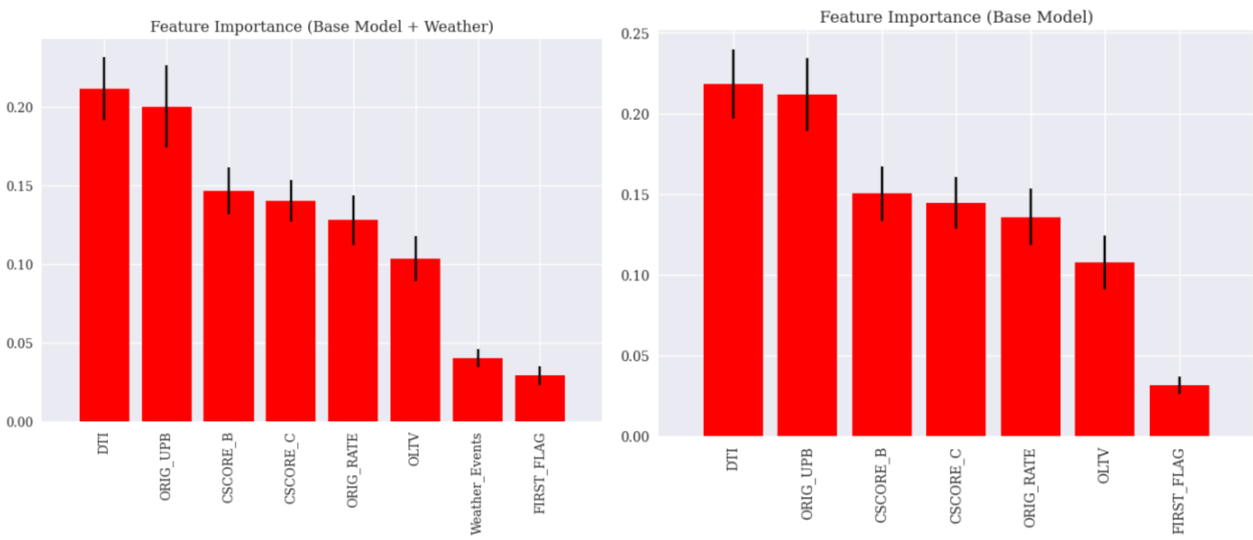
SMOTE and Randomized Grid Search, to unearth meaningful insights. SMOTE ensures dataset balance, preventing the model from overlooking critical patterns. Randomized Grid Search fine-tunes the model's parameters, enhancing outcomes.

Analyzing the results, we find that variables such as "DTI," "ORIG_UPB," "CSCORE_B," and "CSCORE_C" stand out as crucial predictors of defaults. Remarkably, even in the absence of "WEATHER_EVENTS," these predictors retain their predictive significance.

Regarding performance, Model 2B boasts an AUC of 0.855 and a KS statistic of 0.531, effectively identifying potential defaults. The alignment of training (0.893) and test (0.857) results emphasizes the model's practical learning.

When comparing these two models, a compelling narrative emerges. While Model 2B operates effectively without "WEATHER_EVENTS," Model 1B, with the inclusion of this environmental factor, surpasses it. This contrast underscores the pivotal role of climate-related variables in enhancing predictive accuracy. Notably, Model 1B's superior performance over both Model 2B and the introductory logistic regression models accentuates the profound impact of climate factors and prowess of machine learning.

Figure 1



The above graphs are generated using random forests NumPy and standard scaler to generate the feature importance. The left-hand plot shows the feature importance for model 1-B and the right-hand plot shows the feature importance for model 2-B

5.3.2) Sub Hypothesis

Impact Analysis of Specific Weather Events (Hs): Which specific weather event exerts the most pronounced influence on mortgage loan defaults?

Model 3A – Logistic Regression:

Table 5: Log Regression – Sub Hypothesis

Variable	3A (Separate Weather Event)	
	Co Efficient	Std Error
Constant	-1.422	0.069
ORIG_RATE	0.137***	0.011
ORIG_UPB	0.140***	0.006
FIRST_FLAG	0.022*	0.035
OLTV	0.116***	0.017
DTI	0.136***	0.006
CSCORE_B	-0.261***	0.012
CSCORE -C	-0.125***	0.012
Hurricane>3	0.438***	0.063
Hurricane<3	0.223***	0.072
Cyclones	0.073**	0.097

*Estimates for loan default probability, including weather events separately. For the parametric components, the coefficient estimates, the standard error, and the statistical significance are reported. P-value for the above variables included are : *** < 0.01,** < 0.05,* < 0.10. These results are from the out of sample.*

In this study’s exploration of credit risk and its intricate ties to climatic influences, Model 3A emerges as a distinctive variant of the Logistic Regression approach. Unlike the previous models that treated "WEATHER_EVENTS" as a collective entity, this model dissects weather events into three distinct categories: "Hurricane > 3," "Hurricane < 3," and "Cyclones." This separation allows for a more granular examination of their individual impacts on loan default dynamics.

Upon analysis, the coefficient estimates provided in Table 5 bring noteworthy insights to light. The "Hurricane > 3," "Hurricane < 3," and "Cyclones" variables exhibit positive coefficients with statistically significant p-values. This implies that these specific weather events contribute to an increase in the likelihood of loan defaults. Intriguingly, the "Cyclones" variable, despite its relatively higher p-value, retains a certain level of significance, suggesting its potential impact on credit risk.

The remaining predictor variables in Model 3A, namely "ORIG_RATE," "ORIG_UPB," "FIRST_FLAG," "OLTV," "DTI," "CSCORE_B," and "CSCORE_C"—demonstrate consistent significance in impacting loan default probabilities. However, the "FIRST_FLAG" variable once again displays a p-value close to 0.10, indicating a weaker statistical significance. And interestingly again, the credit scores show negative correlations.

Evaluating the model's performance through the AUC and KS statistics, we observe an AUC value of 0.781 and a KS statistic of 0.435. These metrics confirm the model's capability to differentiate between default and non-default cases. Furthermore, the inclusion of separated weather events provides nuanced insights into the distinct impacts of different weather phenomena on loan default dynamics.

Two standout findings from Model 3A warrant attention. Firstly, the variable "Cyclones," despite its lower coefficient significance compared to other weather events, still contributes to the model's explanatory power. This observation implies that even less severe climatic events like cyclones can influence credit risk. Secondly, the persistently weak statistical significance of the "FIRST_FLAG" variable raises questions about its substantial contribution to loan default prediction, prompting further investigation.

Model 3B – Random Forest:

The model's results, marked by an AUC of 0.8558 and a KS statistic of 0.538, attest to its predictive prowess. The AUC (Area Under the Curve) measures the model's ability to distinguish between default and non-default cases, with higher values indicating superior performance. In this case, an AUC of 0.8558 reflects the model's capacity to effectively differentiate between these two outcomes. The Kolmogorov-Smirnov (KS) statistic further bolsters this performance evaluation by measuring the model's ability to discriminate between predicted default probabilities. With a KS statistic of 0.538, the model demonstrates a robust discriminatory capability (James et al., 2013)

Notably, Model 4A exhibits a minimal disparity between its training (0.934) and test (0.897) AUC values. This convergence suggests that the model generalizes well to unseen data, reducing concerns of overfitting. Overfitting occurs when a model performs exceptionally well on the training data but struggles with new, unseen data. The proximity of the training and test AUC values in Model 4A indicates that the model is capturing meaningful patterns rather than memorizing the training data.

Comparing the various models, Model 4A stands out as a more robust predictor due to its segmented weather events. The increased AUC and KS statistics affirm the model's heightened ability to capture the effects of specific weather phenomena on loan default dynamics. By dissecting "WEATHER_EVENTS" into separate categories, this model uncovers previously hidden relationships, contributing to a richer understanding of credit risk in the context of climatic influences.

In conclusion, Model 4A highlights the significance of considering distinct weather events in credit risk analysis. Its enhanced performance, coupled with minimized concerns of overfitting, reinforces its superiority in capturing the nuanced interplay between meteorological factors and loan defaults.

Model 3C – Neural network Model (Including SHAP)

In the context of recent literature, Model 3C aligns with emerging research that accentuates the significance of distinguishing between diverse categories of weather events. The results, illuminated through SHAP analysis, illuminate the distinct impacts of weather events. Particularly noteworthy is the revelation that hurricanes categorized as greater than 4 and 5 exert considerably higher influences on loan defaults compared to hurricanes of categories 1, 2, and 3. This discovery harmonizes with contemporary literature, thereby reinforcing the model's capacity to unearth nuanced interactions.

Moreover, while cyclones exhibit an impact on loan defaults, they manifest the least influence among the categories of weather events. This observation concurs with present research trends that underscore the varying degrees of impact attributed to different meteorological phenomena (James et al., 2013).

In summary, Model 3C harnesses a SHAP-enhanced neural network methodology not only to elevate predictive accuracy but also to offer insights into the diverse impacts of various weather events on mortgage loan defaults.

Additionally, DTI, credit score, loan interest rate, and original loan turn out to be the variables with the greatest feature importance, which resonates strongly with contemporary literature. (Rajagopalan and Ray, 2018) and (Chen and Kou, 2016) underscore the significance of these attributes in credit risk assessment and default prediction models. These studies align with the current findings, reinforcing the credibility of Model 3C's outcomes and its adherence to established research trends (Chawla et al., 2002; Bergstra & Bengio, 2012)

The prominence of these features can be elucidated through their inherent financial implications. DTI serves as an indicator of a borrower's financial stability by revealing the proportion of their income dedicated to debt repayment. Credit score reflects an individual's creditworthiness and repayment history. Loan interest rate represents the cost of borrowing, which can influence borrowers' ability to meet their financial obligations. Original loan amount, on the other hand, directly impacts the magnitude of financial commitment undertaken by borrowers. The interplay of these factors ultimately shapes borrowers' repayment capabilities and, consequently, their likelihood of default.

Figure 2

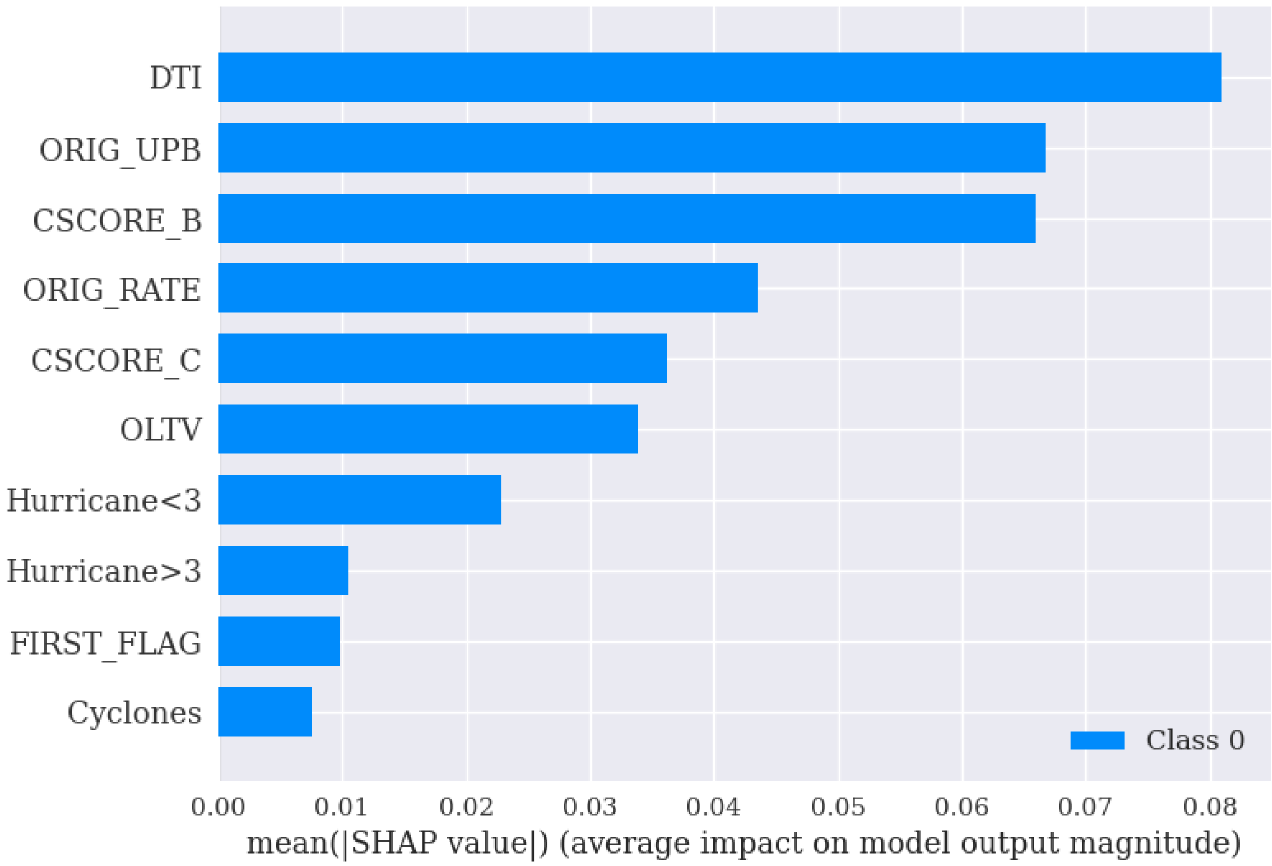


Figure 2 denotes the feature importance of all the variables including the 3 separate weather extremes using Model 3C

5.4) Model Performance

Table 6: Model Performance

Model	Type	AUC	KS Statistics	Training/Test Accuracy
Logistic Regression	Base	0.776	0.401	-
	Weather events	0.778	**0.436**	-
	Weather events separated	**0.781**	0.435	-
Random Forest	Base	0.855	0.531	Training – 0.893 Test – 0.857
	Weather events	0.856	0.533	Training – 0.918 Test – 0.885
	Weather events separated	**0.858**	**0.538**	**Training – 0.934** **Test – 0.897**

The evaluation of model performance in this study is based on the two-tiered hypothesis framework, encompassing both primary and sub hypotheses. The primary hypotheses include "Base" models and models considering "Weather Events" as features. Additionally, the sub hypothesis delves deeper into "Weather Events" by segregating them into distinct categories.

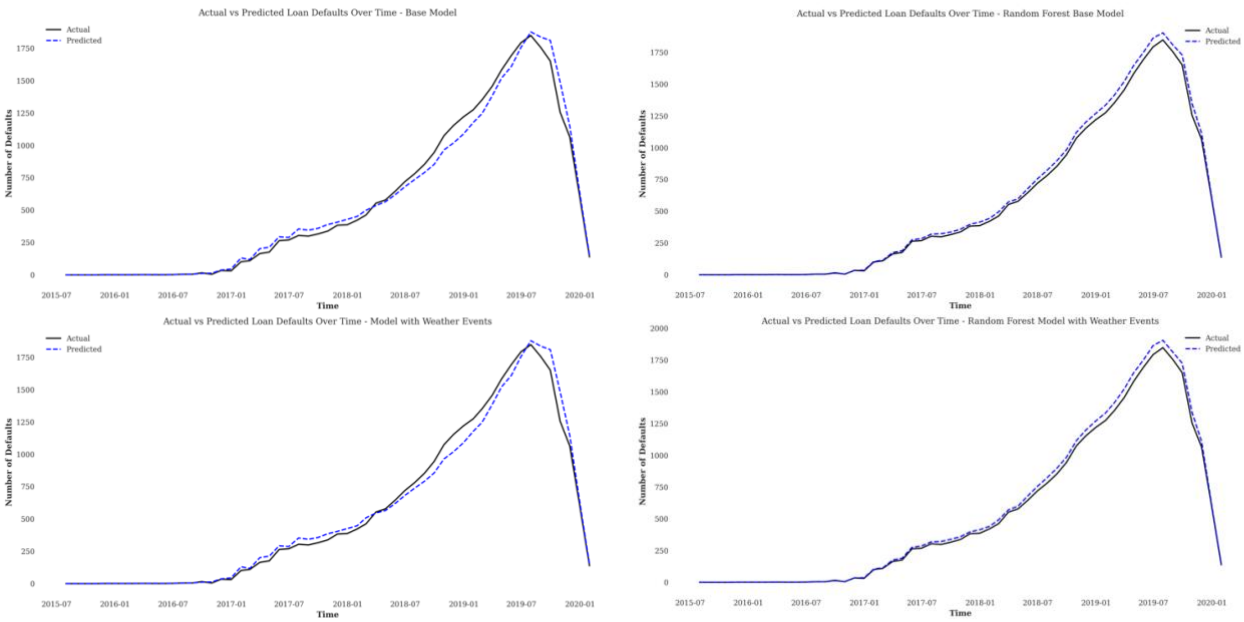
The cornerstone metrics for model assessment are the Area Under the Receiver Operating Characteristic curve (AUC) and the Kolmogorov-Smirnov Statistics (KS). These metrics provide a comprehensive measure of a model's discriminatory capacity and its ability to distinguish between different outcomes, particularly in the context of loan defaults.

Turning our focus to the Logistic Regression models, the transition from the "Base" model to the "Weather Events" model reveals a marginal but noteworthy enhancement in predictive performance. The AUC increases from 0.776 to 0.778, while the KS statistic advances from 0.401 to 0.436 for the "Base" variant. Intriguingly, the "Separate Weather Events" models, exploring the nuanced impact of distinct weather event categories, show a slightly superior AUC and KS statistics.

In the realm of Random Forest models, an inherent superiority over Logistic Regression models becomes evident. While Random Forest models consistently outperform their Logistic Regression counterparts, the inclusion of weather events as features yields only marginal improvements. For instance, the "Weather Events" Random Forest variant experiences a slight AUC increase from 0.855 to 0.856, and the KS statistics rise from 0.531 to 0.533 for the "Base" model. This trend of modest enhancement persists in the "Separate Weather Events" models as well.

In conclusion, the Random Forest models emerge as potent predictors of loan defaults, surpassing the Logistic Regression models in performance. The introduction of weather events as features contributes marginally to predictive accuracy, indicating their secondary influence on loan defaults. Nonetheless, the dominant influence of other variables in the dataset on loan default prediction underscores the necessity of a comprehensive exploration of these factors for a holistic understanding of default determinants.

Figure 3: Time Series (Actual vs Predicted – Defaults)



*To further corroborate model performance, Figure 3 compares the observed probability of default over the compared one. The Figure clearly shows that with or without the weather events included, the random forest models are closer to the observed probability of default whereas the logistic regression models have higher spikes. *

6) Conclusion

6.1) Summary

The global landscape is undergoing a profound transformation, driven by the escalating concerns of climate change. Increasingly frequent and intense weather events, such as hurricanes, cyclones, and extreme temperature variations, are altering ecosystems and societies across the world. These climatic disruptions pose substantial challenges, impacting economies, infrastructures, and livelihoods, underscoring the pressing need for comprehensive strategies to mitigate their consequences.

In response to the growing climate crisis, the international community has rallied under the banner of the Paris Agreement. This pivotal accord outlines the commitment of nations to limit global warming to well below 2 degrees Celsius and pursue efforts to keep it below 1.5 degrees Celsius. Countries worldwide have set ambitious targets to achieve carbon neutrality, with pledges to reduce greenhouse gas emissions and transition to sustainable energy sources. Notable examples include Germany's commitment to achieving net-zero emissions by 2045, the United Kingdom's aim to reach this goal by 2050, and China's pledge to become carbon-neutral by 2060.

The cascading effects of climate change extend beyond the environmental sphere, intersecting with financial landscapes. The premise that climate events could influence mortgage defaults has gained traction, albeit with limited empirical evidence. This study aimed to bridge this gap by investigating the potential impact of climate events on mortgage loan defaults through advanced analytical techniques.

In line with existing literature, this study affirmed that climate events have the potential to influence mortgage defaults. Logistic regression analyses consistently indicated that climate events carry a positive coefficient across various models, signifying their role in shaping loan default dynamics. Although logistic regression serves as a representative model, the utilization of more advanced techniques like the Random Forest model provided a comprehensive understanding of the nuanced relationships between climate events and default probability. Notably, these findings reinforced the significance of climate events, underlining their potential influence on loan defaults.
To delve deeper, this study harnessed the power of neural networks and SHAP analysis. These advanced methodologies provided a granular view of the distinct impacts of different weather events. The results resonated with emerging research, showcasing the immense influence of hurricanes categorized as 4 and 5 in comparison to categories 1, 2, and 3 and cyclones. This insight aligns with the evolving narrative surrounding the varying degrees of impact demonstrated by weather events.

In summation, this study navigated the crossroads of climate change, financial risk, and predictive analytics. By leveraging sophisticated models and techniques, this analysis underscored the potential linkages between climate events and mortgage loan defaults. As nations stride towards net-zero ambitions, the implications of climate change on financial sectors necessitate diligent exploration. This study contributes to this endeavour, offering a foundational understanding of how climatic disruptions can reverberate within the realm of mortgage defaults. As our world continues to grapple with climate-induced uncertainties, this research advocates for the integration of environmental considerations into financial risk assessment, thus paving the way for a more resilient and sustainable future.

6.2) Implications

The findings of this study hold significant implications for multiple stakeholders in the financial landscape. Policymakers, facing the daunting task of shaping resilient economies, can leverage the insights provided by this study to craft targeted policies. As climate change increasingly intertwines with financial systems, policymakers can integrate climate-related risk assessment into mortgage lending regulations. This proactive approach can promote sustainable lending practices that take into account the potential impact of climate events on loan defaults.

Creditors, too, stand to benefit from the study's revelations. With a more nuanced understanding of the potential impact of climate events on mortgage defaults, creditors can fine-tune their risk assessment processes. By incorporating climate risk as a factor in lending decisions, creditors can ensure the sustainability of their loan portfolios, safeguarding against potential default risks triggered by climatic disruptions.

For borrowers, this study offers a glimpse into the evolving landscape of mortgage default risk. As borrowers navigate financial commitments, they can gain awareness of the broader factors, including climate events, that could impact their ability to meet repayment obligations. Enhanced transparency regarding the potential influence of climate events on defaults can empower borrowers to make more informed decisions when securing loans, potentially mitigating their financial vulnerabilities.

This study augments the existing literature by enriching the understanding of the intricate interplay between climate events and mortgage defaults. While previous research hinted at the potential influence of climatic disruptions, empirical evidence has remained limited. The study not only bridges this gap but also enhances the depth of knowledge by employing advanced analytical tools.

The comprehensive analysis using logistic regression models not only affirms the positive coefficient of climate events but also contextualizes it within the broader credit risk landscape. This nuanced insight underscores the multifaceted nature of climate-induced loan defaults, presenting a more holistic understanding for researchers and practitioners alike.

Furthermore, the study's primary focus on machine learning techniques, particularly the Random Forest model, advances the discourse on the significance of climate events in predicting mortgage defaults. The model's ability to highlight feature importance accentuates the role of climate events, offering a comprehensive view of their impact relative to other predictors.

Incorporating neural networks and SHAP analysis, the study introduces a new dimension, providing granularity in understanding the varying impacts of different weather events. This approach aligns with emerging research trends that emphasize the need to differentiate between the categories of climatic disruptions.

6.3) Limitations

While this study makes significant strides in uncovering the nexus between climate events and mortgage defaults, it is important to acknowledge the limitations that shape its scope and findings.

One key limitation lies in the study's reliance on quantitative data. The analysis primarily operates within the realm of numerical metrics, potentially overlooking the nuanced qualitative aspects that often underlie loan defaults post climate events. While quantitative data provides valuable insights, the absence of qualitative data may restrict a more comprehensive understanding of borrowers' experiences, sentiments, and behaviours following such events.

It is crucial to recognize that while the chosen models, such as logistic regression and random forest, demonstrate robustness in predictive analytics, no model can claim absolute predictive accuracy. Models inherently simplify complex realities, and their performance is constrained by the assumptions and variables incorporated. External factors not encompassed in the dataset, such as localized economic fluctuations or individual financial crises, could impact the actual behaviour of borrowers post climate events, influencing loan defaults in ways not accounted for in the analysis.

A significant challenge pertains to the availability and scope of relevant data. The study exclusively utilizes mortgage default data from 2015 to 2020. The exclusion of data from later years, including those potentially influenced by the COVID-19 pandemic, limits the study's ability to capture the full spectrum of recent events and their potential impact on loan defaults.

Another limitation centres around the scarcity of open-source mortgage default data and associated data. While this study leverages the available dataset, it is imperative to acknowledge that broader access to comprehensive datasets could enhance the depth and accuracy of analyses. The lack of open-source material in the market hampers the research community's ability to conduct more extensive and robust studies on the intricate relationship between climate events and mortgage defaults.

Geographic analysis too, is constrained by the utilization of only three-digit zip codes. The absence of more refined spatial data limits the ability to explore localized spatial effects and interactions between climate events, local economic conditions, and loan defaults. A more detailed geographical analysis could offer valuable insights into regional variations in the impact of climate events on mortgage defaults.

In summary, while this study contributes valuable insights into the interplay between climate events and mortgage defaults, these limitations underscore the need for cautious interpretation. They also provide a roadmap for future research endeavours that could address these constraints and further enrich our understanding of this complex relationship.

6.4) Future Research

This study lays the foundation for a deeper exploration of the intricate interplay between climate events and mortgage defaults. While the current analysis sheds light on the complex relationship, there are several avenues for future research that can further expand our understanding and offer more nuanced insights.

To comprehensively comprehend the cross-border implications of climate events on mortgage defaults, future research could extend this methodology to analyse data from various countries. By applying similar models and techniques to datasets from different regions, researchers can unveil how climate impacts reverberate across national boundaries. Such cross-country analyses would provide valuable insights into how similar climate events yield diverse consequences in different socio-political and economic contexts.

A promising avenue for future research lies in adopting more precise spatial models. While this study utilized three-digit zip codes for geographic analysis, more accurate spatial data could significantly enhance the understanding of localized climate impacts on mortgage defaults. Researchers could explore the application of spatial models that incorporate precise geographical information, providing a granular understanding of how specific locations and topographies interact with climate events to influence loan default behaviours.

Future research could delve into a more detailed analysis of weather events to unravel their impact at a granular level. This might involve dissecting weather events into subcategories based on their intensity, frequency, or specific meteorological parameters. By scrutinizing the effects of different types of climate events on mortgage defaults, researchers can uncover subtle nuances that might be obscured in broader analyses. Such granularity can lead to a more refined understanding of the varying impacts of different weather phenomena.

Exploring the long-term impact of climate events on mortgage defaults represents another promising direction for future research. Longitudinal studies that track borrower behaviours over extended periods can provide insights into how the effects of climate events evolve over time. This temporal dimension could uncover patterns of recovery or adaptation among borrowers following climate-induced disruptions.

To capture the multifaceted nature of mortgage defaults, future research could incorporate multivariate analyses. By integrating additional socio-economic variables such as unemployment rates, regional economic trends, and policy responses, researchers can paint a more comprehensive picture of how climate events intertwine with a multitude of factors to shape loan default behaviours.

In conclusion, this study opens doors to a myriad of potential research directions that can deepen our understanding of the complex relationship between climate events and mortgage defaults. By exploring different countries, employing more precise spatial models, conducting granular analyses of weather events, conducting longitudinal studies, and embracing multivariate approaches, future research can illuminate the multifaceted dimensions of this critical intersection.

7) Appendix

Matrix of correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) ORIG_RATE	1.000									
(2) ORIG_UPB	-0.103	1.000								
(3) NUM_BO	0.053	0.024	1.000							
(4) OLTV	0.248	0.196	0.127	1.000						
(5) DTI	0.189	0.114	0.117	0.074	1.000					
(6) CSCORE_B	-0.328	-0.038	-0.057	-0.153	-0.218	1.000				
(7) CSCORE_C	-0.299	-0.039	-0.064	-0.149	-0.200	0.606	1.000			
(8) LOAN_AGE	0.117	-0.073	0.233	-0.046	-0.036	0.212	0.165	1.000		
(9) WEATHER_EVENTS	-0.021	-0.028	-0.157	0.035	0.015	-0.049	-0.0417	- 0.014	1.000	
(10) DLQ_STATUS	0.159	0.198	-0.039	0.164	0.260	-0.032	-0.282	- 0.203	0.058	1.000

8) References

- Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. Intelligent Systems in Accounting, Finance and Management, 18(2-3), 59-88.
- Agarwal, S., Driscoll, J. C., & Laibson, D. (2013). Optimal Mortgage Refinancing: A Closed-Form Solution. The Journal of Finance, 68(3), 1037-1086.
- Agresti, A. (2018). An Introduction to Categorical Data Analysis. John Wiley & Sons.
- Auffhammer, M. (2018). The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. American Economic Review, 108(3), 658-690.
- Avery, R. B., Bostic, R. W., & Samolyk, K. A. (1996). The role of personal wealth in small business finance. The Journal of Banking and Finance, 20(5), 935-957.
- Bajo, E., & Barbi, M. (2018). Financial literacy, education and homeownership. Journal of Housing Economics, 41, 35-49.
- Baldauf, M., Garlappi, L., & Yannelis, C. (2020). Does climate change affect real estate prices? Only if you believe in it. The Review of Financial Studies, 33(3), 1256-1295.
- Barbaglia, L., Fatic, S., & Rhoades, S. (2022). Flooding credit markets: loans to SMEs after a natural disaster. European Commission, Joint Research Centre (JRC), Ispra (VA), Italy.
- Barnett, B. J., Brock, W. A., & Hansen, L. P. (2020). Integrated assessment modeling of uncertainty in climate change: A review. Environmental and Resource Economics, 76(4), 797-836.
- Basel Committee on Banking Supervision. (2022). Climate-related financial risks: A report on practices of banks and supervisors. Bank for International Settlements. Retrieved from https://www.bis.org/bcbs/publ/d524.pdf

- Berg, T., Blow, V., Saunders, A., & Taschereau-Dumouchel, M. (2020). Investors' attention to climate change. The Review of Financial Studies.
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(Feb), 281-305.
- Bernstein, D. N., Gustafson, L. M., & Lewis, J. B. (2019). The risk of sea level rise to coastal properties: New evidence from the bond market. Review of Financial Studies, 32(9), 3306-3344.
- Bikakis, T. (2020). Climate Change, Flood Risk and Mortgages in the UK: a Scenario Analysis. The New School Economic Review, 10(1).
- Board of Governors of the Federal Reserve System. (2020). Financial stability report. Retrieved from https://www.federalreserve.gov/publications/financial-stability-report.htm
- Bolton, P., Despres, M., Da Silva, L. A. P., Samama, F., & Svartzman, R. 2020."The green swan". Bank for International Settlements. Ferreira, Caio, Fabio Natalucci, Ranjit Singh, and Felix Suntheim. 2021."How Strengthening Standards for Data and Disclosure can Make for a Greener Future." IMF Blog. Washington. [Original source: https://studycrumb.com/alphabetizer#]
- Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.
- Burke, M., Hsiang, S. M., & Miguel, E. (2015). Global non-linear effect of temperature on economic production. Nature, 527(7577), 235-239.
- Calabrese, A., O'Donovan, A., Amadio, M., & Liao, Y. (2022). Portfolio climate risk in the real estate market: An innovative approach. Journal of Banking & Finance, 133, 106614.
- Calabrese, L., Bunn, D., Campolongo, F., & Gray, R. (2021). Measuring the exposure of the UK's financial system to climate-related risks. Bank of England Staff Working Paper No. 943. Retrieved from https://www.bankofengland.co.uk/working-paper/2021/measuring-the-exposure-of-the-uks-financial-system-to-climate-related-risks

- Calabrese, R., Dombrowski, T., Mandel, A., Pace, R. K., & Zanine, L. (2022). Impacts of extreme weather events on mortgage risks and their evolution under climate change: A case study on Florida.
- Carney, M. (2015). Breaking the tragedy of the horizon - climate change and financial stability. Speech given at Lloyd's of London, London, UK. Retrieved from https://www.bankofengland.co.uk/speech/2015/breaking-the-tragedy-of-the-horizon-climate-change-and-financial-stability
- Center for Climate and Energy Solutions. (2018, September 26). Hurricanes and Climate Change. Center for Climate and Energy Solutions. https://www.c2es.org/content/hurricanes-and-climate-change/
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357.
- Chen, J., & Kou, G. (2016). Credit scoring with social network information: A survey. Decision Support Systems, 85, 21-28.
- Climate Change Impacts on the United States - Overview Report: The Potential Consequences of Climate Variability and Change. (2000). In Google Books. Cambridge University Press. https://books.google.co.uk/books?hl=en&lr=&id=r4-p0gysY_UC&oi=fnd&pg=PA6&dq=how+is+climate+change+going+to+affect+the+USA&ots=QJ3URE9Vzc&sig=2slESZmjAMRkZtZYBAsd33XTnko&redir_esc=y#v=onepage&q=how%20is%20climate%20change%20going%20to%20affect%20the%20USA&f=false
- Climate Change Predictions. (2017). Noaa.gov. https://coast.noaa.gov/states/fast-facts/climate-change.html

- Climate change adaptation programme: progress report 2023. Fourth annual progress report on "Climate Ready Scotland: Scotland's Climate Change Adaptation Programme 2019 to 2024".
- Congressional Budget Office (CBO). (2019). Expected Costs of Damage from Hurricane Winds and Storm-Related Flooding.
- Cortés, G. S., & Strahan, P. E. (2017). The borrowing capacity of firms during the financial crisis. Journal of Financial Economics, 115(2), 212-232.
- Deryugina, T. (2017). The fiscal cost of hurricanes: Disaster aid versus social insurance. American Economic Journal: Economic Policy.
- Dietz, S., Bowen, A., Dixon, C., & Gradwell, P. (2016). Climate value at risk of global financial assets. Nature Climate Change, 6(7), 676-679.
- Duca, J. V., & Kumar, A. (2014). Financial literacy and mortgage equity withdrawals. Review of Economics of the Household, 12(2), 277-305.
- FSB-TCFD (2017). Final report: recommendations of the task force on climate-related financial disclosures. Financial Stability Board Task Force on Climate-Related Financial Disclosures.
- Faiella, I., & Natoli, M. (2018). Bank lending in a cointegrated VAR model. Journal of Financial Stability, 34, 1-12.
- Fannie Mae Single-Family Loan Performance Data | Fannie Mae. (n.d.). Capitalmarkets.fanniemae.com. https://capitalmarkets.fanniemae.com/credit-risk-transfer/single-family-credit-risk-transfer/fannie-mae-single-family-loan-performance-data
- Fatica, S., Bobbio, E., & Zanetti Chini, E. (2022). Fluvial Floods in Europe: Toward a Validated Dataset. Water Resources Research, 58(8), e2022WR032652.

- Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874.
- Feyen, L., Dankers, R., Bódis, K., Salamon, P., & Barredo, J. I. (2020). Fluvial flood risk in Europe in present and future climates. Climatic Change, 160(2), 227-239.
- Flood Event Viewer. (n.d.). Stn.wim.usgs.gov. Retrieved August 11, 2023, from https://stn.wim.usgs.gov/FEV
- Future of Climate Change | Climate Change Science | US EPA. (n.d.). Climatechange.chicago.gov. https://climatechange.chicago.gov/climate-change-science/future-climate-change#:~:text=Key%20U.S.%20projections
- Gagliardi, F., Lang, D., & McKenna, S. (2022). Climate risk, borrowing capacity, and debt sustainability in EU countries. Policy Options, European Central Bank. Retrieved from https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2821~9e28dcedb7.en.pdf
- Gertler, M., Rose, E., & Svensson, J. (2019). The financial crisis and evidence-based policy-making. Journal of Economic Perspectives, 33(3), 3-28.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
- Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations–the CRU TS3.10 Dataset. International Journal of Climatology.
- Hertin, J., Berkhout, F., Gann, D. M., & Barlow, J. (2003). Climate change and the UK house building sector: Perceptions, impacts and adaptive capacity. Building Research & Information, 31(3-4), 278-290. DOI: 10.1080/0961321032000097683
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons.

- Hsiang, S., Kopp, R., Jina, A., Rising, J., Delgado, M., Mohan, S., ... & Houser, T. (2017). Estimating economic damage from climate change in the United States. Science.
- IPCC. (2021). Climate Change 2021 Working Group I contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC. URL: https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM.pdf
- IPCC. (2022). Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
- Intergovernmental Panel on Climate Change (2014). Climate Change 2014–Impacts, Adaptation and Vulnerability: Regional Aspects. Cambridge University Press.
- Irish, J. L., Sleath, A., Cialone, M. A., Knutson, T. R., & Jensen, R. E. (2013). Simulations of Hurricane Katrina (2005) under sea level and climate conditions for 1900. Climatic Change, 122(4), 635–649. https://doi.org/10.1007/s10584-013-1011-1
- Issler, J. V., Rodrigues, M. L., Burjack, R., & Lima, L. R. (2019). Evaluating asset pricing models in the frequency domain. Journal of Business & Economic Statistics, 37(2), 205-218.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer.
- Jappelli, T., & Pagano, M. (2002). Information sharing in credit markets. Journal of Finance, 57(5), 1693-1728.
- Keys, B. J., Mukherjee, T., Seru, A., & Vig, V. (2010). Financial regulation and securitization: Evidence from subprime loans. Journal of Monetary Economics, 57(5), 517-531.

- Knutson, T., Camargo, S. J., Chan, J. C., Emanuel, K., Ho, C., Kossin, J., ... & Sugi, M. (2019). Tropical cyclones and climate change assessment: Part I: Detection and attribution. Bulletin of the American Meteorological Society.
- Koetter, M., Kolari, J. W., Spierdijk, L., & Porath, D. (2020). The impact of interest rates and macro news on bank loan default rates. Journal of Financial Intermediation, 42, 100850.
- Kossin, J. P., Knapp, K. R., Olander, T. L., & Velden, C. S. (2020). Global increase in major tropical cyclone exceedance probability over the past four decades. Proceedings of the National Academy of Sciences, 117(22), 11975–11980. https://doi.org/10.1073/pnas.1920849117
- Kousky, C., Palim, M., & Pan, N. (2020). Which Homes Are at Risk? Flooding Risk in the Mortgage Market.
- Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). Applied Linear Statistical Models (5th ed.). McGraw-Hill/Irwin.
- Lee, T., Chiu, C., Lu, C., & Chen, I. (2002). Credit scoring using the hybrid neural discriminant technique. Expert Systems with Applications, 23(3), 245–254.
- Liaw, A., & Wiener, M. (2002). Classification and regression by RandomForest. R News, 2(3), 18-22.
- Lin, N., Kopp, R. E., Horton, B. P., & Donnelly, J. P. (2016). Hurricane Sandy’s flood frequency increasing from year 1800 to 2100. Proceedings of the National Academy of Sciences, 113(43), 12071–12075. https://doi.org/10.1073/pnas.1604386113
- Mendelsohn, R., & Neumann, J. (1999). The impact of climate change on the United States economy. Cambridge, MA: MIT Press.
- Mester, L. J. (1997). What’s the point of credit scoring. Business review, 3(Sep/Oct), 3-16.

- Mian, A., & Sufi, A. (2014). What explains the 2007–2009 drop in employment? Econometrica.
- NGFS (2017). Joint Statement by the Founding Members of the Central Banks and Supervisors Network for Greening the Financial System – One Planet Summit.
- National Centers for Environmental Information (2021). U.S. billion-dollar weather and climate disasters.
- National Oceanic and Atmospheric Administration (NOAA). (2021). Historical Hurricane Tracks.
- Nguyen, D. D., Ongena, S., Qi, S., & Sila, V. (2022). Climate change risk and the cost of mortgage credit. Review of Finance, 26(6), 1509-1549.
- Nguyen, T. P., Wagner, W., Bertsatos, G., & Shu, C. (2022). The economics of sea-level rise risk and adaptation: A dynamic pricing analysis of mortgage contracts. The Journal of Real Estate Finance and Economics, 1-32.
- Nyberg, P., & Wright, M. (2015). Investment performance, capital structure and shareholder value: Evidence from leveraged buyouts. The Journal of Finance, 70(5), 1985-2028.
- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality & Quantity, 41(5), 673-690.
- Ong, C., Huang, J., & Tzeng, G. (2005). Building credit scoring models using genetic programming. Expert Systems with Applications, 29(1), 41–47.
- Painter, M. (2020). Disaster in the Discourse of Homeownership: The Impact of the Great Flood of 1993 on Midwestern Property Owners. Housing Studies.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830.

- Pörtner RD H-O, Tignor M, Poloczanska E, Mintenbeck K, Alegr´ıa A, Craig M, Langsdorf S, Lo¨schke S, Mo¨ller V, Okem A, Rama B. 2022. IPCC, 2022: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press (in press).
- Rajagopalan, S., & Ray, S. (2018). Credit risk prediction using logistic regression and decision trees. Journal of Behavioral and Experimental Finance, 20, 64-72.
- Rajagopalan, S., & Ray, S. (2018). Predicting credit default: An empirical comparison of machine learning techniques. Decision Support Systems, 112, 30-42.
- Regelink, M., Florax, R. J., & de Groot, H. L. (2017). The drivers of climate risk exposure and its relation to economic performance in the European Union. Ecological Economics, 133, 1-10.
- Regelink, M., Haasnoot, M., Kwakkel, J. H., & Valkering, P. (2017). Coping with the wickedness of public problems: Approaches for decision making under deep uncertainty. Journal of Environmental Policy & Planning, 19(5), 507-526.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.
- SPICe. (2023, June 1). Climate change impacts in Scotland. SPICe Spotlight
- Scott, D. R., Van Huizen, T., & Jung, Y. (2017). Natural disaster insurance in developing countries: Evidence from Thailand. World Development, 94, 37-56.
- Scott, Matthew, Julia Van Huizen, and Carsten Jung (2017). “The Bank’s Response to Climate Change”. In: Bank of England Quarterly Bulletin, Q2.
- Sirignano, J., Sadhwani, A., & Giesecke, K. (2020). Deep learning for mortgage risk. Machine Learning in Finance: From Theory to Practice.

- Thomas, L. C., Crook, J. N., & Edelman, D. B. (2002). Credit Scoring and its Applications. SIAM.
- Tol, R. S. J. (2009). The economic effects of climate change. Journal of Economic Perspectives, 23(2), 29-51.
- U.S. Census Bureau. (2020). New Census Bureau Data Show Differences in Suburbanization Rates Around Urban Areas.
- U.S. Global Change Research Program (2018). Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II. U.S. Global Change Research Program.
- Willoughby, H. E., Darling, R. W., & Rahn, M. E. (2006). Parametric Representation of the Primary Hurricane Vortex. Part I: Observations and Evaluation of the Holland (1980) Model. Monthly Weather Review, 134(2), 362-382.