Machine Learning for Mortgage Risk: An Interpretable Approach

Vincent Zong Stone, Maggie Chen, Yao Zhang, Ming Xia

A Thesis in the Field of Data Science
for the Degree of Master of Liberal Arts in Extension Studies
Harvard University
December 2024

Abstract

The increasing complexity of mortgage risk assessment demands sophisticated yet transparent analytical methods. This study presents a novel framework for predicting mortgage delinquency risks by combining advanced time-series modeling with interpretable machine learning. Leveraging diverse datasets comprising mortgage, economic, and weather-related variables, the framework incorporates SHAP (SHapley Additive exPlanations). It introduces feature-wise fidelity R^2 scores for robust feature selection in multivariate time-series models.

Through this proposed framework, the Temporal Fusion Transformer model, streamlined to six critical features including *Consumer Confidence*, *Loan-to-Value Ratio*, and *Unemployment Rate*, achieved an exceptional R^2 score of 0.9961 and near-zero MSE on test data, with SHAP-reconstructed predictions attaining a fidelity R^2 score of 0.9653. The analysis highlights individual financial conditions and borrower-level variables as the strongest predictors of delinquency risk, with broader economic indicators providing complementary insights. This research advances financial risk assessment by combining high-performance predictive modeling with interpretable insights, supporting improved underwriting processes and risk mitigation in the mortgage industry.
Keywords: Machine Learning, Mortgage Risk, Temporal Fusion Transformer, SHAP, Interpretable AI, Feature Selection, Time Series Analysis

Acknowledgments

We extend our sincere gratitude to Dr. Bruce Huang for his invaluable insights into interpretability methods. We also thank Dr. Stephen Elston for his extensive guidance on time-series modeling approaches. Our mentor and project sponsor, Venkataraman Muthuramalingam, provided expert professional advice on mortgage modeling and valuable feedback that shaped our work. We deeply appreciate their collective support and contributions throughout this research project.

Lastly, we are grateful to everyone who supported us, making this research a meaningful and rewarding experience.

On a personal note, I, Vincent Zong Stone, would like to express my heartfelt gratitude to my wife, Danni, for her unwavering support and encouragement, as well as to my children, Vincent and Longma, for their understanding and patience. To our soon-to-arrive baby, thank you for being my inspiration throughout this journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The mortgage industry faces increasing challenges in risk assessment as financial markets become more complex and data-driven. While machine learning models offer powerful predictive capabilities, their "black box" nature often makes it difficult for stakeholders to trust and act upon their predictions. This study addresses this challenge by developing an interpretable machine learning framework that combines high predictive accuracy with transparent insights into the decision-making process.

1.1 Motivation

Motivation Traditional predictive models often prioritize accuracy but lack interpretability, leaving stakeholders with a limited understanding of the factors driving model predictions. This gap is particularly problematic in high-stakes domains such as mortgage risk assessment, where decision-makers require accurate forecasts and actionable insights to mitigate risks effectively. The emergence of advanced time-series models, particularly the Temporal Fusion Transformer (TFT), coupled with interpretable machine learning methods like SHAP (SHapley Additive exPlanations), offers new opportunities to bridge this gap between accuracy and interpretability.

This study focuses on predicting mortgage delinquency rates in Florida by developing a robust predictive framework that combines TFT's powerful temporal modeling capabilities with the transparent insights provided by SHAP. While we explore various potential influencing factors, including mortgage characteristics, economic indicators, and weather-related variables, our analysis reveals that traditional financial and economic features remain the most reliable predictors of delinquency risk. By leveraging both TFT's predictive power and SHAP's interpretability features, we aim to provide stakeholders with accurate predictions and a clear understanding of how key economic and mortgage-specific factors drive these predictions.

1.2 Research Objectives

The primary objectives of this study are:

- Develop a robust and interpretable machine learning framework for mortgage delinquency prediction by combining the predictive power of the Temporal Fusion Transformer with SHAP-based interpretability
- Design a systematic approach for feature selection and importance ranking in time-series models through feature-wise fidelity R^2 scores and adjusted SHAP values
- Evaluate the relative importance of various predictive factors, with a particular focus on traditional financial indicators, including economic and mortgage-specific variables
- Create a practical and scalable framework that provides both accurate predictions and actionable insights to support decision-making in mortgage risk assessment

1.3 Contributions

This research advances the field through the following key innovations:

- A novel framework integrating Temporal Fusion Transformer with SHAP-based interpretability for mortgage delinquency prediction, demonstrating that complex time-series models can provide both high accuracy and transparent insights
- An innovative feature selection methodology using feature-wise fidelity R^2 scores and adjusted SHAP values to identify and rank essential features in time-series models
- A systematic evaluation of mortgage delinquency predictors that empirically confirms the dominance of traditional financial and borrower-level variables
- A practical blueprint for financial institutions to implement interpretable machine learning in risk assessment, balancing predictive power with stakeholder understanding

1.4 Thesis Overview

The remainder of this thesis is organized as follows:

Chapter 2 reviews relevant literature and establishes the theoretical foundations for mortgage risk assessment, time-series modeling, and machine learning interpretability. Particular attention is paid to recent developments in the Temporal Fusion Transformer architecture and SHAP-based interpretation methods.

Chapter 3 describes our comprehensive data methodology, detailing the collection and preprocessing of mortgage, economic, and supplementary datasets. This chapter also introduces our novel approach to feature selection and the framework for evaluating feature importance.

Chapter 4 presents model development and results, introducing our integrated framework, which combines TFT modeling with SHAP-based interpretability. The chapter details the model architecture, training process, and performance metrics achieved.

Chapter 5 provides an in-depth analysis of our findings, focusing particularly on the identified key predictors of mortgage delinquency and their relative importance. This chapter also examines the effectiveness of our feature selection methodology and the model's interpretability.

Chapter 6 concludes by synthesizing our key findings and contributions and discussing their implications for academic research and industry practice. The chapter also addresses limitations and proposes promising directions for future research in interpretable machine learning for financial risk assessment.

Chapter 2

Literature Review

2.1 Previous Work

In this chapter, the state of the art related to this work is reviewed and discussed, focusing on its application to this work. The topics include the importance of research gaps concerning mortgage risk when exposed to extreme weather events, effective statistical methods for mortgage risk modeling, Bayesian methods for modeling and forecasting extreme weather events, and combining extreme weather events with mortgage risk modeling. Each topic's state of the art and current work are critically assessed. Highlights from this literature review inform the primary methodology employed in this project.

2.1.1 Extreme Weather Events and Mortgage Risk

Exploring the impact of extreme weather events on mortgage risk represents a novel and crucial topic in mortgage risk modeling. While extensive research has been conducted on mortgage default and prepayment risks, the increasing frequency and severity of extreme weather events in recent years have heightened the relevance of this issue. Climate events can lead to significant damages and losses, undermining the value of mortgage securities and destabilizing the housing market.

Several studies have documented the impacts of extreme weather events on mortgage defaults and prepayments. Issler et al. [15] highlight the significant impact of wildfires on mortgage risk. Similarly, Kousky et al. [17], and Vigdor [33] examine the effects of hurricanes on mortgage performance. Rossi [28] finds a substantial impact of hurricane exposure on mortgage defaults, while Ouazad and Kahn [24] demonstrate how tropical cyclones influence lenders' risk perceptions and securitization behaviors. These studies emphasize the critical role of insurance and geographic variability in mortgage risk.

2.1.2 Combining Extreme Weather with Mortgage Risk Modeling

Integrating extreme weather events with mortgage risk modeling is a relatively recent focus in the field. Calabrese et al. [6] examined the influence of heavy rains and tropical cyclones on Florida's mortgage risks, using an additive Cox proportional hazard model combined with GAMs to handle nonlinear relationships and time-varying effects. This methodology enhances flexibility and provides valuable insights into the interaction between weather events and mortgage performance.

Du and Zhao [12] utilized a difference-in-differences approach to assess the impacts of Hurricanes Harvey and Maria on residential mortgages, employing loan-level data from Fannie Mae and Freddie Mac. Their analysis revealed significant differences in mortgage performance due to geographic variability and access to federal assistance. Pastor-Paz et al. [26] modeled the relationship between extreme precipitation and property damages in New Zealand, integrating climate projections into econometric analyses to estimate future financial liabilities. These methods hold promise for application in this project, particularly in understanding the complex dynamics of climate events on mortgage risk.

2.1.3 Statistical Methods for Mortgage Risk Modeling

A crucial component of this project involves analyzing and modeling mortgage risk using effective statistical methods. Research indicates that combining survival analysis and spatial statistics is a practical methodology for analyzing mortgage risk. This approach predicts outcomes over time while accounting for spatial dependencies and relationships between locations, improving prediction accuracy and capturing specific dataset patterns.

Calabrese et al. [7] developed a binary choice model incorporating neighborhood effects into disturbances to enhance mortgage default predictions. This model uses a generalized extreme value distribution to handle non-normal errors and spatial dependencies, estimated via the Geweke-Hajivassiliou-Keane algorithm. Berg et al. [3] documented the effectiveness of generalized additive models (GAMs) for mortgage risk modeling. GAMs extend generalized linear models (GLMs) by replacing the linear predictor with a sum of smooth functions, enabling the modeling of complex, nonlinear relationships between variables and outcomes.

Bayesian methods further complement these statistical approaches by offering a probabilistic framework that accounts for uncertainties and dependencies in the data. Bhattacharya et al. [4] applied Bayesian competing risk proportional hazards models to assess mortgage defaults and prepayments simultaneously. This approach captures the dynamics of competing events and integrates prior information, providing a more comprehensive understanding of mortgage risk. Köhler et al. [16] demonstrated similar methods for predicting default and prepayment risks in consumer loans, showcasing the adaptability of Bayesian approaches across different financial domains.

In addition to mortgage risk modeling, Bayesian methods have proven effective in forecasting extreme weather events. Chu and Zhao [9] reviewed hierarchical Bayesian models for event frequency prediction and generalized extreme value distributions for intensity analysis. These models have been successfully applied to forecast seasonal tropical cyclones and detect shifts in hurricane frequencies. The ability of Bayesian frameworks to continuously update predictions with new data makes them particularly relevant for integrating climate and economic factors into mortgage risk assessments.

2.1.4 Machine Learning Models for Mortgage Risks

Applying machine learning (ML) techniques have markedly advanced mortgage risk modeling, facilitating the development of predictive models that adeptly capture complex, non-linear relationships within financial datasets [8, 22]. Among ML approaches, gradient boosting and random forest algorithms are noted for their effectiveness due to their ensemble-based structures that improve predictive accuracy and robustness by aggregating multiple decision trees.

Random forest, an ensemble method based on bagging (bootstrap aggregating), leverages random subsets of data to train individual classifiers, which are combined to form a more accurate prediction model [5]. This method is particularly adept at handling overfitting, making it well-suited for mortgage default prediction where borrower characteristics, macroeconomic conditions, and geographic factors interplay. In contrast, boosting adjusts the weight of each observation, increasing it for misclassified data in each iteration, ensuring that subsequent classifiers focus more on these challenging cases. Gradient Boosting, a specific type of boosting developed by Friedman, optimizes the gradient of the loss function to minimize prediction errors [13]. This approach efficiently improves the model's accuracy by focusing iteratively on reducing the errors of previous predictions.

Previous studies validate the superiority of these methods in predicting loan defaults. In his work, Lai experimented with machine learning techniques, including boosting, random forest, k-nearest neighbor, and multilayer perceptrons. The results indicate that boosting methods exhibit better accuracy, and among all the bagging methods, random forest demonstrates the highest accuracy for loan default predictions [19]. Zhu et al. found that the Random Forest algorithm outperformed logistic regression, decision trees, and support vector machines in predicting loan defaults on large datasets, highlighting its robust generalization capabilities [34].

Furthermore, Chang et al. demonstrated the exceptional accuracy of the XGBoost model in assessing credit risk, suggesting its potential as a powerful tool for mortgage risk prediction [8]. This demonstration of the XGBoost method's excellence in default prediction is further confirmed by Odegua et al., who showcased comprehensive evaluation metrics including accuracy, recall, precision, F1-Score, and ROC area to illustrate XGBoost's performance [23].

Recent studies have also explored the application of neural networks in mortgage risk prediction, highlighting the capability of deep learning models to delineate nonlinear dynamics in mortgage borrower behavior across various economic cycles [10, 18, 2, 29]. Sadhwani et al. provide a detailed account of the neural network architecture employed, including the number of layers and the specific types of data input. Utilizing data from over 120 million U.S. mortgages spanning two decades, their work emphasizes the influence of economic factors on mortgage performance. It confirms the potential of neural networks to exceed the predictive accuracy of traditional linear models [29].

These insights collectively suggest that advanced machine learning methods, particularly those leveraging ensemble techniques and deep learning architectures, are powerful tools for enhancing the accuracy of mortgage default risk predictions. Such methods align well with this study's objectives, which seek to forecast mortgage risk predictions using data sources as comprehensive as those analyzed in the referenced research.

2.1.5 Neural Network-Based Time-Series Models

Time-series prediction models are indispensable in finance because they capture temporal dependencies and provide accurate forecasts in high-stakes domains. The Temporal Fusion Transformer (TFT) and similar architectures integrate advanced techniques, such as attention mechanisms and recurrent networks, enabling models to handle complex patterns and uncertainties inherent in financial datasets. This subsection reviews the literature on time-series prediction models, focusing on their use in financial applications.

Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks, have been widely used for time-series prediction due to their ability to model sequential data effectively. Hochreiter and Schmidhuber [14] introduced LSTMs to address the vanishing gradient problem, enabling these models to capture long-term dependencies. LSTMs have demonstrated their robustness in financial applications by accurately predicting stock prices, currency exchange rates, and credit risks, showcasing their ability to handle sequential patterns.

Attention mechanisms, first introduced by Vaswani et al. [32], have significantly enhanced the predictive power of time-series models. These mechanisms allow models to focus on specific time steps or features, improving interpretability and performance, particularly in multivariate financial datasets. Extensions such as the Multi-Head Attention mechanism have been employed to capture diverse temporal dependencies. Notably, Lim et al. [20] introduced the Temporal Fusion Transformer (TFT), a hybrid architecture that combines LSTM layers with attention mechanisms for interpretable and accurate time-series forecasting.

2.1.6 Interpretable Machine Learning

Interpretable Machine Learning (IML) is critical for understanding model behavior in high-stakes finance applications where trust, accountability, and actionable insights are paramount. This subsection reviews the two primary interpretability methods employed in this study—Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP)—and evaluates their fidelity, robustness, and limitations in financial time series contexts.

LIME LIME [27] is a model-agnostic technique designed to approximate complex models locally by training interpretable surrogate models, such as linear regressions, around individual predictions. LIME provides localized insights into feature contributions, making it particularly useful for anomaly detection or single-instance diagnostics.

However, its reliance on random perturbations to simulate feature importance often leads to inconsistent explanations, especially in multivariate or time-series data where feature dependencies are significant [1]. LIME's fidelity, or its ability to approximate the original model's behavior, is often limited, particularly for highly non-linear models. Furthermore, the lack of robustness in LIME explanations—manifested as sensitivity to input perturbations—restricts its utility in applications requiring stable and reproducible interpretations [31].

SHAP SHAP [21] offers a mathematically grounded framework based on Shapley values from cooperative game theory, providing local and global explanations. Unlike LIME, SHAP quantifies the contribution of each feature across all predictions, offering a comprehensive view of feature importance.

In multivariate time-series models, SHAP excels in fidelity by accurately capturing the model's behavior, as evidenced by high R^2 scores in not yet published studies. However, its computational demands pose challenges, particularly for large datasets or highly correlated features. While computationally efficient, the SHAP KernelExplainer [1] relies on the assumption of feature independence, which may not hold in correlated time-series contexts.

Evaluation Metrics Evaluating interpretability methods requires metrics that assess both fidelity and robustness. Fidelity measures how accurately the explanations reflect the model's decision-making [31]. Robustness quantifies the stability of explanations under perturbations, with methods such as pairwise L2-norm differences used to assess this aspect [31]. Studies show that balancing these metrics is essential for reliable and actionable interpretations, especially in dynamic contexts like financial forecasting.

Strengths and Weaknesses LIME and SHAP each exhibit unique strengths and limitations. LIME's computational efficiency and ease of use make it suitable for quick, instance-level diagnostics, but its lack of fidelity and robustness in complex data limits its applicability. SHAP provides high-fidelity, consistent explanations, but its computational cost and reliance on independence assumptions pose challenges for large datasets.

2.2 Gaps in the Literature

Despite their success, advanced models like the TFT in financial applications often face issues related to trustworthiness and interpretability. Black-box models, while delivering superior predictive performance, are frequently met with skepticism from decision-makers who are reluctant to adopt systems they cannot fully comprehend. This hesitancy underscores the importance of interpretability methods.

Models in financial time series applications must adapt to dynamic economic conditions, necessitating periodic parameter updates. Decision-makers and stakeholders naturally seek explanation methods that clarify model predictions and identify scenarios where model performance may degrade or require recalibration. Adequate explanations can enhance trust, support practical interventions, and improve decision-making.

Future directions for research in this field include developing hybrid architectures that combine predictive accuracy with interpretability and scalability. Additionally, leveraging post hoc interpretability methods such as SHAP KernelExplainer can provide actionable insights into model decisions, fostering greater transparency and stakeholder confidence in financial applications.

Despite significant advancements in interpretable machine learning (IML), key challenges remain, particularly in applying methods like SHAP KernelExplainer to multivariate time series data. This section highlights critical gaps identified in the literature and emphasizes areas where further research is needed to improve the fidelity and robustness of interpretability methods.

Feature Correlations and Independence Assumptions A primary limitation of SHAP KernelExplainer is its reliance on the assumption of feature independence. In multivariate time-series datasets, features often exhibit strong correlations and temporal dependencies. This assumption can lead to inaccuracies in feature attribution, as the interactions among features are not fully accounted for [1]. Addressing this limitation requires exploring alternative methods or enhancements to SHAP that incorporate feature dependencies.

Fidelity and Robustness Trade-offs Fidelity and robustness are essential metrics for evaluating interpretability methods. Fidelity measures how accurately the explanations align with the model's predictions, while robustness assesses the stability of explanations under input perturbations. Studies have shown a trade-off between these metrics, where high-fidelity features often exhibit lower robustness, particularly in dynamic time-series contexts [31]. This trade-off underscores the need for balanced metrics that evaluate interpretability methods holistically.

Temporal Dynamics and Explanation Drift In time-series applications, the importance of features can vary over time, necessitating methods that can capture and adapt to these temporal dynamics. Current interpretability techniques, including SHAP, often fail to account for "explanation drift," where feature contributions change due to shifts in data distributions or model behavior over time [25]. Developing metrics and methods to track and adapt to explanation drift remains an open area of research.

Computational Complexity The computational demands of SHAP, particularly in large-scale and high-dimensional datasets, present a significant challenge. While approximations like Kernel SHAP reduce complexity, they introduce trade-offs in accuracy and efficiency. Scaling SHAP-based methods to handle complex, multivariate datasets without compromising fidelity remains a critical gap [21, 11].

Evaluation Frameworks for Time-Series Data Most evaluation frameworks for interpretability methods are designed for static datasets and do not fully address the unique challenges of time-series data, such as temporal dependencies and the importance of evolving features. While useful, metrics such as fidelity and robustness need to be adapted or extended to evaluate methods in dynamic contexts better.

Addressing these challenges will advance interpretable machine learning toward more reliable and actionable methods, particularly in high-stakes applications like financial forecasting and climate risk assessment.

Chapter 3

Data

3.1 Data Collection

The study integrates mortgage data, climate records, and economic indicators to explore the interplay between natural and economic factors influencing mortgage risks. The primary data sources are proprietary databases such as **Bloomberg**, which provide financial and economic datasets, and **Black Knight's McDash servicing database**. The McDash database is a comprehensive repository of loan-level mortgage performance data, covering over 175 million unique loans, including mortgages and home equity loans. This database spans three decades and is updated within 15 days after each month's end.

Additionally, the study utilizes public data sources such as the **Fannie Mae** and **Freddie Mac** Single-Family Loan-Level Dataset, Federal Emergency Management Agency **(FEMA)** disaster declarations, and climate datasets from the National Oceanic and Atmospheric Administration **(NOAA)**.

3.2 Summary of Variables

The datasets comprise diverse features critical for modeling mortgage delinquency risk. These features are categorized into three main groups: mortgage-related variables, weather and natural disaster-related variables, and economic indicators. The final variable naming conventions and their abbreviations, as used in this study, are summarized below. These variables will be referenced throughout the Section 3.5, 4 and 5 for preliminary analysis modeling and discussion.

3.2.1 Mortgage-Related Variables

This category includes features directly tied to mortgage characteristics and borrower financial profiles:

- Florida 60-day-plus Mortgage Delinquency Rate: `FL_60`
- Florida 30-Year Mortgage Rate: *30ymtgR_fl*

- Florida Avg Escrow Monthly Payment: *FL_mnth_escrow_pmt_amt_avg*
- Florida Original Loan-to-Value Ratio: *FL_origltv*
- Florida Current Loan-to-Value Ratio: *FL_currltv*
- Florida Current to Original Loan-to-Value Ratio: *FL_curr2orgltv*
- Florida Outstanding Mortgage Balance: *FL_Bal*
- Florida Original Debt-to-Income Ratio: *FL_orig_dti*
- U.S Effective Mortgage Rate: *effmtgR_us*

3.2.2 Weather and Natural Disaster-Related Variables

This category captures climatic conditions and the presence of extreme weather events:

- Average Monthly Temperature: *AvgMonthTemp*
- Average Monthly Precipitation: *AvgMonthPrecipitation*
- Average Monthly Wind Speed: *AvgMonthWindSpeed*
- CO_2 Monthly Average Emission: *CO2MonthAvgEmission*
- Disaster Frequency: *Disaster_Count*

3.2.3 Economic Variables

This group includes macroeconomic and regional financial conditions:

- 10-Year Inflation Rate: *10yinfR*
- Florida Consumer Confidence Index: *cnsmconf_fl*
- Florida Unemployment Rate (%): *unempl_fl*
- U.S Generic 10-Year Real Yield: *10ry_us*
- U.S Consumer Credit Outstanding: *cnsmcrdtout_us*
- U.S Overnight Indexed Swap Rate: *ois_us*
- U.S Payroll Employment: *payroll_us*
- U.S Household Debt Service Ratio (Seasonality Adjusted): *serv2debtsa_us*
- U.S. Treasury Yield Difference (30-year Minus 5-year): *yc530*

These categorizations provide a structured overview of the features used in the modeling process, facilitating a clear understanding of their roles in analyzing mortgage risk. By integrating traditional financial data with environmental and economic factors, this work highlights the interdisciplinary nature of the study, offering a holistic approach to understanding and predicting mortgage delinquency.

3.3 Data Integration

Data Frequency The study addresses the misaligned frequency of data collected by various vendors by restructuring higher-frequency data (e.g., daily or weekly) into a lower frequency (monthly) to align with the least frequent data available across all sources. The average of the higher-frequency data is calculated for the period from the first to the last day of the month. This averaged data is then combined with monthly frequency data under a new date label in the format of YYYY-MM-01, representing the data collected within that month.

Data Length To ensure consistency across data sources, data with longer historical records are truncated to match the start dates of datasets with shorter histories. This alignment facilitates a uniform analysis period across all datasets.

Data Normalization Data inputs observed across various ranges are z-score normalized to scale model inputs to consistent ranges. This normalization prevents significant feature differences from dominating gradient updates in the modeling process. For instance, models like LSTM, which utilize sigmoid and tanh activations within their gates, are sensitive to input scales. Normalized inputs help these functions operate within their effective range (around 0 for sigmoid and between -1 to 1 for tanh), resulting in more accurate gate outputs.

Missing Data In cases of missing data, the nearest available data point to the current date is utilized as a proxy. This approach assumes that the most recent data provides the best estimate for missing entries, barring additional information.

Outliers Outliers are identified using a z-score threshold (e.g., $|z| > 3$). Once identified, outliers are either capped or removed depending on their impact on the dataset. For significantly large values, logarithmic scaling is applied to compress the range and mitigate the effect of right-skewed distributions commonly found in financial datasets.

3.4 Feature Overview

Geographic Integration Zip codes and FEMA disaster zones link mortgage, economic, and climate datasets to ensure accurate geographic mapping to the appropriate disaster events. This integration facilitates analyzing how localized natural disasters impact mortgage performance within specific regions.

Temporal Alignment Regional delinquency rates and mortgage data are aligned with disaster timelines to analyze such events' short-term and long-term effects on mortgage stability. Weather data is synchronized with mortgage timelines based on the first payment date, allowing the study to capture climatic conditions that prevailed throughout the loan.

Forward Looking Bias The target variable, specifically the *60-day-plus Mortgage Delinquency Rate*, is shifted by one month to mitigate forward-looking bias inherent in time series data. This adjustment helps maintain the integrity of the cause-and-effect analysis in the predictive modeling process.

Data Categorization Data types such as zip codes and station IDs are transformed into categorical variables to serve as binary inputs in the modeling process. This transformation is crucial for models that require nominal rather than numeric inputs, ensuring more precise data handling and analysis.

3.5 Data Preliminary Analysis

This section, detailed further in Sec. 3.5.1, 3.5.2, and 3.5.3, conducts preliminary analyses of mortgage, climate, and economic datasets to identify key variables and discern patterns. These analyses lay the groundwork for feature engineering and selection, which is critical for building robust predictive models. Each segment explores initial findings, transforms data into usable formats, and discusses the rationale for feature selection. This approach ensures that only the most accurate and impactful data are used in the modeling process.

3.5.1 Mortgage Data

The mortgage dataset is obtained from Black Knight's McDash servicing database and the Fannie Mae and Freddie Mac Single-Family Loan-Level Dataset, providing a comprehensive view of loan performance. This data encompasses a range of borrower characteristics and geographic vulnerabilities critical for assessing risk profiles. Key features extracted from this dataset include *Florida Avg Escrow Monthly Payment* included in Mortgage Payments, which indicates that higher taxes and insurance costs are correlated with increased mortgage delinquencies as shown in Fig. 3.1. Analysis reveals that borrowers experiencing a significant increase in escrow payments tend to exhibit a higher rate of delinquency, particularly 1.5% of borrowers with a 40-50% increase in escrow payments become delinquent by the 6-month mark compared to only 1% in the 5-10% increase bracket.

The analysis also considers additional mortgage features such as *Florida Original Debt-to-Income Ratio*, *Florida Outstanding Mortgage Balance*, *Florida Original Loan-to-Value Ratio*, and *Florida Current Loan-to-Value Ratio*. These variables underwent log transformation to stabilize variance across the dataset, ensuring that the model could effectively interpret and utilize the range of data without distortion and to help keep the model performance much more stable over time. The results after log transformation are shown in Fig. 3.2.

As shown in Figs. 3.3 and 3.4, an Ordinary Least Squares (OLS) regression analysis with log-transformed features reveals a strong model fit, with an R-squared value of 88.9%, indicating that these features significantly predict the *Florida 60-day-plus Mortgage Delinquency Rate*. Initial regression outcomes guided the selection of the most predictive features using forward stepwise selection, notably omitting the less significant *Current to Original LTV Difference* variable due to its crossing the zero confidence interval. The retained features substantiate the analytical framework that underpins the risk assessment model.

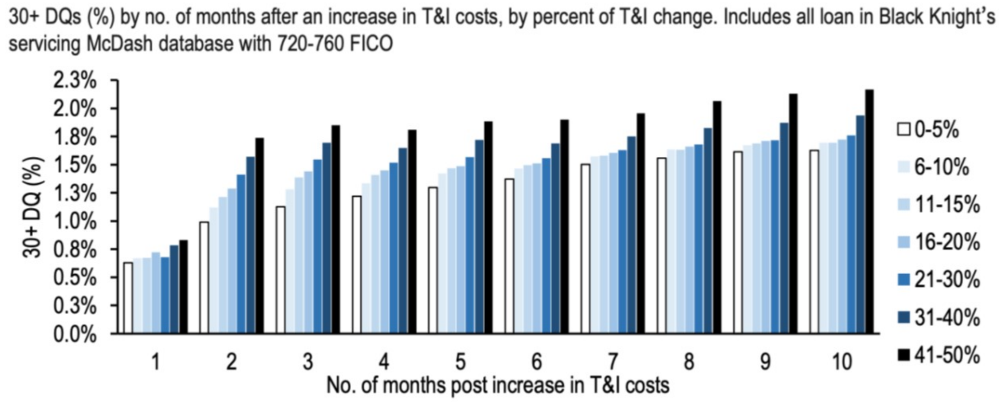

Figure 3.1: Delinquency rates increase with higher escrow payment increments. (Data source: Black Knight)

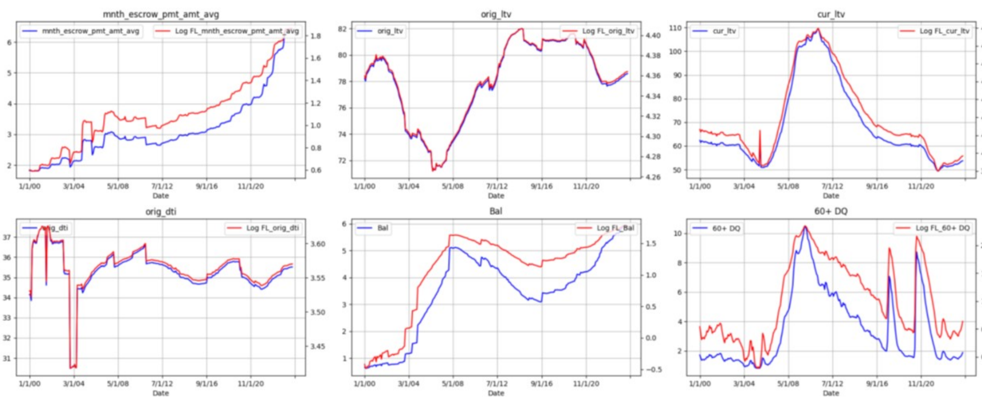

Figure 3.2: Log transformation of mortgage features to cap the numbers within a tighter range without changing the underlying meaning of the features. (Data source: Bloomberg)

3.5.2 Climate Data

This analysis examines the influence of climate variables on mortgage risk, focusing on temperature, carbon dioxide levels, precipitation, wind speed, and disaster frequency. These climate factors are considered for their potential impact on mortgage delinquency rates, reflecting that environmental conditions can significantly affect mortgage stability.

Climate data preparation involved compiling daily temperature measurements from The **Florida Climate Center** to calculate the *Average Monthly Temperature*. This metric provides insights into temperature fluctuations across the state, which are crucial for understanding seasonal impacts on mortgage delinquency. Similarly, *Average Monthly Precipitation* and *Average Monthly Wind Speed* were derived from daily data to ensure a comprehensive analysis of climatic conditions. In contrast, global CO_2 emission averages were sourced directly from the **Our World in Data Website**, based on the assumption that climate change impacts are globally interrelated and influence local mortgage markets. Additionally, a *Disaster Count* was developed by aggregating total disaster events reported monthly by FEMA and NOAA, which helps correlate significant weather events with changes in mortgage delin-

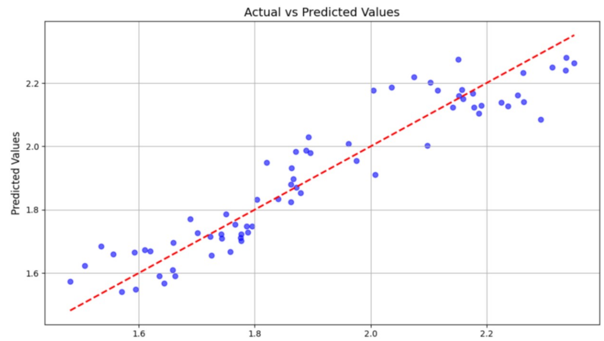

Figure 3.3: Preliminary OLS regression analysis demonstrating model fit.

OLS Regression Results

Dep. Variable:	FL_60+ DQ	R-squared:	0.889
Model:	OLS	Adj. R-squared:	0.878
Method:	Least Squares	F-statistic:	83.81
Date:	Sun, 08 Dec 2024	Prob (F-statistic):	4.03e-28
Time:	18:04:19	Log-Likelihood:	76.984
No. Observations:	70	AIC:	-140.0
Df Residuals:	63	BIC:	-124.2
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	71.0679	16.157	4.399	0.000	38.781	103.355
FL_mnth_escrow_pmt_amt_avg	6.6036	0.755	8.746	0.000	5.095	8.112
FL_orig_ltv	-6.5131	1.914	-3.403	0.001	-10.338	-2.688
FL_cur_ltv	3.5898	0.361	9.937	0.000	2.868	4.312
FL_orig_dti	-17.1774	3.538	-4.855	0.000	-24.248	-10.107
FL_Bal	-1.7587	0.629	-2.798	0.007	-3.015	-0.503
FL_curr2orgltv	0.0253	0.026	0.958	0.342	-0.027	0.078

Omnibus:	1.818	Durbin-Watson:	0.616
Prob(Omnibus):	0.403	Jarque-Bera (JB):	1.633
Skew:	-0.252	Prob(JB):	0.442
Kurtosis:	2.446	Cond. No.	1.31e+04

Figure 3.4: Insignificance of the FL_curr2orgltv feature in model fitting.

quency rates.

Feature selection revealed that average monthly temperature, wind speed, precipitation, and disaster frequency are directly associated with mortgage delinquency rates. This association is particularly relevant in regions like Florida, where hurricanes and floods frequently cause substantial property damage and thus impact mortgage defaults. Although *CO_2 Monthly Average Emissions* showed a weaker correlation with delinquency rates, they are considered valuable for enhancing model performance when combined with other variables or through advanced analytical techniques.

The findings underscore the significant impact of precipitation, *CO_2 Monthly Average Emissions*, and *Disaster Frequency* on mortgage delinquency rates, advocating for their prioritization in predictive models. While *Average Monthly Wind Speed* exhibited a lower correlation, its inclusion could improve model performance through complex transformations. This detailed analysis of climate variables enables fine-tuning predictive models, allowing for more accurate predictions of climate-related mortgage risks and facilitating effective risk management strategies.

3.5.3 Economic Data

The analysis incorporates a range of economic indicators to understand how market and localized economic conditions affect mortgage risks. Macroeconomic data, including *Florida Consumer Confidence*, *U.S. Overnight Indexed Swap Rate*, *10-Year Inflation Rate*, and *10-Year Treasury Yield*, are utilized to gauge the impact of inflation on mortgage affordability and delinquency. Additionally, financial data such as regional *Unemployment Rate*, *Original Debt-to-Income Ratio*, non-farm *Payroll Employment*, and *Treasury Yield Curve Spread (30-year minus 5-year)* are examined to explore the relationship between economic stressors and mortgage delinquency, particularly following extreme weather events.

Several techniques were employed to refine the data for analysis. For instance, first-order differencing was applied to several macroeconomic features to reduce non-stationarity, rendering them more suitable for model inputs as demonstrated in Fig. 3.5. Logarithmic scaling was also used on the U.S. service-to-debt (seasonally adjusted) feature to normalize the trends within a manageable range.

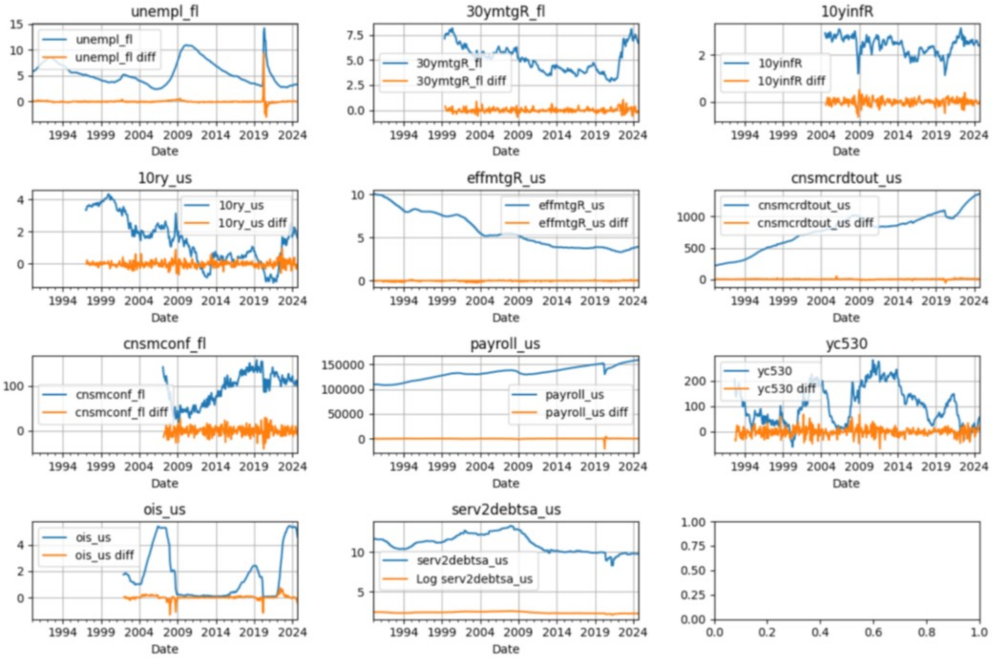

Figure 3.5: After applying the first difference, macroeconomic features exhibit increased stationarity, enhancing their utility in predictive models.

Feature selection involved exploring the correlations between each macroeconomic feature and the target variable, mortgage delinquency rates. It was found that Florida's unemployment rate and the treasury yield spread between 5-year and 30-year bonds show strong positive correlations with Florida's delinquency rates. In contrast, the Florida consumer confidence index shows a strong negative correlation (Fig. 3.6). These findings underscore the direct impact of economic conditions on mortgage performance. Features with weaker correlations, such as national outstanding consumer credit, demonstrate limited direct influence on delinquency rates but may enhance model accuracy when combined with other variables or through sophisticated analytical techniques.

These analyses inform the modeling process, identifying strongly correlated economic features as key predictors of delinquency rates. This nuanced understanding of financial impacts allows for fine-tuning predictive models, enhancing their accuracy and effectiveness in forecasting mortgage delinquencies.

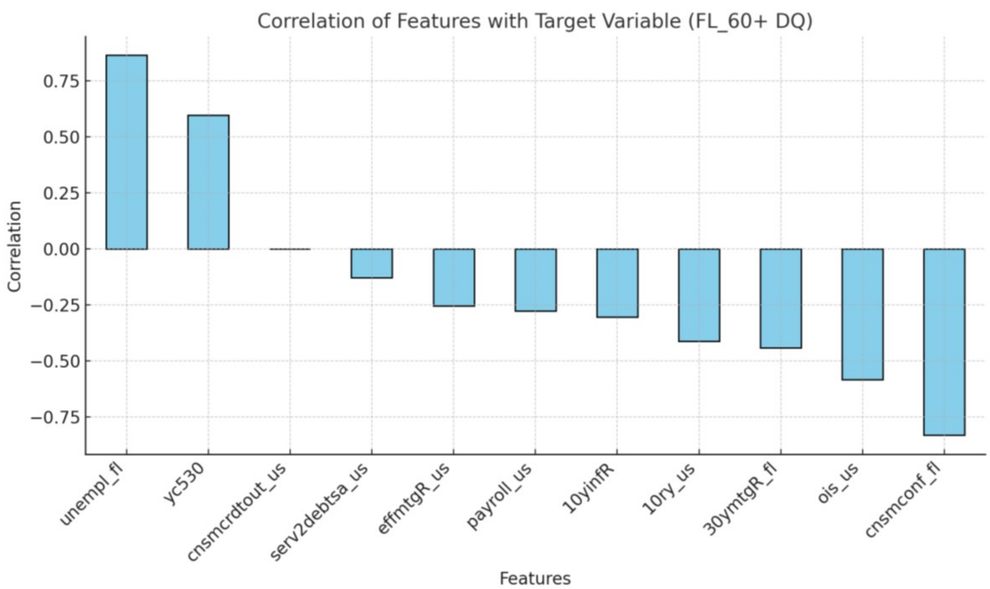

Figure 3.6: Correlation analysis showing relationships between macroeconomic indicators and Florida's 60d+ Mortgage Delinquency Rate. (Data source: Bloomberg)

Chapter 4

Model Development and Results

This section details the exploration and application of various predictive models ranging from traditional linear regression to advanced machine learning and neural network approaches. Initially, a linear regression model was employed to establish a baseline understanding of the relationships between mortgage delinquency rates and the selected features. Subsequently, more complex non-linear models, including gradient boosting and random forest algorithms, were utilized to capture non-linear interactions and improve predictive accuracy. In addition, neural network (NN) models and Temporal Fusion Transformer (TFT) were implemented to leverage their ability to process sequences and capture temporal dependencies inherent in economic and climate data. Lastly, a SHAP (SHapley Additive exPlanations) model was developed to provide interpretability, allowing for a deeper understanding of the contribution of each feature to the predictive outcomes, which is crucial for actionable insights and strategic decision-making.

4.1 Autocorrelation Analysis

Autocorrelation analysis, using the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), was conducted to examine temporal dependencies of the data, are shown in Fig. 4.1 and 4.2. These analyses are crucial for understanding the persistence of variables over time and identifying the appropriate lag structures for predictive models. ACF measures how strongly current values are related to their past values across multiple lags, while PACF isolates the direct correlation at each lag, controlling for intermediate effects. This information aids in selecting the optimal temporal features and ensuring that the models adequately capture time-dependent patterns within the dataset.

The ACF analysis was performed on all variables in the dataset, with lags set to 20. The results, visualized in the accompanying plot, reveal varying degrees of autocorrelation among the variables. Variables such as *Average Monthly Temperature* and *Payroll Employment* exhibit a slow decline in autocorrelation, indicating strong persistence and long-term dependencies. Conversely, variables like *Average Monthly Wind Speed*, *Disaster Frequency* display weaker autocorrelations, suggesting that past observations less influence their values. Mortgage-related variables, including loan-to-value ratios and fund measures, demonstrate diverse autocorrelation patterns, potentially reflecting economic cycles and the broader financial conditions that affect these metrics over time.

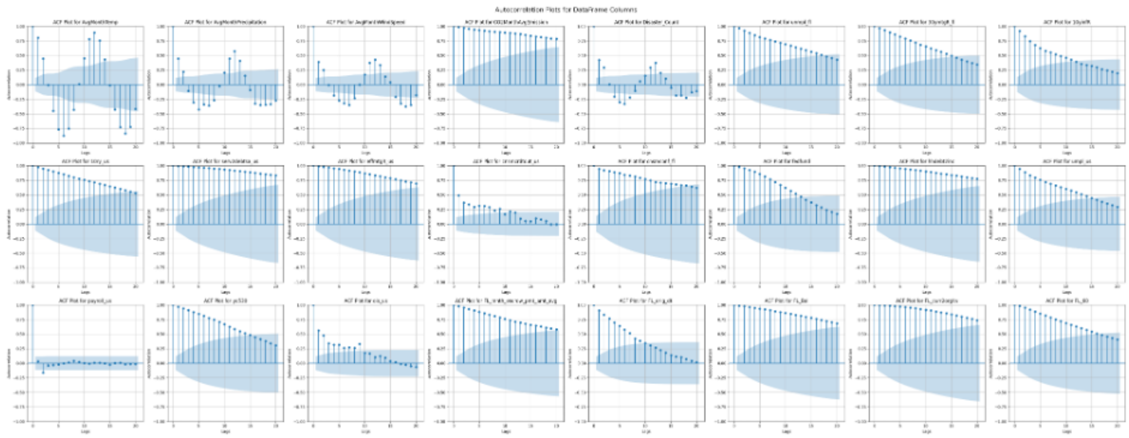

Figure 4.1: Autocorrelation plots for variables in dataframe

The PACF analysis, conducted with lags set to 36, provides further insights into the direct dependencies of variables. Variables like *Average Monthly Temperature* and *Payroll Employment* show relatively significant partial autocorrelations beyond the immediate lags, suggesting that more recent values are directly influential, while older values are less impactful. Most variables stabilize as partial autocorrelations approach zero after the initial few lags, indicating diminishing direct effects from older observations.

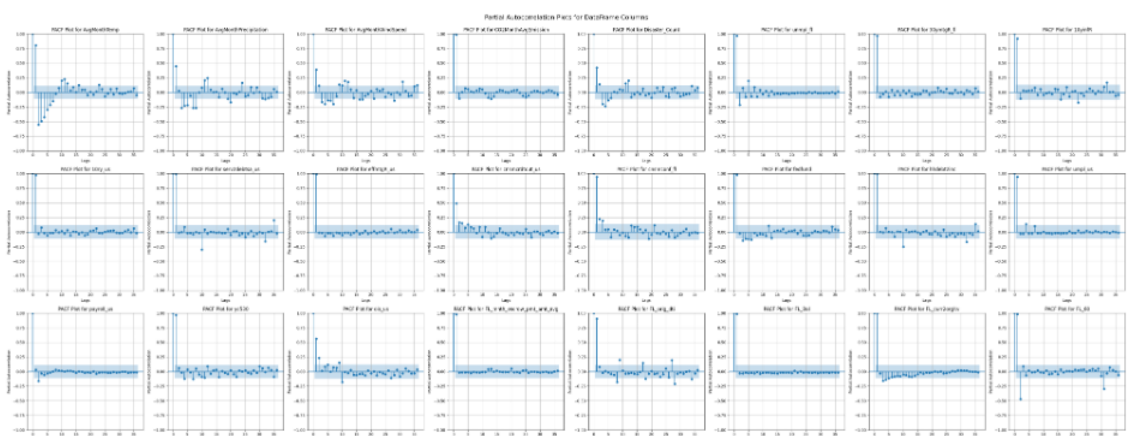

Figure 4.2: Partial autocorrelation plots for variables in dataframe

These findings underscore the importance of accounting for temporal structures in modeling. Variables with strong autocorrelations, such as *Average Monthly Precipitation*, require additional treatment for lag terms to capture their persistence. In contrast, variables with weaker temporal dependencies, like *Disaster Frequency*, may not benefit from extensive lagged features. By leveraging ACF and PACF analyses, the temporal dynamics of the dataset are adequately incorporated into the predictive modeling framework.

4.2 Variables and Modeling Strategy

The primary dependent variable in this study is *Florida 60-day-plus Mortgage Delinquency Rate*, which serves as the target variable (y) for predictive models. To enhance interpretability and performance, the delinquency rate was normalized using min-max normalization, scaling values to a 0–1 range. This normalization addresses skewness in delinquency durations and makes the data suitable for machine learning algorithms. Independent variables include mortgage, economic, and weather data, encompassing many factors influencing mortgage risks. All variables are listed in Section 3.2.

In preparation for modeling, the data from the mortgage and weather datasets were augmented and merged based on the "Year" and "Month" columns to ensure chronological alignment and consistency. This careful alignment is crucial for accurately analyzing the temporal dynamics of mortgage delinquencies. Additionally, the normalized delinquency score was shifted by one month to account for economic lags potentially influencing financial outcomes. This shift optimizes the dataset for predictive modeling, enhancing the evaluation of how past economic and environmental conditions impact future delinquency rates.

Each modeling method implemented in this study constructs two models for comparative analysis: one utilizing variables solely from the mortgage and economic datasets and another incorporating additional variables from the weather dataset. This dual-model approach allows for a direct assessment of the impact of weather-related variables on the predictability of mortgage delinquencies, aiming to delineate the contribution of environmental factors relative to traditional economic and mortgage-related factors in predicting mortgage delinquency rates.

4.3 Linear Regression: Baseline Analysis

Linear regression was the foundational model for understanding fundamental relationships within the dataset. The analysis included two models: one that exclusively utilized financial and macroeconomic predictors and another incorporating additional weather-related factors. These models were trained on an 80/20 data split and standardized to address scale biases. Their performances were evaluated using Mean Squared Error (MSE) and R-squared values.

The model excluding weather variables recorded an MSE of 0.0073 and an R-squared of 0.85. However, incorporating weather variables slightly increased the MSE to 0.0074 with no change in the R-squared value. This result suggests that linear regression, including weather-related variables, did not enhance predictive accuracy. However, it highlights the need to explore more advanced modeling techniques to integrate environmental factors into mortgage risk assessments effectively.

Feature importance analysis revealed that, in the absence of weather features, the primary determinants of delinquency risk were the regional *30-Year Mortgage Rate* and *Unemployment Rate*, the U.S *10-Year Treasury Yield*, and the *Servicer To Debt Service Ratio*. However, with the inclusion of weather variables, *CO_2 Monthly Average Emission*, ranked as the eighth most significant contributor, emerged as important. The top seven parameters remained economic and mortgage-related. The basic linear regression model underscores the significance of traditional economic indicators in predicting mortgage delinquency, and the addition of weather variables illustrates their potential to enhance predictive accuracy by accounting for environmental factors that also impact delinquency risks.

4.4 Machine Learning Models: Gradient Boosting and Random Forest

Building on the insights gained from linear regression, advanced machine learning models such as Gradient Boosting, implemented via XGBoost, and Random Forest were employed to capture non-linear relationships better and improve predictive accuracy. These ensemble-based approaches are ideally suited for handling datasets that exhibit complex interactions among predictors. By leveraging multiple decision trees to form a more robust model, both Gradient Boosting and Random Forest techniques enhance the ability to discern subtle patterns and variations within the data, thereby offering a more nuanced understanding of the factors influencing mortgage delinquency rates.

For the Gradient Boosting model, the configuration excluding weather features achieved a Mean Squared Error (MSE) of 0.0054 and an R-squared value of 0.89, demonstrating robust predictive accuracy. When weather variables were included, there was a notable improvement in performance, with the MSE decreasing to 0.0044 and the R-squared rising to 0.91. This suggests that environmental factors contribute significantly to the predictability of mortgage delinquency rates. Feature importance analysis for the XGBoost model without weather features highlighted the *Unemployment Rate*, *Servicer To Debt Service Ratio*, and *Florida Original Debt-to-Income Ratio* as the top predictors of delinquency risk. Conversely, the model with weather features showed that along with these economic indicators, the *CO_2 Monthly Average Emission* and *Disaster Frequency* ranked as the eighth and tenth, and also emerged as significant predictors, underlining the added value of including environmental data in the risk assessment models.

Like the Gradient Boosting model, Random Forest demonstrated strong predictive performance with slightly lower accuracy than XGBoost. The configuration without weather features achieved an MSE of 0.0063 and an R-squared value of 0.86, while incorporating weather features improved the MSE to 0.0062 and the R-squared to 0.87. Feature importance analysis revealed that Florida *Unemployment Rate*, *Consumer Confidence*, *Original Debt-To-Income Ratio*, and *Outstanding Mortgage Balance* were the most significant predictors without weather variables. When weather features were included, *CO_2 Monthly Average Emission*, *Disaster Frequency*, and *Average Monthly Wind Speed* were ranked between 10th and 12th, highlighting their modest but notable contribution to predicting mortgage delinquency risk.

4.5 Neural Network

To complement traditional machine learning techniques, neural networks (NNs) were employed to assess their effectiveness in predicting mortgage delinquencies. Neural networks are particularly adept at capturing complex, non-linear relationships within data, a characteristic often observed in financial datasets. Similar to previous modeling configurations, two models were implemented: one excluding weather-related variables, focusing solely on economic and macroeconomic features, and another incorporating weather-related variables to evaluate their additional predictive value.

The NN architecture included multiple dense layers to capture feature interactions, with dropout layers added to mitigate overfitting, a common challenge in deep learning. Models were trained over 100 epochs with a batch size of 32, using an 80/20 training-testing split and an additional validation split during training to optimize generalization. Standardization was applied to ensure that the scale of input features was consistent across the dataset, preventing any disproportionate influence on the training process.

Model performance was evaluated using Mean Squared Error (MSE). The NN model without weather features achieved an MSE of 0.0064, while the model including weather features resulted in a slightly higher MSE of 0.0089. This increase suggests that the added complexity of weather variables introduced non-informative variance or noise, potentially overshadowing their predictive contribution. This outcome underscores a limitation in integrating weather data into neural networks, especially when the predictive signal of these variables is weaker compared to dominant economic indicators.

In Fig. 4.3, we visually compare the actual delinquency rates alongside predictions from both neural network models. The figure highlights that both models successfully capture the overall trends in delinquency rates. However, the more complex model, which includes weather-related features, appears to better account for sharp changes and more extreme trends in the data. These findings suggest that while neural networks are adept at modeling intricate patterns within the dataset, including additional variables—particularly those with weaker predictive signals—must be carefully managed to avoid introducing unnecessary complexity or noise that may hinder overall performance.

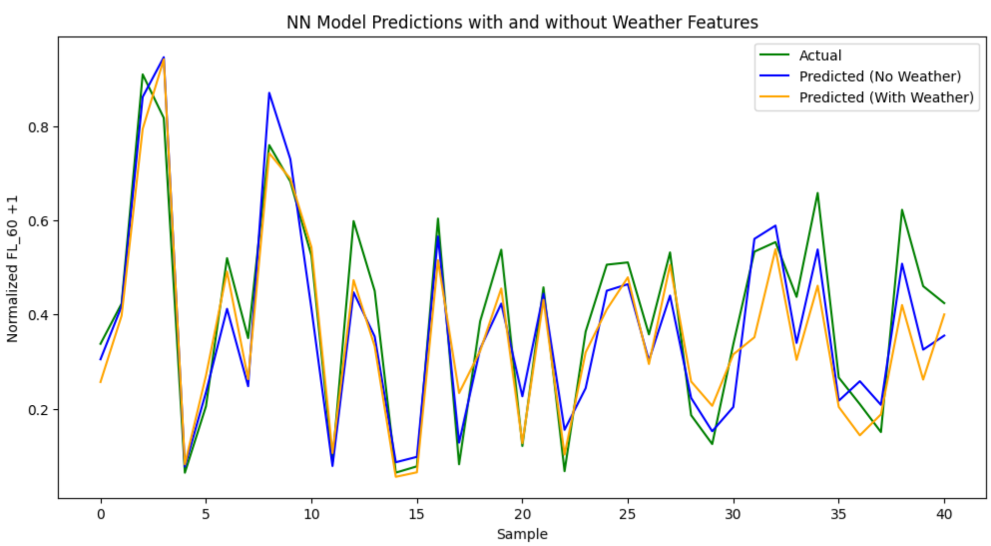

Figure 4.3: NN model predictions with and without weather features

4.6 SHAP-Enhanced Temporal Fusion Transformer

In this study, a novel framework is proposed to aid in developing a Temporal Fusion Transformer (TFT)-like model by leveraging interpretable methods, specifically SHAP KernelExplainer, for feature selection. This framework refines the feature selection process by addressing the limitations of the traditional SHAP KernelExplainer in the context of multivariate time series data, where feature correlations are prevalent.

4.6.1 TFT Model Architecture

The TFT-like model used in this study combines several advanced neural network components to effectively capture temporal dependencies, feature interactions, and seasonality effects in time-series data. The architecture is visualized in Figure 4.4 and described below.

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 12, 7)	0	-
time_distributed (TimeDistributed)	(None, 12, 64)	512	input_layer[0][0]
multi_head_attention (MultiHeadAttention)	(None, 12, 64)	66,368	time_distributed[0][0], time_distributed[0][0]
add (Add)	(None, 12, 64)	0	time_distributed[0][0], multi_head_attention[0][0]
layer_normalization (LayerNormalization)	(None, 12, 64)	128	add[0][0]
lstm (LSTM)	(None, 12, 64)	33,024	layer_normalization[0][0]
dropout_1 (Dropout)	(None, 12, 64)	0	lstm[0][0]
lstm_1 (LSTM)	(None, 32)	12,416	dropout_1[0][0]
dropout_2 (Dropout)	(None, 32)	0	lstm_1[0][0]
dense_1 (Dense)	(None, 16)	528	dropout_2[0][0]
dense_2 (Dense)	(None, 1)	17	dense_1[0][0]

Total params: 338,981 (1.29 MB)
Trainable params: 112,993 (441.38 KB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 225,988 (882.77 KB)

Figure 4.4: Architecture of the TFT-like model. The model processes sequences of length 12 with seven features.

Key Components of the Model

- **Time-Distributed Dense Layer**: This layer applies a fully connected network independently to each time step of the input sequence. Encoding temporal information at each time step transforms the original seven features into 64-dimensional feature vectors.
- **Multi-Head Attention**: Multi-head attention is applied to model feature interactions and temporal dependencies across time steps. Focusing on specific time steps and relationships improves both interpretability and performance.

- **Residual Connection and Layer Normalization**: A residual connection between the time-distributed layer and the attention output ensures stable gradient flow. In contrast, layer normalization standardizes the input to subsequent layers, improving training efficiency.
- **LSTM Layers**: Two LSTM layers, with 64 and 32 units, respectively, capture long-term dependencies and sequential patterns. These layers are essential for modeling time-series data, allowing the model to learn temporal structures effectively.
- **Dropout Layers**: Dropout layers are applied after each LSTM layer to prevent overfitting by randomly deactivating a fraction of the neurons during training.
- **Dense Layers**: The final fully connected layers map the learned features to the target variable. The first dense layer reduces the feature space to 16 dimensions, while the output layer provides the final prediction.

Seasonality Considerations The model uses a 12-input sequence length determined by autocorrelation (ACF) and partial autocorrelation (PACF) analyses. This sequence length is sufficient to capture the effects of seasonality and temporal dependencies in the dataset, ensuring the model can fully consider patterns recurring over a 12-time-step horizon.

4.6.2 Fidelity-Driven Feature Ranking and Selection

To address the limitations of traditional SHAP KernelExplainer, this study adopts a novel feature ranking method inspired by recent insights from Stone (2024), an unpublished work[30]. The framework evaluates feature importance by isolating individual features and mitigating the effects of correlations through a masking strategy. The steps are as follows:

1. For each feature in the dataset:
   - Preserve the original values of the selected feature.
   - Mask all other features by replacing their values with a baseline value (e.g., their mean).
2. Run the SHAP KernelExplainer on the modified dataset to compute feature attributions.
3. Evaluate the fidelity of the explanations by calculating the R^2 score, which measures how well the SHAP values reconstruct the model's predictions on the manipulated dataset.

This approach iterates over all features and effectively removes the correlations between them, allowing a more accurate assessment of individual feature importance. The features are then ranked based on their $R2$ fidelity scores, providing a more precise and reliable feature ranking.

The proposed framework offers several advantages:

- **Improved Interpretability**: By isolating features, the method avoids the confounding effects of correlations, leading to more accurate and interpretable rankings.

- **Fidelity as a Metric**: Using R^2 fidelity scores provides a quantitative and robust measure of feature importance, which aligns better with domain knowledge and model behavior, as also suggested by Stone in a recent unpublished study (2024)[30].

- **Enhanced Feature Selection**: The resulting rankings help shortlist relevant features for the final model, reducing noise and improving model performance.

As an illustration, Figure 4.5 displays the R^2 scores for more features trained with the model, and Figure 4.6 highlights the final shortlisted features.

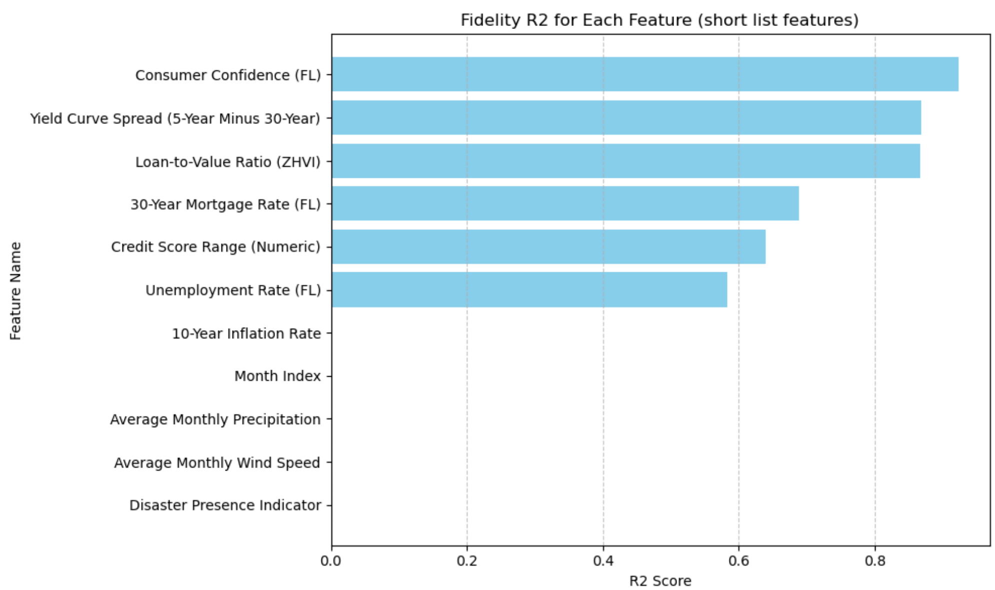

Figure 4.5: Fidelity R^2 Scores for All Features in the Dataset.

As shown in Figure 4.5, some features, such as the *Month Index*, exhibited near-zero R^2 fidelity scores, indicating no significant role in the model's predictions. This is consistent with the role of *Month Index* as a seasonal indicator in other models; however, for the TFT model, the sliding window of sequence length 12 inherently captures seasonality, rendering this feature redundant.

In contrast, Figure 4.6 displays the six shortlisted features with the highest fidelity R^2 scores. These features are critical to the model's predictions and include:

- Loan-to-Value Ratio (ZHVI)

- Consumer Confidence (FL)

- Yield Curve Spread (5-Year Minus 30-Year)

- 30-Year Mortgage Rate (FL)

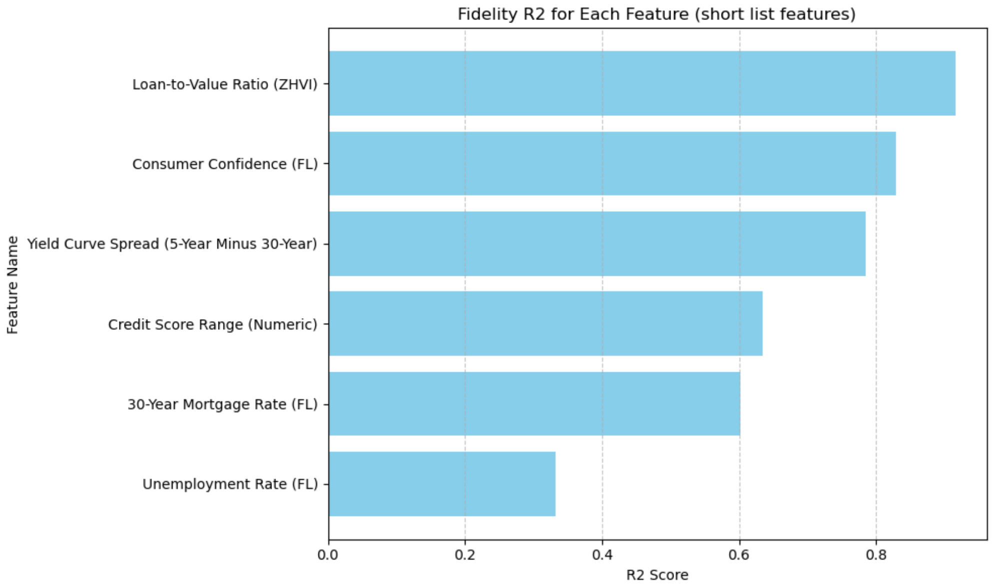

Figure 4.6: Fidelity R^2 Scores for Shortlisted Features.

- Credit Score Range (Numeric)
- Unemployment Rate (FL)

The feature selection process ensures that only relevant features are retained, simplifying the model without sacrificing predictive accuracy. Notably, after pruning the features to the final six, the model's performance metrics remained consistent, confirming the effectiveness of the selection process.

4.6.3 Performance Evaluation of the TFT Model and SHAP KernelExplainer

The performance evaluation of the TFT model under SHAP KernelExplainer enhancement is assessed through the following two key aspects:

- Accuracy of the TFT Model Predictions: This evaluates how closely the TFT model predictions align with the observed target variable, i.e., the delinquency rate in Florida. Metrics such as Mean Squared Error (MSE) and R^2 score measure the prediction model's performance. Additionally, residual analysis is implemented as a complementary tool to detect potential biases and validate performance.
- Effectiveness of the SHAP KernelExplainer: This determines how accurately the SHAP KernelExplainer reflects the decision-making process of the TFT model. Fidelity is assessed by calculating the MSE and R^2 score between the model predictions and SHAP-reconstructed target values to ensure reliable and trustworthy explanations.

TFT Prediction Model Performance

As confirmed by the evaluation metrics, the TFT model, enhanced by the novel feature selection procedure, demonstrates superior performance with only six critical features. The residual plot (Figure 4.7) further validates the model's performance by showcasing the homogeneity of residuals and the absence of systematic bias.

- **Performance of the TFT Model on Test Data**:
  - R^2 Score: 0.9961
  - Mean Squared Error (MSE): 0.0000

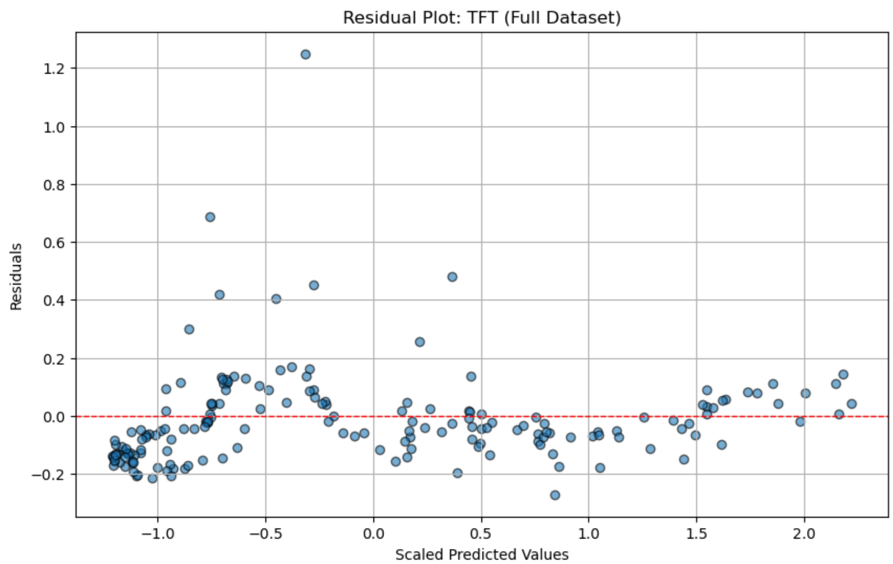

Figure 4.7: Residual Plot: TFT Model (Full Dataset).

SHAP KernelExplainer-Based Explanation Method

The effectiveness of the SHAP KernelExplainer was evaluated to ensure alignment between the TFT model predictions and the SHAP-reconstructed predictions. The fidelity evaluation demonstrates that the SHAP KernelExplainer can explain 96.53% of the variance in the model predictions, providing a reliable and trustworthy explanation of the model's decision-making process.

- **Fidelity of SHAP KernelExplainer**:
  - R^2 Score: 0.9653
  - Mean Squared Error (MSE): 0.0003

Finally, Figure 4.8 compares the TFT model predictions, SHAP-reconstructed delinquency rates, and observed delinquency rates. As shown in the plot, all three lines align closely, demonstrating the superior performance of both the prediction model and the explanation method.

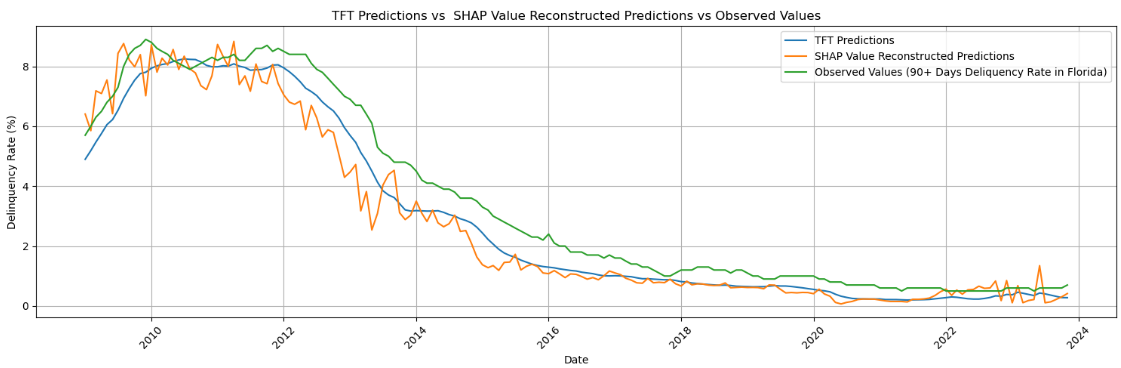

Figure 4.8: Comparison of TFT Predictions, SHAP-Reconstructed Predictions, and Observed Delinquency Rates.

Chapter 5

Analysis and Discussion

5.1 Model Performance Overview

The performance of various models used in Chapter 4 highlights their ability to predict mortgage delinquency rates with differing levels of accuracy and complexity. The performance of each model is summarized in Table 5.1.

Table 5.1: Model Performance Summary

Method	**R-squared**	**MSE**
Excluding Weather Variables		
Linear Regression	0.85	0.0073
Gradient Boosting	0.89	0.0054
Random Forest	0.86	0.0063
Neural Networks	N/A	0.0064
TFT Model	0.9961	0.0000
Including Weather Variables		
Linear Regression	0.85	0.0074
Gradient Boosting	0.91	0.0044
Random Forest	0.87	0.0062
Neural Networks	N/A	0.00809
TFT Model	0.9945	0.0001

Linear regression, the baseline model, achieved an R-squared of 0.85 with and without weather features, indicating a strong but limited capacity to capture non-linear relationships in the data. Gradient Boosting, implemented via XGBoost, showed significant improvement, with an R-squared increasing from 0.89 to 0.91 and the Mean Squared Error (MSE) decreasing from 0.0054 to 0.0044 when weather features were included. Random Forest models similarly exhibited strong performance, with an R-squared improving from 0.86 to 0.87 and an MSE reduction from 0.0063 to 0.0062 upon incorporating weather-related variables. Neural Networks demonstrated the ability to model complex interactions, achieving an MSE of 0.0507 without weather variables and a slightly higher MSE of 0.0809 when weather features were added, suggesting challenges with integrating weather-related noise.

The Temporal Fusion Transformer (TFT) model provided additional insights into the importance of features under the SHAP-based feature selection framework. Based on six key features, the final shortlisted version demonstrated near-perfect alignment with test data predictions, achieving an R^2 score of 0.9961 and an MSE close to zero. These shortlisted features, comprising economic and mortgage-related variables such as *Consumer Confidence*, *Yield Curve Spread (5-year Minus 30-year)*, and *30-Year Mortgage Rate*, underscore the dominance of traditional financial indicators in driving delinquency risk predictions.

Advanced interpretability methods, such as SHAP, by definition, guarantee perfect fidelity. However, approximate methods like SHAP KernelExplainer may involve a practical trade-off between fidelity and complexity. As discussed in the previous section, SHAP KernelExplainer requires the feature independence assumption, which does not always hold in multivariate time-series applications. However, with the assistance of the novel masking approach discussed earlier, the explanations provided by SHAP KernelExplainer demonstrated high fidelity in reconstructing model predictions. Specifically, for the entire dataset, the SHAP KernelExplainer achieved an R^2 score of 0.9653 and an MSE of 0.0003 when comparing SHAP-reconstructed predictions to the original model predictions.

This evaluation reflects the credibility and reliability of the SHAP KernelExplainer, which was guided by our feature selection procedure. SHAP's interpretable setup enhances the model by grounding it in meaningful feature contributions. The next chapter will present the results from the proposed interpretable method framework.

5.2 Insights into the Important Features

Feature importance analysis across various models provides valuable insights into the factors driving mortgage delinquency predictions. The results consistently highlight the significant role of economic and mortgage-related variables, with weather-related features contributing to a lesser but non-negligible extent in some cases.

5.2.1 Interpreting SHAP KernelExplainer Results for the TFT Model

This section interprets the SHAP KernelExplainer results for the TFT model applied to predicting the Florida delinquency rate. The SHAP summary plot (Figure 5.1) illustrates the impact of six critical features on the model's output, providing insights into their contributions to delinquency risk predictions.

Feature Contributions to the TFT Model

The SHAP summary plot provides the following key insights into the contributions of each feature to the TFT model's predictions:

- **Loan-to-Value Ratio (ZHVI)**: Higher loan-to-value ratios (red points) are associated with increased delinquency risk, as indicated by positive SHAP values. This aligns with financial theory, where higher leverage increases the likelihood of mortgage default.

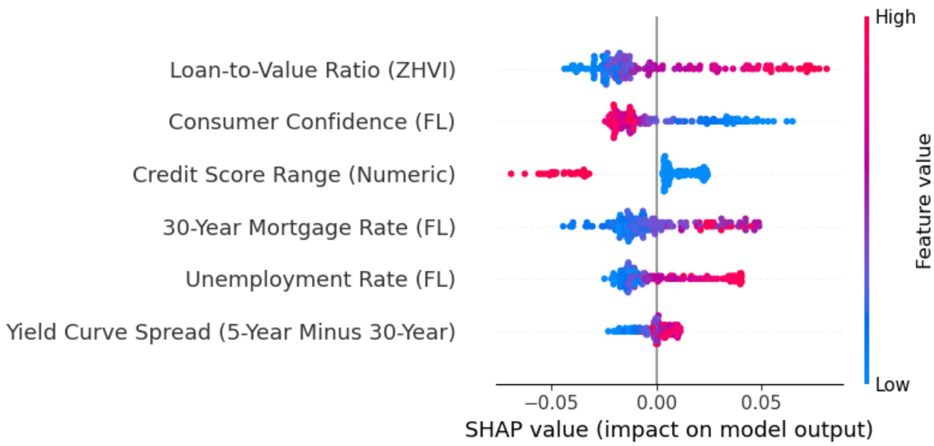

Figure 5.1: SHAP Summary Plot for Critical Features Impacting Florida Delinquency Rate. The x-axis represents the SHAP value (impact on model output), while the color gradient indicates feature values (blue for low and red for high values).

- **Consumer Confidence (FL)**: Lower consumer confidence (blue points) strongly correlates with higher delinquency risk, as reflected by negative SHAP values. This captures the economic stress experienced by households during periods of low confidence.
- **Credit Score Range (Numeric)**: Low credit scores (blue points) significantly contribute to an increased delinquency rate, highlighting the vulnerability of borrowers with lower creditworthiness.
- **30-Year Mortgage Rate (FL)**: Higher mortgage rates (red points) positively impact delinquency risk, as they increase the cost of borrowing and strain household finances.
- **Unemployment Rate (FL)**: Elevated unemployment rates (red points) are strongly associated with higher delinquency risk, reflecting the direct economic impact of job loss on mortgage repayment ability.
- **Yield Curve Spread (5-Year Minus 30-Year)**: A lower yield curve spread (blue points) correlates with increased delinquency risk. This reflects macroeconomic conditions, as an inverted or flat yield curve often signals economic downturns.

Domain-Specific Insights from SHAP Results

The SHAP results align with theoretical expectations while offering practical insights specific to the Florida delinquency rate:

- **Economic Indicators**: Features such as *Consumer Confidence*, *Unemployment Rate*, and *Yield Curve Spread* reflect broader economic conditions that influence mortgage delinquencies. These indicators become critical during periods of economic uncertainty.
- **Mortgage-Related Variables**: Features like *Loan-to-Value Ratio*, *Credit Score Range*, and *30-Year Mortgage Rate* directly measure financial stressors on individual borrowers, highlighting the model's ability to detect household-level risk factors.

- **Florida-Specific Trends**: Florida's economy is susceptible to economic cycles and housing market dynamics. The strong impact of consumer confidence and mortgage-related variables illustrates the state's vulnerability to economic downturns and housing instability.

Temporal Dynamics in SHAP Contributions

The SHAP summary plot also highlights the temporal variation in feature contributions, a common phenomenon in multivariate applications. For instance, while the *Loan-to-Value Ratio* remains the most influential predictor overall, the long tails on the right side of the plot indicate that its impact fluctuates significantly over time.

To account for both average importance and stability over time, we introduce an adjusted SHAP value metric inspired by recent research by Stone (2024)[30]. This metric adjusts the SHAP value by dividing the mean absolute SHAP value by the standard deviation of the SHAP values for each feature:

$$\text{Adjusted SHAP Value} = \frac{\text{Mean Absolute SHAP Value}}{\text{Standard Deviation of SHAP Values}}$$

This approach ranks features not only by their average contribution but also by the consistency of their impact over time. The adjusted SHAP rankings for the six critical features are shown below:

Table 5.2: Adjusted SHAP Values for Critical Features in TFT Model (Full Dataset).

Feature	**Adjusted SHAP Value**
Consumer Confidence (FL)	0.849
Unemployment Rate (FL)	0.841
Loan-to-Value Ratio (ZHVI)	0.822
30-Year Mortgage Rate (FL)	0.799
Credit Score Range (Numeric)	0.769
Yield Curve Spread (5-Year Minus 30-Year)	0.759

Key Insights from Adjusted SHAP Analysis

The adjusted SHAP results illustrate the following:

- Features directly tied to individual financial conditions, such as *Consumer Confidence* and *Unemployment Rate*, are the most important and stable over time.
- Localized borrower-level variables, like *Loan-to-Value Ratio* and *Credit Score Range*, retain significant predictive power but exhibit slightly greater variability.
- Broader economic indicators, such as the *Yield Curve Spread*, play a secondary but complementary role in predicting delinquency risk.

This analysis underscores the importance of balancing interpretability with stability in multivariate time-series applications. By prioritizing features with high average contributions and temporal consistency, the adjusted SHAP framework enhances the robustness and reliability of the TFT model.

5.3 Discussion

This section discusses the strengths, limitations, and potential future developments of the SHAP-enhanced Temporal Fusion Transformer (TFT) model for predicting Florida delinquency rates.

5.3.1 Strengths

The SHAP-enhanced TFT model demonstrates the following key strengths:

- **Superior Predictive Performance**: Achieving an R^2 score of 0.9961 and near-zero MSE, the model accurately captures complex temporal and multivariate relationships, making it highly reliable for financial forecasting.

- **High-Fidelity Explanations and Stability**: SHAP KernelExplainer provides interpretability by revealing the contributions of critical features such as *Consumer Confidence* and *Loan-to-Value Ratio*. Furthermore, introducing the adjusted SHAP metric ensures that features are ranked by average importance and stability over time. This innovative metric highlights the most consistent and impactful features, such as *Consumer Confidence* and *Unemployment Rate*, which play critical roles in driving predictions.

5.3.2 Limitations

Despite its strengths, the model has some limitations:

- **Sensitivity to Outliers**: The residual plot reveals a few outliers during special periods, such as the COVID-19 pandemic, where the model's performance may degrade. While the impact of outliers has been minimized through careful feature selection and engineering, further refinement is needed to handle such anomalies more effectively.

5.3.3 Future Developments

To address these limitations and enhance the model further, the following directions are proposed:

- **Hybrid Modeling Approach**: Integrate Bayesian frameworks to develop a hybrid model that can switch between tailored sub-models for normal and abnormal periods, improving robustness during economic shocks.

- **Dynamic Adjustment Mechanisms**: Explore methods to dynamically adjust the model for evolving economic conditions, leveraging external real-time indicators.

Chapter 6

Conclusion

This study explored multiple predictive models, including linear regression, Gradient Boosting, and Random Forest, and finally presented a comprehensive framework combining the Temporal Fusion Transformer (TFT) model with SHAP-based interpretability to predict and analyze delinquency rates in Florida. We experimented with various datasets, including macroeconomic, mortgage, and weather data, to develop a highly predictive model for delinquency rates. By leveraging the proposed framework and the presented TFT model, we achieved high predictive accuracy and transparent insights into the contributions of key features, advancing the application of interpretable machine learning in mortgage risk prediction.

6.1 Main Contributions

The primary contributions of this project are as follows:

- **Development of a SHAP-Enhanced TFT Model**: We proposed a robust TFT model guided by a SHAP-based feature selection framework, integrating interpretable machine learning with advanced time-series modeling.
- **Novel Use of Enhanced SHAP Explainer for Feature Selection**: By introducing the feature-wise R^2 fidelity score as a metric, we provided a framework and pipeline for time-series (black-box) model development under the guidance of SHAP explanations. The framework also serves as a debugging tool to iterate and refine the development cycle.
- **High-Fidelity Explanations**: Despite the limitations of SHAP KernelExplainer regarding feature independence, our framework demonstrated its ability to provide reliable and trustworthy explanations, achieving an R^2 fidelity score of 0.9653 and an MSE of 0.0003 for SHAP-reconstructed predictions.
- **Streamlined Feature Set**: The SHAP-guided feature selection process reduced the feature set to six critical variables while maintaining the model's performance. Additionally, the novel adjusted SHAP metric provided insights into features that impact the model and exhibit stable contributions over time.

6.2 Main Findings

The results of this study revealed the following key insights:

- Economic and mortgage-related variables, such as *Consumer Confidence*, *Unemployment Rate*, and *Loan-to-Value Ratio*, were identified as the most influential and stable predictors, aligning closely with domain knowledge and their correlation with delinquency risks.

- Weather-related variables showed limited predictive impact. This suggests that while weather events may cause localized disruptions, their influence on aggregate delinquency rates is overshadowed by broader economic and financial indicators. Additionally, including lagged macroeconomic features in the model may have indirectly captured the downstream effects of weather-related risks.

6.3 Weather Impact on Mortgage Delinquency Rates

The limited impact of weather variables can be attributed to several factors:

- **Aggregation Effects**: Using state-level aggregate delinquency rates may dilute localized impacts of extreme weather events, reducing their contribution to overall model predictions.

- **Temporal Overlap with Economic Indicators**: Weather-related disruptions may indirectly manifest through macroeconomic variables, such as unemployment or consumer confidence, which already dominate the model's predictions.

- **Weather Data Quality**: The weather data used in this study, such as global-level CO_2 emissions, average temperature, average wind speed, and monthly precipitation, may not effectively gauge extreme weather events. Future research could experiment with more granular indicators or consult domain experts in meteorology.

6.4 Future Directions

Building on these findings, the following directions can be pursued for further investigation:

- Incorporating more granular, localized weather data to capture region-specific impacts better and refine weather-related risk assessments.

- Developing hybrid models that separately address normal and abnormal periods, such as those influenced by extreme weather events or economic shocks.

- Investigating alternative methods to disentangle the interactions between weather and macroeconomic variables, allowing a more precise understanding of their unique contributions to delinquency risks.

6.5 Conclusion

In conclusion, the SHAP-enhanced TFT model demonstrates significant potential for advancing financial risk prediction by combining state-of-the-art time-series modeling with robust interpretability frameworks. While economic and mortgage-related variables remain the primary drivers of delinquency risk, integrating advanced interpretability tools ensures that the model's insights are accurate and actionable. These contributions pave the way for more informed decision-making in financial forecasting and risk assessment, providing a strong foundation for future research and application.

Bibliography

[1] Kjersti Aas, Martin Jullum, and Anders Løland. Explaining individual predictions when features are dependent: More accurate approximations to shapley values. *arXiv preprint arXiv:1903.10464*, 2019.

[2] Selçuk Bayraci, Orkun Susuz, et al. A deep neural network (dnn) based classification model in application to loan default prediction. *Theoretical and Applied Economics*, 4(621):75–84, 2019.

[3] Daniel Berg. Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, 23(2):129–143, 2007.

[4] Arnab Bhattacharya, Simon P Wilson, and Refik Soyer. A bayesian approach to modeling mortgage default and prepayment. *European Journal of Operational Research*, 274(3):1112–1124, 2019.

[5] Leo Breiman. Bagging predictors. *Machine learning*, 24:123–140, 1996.

[6] Raffaella Calabrese, Timothy Dombrowski, Antoine Mandel, R Kelley Pace, and Luca Zanin. Impacts of extreme weather events on mortgage risks and their evolution under climate change: A case study on florida. *European Journal of Operational Research*, 314(1):377–392, 2024.

[7] Raffaella Calabrese, Meagan McCollum, and R Kelley Pace. Mortgage default decisions in the presence of non-normal, spatially dependent disturbances. *Regional Science and Urban Economics*, 76:103–114, 2019.

[8] Yung-Chia Chang, Kuei-Hu Chang, and Guan-Jhih Wu. Application of extreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Applied Soft Computing*, 73:914–920, 2018.

[9] Pao-Shin Chu and Xin Zhao. Bayesian analysis for extreme climatic events: A review. *Atmospheric research*, 102(3):243–262, 2011.

[10] Collins, Ghosh, and Scofield. An application of a multiple neural network learning system to emulation of mortgage underwriting judgements. In *IEEE 1988 International Conference on Neural Networks*, pages 459–466. IEEE, 1988.

[11] Junlong Dai, David Vigouroux, Pieter Kleynhans, and Julien Moyse. Autoxai: A framework to automatically select the most adapted xai solution. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 4423–4433, 2022.

[12] Ding Du, Xiaobing Zhao, and Drive NAU Box. Hurricanes and residential mortgage loan performance. *Office of the Comptroller of the Currency (OCC) Working Paper*, 2020.

[13] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[14] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[15] Paulo Issler, Richard Stanton, Carles Vergara-Alert, and Nancy Wallace. Mortgage markets with climate-change risk: Evidence from wildfires in california. *Available at SSRN 3511843*, 2020.

[16] William Köhler. A bayesian approach to predicting default, prepayment and order return in unsecured consumer loans, 2023.

[17] Carolyn Kousky, Mark Palim, and Ying Pan. Flood damage and mortgage credit risk: A case study of hurricane harvey. *Journal of Housing Research*, 29(sup1):S86–S120, 2020.

[18] Håvard Kvamme, Nikolai Sellereite, Kjersti Aas, and Steffen Sjursen. Predicting mortgage default using convolutional neural networks. *Expert Systems with Applications*, 102:207–217, 2018.

[19] Lili Lai. Loan default prediction with machine learning techniques. In *2020 International Conference on Computer Communication and Network Security (CCNS)*, pages 5–9. IEEE, 2020.

[20] Bryan Lim, Sercan Ö. Arık, Nicolò Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

[21] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, pages 4765–4774, 2017.

[22] Ionuț Nica, Daniela Blană Alexandru, Simona Liliana Paramon Crăciunescu, and Ștefan Ionescu. Automated valuation modelling: analysing mortgage behavioural life profile models using machine learning techniques. *Sustainability*, 13(9):5162, 2021.

[23] Rising Odegua. Predicting bank loan default with extreme gradient boosting. *arXiv preprint arXiv:2002.02011*, 2020.

[24] Amine Ouazad and Matthew E Kahn. Mortgage finance and climate change: Securitization dynamics in the aftermath of natural disasters. *The Review of Financial Studies*, 35(8):3617–3665, 2022.

[25] Pranoy Panda, K. S. Srinivas, Vineeth N. Balasubramanian, and Gaurav Sinha. Interpretable model drift detection. In *7th Joint International Conference on Data Science & Management of Data (CODS-COMAD 2024)*, 2024.

[26] Jacob Pastor-Paz, Ilan Noy, Isabelle Sin, Abha Sood, David Fleming-Munoz, and Sally Owen. Projecting the effect of climate change on residential property damages caused by extreme weather events. *Journal of Environmental Management*, 276:111012, 2020.

[27] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should i trust you? explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144, 2016.

[28] Clifford V Rossi. Assessing the impact of hurricane frequency and intensity on mortgage delinquency. *Journal of Risk Management in Financial Institutions*, 14(4):426–442, 2021.

[29] Apaar Sadhwani, Kay Giesecke, and Justin Sirignano. Deep learning for mortgage risk. *Journal of Financial Econometrics*, 19(2):313–368, 2021.

[30] Vincent Zong Stone. Temporal interpretability dynamics exploration: A case study of shap explanations in multivariate time series data. *Unpublished Manuscript*, 2024. Manuscript in preparation.

[31] Varshini Subhash, Zixi Chen, Marton Havasi, Weiwei Pan, and Finale Doshi-Velez. What makes a good explanation? a harmonized view of properties of explanations. *arXiv preprint arXiv:2211.05667v3*, 2024.

[32] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30:5998–6008, 2017.

[33] Jacob Vigdor. The economic aftermath of hurricane katrina. *Journal of Economic Perspectives*, 22(4):135–154, 2008.

[34] Lin Zhu, Dafeng Qiu, Daji Ergu, Cai Ying, and Kuiyi Liu. A study on predicting loan default based on the random forest algorithm. *Procedia Computer Science*, 162:503–513, 2019.

Appendix A: Model Card

Model Details

- **Model Name**: SHAP-Enhanced Temporal Fusion Transformer (TFT).
- **Model Version**: v11.5.
- **Model Type**: Neural network-based time-series prediction model.
- **Developers**:
- **Model Creation Date**: October 2024.
- **Last Updated**: December 2024.

Model Purpose

- **Intended Use**: Predict mortgage delinquency rates in Florida, providing interpretable insights into key risk factors using SHAP-based feature selection.
- **Primary Users**: Researchers, financial institutions, and policymakers interested in mortgage risk assessment and forecasting.
- **Out-of-Scope Use Cases**: This model is not designed for predicting delinquency rates outside the scope of mortgage portfolios or for non-financial time-series datasets.

Data

- **Target Variable**:
  - *FL 90+ Days Delinquency Rate*: The proportion of mortgage loans in Florida delinquent for 90 days or more, representing the target variable for predictions.
- **Features**: The dataset includes six critical features selected through SHAP-based feature selection:
  - *Consumer Confidence (FL)*: A measure of consumer optimism regarding economic conditions in Florida.

  - *Unemployment Rate (FL)*: The proportion of the labor force in Florida currently unemployed.
  - *Loan-to-Value Ratio (ZHVI)*: The ratio of the loan amount to the appraised property value, reflecting leverage levels.
  - *30-Year Mortgage Rate (FL)*: The average interest rate for a 30-year fixed mortgage in Florida.
  - *Credit Score Range (Numeric)*: A numeric representation of borrower creditworthiness, indicating the likelihood of default.
  - *Yield Curve Spread (5-Year Minus 30-Year)*: The difference between the 5-year and 30-year Treasury yields, often used as an economic indicator.

- **Data Preprocessing**:
  - *Differencing*: Features and target variables were differenced to remove trends.
  - *Scaling*: Data was scaled using a modified Python StandardScaler to normalize values with a mean of 0 and a standard deviation of 0.1, optimized for both the TFT model and SHAP KernelExplainer.
  - *Missing Data Handling*: Missing values were imputed using the nearest available data points to ensure consistency.
  - *Sequence Adjustment*: Data was structured into sequences of length 12, determined based on autocorrelation and partial autocorrelation analyses to capture temporal dependencies effectively.

- **Data Split**: Data was randomly split into 80% for training and 20% for testing. The training dataset was further divided into 80% training and 20% validation subsets.

Model Architecture and Training

- **Architecture Overview**:
  - Time-distributed dense layers for feature encoding.
  - Multi-head attention mechanisms for capturing feature interactions.
  - Residual connections and layer normalization for training stability.
  - LSTM layers for modeling sequential patterns.
  - Fully connected layers for generating predictions.

- **Training Process**: Trained using the Adam optimizer with a learning rate of 0.001 and dropout layers for regularization.

- **Evaluation Metrics**: R^2 score and MSE were used for model evaluation.

Performance

- **Model Performance Summary**:
  - Test Data: $R^2 = 0.9961$, MSE = 0.0000.
  - SHAP Fidelity: $R^2 = 0.9653$, MSE = 0.0003.
- **Performance by Data Subsets**: Not applicable.

Fairness and Bias

- **Fairness Criteria**: The model emphasizes feature stability and interpretability, ensuring that predictions are not biased toward specific time periods or outliers. Features with consistent contributions over time are prioritized using adjusted SHAP metrics.
- **Addressing Potential Bias**: Feature selection incorporates a systematic evaluation to prevent over-reliance on any single predictor. For instance, weather-related variables were deprioritized due to their limited impact, avoiding misleading predictions.
- **Ethical Implications**: Bias in climate-related risk assessments may disproportionately impact vulnerable populations. By ensuring transparency and robust feature selection, the model mitigates unfair biases in decision-making processes.

Transparency and Explainability

- **Ensuring Transparency**: SHAP KernelExplainer provides a clear breakdown of how each feature contributes to the model's predictions, enabling stakeholders to understand the reasoning behind the risk scores.
- **Actionable Insights**: The interpretable framework allows users to assess the implications of risk scores on mortgage approval or terms. For example, high delinquency risks associated with borrower-level variables like *Loan-to-Value Ratio* and *Credit Score Range* are clearly communicated.
- **Monitoring for Explainability**: Regular fidelity checks are performed using R^2 and MSE metrics to ensure that SHAP explanations remain accurate and aligned with the model's behavior over time.

Ethical Considerations

- **Ethical Risks and Impact**: Potential misuse in decision-making if explanations are misunderstood or over-reliance on predictions during economic uncertainty.
- **Mitigation Strategies**: The SHAP-enhanced framework ensures transparency, while recommendations include consulting domain experts for decisions influenced by model outputs. Regular fidelity checks help maintain the trustworthiness of explanations.

Limitations

- **Known Limitations**: The model may underperform during abnormal periods (e.g., the COVID-19 pandemic) and is sensitive to the quality of input data.
- **Impact of Weather Data**: Aggregated weather variables showed limited predictive power, necessitating more granular and localized data for future use.

Use and Deployment

- **Model Access**: Currently used for academic research and accessible upon request.
- **Deployment Guidelines**: Users are encouraged to validate predictions with external benchmarks and interpret results alongside domain expertise.
- **Monitoring and Maintenance**: Regular updates and retraining are recommended to adapt to changing economic conditions. Fidelity scores for SHAP explanations should be monitored periodically to ensure interpretability remains accurate.

References and Additional Information

- **References**: Machine Learning for Mortgage Risk: An Interpretable Approach by Vincent Zong Stone, Maggie Chen, Yao Zhang, Ming Xia.
- **Additional Resources**: Code repository and documentation available at `https://github.com/VSTONE9/shap-enhanced-tft`.

Appendix B: Project Repository

The full project report, source code, and related resources for this study are available in the following GitHub repository:

`https://github.com/VSTONE9/ml-mortgage-interpretability`

The repository is a comprehensive resource for replicating and extending this study. Users are encouraged to explore it to better understand the methodology and results or adapt the code for related financial forecasting applications.