Can Machine Learning Algorithms Better Help Predict Fiscal Stress in Local Governments?

Can Chen
Associate Professor
Andrew Young School of Policy Studies
Georgia State University
cchen64@gsu.edu

Yilun Zha
PhD Candidate
School of Architecture
Georgia Institute of Technology
yilunzha@gatech.edu

Chaowang Ren
PhD Student
Andrew Young School of Policy Studies
Georgia State University
cren2@student.gsu.edu

Yuxiang Zhao
Graduate Research Assistant
School of City & Regional Planning
Georgia Institute of Technology
yzhao758@gatech.edu

Working Paper Draft. Comments Welcome. Please Do Not Cite Without Permission.

Abstract
Fiscal stress poses a recurring challenge to local governments in the United States, often threatening the provision of essential public services. Traditional early warning models based on linear assumptions may struggle to capture the complex and nonlinear nature of fiscal stress. In this study, we assess whether machine learning (ML) techniques can improve the out-of-sample prediction of fiscal stress in local governments, using a comprehensive dataset covering economic, social, fiscal, demographic, geographic, and environmental variables for over 900 localities in New York from 2013 to 2022. Comparing the predictive performance of traditional logistic regression with several ML models—including random forests, gradient boosting, and neural networks—we find that ML methods significantly outperform traditional approaches in both accuracy and robustness. Our findings offer important implications for designing more effective early warning systems to assist fiscally distressed localities.

Keywords:
machine learning, fiscal stress, local governments, early warning systems, budget deficits

Evidence for Practice

- Machine learning models significantly outperform traditional regression-based approaches in predicting fiscal stress among local governments.
- Key predictors of fiscal distress include lagged deficits, unemployment, and housing vacancy—factors that reflect both financial and socioeconomic conditions.
- State fiscal monitoring systems can improve early detection and responsiveness by incorporating machine learning into existing oversight frameworks.
- Advanced analytics offer public managers a more adaptive, accurate, and scalable approach to anticipating and addressing local fiscal crises.

INTRODUCTION

Fiscal stress is a recurring challenge for local governments in the United States, with serious implications for public service delivery and community well-being. From declining tax revenues and mounting service demands to debt burdens and unfunded liabilities, local governments often find themselves in precarious financial positions. Between 2001 and 2017, at least 123 local governments filed for bankruptcy in the U.S. (Murphy & Cook, 2018), with the most notable case being the City of Detroit's Chapter 9 bankruptcy in 2013—the largest municipal bankruptcy in U.S. history. These high-profile cases underscore the urgent need for tools that can anticipate fiscal stress before they escalate into full-blown crises.

Governments and researchers alike have long sought to build early warning systems (EWS) to detect fiscal distress in advance (Pew Foundation, 2016; Justice and Scorsone, 2012; Justice et al. 2019). Traditional models—primarily linear regression-based approaches and the signal detection method—have been the mainstay of these efforts. While these models have offered useful insights, they often rest on restrictive assumptions about linearity, independence, and limited interaction among predictors. As a result, they may fall short in capturing the complex, dynamic, and nonlinear factors that often underlie local fiscal crises (Moreno Badia et al., 2022; Hellwig, 2021).

In response to rising fiscal uncertainty, many U.S. states have developed their own fiscal monitoring systems to track local government financial health and flag jurisdictions at risk of distress (Pew Foundation, 2016; Nahmurina, 2024). Examples include New York's Fiscal Stress Monitoring System (FSMS), as well as similar initiatives in Michigan, Ohio, Pennsylvania, and New Jersey. These systems typically rely on threshold-based indicators and traditional statistical models. While valuable, such approaches may overlook early warning signs or fail to adapt to changing conditions—highlighting the need for more flexible, data-driven tools.

In recent years, machine learning (ML) has emerged as a promising alternative for predictive analytics in the public sector. ML techniques can process large volumes of data from heterogeneous sources, detect intricate patterns and interactions among variables, and optimize model performance through regularization and cross-validation. These capabilities are particularly advantageous when the predictors of fiscal stress span a wide range of economic, social, political, demographic, and environmental domains (Moreno Badia et al., 2022; Hellwig, 2021). Despite this potential, the application of ML in forecasting local government fiscal stress—particularly within the United States—remains underexplored. To the best of our knowledge, no existing study has systematically compared traditional econometric models and ML algorithms in predicting local fiscal stress using real-world administrative data from U.S. municipalities.

This study addresses this gap by assessing whether and how machine learning methods can improve the prediction of fiscal stress in local governments. Our research questions*: How do ML algorithms compare to traditional econometric models in predicting local fiscal stress? What are the most important predictors of fiscal stress across fiscal, economic, social, and environmental domains?* Specifically, we examine the out-of-sample predictive performance of both traditional econometric models and a suite of ML algorithms—including random forests, gradient boosting machines, extremely randomized trees, artificial neural networks, and support vector machines—using data from 988 local governments in New York State between 2013 and 2022. We leverage a unique collaboration with the Office of the State Comptroller (OSC), which has maintained the FSMS since 2013. Our dataset integrates a rich array of fiscal, socio-economic, demographic, political, geographic, and environmental indicators at the municipal level.

This study contributes to literature in several meaningful ways. First, it introduces a novel application of machine learning to the prediction of local government fiscal stress in the U.S.—a context where these tools have seen limited use despite their growing popularity in private finance and national-level forecasting. Second, it provides a rigorous, empirical comparison of machine learning models and traditional econometric approaches using real-world administrative data. This comparison is critical for informing both the scholarly debate over methodological advancements and the practical question of how governments can improve their predictive tools. Third, the study generates policy-relevant insights by identifying the most important predictors of fiscal stress and demonstrating how data-driven modeling can enhance existing early warning systems. As more states implement or revise their fiscal monitoring systems, these findings offer a roadmap for how advanced analytics can be embedded into public financial oversight.

The rest of the paper proceeds as follows. Section 2 reviews the relevant literature on fiscal stress prediction and the emerging role of machine learning in public sector forecasting. Section 3 presents our data and modeling approach. Section 4 compares the predictive performance of traditional and machine learning models and identifies the most important predictors of fiscal stress. Section 5 discusses the implications of our findings for research and practice. Section 6 concludes with a summary of contributions, limitations, and directions for future research.

REVIEW OF RELEVANT LITERATURE

Traditional Approaches to Predicting Fiscal Stress

Scholars have long sought to develop early warning systems to detect fiscal stress before it escalates into crisis. Two dominant approaches in the literature are the signal approach and traditional linear regression models. The signal approach, introduced by Kaminsky et al. (1998), monitors a set of economic or financial indicators and triggers an alert when any indicator exceeds a predefined threshold. This method has been widely used in the context of sovereign financial crises due to its simplicity and intuitive appeal.

The second dominant approach is based on traditional econometric models, including logit, probit, and ordinary least squares (OLS) regression. These models estimate the likelihood of fiscal crises based on historical data and are frequently used in the prediction of sovereign defaults (Fioramanti, 2008; Sarlin, 2014). Linear regression models offer quantitative assessments of fiscal risk by estimating the conditional probability of crisis events given certain predictor values.

While these methods provide useful insights, they have notable limitations. Their reliance on linear assumptions and additive relationships constrains their ability to capture the complex, nonlinear, and interactive dynamics often inherent in fiscal systems. Furthermore, these models often underperform in out-of-sample prediction tasks, due to overfitting and limited generalizability (Demyanyk & Hasan, 2009; Hellwig, 2021). As fiscal environments become increasingly data-rich and multifaceted, traditional models struggle to incorporate large volumes of heterogeneous predictors, thereby limiting their effectiveness in real-world early warning systems (Liu et al., 2021; Antulov-Fantulin et al., 2021).

The Rise of Machine Learning in Fiscal Prediction

Machine learning (ML) techniques have gained significant traction as powerful alternatives to traditional econometric models in the prediction of financial crises, including fiscal stress. Unlike linear models, ML algorithms can process high-dimensional data, uncover complex and nonlinear relationships, and adaptively improve their predictions through iterative learning. These strengths are especially valuable in fiscal contexts where multiple economic, social, political, and environmental factors may interact in unpredictable ways.

A growing number of studies document the superior performance of ML models in fiscal and financial forecasting. For example, Belly et al. (2023) show that XGBoost outperforms Bayesian Model Averaging in capturing sovereign risk dynamics across the Euro area. Arakelian et al. (2019) demonstrate that regression trees and random forests provide more accurate predictions than OLS fixed effects models, particularly in the presence of macroeconomic shocks and market contagion. Similarly, artificial neural networks and recurrent neural networks (RNNs)—including LSTM and GRU architectures—excel at modeling temporal dependencies in financial systems, improving the accuracy of crisis forecasts (Fioramanti, 2008; Tölö, 2020).

Another key strength of ML lies in its ability to incorporate a broader and more diverse set of indicators. While traditional models tend to rely on a narrow set of financial or fiscal ratios, ML approaches have successfully included macroeconomic variables, institutional characteristics, market indicators, and even textual data. For instance, Bluwstein et al. (2023) highlight the predictive value of credit growth and yield curves, while Chen et al. (2023) use natural language processing to analyze textual data from financial reports and media sources, reducing false positives and negatives in crisis prediction. Other studies emphasize the importance of inflation, net foreign assets (Liu et al., 2022), fiscal rules compliance (Baret et al., 2024), and public external debt (De Marchi & Moro, 2023) as key predictors.

Despite these advantages, applying ML in fiscal forecasting also presents methodological and practical challenges. A major concern is model interpretability. Ensemble models like random forests and gradient boosting, as well as deep learning techniques, often function as "black boxes," making it difficult for policymakers to understand or justify their outputs (Tölö, 2020). To address this, researchers increasingly use model-agnostic interpretability tools such as SHAP values to evaluate feature importance and enhance model transparency (Liu et al., 2021). Moreover, integrating domain knowledge with ML outputs—such as those from public budgeting or financial oversight—can make insights more actionable (Piermarini et al., 2023). Another technical challenge is overfitting, where models learn patterns specific to the training data but fail to generalize. Regularization techniques, cross-validation, and ensemble averaging help address this risk (Jarmulska, 2022). Additionally, model performance is sensitive to algorithm selection and hyperparameter tuning. As Bluwstein et al. (2023) note, different ML models capture interactions and nonlinearities in distinct ways, making systematic experimentation essential for optimizing predictive accuracy.

Research Gaps in the Current Literature

While machine learning has gained momentum in the study of sovereign debt crises and macro-financial instability, its application to local government fiscal stress—particularly in the United States—is strikingly limited. Most empirical research in this domain focuses on national or cross-country contexts (e.g., Hellwig, 2021; Claessens & Kose, 2013), where institutional structures and fiscal rules differ substantially from those governing U.S. municipalities.

Some recent studies have extended ML approaches to the subnational level outside the U.S. For instance, Antulov-Fantulin et al. (2021) employ gradient boosting machines to assess fiscal stress in Italian municipalities, highlighting the role of non-financial indicators such as geographic and demographic characteristics. Gallardo Del Angel (2019) applies artificial neural networks to Mexican local governments, finding that operational expenditures have a greater influence on deficits than capital investment. These studies demonstrate the value of applying ML tools to local fiscal contexts—but they remain internationally focused.

To the best of our knowledge, there is no published study that systematically applies machine learning models to predict fiscal stress among U.S. local governments. This gap is critical, given the financial autonomy, institutional diversity, and legal constraints that characterize American municipalities. Moreover, local governments in the U.S. are often on the frontlines of service provision and revenue generation, making them especially vulnerable to fiscal instability and in need of reliable early warning systems. Our study aims to fill this gap by comparing the predictive performance of traditional econometric models and multiple ML algorithms in forecasting fiscal stress across over 900 localities in New York State from 2013 to 2022. In doing so, we contribute to both the methodological literature on fiscal forecasting and the practical field of local government financial management.

METHODOLOGY

Study Area and Data Sources

To examine the potential of machine learning (ML) techniques in predicting fiscal stress in local governments, this study focuses on the State of New York (Figure 1). New York is an ideal case for analysis due to its fiscal diversity, institutional heterogeneity, and the availability of comprehensive financial and demographic data. The unit of analysis is the municipality, defined as a general-purpose local government with substantial administrative and fiscal autonomy (Office of the New York State Comptroller, 2025).

Under this definition, New York State contains four primary forms of municipalities: counties, cities (including the five boroughs of New York City), towns, and villages—totaling 1,584 general-purpose local governments. However, for analytical clarity and to avoid overlapping jurisdictions, we exclude counties (which operate at a higher tier) and villages (which are typically nested within towns). Our final sample includes only cities (N = 60) and towns (N = 928), yielding a study population of 988 municipalities. This approach ensures comparability across units of government that perform similar service functions and operate with independent fiscal authority.

Towns and cities were chosen due to their relative data completeness, autonomy, and importance in local public service delivery.

Table 1 provides summary statistics on key demographic and socioeconomic characteristics for the top 10 cities and towns by population size. These variables serve as contextual inputs for the machine learning models and as controls in robustness checks.

Figure 1 Cities and Towns in the State of New York

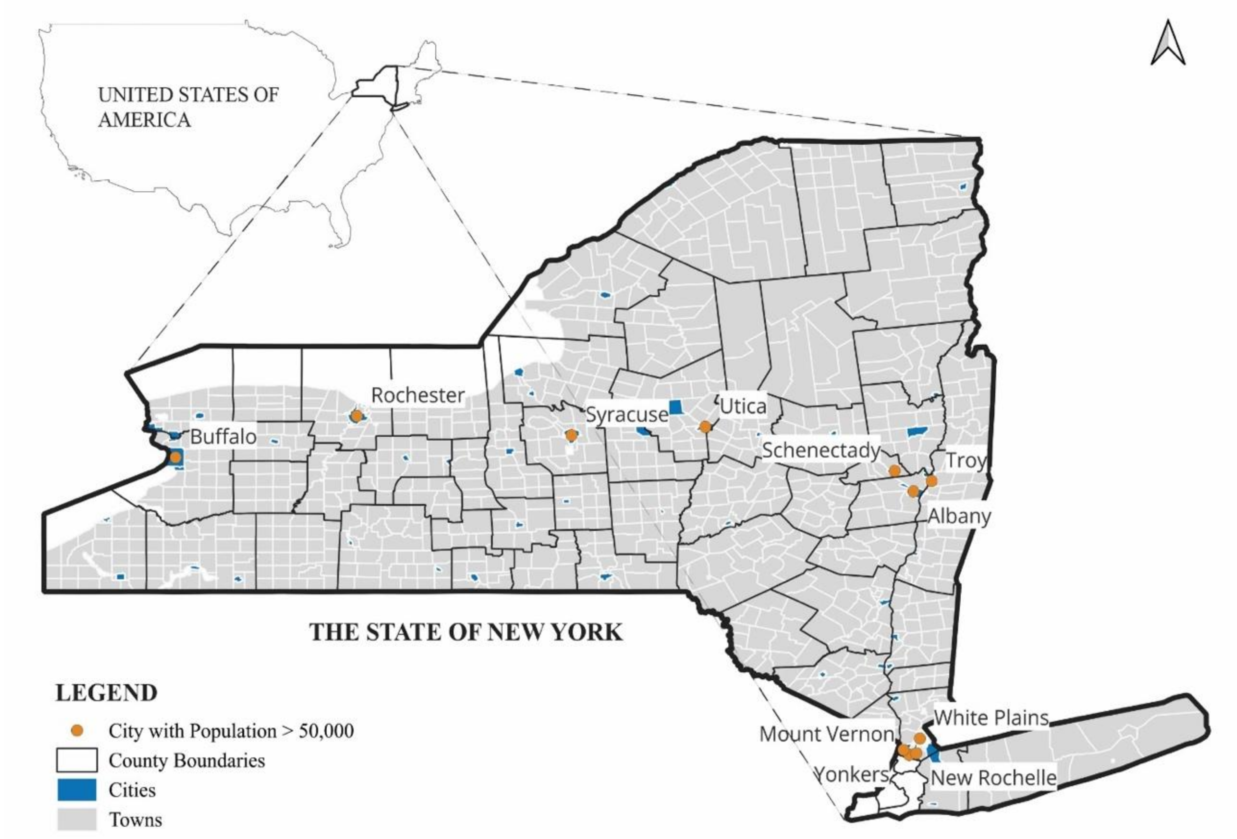


Table 1 Selected Demographic and Socio-Economic Indicators of the Top 10 Cities and Towns in New York State

Name	Population	Median Household Income	Unemployment Rate (%)	Name	Population	Median Household Income	Unemployment Rate (%)
City				**Town**			
Buffalo	276,688	$46,184	7.0	Hempstead	789,763	$132,468	5.1
Rochester	210,992	$44,156	8.5	Brookhaven	487,162	$114,845	5.2
Yonkers	209,780	$78,208	6.6	Islip	339,123	$122,726	4.2
Syracuse	146,134	$43,584	8.6	Oyster Bay	299,958	$152,952	4.4
Albany	99,692	$54,736	8.0	North Hempstead	236,573	$148,263	3.8
New Rochelle	80,828	$100,542	7.0	Babylon	217,830	$115,992	5.2
Mount Vernon	72,817	$75,511	8.0	Huntington	203,808	$153,782	4.9
Schenectady	68,476	$54,650	9.4	Ramapo	148,558	$80,955	6.8
Utica	64,728	$48,212	7.0	Amherst	129,577	$87,280	3.6
White Plains	59,421	$109,551	5.4	Smithtown	116,157	$147,104	3.9

Note: Data was retrieved from US Census American Community Survey 5-year estimate (2018-2022).

We obtained detailed financial records for local governments from Open Book New York (2025), a comprehensive public database maintained by the Office of the New York State Comptroller. The dataset includes itemized records of expenditures, revenues, balance sheets, and outstanding debt for cities and towns, covering a ten-year period from 2013 to 2022. These data serve as the foundation for constructing annual fiscal indicators used in the machine learning models.

To account for environmental and socioeconomic factors, we integrated data from the American Community Survey (ACS), administered by the U.S. Census Bureau (2025). The ACS provides a rich set of demographic, economic, and social characteristics at the municipal level, allowing us to control for exogenous community-level influences on fiscal outcomes. Detailed descriptions of each independent variable are provided in the following section.

Recognizing the impact of major external shocks—such as the COVID-19 pandemic and Hurricane Ida—on local government finances, we also included measures of federally declared disasters. These data were obtained from OpenFEMA, the open data platform of the Federal Emergency Management Agency (US FEMA, 2024). Disaster declarations are used to construct binary or intensity-adjusted variables to capture the influence of natural and public health emergencies on fiscal stress.

Variables and Measures

Dependent Variable: Predicting Outcomes

There is limited consensus in the literature on a single best measure of fiscal stress for local governments. Both objective indicators (e.g., budget deficits, debt burdens) and subjective assessments (e.g., credit ratings, expert surveys) have been employed, though each faces challenges related to measurement error, comparability, and data availability.

Given the focus of this study on early warning and predictive modeling, we selected budget deficit as the primary outcome variable. Budget deficits are widely used in fiscal stress studies and offer a clear, observable metric for identifying early fiscal imbalance. Two versions of the dependent variable were constructed based on this measure.

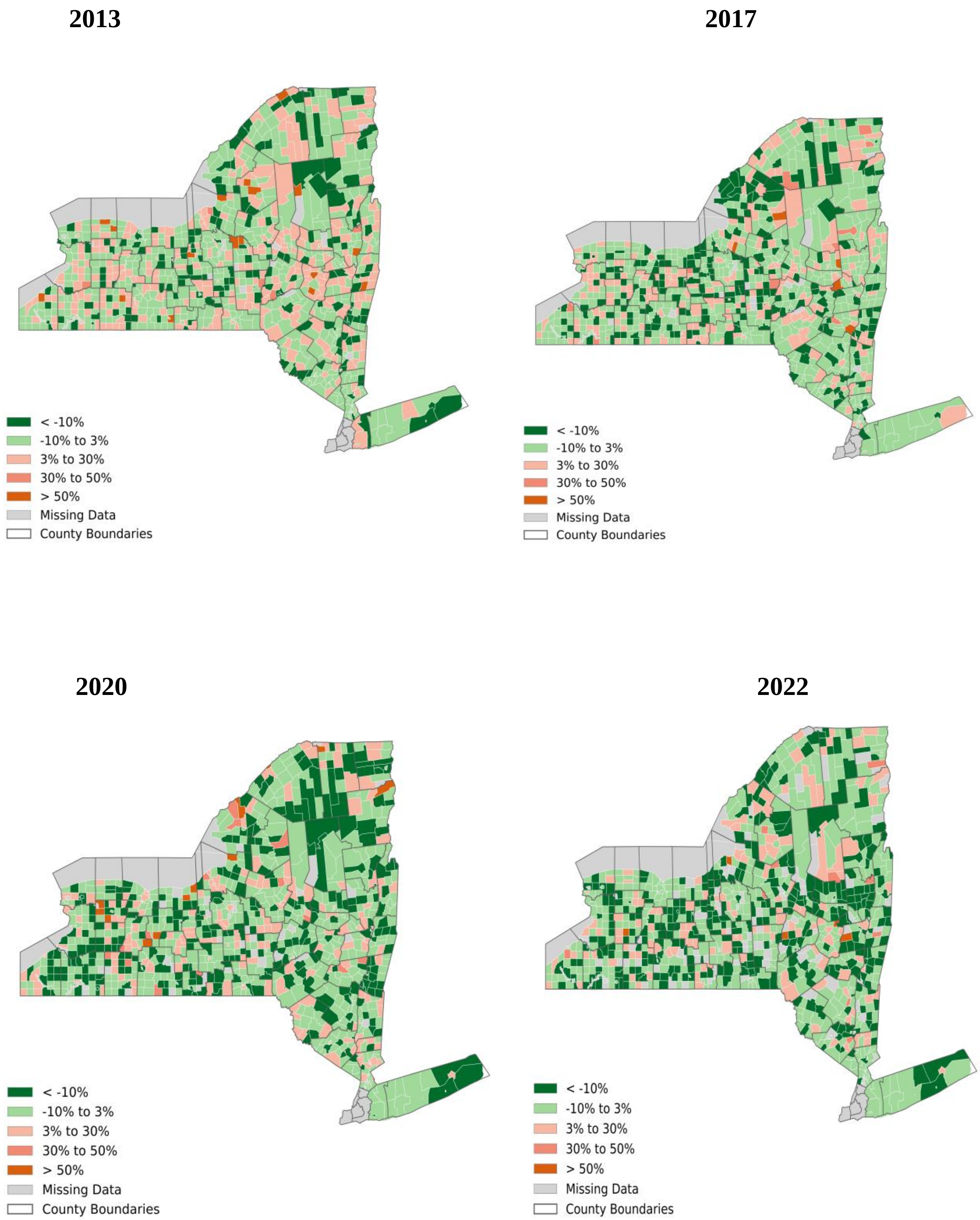
For each municipality i in year t, the budget deficit ratio BD_{it} is defined as follows:

$$Budget\ Deficit\ Ratio(BD)_{i,t} = \frac{Total\ Revenue\ (TR)_{i,t} - Total\ Expenditure\ (TE)_{i,t}}{Total\ Revenue\ _{i,t}} \quad (1)$$

In addition to the continuous budget deficit ratio, we created a binary fiscal stress indicator to facilitate classification modeling. A municipality is coded as fiscally stressed (1) if its deficit exceeds a pre-defined threshold, and not stressed (0) otherwise. We use a +3% (0.03) threshold as the baseline, consistent with regulatory benchmarks and prior studies, while also testing alternative thresholds (ranging from 2% to 10%) for robustness.

Figure 2 displays the distribution of fiscal deficits across municipalities in selected years, illustrating temporal variation and the proportion exceeding key deficit ceilings. As shown in Figure 2, a positive budget deficit ratio (shown in red) indicates a real fiscal deficit. A negative budget deficit ratio (shown in green) indicates a real fiscal surplus.

Figure 2 Visualization of Fiscal Deficit in New York Municipalities: 2013, 2017, 2020, and 2022



Predicting Variables and Measures

This study utilizes a comprehensive set of indicators to predict fiscal stress, grouped into four domains: financial, socioeconomic, demographic, and housing. Detailed definitions, sources, and descriptions of each variable are provided in Table A1 of Appendix I. While most variables are drawn directly from public data sources without further transformation, the financial indicators are constructed following standard practices in the fiscal health literature.

In the financial domain, we include 21 indicators that capture the core components of municipal fiscal operations—revenues, expenditures, fund balances, and debt—reflecting the central role these dimensions play in shaping local government financial stability (Luitel & Tosun, 2014; Sohl et al., 2009; Stone et al., 2015; Skidmore & Scorsone, 2011; Gorina et al., 2018).

- **Revenue indicators** include transfer revenues, federal and state aid, sales tax, and property tax, each expressed as a proportion of total revenue.
- **Expenditure indicators** reflect the allocation of municipal spending across key functions such as personnel, economic development, disaster response, policing, and transportation—measured as shares of total expenditure.
- **Fund balance indicators** include the ratios of total fund balance and cash reserves to total expenditure, serving as proxies for fiscal reserves and liquidity.
- **Debt-related indicators** comprise nine measures, including debt service as a percentage of total revenue, total bond issuance and anticipated bond note issuance (and their ratios to total expenditure), as well as bond repayments and anticipated note repayments (and their respective ratios to total revenue).

The finalized set of financial predictors and their calculation formulas are summarized in **Table 2**. Collectively, these indicators provide a multidimensional view of municipal fiscal operations and form the foundation for building robust, data-driven models to forecast fiscal stress

Table 2 Financial Predicting Indicators and Authors' Calculation

Indicators	Authors' Calculation
Revenue Variables	
TRANSFER_REVENUE	TransRev/TR
AID_REVENUE	FSAidRev/TR
SALES_REVENUE	AssetSaleRev/TR
SALES_TAX_REVENUE	SalesUseRev/TR
PROPERTY_TAX_REVENUE	PropTaxRev/TR
Expenditure Variables	
ECON_DEV_EXP_RATIO	EconDevExp/TE
PERSONNEL_REV_RATIO	PerExp/TR
DISASTER_EXP_RATIO	DisasterExp/TE
POLICE_EXP_RATIO	PolExp/TE
TRANSPORT_EXP_RATIO	TransportExp/TE
Fund Balance Variables	
FUND_BALANCE_EXP_RATIO	TotFundBal/TE
CASH_EXP_RATIO	TotCash/TE
Debt Variables	
DEBT_REV_RATIO	DebtServExp/TR
BOND_ISSUE	IssuedBnd
BOND_PAID	PaidBnd
BOND_ANTICIPATION_ISSUE	BANIssued
BOND_ANTICIPATION_PAID	BANPaid
BOND_ISSUE_EXP_RATIO	BndIssuedCY/TE
BOND_ANT_NOTE_ISSUE_EXP_RATIO	BANIssuedCY/TE
BOND_PAID_REV_RATIO	BndPaidCY/TR
BOND_ANT_NOTE_PAID_REV_RATIO	BANPaidCY/TR

Note: Abbreviations appearing in the financial indicator calculations are defined in Table 1 of the Appendix.

As part of the **demographic domain**, we include 11 indicators to capture population composition and dynamics. Prior research suggests that demographic shifts—particularly population aging—can reduce income and sales tax revenue per capita (Felix & Watkins, 2013), increase public expenditure and debt burdens (Hondroyiannis & Papapetrou, 2000), and thereby contribute to heightened fiscal stress. Accordingly, one key indicator is the percentage of the population aged 65 and over. Additional demographic variables include total population, sex ratio, average household size, percentage of the population over age 18, percentage aged 25 and over with at least a high school diploma, disability rate, percentage of households with minors, and the proportions of African American and Hispanic residents.

To assess broader **socioeconomic conditions**, we incorporate 10 indicators reflecting income distribution, employment structure, and economic vulnerability. Edinak (2021) finds that shifts in employment—especially the decline in low-skilled labor—can influence household consumption and, in turn, local sales tax revenues. Reflecting this dynamic, we include the percentage of jobs in the service, managerial, and public sectors. Additional variables include the Gini index, median household income, unemployment rate, public assistance receipt, poverty rate, uninsured population rate, and the number of federally declared disasters.

Last, the **housing domain** comprises five indicators that capture housing market characteristics and their fiscal implications. Prior studies have shown that housing prices (Alm, Buschman, & Sjoquist, 2011) and foreclosure activity (Alm, Buschman, & Sjoquist, 2014) significantly affect local property tax bases. We include median home value as a proxy for housing price levels. Two affordability measures—gross rent and selected monthly owner costs as percentages of household income—capture housing cost burdens. Additionally, we include the percentage of vacant housing units and renter-occupied units to reflect broader housing stock conditions.

Methods

This section outlines the methodological framework of the study, including the datasets used, preprocessing procedures, model architecture, and the strategies employed for both prediction and interpretation. The study adopts a dual-framework approach that distinguishes between two modeling objectives: (1) an explanatory model and (2) a predictive (early warning) model.

The explanatory model is designed to identify and interpret the contemporaneous relationships between fiscal stress and its potential determinants within the same fiscal year. This approach enables the investigation of structural and contextual factors that are associated with fiscal outcomes at a given point in time. In contrast, the predictive model functions as an early warning system, leveraging lagged independent variables to forecast future fiscal stress events. By using past information to anticipate future outcomes, this model is better suited for informing proactive fiscal management and risk mitigation strategies.

This dual-modeling structure, illustrated in **Figure 3**, allows for a comprehensive analysis that serves both diagnostic and prognostic purposes. It enables a direct comparison of model performance (e.g., predictive accuracy) and the relative importance of fiscal, demographic, and socioeconomic indicators across different temporal frameworks. Through this design, the study bridges the gap between explanation and prediction in the context of local government fiscal health.

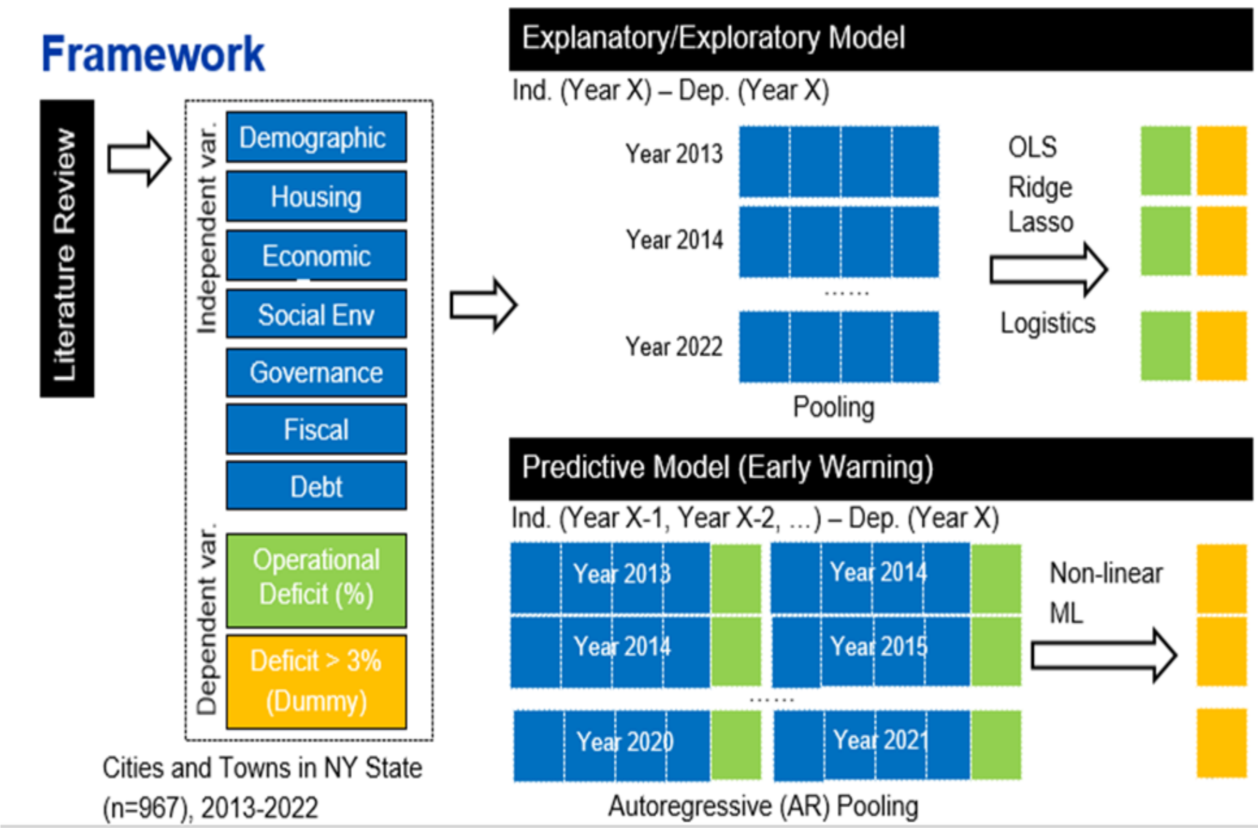


Figure 3. Analytical Framework: Structure of Explanatory and Predictive Models for Identifying Municipal Budget Deficits

The explanatory models utilize pooled panel data from 988 cities and towns in New York State, covering the period from 2013 to 2022. Fiscal indicators and outcomes were measured contemporaneously to assess their direct associations. We applied Ordinary Least Squares (OLS) regression to estimate the budget deficit ratio and used logistic regression to classify municipalities into binary fiscal stress categories—defined as having a budget deficit ratio exceeding 3%. These traditional econometric techniques are widely adopted in fiscal distress research due to their transparency, ease of interpretation, and well-established statistical properties (Ashraf, Félix, & Serrasqueiro, 2019). To enhance model robustness and mitigate overfitting and multicollinearity, we incorporated regularized linear models, including Ridge regression and the Least Absolute

Shrinkage and Selection Operator (LASSO). Ridge regression introduces an L2-norm penalty to shrink coefficients, while LASSO applies an L1-norm penalty to promote coefficient sparsity and perform variable selection.

The objective of the explanatory modeling phase was twofold: (1) to evaluate the predictive performance of traditional econometric methods and (2) to identify key fiscal, socioeconomic, and demographic indicators associated with budget deficits. Initially, OLS, Ridge, and LASSO models were compared using standard performance metrics. OLS results served as a benchmark for identifying statistically significant predictors. For classification, logistic regression assigned municipalities a value of "1" if their budget deficit ratios exceeded 3%, and "0" otherwise. To test model sensitivity, the deficit threshold was varied across levels (0%, 2%, 4%, 6%, 8%, and 10%).

For predictive modeling, we adopted an autoregressive framework, using predictors from years *t–2* and *t–1* to forecast fiscal stress in year *t*. This approach incorporated lagged fiscal indicators, including a municipality's deficit status in the preceding two years, to account for fiscal persistence and temporal dynamics.

To capture nonlinear relationships and complex interactions, we applied four machine learning (ML) algorithms:

1. **Random Forests** – An ensemble method that averages predictions from multiple decision trees built via bootstrap aggregation to reduce variance.
2. **Extremely Randomized Trees (Extra-Trees)** – A variant of Random Forests that introduces randomized split thresholds to further reduce variance and improve computational efficiency.
3. **Gradient Boosting Machines (GBMs)** – A sequential learning algorithm that combines weak learners to correct residual errors iteratively.

4. **Artificial Neural Networks (ANNs)** – Deep learning models that capture nonlinear interactions through multi-layer weight adjustments and backpropagation.

These models were selected based on prior evidence demonstrating their superior performance in fiscal forecasting tasks compared to linear models, particularly due to their ability to model high-order interactions and nonlinearity (Jarmulska, 2021). All ML models were trained and tuned using cross-validation, emphasizing out-of-sample predictive accuracy.

Although machine learning models offer improved predictive performance, they often function as “black boxes” with limited interpretability (Lipton, 2018). To address this limitation, we employed model-agnostic interpretation tools, particularly SHAP (Shapley Additive Explanations) values and dependence plots, to explain individual predictor contributions and improve model transparency.

Data Preprocessing

Following model specification, a key challenge involved preparing the dataset for robust and reliable analysis. To address common data quality issues—including missing values, variable scaling, distributional skewness, and class imbalance—we applied a systematic series of preprocessing techniques.

First, missing values in predictor variables were imputed using column means, ensuring the full dataset could be retained for analysis. Next, all predictors were standardized using z-score transformation, producing variables with a mean of zero and a standard deviation of one. Standardization is especially important for models sensitive to feature magnitudes, such as regularized regressions and neural networks.

To address skewness in the budget deficit outcome variable, we applied the Yeo–Johnson power transformation, a flexible alternative to Box–Cox that accommodates both zero and negative values. The Yeo–Johnson transformation is defined as follows:

$$\psi(\lambda, y) = \begin{cases} \frac{(y+1)^{\lambda}-1}{\lambda}, & \text{if } y \geq 0, \lambda \neq 0 \\ \log(y+1), & \text{if } y \geq 0, \lambda = 0 \\ -\frac{(-y+1)^{2-\lambda}-1}{2-\lambda}, & \text{if } y < 0, \lambda \neq 2 \\ -\log(-y+1), & \text{if } y < 0, \lambda = 2 \end{cases}$$

This transformation normalized the distribution of fiscal deficits, improving model fit and meeting assumptions of linearity and homoscedasticity for regression analysis.

After transformation and scaling, the dataset was randomly partitioned into training (80%) and testing (20%) subsets. The training data were used for model estimation and cross-validation, while the testing data provided an independent benchmark for evaluating out-of-sample predictive performance.

A significant issue identified during preprocessing was class imbalance: only about 20% of municipalities recorded a budget deficit ratio above the threshold, posing challenges for supervised learning. To mitigate this, we tested three resampling strategies:

1. **Class Weighting (CW):** Adjusts the model's loss function by assigning higher penalties to misclassified minority-class observations. This method preserves the original data distribution while enhancing model sensitivity to rare events.
2. **Random Over-Sampling Examples (ROSE):** Balances the dataset by duplicating minority-class cases. While simple and effective, this method risks overfitting due to repeated data points.

3. **Synthetic Minority Over-Sampling Technique (SMOTE):** Synthesizes new examples by interpolating existing minority-class observations, introducing diverse and plausible patterns that improve the model's ability to detect fiscal stress.

Each resampling method was evaluated during the exploratory modeling phase using logistic regression performance metrics (e.g., F1-score, AUROC). The most effective strategy was then applied consistently within the predictive modeling framework, ensuring a balanced and reliable classification of fiscally distressed municipalities.

EMPIRICAL RESULTS

This section presents the key empirical findings. The analysis proceeds in four stages:

1. **Descriptive Comparisons:** We begin by reporting results from independent samples t-tests to examine statistically significant differences in key indicators between municipalities with budget deficit ratios exceeding 3% and those at or below this threshold.
2. **Exploratory Modeling:** Next, we present findings from the exploratory modeling phase, where three resampling strategies—Class Weighting, ROSE, and SMOTE—are compared to determine the most effective method for addressing class imbalance in binary classification.
3. **Predictive Model Evaluation:** We then evaluate the performance of machine learning models using multiple metrics, including predictive accuracy, Precision–Recall Curves (PRC), and Receiver Operating Characteristic (ROC) curves. These metrics offer complementary perspectives on each model's ability to detect fiscal stress, particularly in the context of imbalanced data.

4. **Model Interpretation:** Finally, we interpret the predictive model outputs using SHAP (Shapley Additive Explanations) values and dependence plots. These tools enable us to assess the relative importance of individual predictors and visualize their marginal effects on the likelihood of fiscal distress.

Together, these results provide robust evidence on the feasibility and interpretability of using machine learning to predict local government fiscal stress and highlight key indicators driving fiscal vulnerability.

Descriptive Comparisons: Independent Samples T-tests

As presented in Table 3, independent samples t-tests were conducted to examine differences in key fiscal indicators between two groups of municipalities: those with budget deficits greater than 3% and those with deficits at or below this threshold. The analysis focused on expenditure and revenue variables, as well as fund balance, cash holdings, and deficit levels.

The results indicate statistically significant differences across several fiscal dimensions. Municipalities experiencing higher budget deficits (>3%) exhibited significantly greater total expenditures ($p = 0.006$), sales revenue ($p < 0.001$), property tax revenue ($p = 0.014$), disaster-related spending ($p = 0.020$), fund balance ($p < 0.001$), and cash holdings ($p < 0.001$) compared to their counterparts with lower deficits. These patterns suggest that higher-deficit municipalities tend to have more intense financial activity, potentially driven by greater service demands, external shocks, or reliance on specific revenue sources.

In contrast, no statistically significant differences were observed in other fiscal variables, including transfer revenue, aid revenue, sales tax revenue, personnel expenditures, economic development spending, police and transportation expenditures, or debt service. These nonsignificant results may imply that such variables are less sensitive to immediate deficit conditions or that their effects are mediated through other financial mechanisms.

Overall, these descriptive comparisons provide an initial empirical foundation for identifying fiscal patterns associated with municipal stress and inform the subsequent modeling analysis.

Table 3 Independent Samples T-Test Comparing Fiscal Variables by Operating Deficit Level

Variable	Budget Deficit Ratio >3%	Budget Deficit Ratio S.D. >3%	Budget Deficit Ratio<=3%	Budget Deficit Ratio SD <=0.03	Difference_of_Means	p_value
TOTAL_EXPENDITURE	15,055,532.32	55,446,615.07	20,074,013.59	119,199,440.20	-5,018,481.27	0.006(**)
TRANSFER_REVENUE	1,877,449.14	10,640,633.41	2,036,590.82	11,543,323.75	-159,141.68	0.54
AID_REVENUE	2,064,067.90	11,304,592.08	2,212,134.71	12,180,507.75	-148,066.81	0.591
SALES_REVENUE	425,014.63	2,877,555.10	1,398,016.96	8,763,634.18	-973,002.33	<0.001(***)
SALES_TAX_REVENUE	1,631,493.83	7,549,595.81	1,716,530.65	7,834,426.19	-85,036.82	0.641
PROPERTY_TAX_REVENUE	4,884,252.32	17,449,466.69	5,952,607.61	20,072,727.86	-1,068,355.30	0.014(**)
PERSONNEL_EXPENDITURE	2,634,979.65	11,928,273.51	2,808,152.90	11,989,595.80	-173,173.25	0.544
ECONOMIC_DEV_EXPENDITURE	213,185.92	1,393,502.58	187,186.85	1,267,231.34	25,999.07	0.426
DISASTER_EXPENDITURE	114,573.30	610,497.93	152,049.08	847,802.77	-37,475.77	0.02(**)
POLICE_EXPENDITURE	1,305,103.98	6,326,130.56	1,313,000.62	6,413,207.02	-7,896.64	0.959
TRANSPORTATION_EXPENDITURE	2,097,707.61	5,254,362.48	2,117,189.40	5,595,814.27	-19,481.80	0.879
FUND_BALANCE	3,781,187.97	20,628,656.42	7,143,405.91	26,119,391.91	-3,362,217.94	<0.001(***)
CASH	6,784,111.23	20,397,102.45	9,178,631.50	38,450,830.07	-2,430,520.26	<0.001(***)
DEBT_SERVICE	1,155,733.37	5,466,912.64	1,348,005.29	7,459,563.89	-192,271.92	0.18
OP_DEFICIT	0.157	0.257	-0.086	0.096	0.24	<0.001(***)

Note: >3% indicates budget deficit ratios are greater than 3%, ≤0.03 indicates budget deficit ratios are less than or equal to 3%.SD represents standard deviations.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

The Exploratory Modeling

During the exploratory modeling phase, we applied three linear regression approaches—Ordinary Least Squares (OLS), Lasso, and Ridge regression—to predict municipal budget deficits using contemporaneous data. These models were implemented without any resampling techniques to establish baseline performance under standard assumptions.

As shown in Table 4, the OLS model achieved a test-set prediction accuracy of 54.2%, serving as the benchmark. The Lasso model, with a regularization parameter λ=0.02, slightly outperformed OLS with an accuracy of 54.8%, while the Ridge model, using α=0.01, yielded a comparable accuracy of 54.4%. Overall, model performance across the three approaches was closely aligned, with only marginal differences in predictive accuracy. The Lasso model's slight edge reflects its capacity for variable selection, which may offer additional interpretive value in identifying key fiscal predictors.

These initial results suggest that while traditional regression techniques provide a reasonable baseline, more complex models—especially those designed to capture nonlinearity and interactions—may be necessary to achieve substantial gains in predictive performance. This finding motivated the subsequent application of machine learning techniques within the predictive modeling framework.

Table 4 Train and Test Accuracy Across OLS, Lasso, and Ridge regression Models

	Train Accuracy	**Test Accuracy**
OLS	0.562	0.542
Lasso (λ =0.02)	0.563	0.548
Ridge (α =0.01)	0.558	0.544

Table 5 reports the significant predictors identified by **the Ordinary Least Squares (OLS) model** for estimating budget deficits within the same fiscal year. The results show that indicators across all four domains—financial, socioeconomic, demographic, and housing—contributed to the model, though financial variables exhibited the strongest predictive influence.

Among financial indicators, the fund balance to total expenditure ratio emerged as a particularly powerful predictor. A one-unit increase in this ratio was associated with a 19.01 percentage point reduction in the probability of incurring a budget deficit, controlling for all other variables ($p < 0.001$). This finding underscores the importance of reserve levels in buffering fiscal stress. Additionally, disaster exposure, captured by the number of FEMA-declared disasters, significantly increased fiscal risk. Each additional disaster declaration was associated with a 2.17 percentage point increase in the likelihood of a deficit, holding other factors constant ($p = 0.011$). This highlights the fiscal strain imposed by natural or public health emergencies on local governments.

The differentiated effects across indicator categories emphasize the value of incorporating multidimensional data in fiscal forecasting. Financial resilience and exposure to external shocks appear to play especially salient roles in driving contemporaneous budget outcomes.

Table 5 OLS Model Results: Significant Variables Predicting Municipal Budget Deficits

Variables	Coefficient	SD	t	P>|t|
PCT_RENTAL	0.0322	0.016	-2.046	0.041
MED_HVALUE	-0.0546	0.021	2.55	0.011
PCT_SMOCAPI	0.0387	0.011	-3.537	<0.001
MED_HHINCOME	-0.0505	0.024	-2.141	0.032
PCT_HSGRAD	-0.0353	0.014	-2.532	0.011
FEMA Count	0.0217	0.009	-2.553	0.011
TRANSFER_REV_RATIO	0.0708	0.009	7.763	<0.001
AID_REV_RATIO	-0.0562	0.01	-5.403	<0.001
SALE_REV_RATIO	-0.0962	0.015	-6.594	<0.001
PERSONNEL_REV_RATIO	0.1289	0.013	10.11	<0.001
POLICE_EXP_RATIO	-0.117	0.013	-9.341	<0.001
FUND_BALANCE_EXP_RATIO	-0.1901	0.015	-12.297	<0.001
CASH_EXP_RATIO	-0.1478	0.015	-9.933	<0.001
BOND_ANTICIPATION_PAID	-0.0234	0.009	-2.59	0.01
BOND_ISSUE_EXP_RATIO	-0.3625	0.015	-23.934	<0.001
BOND_ANT_NOTE_ISSUE_EXP_RATIO	0.3199	0.01	33.56	<0.001
BOND_PAID_REV_RATIO	0.1367	0.013	10.544	<0.001
BOND_ANT_NOTE_PAID_REV_RATIO	-0.1447	0.012	-11.791	<0.001
No. Observations:	6474			
F-statistic:	183.1			
Log-Likelihood:	-6516.3			

Table 6 presents the results of the **logistic regression model**, which was used to predict the likelihood of a municipality experiencing a budget deficit ratio exceeding 3%—coded as 1—versus a deficit at or below that threshold—coded as 0. The model achieved a pseudo-R-squared of 72.4% on the training dataset and an accuracy of 72.3% on the test dataset, indicating strong and consistent predictive performance.

Compared to the continuous regression models examined earlier, the logistic model demonstrated substantially higher accuracy, outperforming the best-performing linear model (Lasso), which did not incorporate a categorical outcome threshold. This suggests that modeling fiscal distress as a binary classification problem may yield more effective results when the goal is early warning or regulatory monitoring.

Interestingly, the set of significant predictors differed somewhat between the OLS and logistic models. Some variables that were significant in the OLS model—such as the percentage of households that rent—were not statistically significant in the logistic specification. Conversely, new variables emerged as important in the logistic model; for example, the percentage of the civilian noninstitutionalized population with disabilities was statistically significant only in the logistic regression.

These variations suggest that different modeling strategies may capture distinct dimensions of fiscal stress and underscore the importance of model selection and specification in applied fiscal analysis.

Table 6 Logistic Model Results: Significant Variables Predicting Municipal Budget Deficit (3% Baseline Threshold)

Variables	Coefficient	SD	t	P>|t|
PCT_DISABILITY	0.0854	0.04	-2.132	0.033
TRANSFER_REV_RATIO	0.2055	0.03	6.826	0
SALE_REV_RATIO	-0.2898	0.055	-5.272	0
SALES_TAX_REV_RATIO	-0.0965	0.043	-2.256	0.024
PERSONNEL_REV_RATIO	0.2751	0.042	6.621	0
POLICE_EXP_RATIO	-0.1582	0.041	-3.867	0
TRANSPORT_EXP_RATIO	0.1653	0.049	3.377	0.001
FUND_BALANCE_EXP_RATIO	-0.392	0.056	-6.971	0
DEBT_REV_RATIO	0.0968	0.046	2.124	0.034
BOND_ANTICIPATION_PAID	-0.1524	0.058	-2.639	0.008
BOND_ISSUE_EXP_RATIO	0.1184	0.065	1.826	0.068
BOND_ANT_NOTE_ISSUE_EXP_RATIO	0.9888	0.061	16.175	0
BOND_ANT_NOTE_PAID_REV_RATIO	-0.1949	0.057	-3.444	0.001
No. Observations:	6474			
F-statistic:	201.1			
Log-Likelihood:	-3959.8			

To further assess the robustness of the logistic regression model, we conducted a sensitivity analysis examining how prediction accuracy varied across different deficit thresholds. As expected, model performance improved as the classification threshold increased. When the threshold was set just below 0%, the prediction accuracy was 68.1%, whereas accuracy rose steadily to 81.3% when the threshold reached 10%. This pattern indicates that the model is increasingly effective at identifying municipalities experiencing more severe budget deficits, as larger fiscal imbalances present clearer signals for classification. These findings suggest that while the model performs adequately under baseline conditions, its predictive strength is particularly robust in detecting extreme cases of fiscal stress.

As shown in Table 7, we further evaluated model performance under three commonly used metrics: area under the precision-recall curve (PRC-AUC), area under the receiver operating characteristic curve (ROC-AUC), and classification accuracy. Among the resampling strategies tested, the Synthetic Minority Over-sampling Technique (SMOTE) demonstrated the best overall performance, achieving a PRC-AUC of 0.636 and an accuracy of 81.7%. These results underscore SMOTE's superior capacity to improve detection of the minority class—municipalities experiencing fiscal stress—while maintaining strong overall accuracy.

While the ROSE method marginally outperformed SMOTE in ROC-AUC (0.819 vs. 0.818), this difference was negligible relative to SMOTE's clear advantages in precision-recall and balanced performance. In contrast, the class weighting (CW) method underperformed across all evaluation metrics, with a notably low PRC-AUC of 0.590, indicating limited effectiveness in identifying fiscally distressed municipalities within an imbalanced dataset.

These findings reinforce the importance of addressing class imbalance in fiscal forecasting and support the use of advanced oversampling techniques—particularly SMOTE—as a best practice in early warning model design.

Table 7 Comparison of Model Performance Using Different Resampling Methods

Datasets	**PRC-AUC**	**ROC-AUC**	**Accuracy**
CW	0.590	0.804	0.737
ROSE	0.631	0.819	0.799
SMOTE	0.636	0.818	0.817

Further supporting these findings, **Figures 4 and 5** present the Receiver Operating Characteristic (ROC) and Precision–Recall (PRC) curves, respectively. As shown in **Figure 4**, all three resampling methods performed above the diagonal baseline, confirming the overall predictive validity of the logistic regression models. Both SMOTE and ROSE exhibited superior performance compared to Class Weighting (CW), particularly at lower false positive rates, where accurate detection of true positives is most critical. **Figure 5** emphasizes performance on the minority class (municipalities with a budget deficit ratio exceeding 3%). It further illustrates these distinctions. While SMOTE and ROSE maintained higher precision across a broad range of recall values, CW consistently underperformed, failing to capture patterns associated with fiscal distress in imbalanced conditions.

Taken together, these visual diagnostics reinforce the earlier quantitative results. SMOTE offers the most reliable and effective approach for identifying municipalities at elevated risk of fiscal stress. Its ability to preserve precision while expanding recall makes it particularly suitable for early warning systems that prioritize accurate classification of rare but critical fiscal events. Considering these results, SMOTE was selected as the preferred resampling strategy for the subsequent predictive modeling phase.

Figure 4 ROC Curve Comparison of Resampling Methods

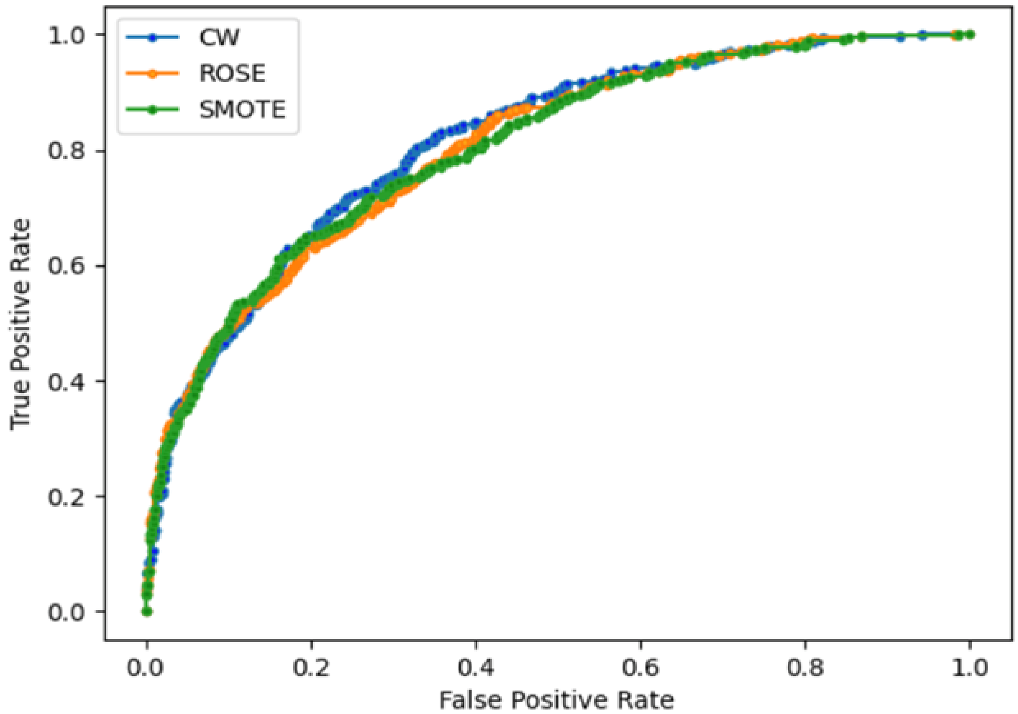
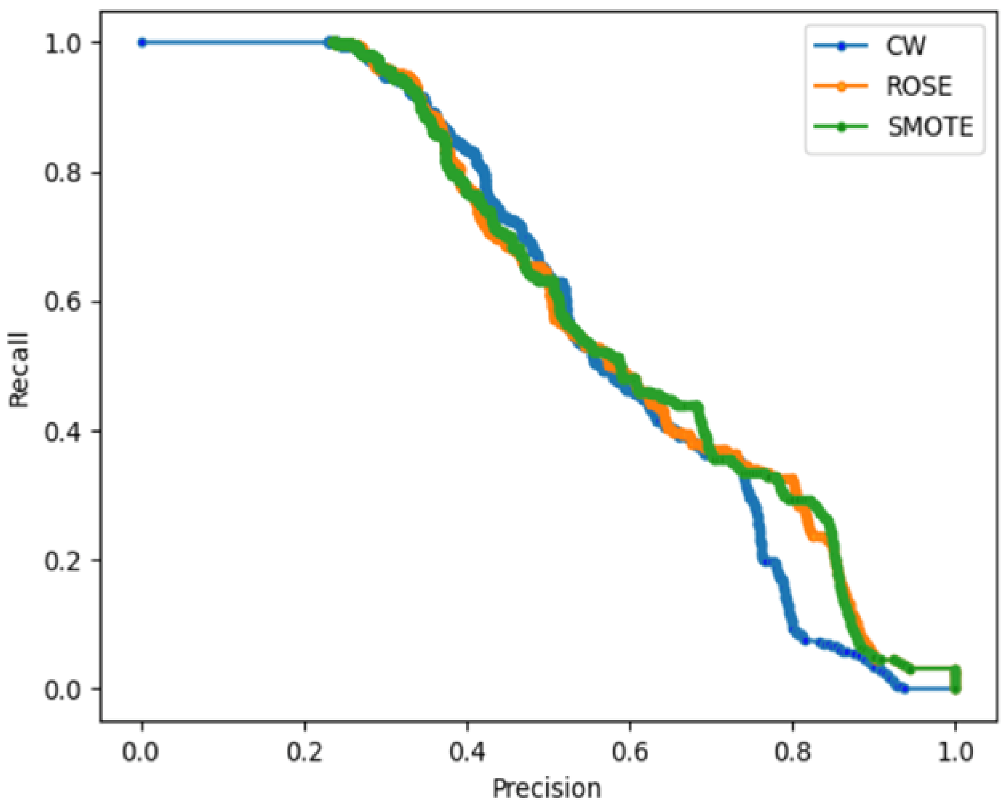


Figure 5 PRC Curve Comparison of Resampling Methods



Predictive Model Evaluation

As shown in Table 8, Extremely Randomized Trees and Random Forest consistently outperformed all other algorithms across the evaluated metrics. The Extremely Randomized Trees model demonstrated the strongest overall performance, achieving a PRC-AUC of 0.491, ROC-AUC of 0.865, and accuracy of 82.5%. Its ability to enhance variance reduction through randomized split thresholds likely contributed to its superior classification capacity, especially in identifying fiscally stressed municipalities.

The Random Forest algorithm closely followed, with a PRC-AUC of 0.448, ROC-AUC of 0.842, and accuracy of 79.4%. This confirms the robustness of ensemble tree-based methods in handling complex, high-dimensional fiscal datasets.

The Neural Network also performed competitively, particularly in ROC-AUC (0.846) and accuracy (0.788). However, its lower PRC-AUC of 0.403 reflects reduced sensitivity to the minority class—an important limitation for early warning applications targeting fiscal distress.

Gradient Boosting displayed modest improvements over logistic regression but remained suboptimal relative to other models. Meanwhile, Logistic Regression, serving as the traditional benchmark, exhibited the weakest performance, with a PRC-AUC of 0.283, ROC-AUC of 0.714, and accuracy of 70.4%, indicating limited ability to discriminate budget deficit cases within an imbalanced dataset.

Table 8 Performance Comparison of Predictive Models

ML algorithms	PRC	ROC	Accuracy
Logistics	0.111	0.642	0.665
Gradient Boosting	0.283	0.714	0.704
Neural Network	0.403	0.846	0.788
Random Forest	0.448	0.842	0.794
Extremely Randomized Trees	0.491	0.865	0.825

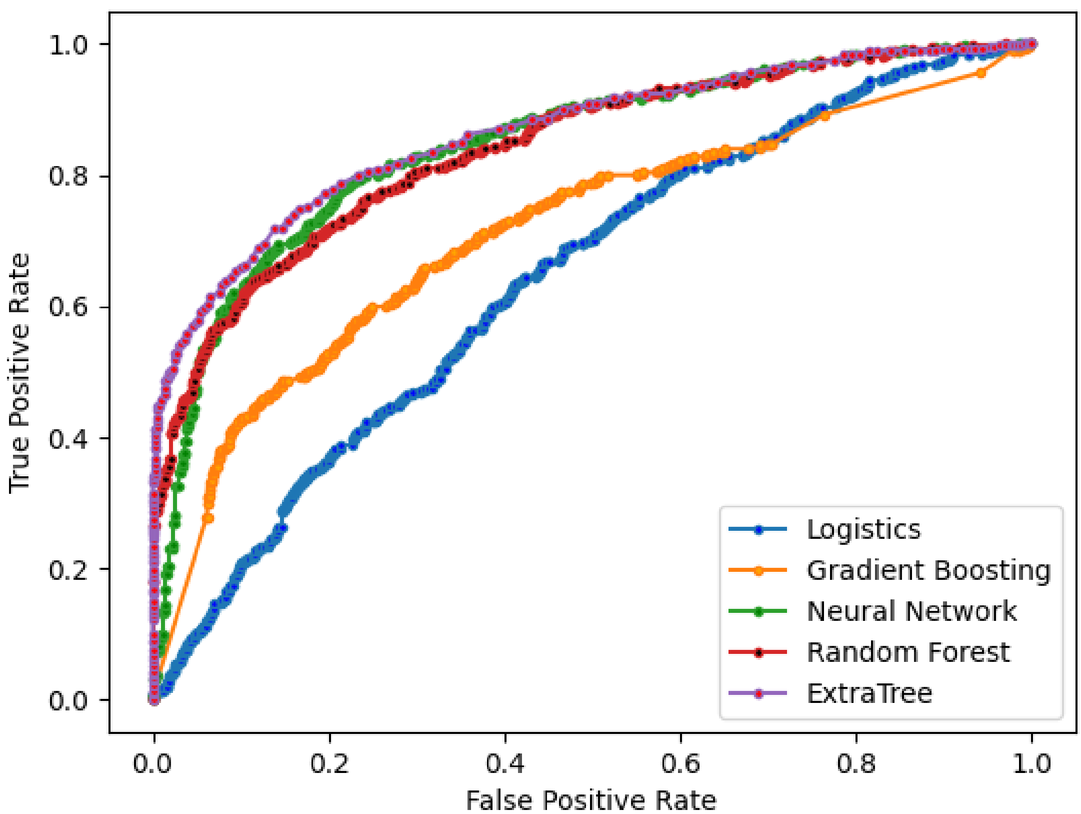
These results underscore the value of machine learning algorithms—particularly Extremely Randomized Trees—in improving predictive accuracy, especially in imbalanced classification contexts. While logistic regression assumes linearity and limited interactions, machine learning techniques effectively model complex, nonlinear relationships, and variable interactions, enhancing their ability to detect high-risk municipalities before deficits escalate.

The ROC curve, presented in **Figure 6**, further reinforces the comparative performance of the predictive models. The Extremely Randomized Trees model consistently outperformed all others across the full spectrum of false-positive rates, with its curve closely approaching the upper-left corner of the plot—a hallmark of strong classification accuracy and high true-positive rates. This dominant performance highlights the model's robustness in distinguishing fiscally distressed municipalities from those in stable condition.

The Random Forest model also demonstrated strong performance, though slightly below that of the Extremely Randomized Trees. The Neural Network model showed competitive results, maintaining a relatively high true-positive rate across a wide range of thresholds. In contrast, the Logistic Regression and Gradient Boosting models exhibited noticeably weaker performance. Their ROC curves remained closer to the diagonal baseline, indicating limited discriminatory power and a higher likelihood of misclassification—particularly in the context of identifying minority-class (deficit-prone) municipalities.

These visual findings corroborate the quantitative metrics presented in **Table 8**, reinforcing the conclusion that ensemble tree-based models—especially Extremely Randomized Trees—offer the most effective approach for predicting fiscal stress in local governments.

Figure 6 ROC Curve Comparison of Predictive Models

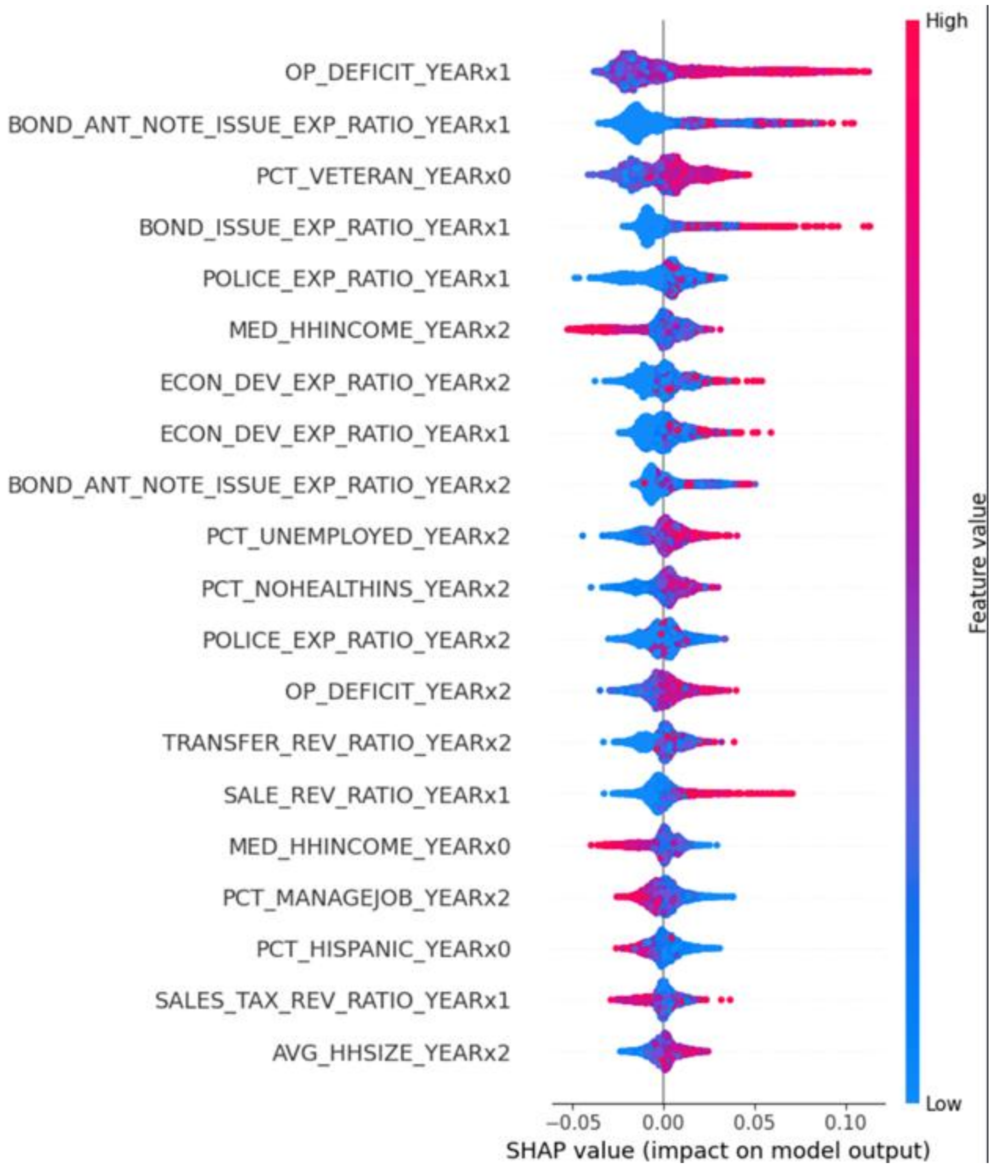


Based on the results presented above, Extremely Randomized Trees emerged as the most accurate and robust model for detecting municipalities at risk of budget deficits. Its superior performance across evaluation metrics—particularly in PRC-AUC—underscores its effectiveness in handling class imbalance and enhancing predictive sensitivity. With this model identified as optimal for forecasting fiscal stress, the following section interprets its predictions using SHAP (Shapley Additive Explanations**)** values to improve model transparency and policy relevance.

Figure 7 displays the most influential predictors, ranked by their mean absolute SHAP values, which quantify each variable's average marginal contribution to the predicted probability of a budget deficit. The horizontal axis represents the SHAP value, with values to the right indicating a positive influence (i.e., increased likelihood of a budget deficit) and values to the left indicating a negative influence (i.e., reduced likelihood of a budget deficit).

Each row corresponds to a specific feature, while each dot represents a SHAP value for an individual observation. The color gradient reflects the magnitude of the original feature values—red indicating higher values and blue indicating lower values—allowing for a nuanced interpretation of how predictor magnitudes affect outcomes. Among all predictors, OP_DEFICIT_YEARx1 (the municipality's budget deficit status in the previous year) stands out as the most influential. Higher values (in red) are strongly associated with increased predicted risk of future budget deficits, highlighting the persistence of fiscal stress over time. Additionally, bond-related indicators—particularly BOND_ANT_NOTE_ISSUE_EXP_RATIO_YEARx1 and BOND_ISSUE_EXP_RATIO_YEARx1—emerge as highly influential, suggesting that debt issuance activity is a critical signal of fiscal vulnerability. These SHAP-based insights not only validate key predictors identified in earlier regression models but also offer an interpretable pathway for policymakers to identify high-risk fiscal behavior and prioritize intervention.

Figure 7 Top 20 Predictors Ranked by SHAP Values



Note: In the variable names used throughout the analysis, the suffix _x0 refers to data from the prediction year, _x1corresponds to data from one year prior, and _x2 denotes data from two years prior to the prediction year.

Socioeconomic variables such as MED_HHINCOME_YEARx2 and housing variables like MED_HVALUE x1 significantly influenced model predictions, highlighting the interplay of financial, socioeconomic, and housing factors in shaping municipal budget deficits. The appearance of indicators across multiple years (e.g., suffixes _x1 and _x2) further underscores the relevance of temporal dynamics and lagged effects in accurately forecasting fiscal stress.

To visualize the marginal effects of key predictors, Dependence Plots were generated for the top ten SHAP-ranked features. Figures 9 through 11 illustrate three representative variables: the unemployment rate (*t –2*), median household income (*t–2*), and budget deficit ratio (*t –1*). In each plot, SHAP values are shown on the vertical axis, representing the feature's impact on the predicted deficit probability, while the horizontal axis reflects the original feature values. The color gradient—ranging from blue (lower values) to red (higher values)—offers additional insight into how the magnitude of each variable relates to model outcomes.

Figure 8 reveals a positive relationship between the unemployment rate and the likelihood of a deficit, particularly when unemployment exceeds 5%, signaling heightened fiscal risk under deteriorating labor market conditions. **Figure 9** shows a negative correlation between median household income and deficit probability, with municipalities below approximately $80,000 in income facing greater fiscal vulnerability. **Figure 10** demonstrates a strong positive association between prior-year budget deficits (*OP_DEFICIT_YEARx1*) and current-year fiscal stress, underscoring the persistence of structural deficits over time. These visualizations reinforce the SHAP summary findings and provide interpretable, policy-relevant insights into how financial and contextual indicators jointly drive fiscal stress at the local level.

Figure 8 Dependence Plot for Unemployment Rate-2

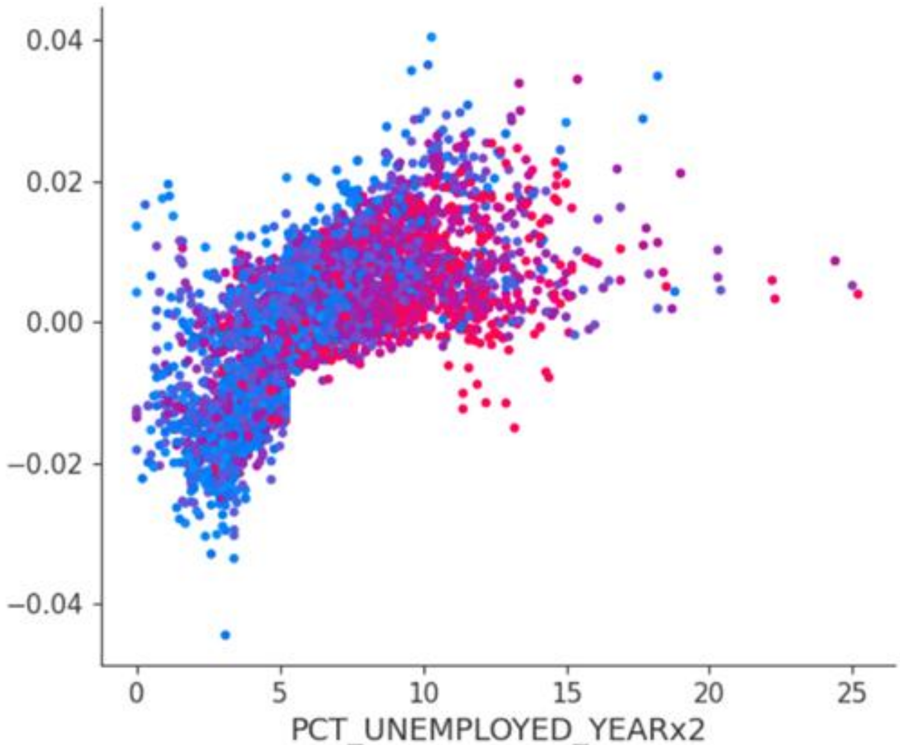


Figure 9 Dependence Plot for Median Household Income Year-2

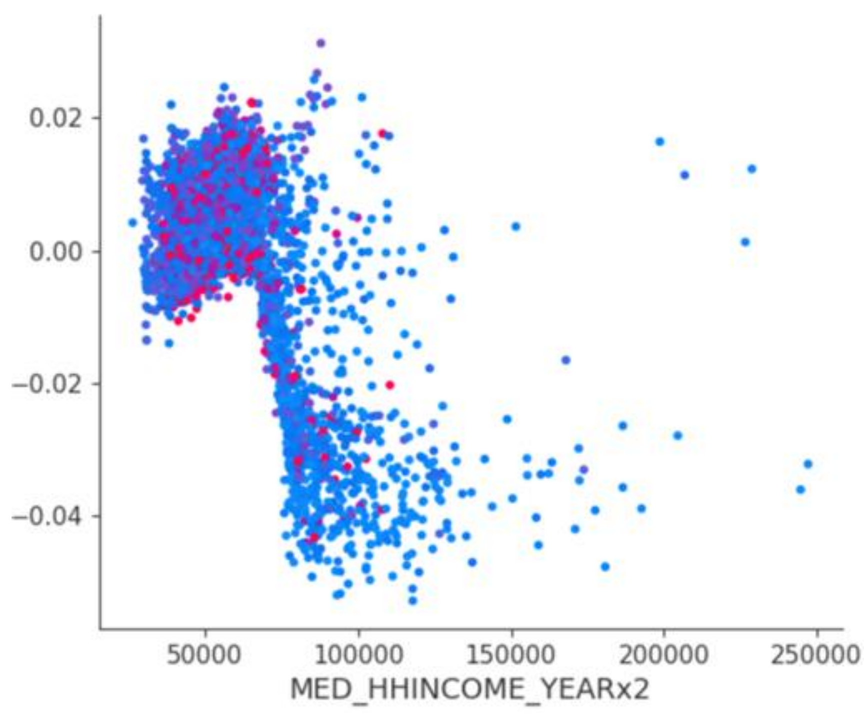
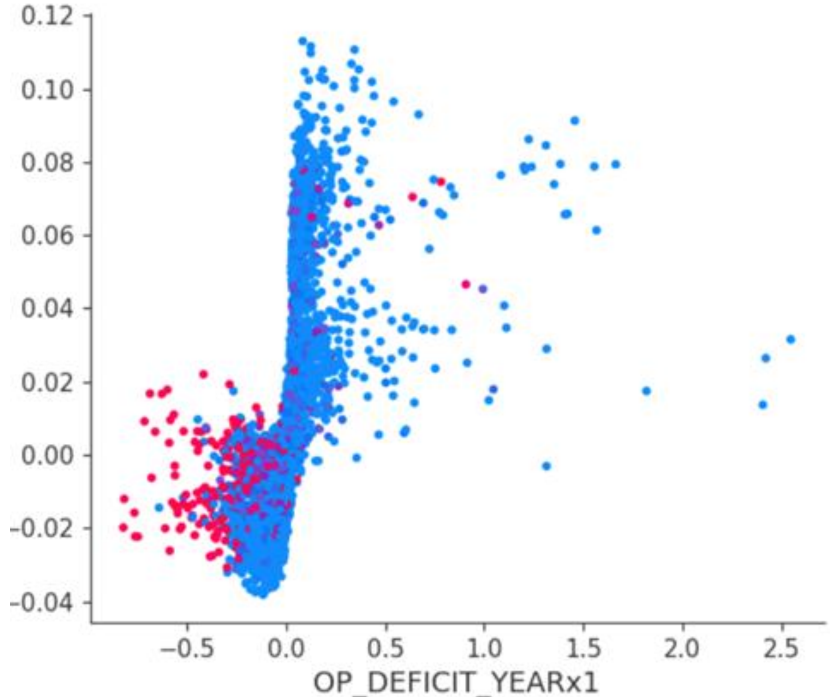


Figure 10 Dependence Plot for Deficit % Year-1



Collectively, the SHAP values and dependence plots provide a powerful solution to the "black box" challenge commonly associated with machine learning algorithms. By quantifying and visualizing the importance, magnitude, and direction of each predictor's influence on model outcomes, these tools substantially enhance interpretability. Beyond improving transparency, this analysis reaffirms the significance of key predictors while offering granular, observation-level explanations of their contributions to predicted fiscal distress. Such insights not only validate the predictive model but also support informed, data-driven decision-making in local government fiscal management.

DISCUSSION

This study assessed whether machine learning (ML) techniques can enhance the prediction of fiscal stress in local governments relative to traditional econometric models. Drawing on a panel dataset of 988 cities and towns in New York State from 2013 to 2022, we compared the predictive performance of several ML algorithms—including random forests, gradient boosting machines, and extremely randomized trees—against conventional approaches such as logistic regression, LASSO, and ridge regression. Our findings demonstrate that non-linear, ensemble-based ML models substantially outperform traditional methods in both predictive accuracy and model robustness.

A key insight from this study is the capacity of ML algorithms to capture complex, nonlinear interactions among fiscal, socioeconomic, demographic, and environmental variables. Models such as extremely randomized trees and random forests not only yielded higher classification performance, but also demonstrated greater resilience to challenges like class imbalance and multicollinearity. These advantages confirm that fiscal stress is often driven by multifaceted and interdependent factors that are inadequately captured by linear models.

Variable importance analyses reinforced the predictive significance of both fiscal and contextual factors. In particular, lagged budget deficits emerged as the most powerful predictors, underscoring the temporal persistence of fiscal distress. Socioeconomic indicators such as unemployment rate, population dynamics, and median housing value rates also played critical roles in identifying vulnerable municipalities. These results suggest that integrating broader structural and community-level variables into fiscal monitoring systems can yield more effective early warning capabilities than frameworks based solely on financial ratios or fund balance thresholds.

These findings are especially relevant considering increasing efforts by U.S. states to develop and enhance local fiscal early warning systems. Over the past decade, states including New York, Michigan, Ohio, Pennsylvania, and New Jersey have implemented formal fiscal oversight programs aimed at identifying distressed municipalities and averting fiscal crises. While these programs have improved oversight capacity, they frequently rely on rule-based thresholds or linear regression models, which may be limited in predictive scope. Our results suggest that ML-based approaches can serve as a next-generation complement, offering higher accuracy, greater adaptability, and the capacity to detect non-obvious risk patterns.

From a usability perspective, concerns about ML model transparency must be addressed to ensure their adoption by public finance practitioners. Although ensemble and neural network models are often viewed as “black boxes,” we show that tools such as SHAP values and dependence plots can provide actionable, interpretable insights. These interpretive tools link model predictions to specific risk indicators, giving decision-makers clearer insight into why certain municipalities are identified as areas of concern.

Methodologically, this study also illustrates the feasibility of implementing ML approaches in practical oversight contexts. Through careful data preprocessing—including missing value imputation, normalization, and resampling techniques such as SMOTE and class weighting—we demonstrate that real-world administrative data, despite its imperfections, can be successfully prepared for advanced modeling. The resulting workflow is replicable and scalable, offering a practical framework for other jurisdictions seeking to modernize their fiscal risk monitoring systems.

In sum, this research offers timely and practical guidance for state governments, oversight agencies, and public financial management professionals. As more states seek to refine or expand their fiscal monitoring systems, machine learning models provide a compelling upgrade to traditional frameworks. These models enable earlier detection of fiscal risk, more targeted intervention, and greater equity in the allocation of oversight resources. By incorporating a richer array of indicators and adjusting dynamically to evolving fiscal conditions, ML-enhanced systems can support a shift from reactive crisis management to proactive risk mitigation. Such a shift holds the potential to strengthen local government resilience, reduce the frequency of fiscal emergencies, and foster more sustainable, data-driven public financial management practices.

CONCLUSION

This study investigates the potential of machine learning (ML) models to enhance the prediction of fiscal stress in local governments, using data from 988 cities and towns in New York State spanning the years 2013 to 2022. By comparing a suite of ML algorithms—including random forests, gradient boosting machines, extremely randomized trees, neural networks, and support vector machines—against traditional econometric models such as logistic regression, LASSO, and ridge regression, we find that ML approaches significantly outperform conventional methods in terms of both predictive accuracy and model robustness.

The research makes three key contributions. First, from a methodological perspective, it introduces and validates the use of machine learning techniques in the context of local fiscal stress prediction—a domain historically dominated by linear regression approaches. Ensemble-based ML models, in particular, demonstrate the capacity to detect nonlinear interactions and complex risk patterns that traditional models may overlook, thereby expanding the analytical toolkit available to public finance researchers and practitioners.

Second, the study provides a substantive empirical contribution by leveraging a rich, multidimensional dataset that includes fiscal, demographic, economic, housing, and environmental indicators at the municipal level. This comprehensive data structure enables a more holistic analysis of fiscal stress, identifying dynamic and interrelated predictors such as lagged deficits, unemployment rates, and housing vacancy levels that jointly shape local fiscal outcomes.

Third, the findings offer direct policy relevance for states seeking to strengthen their local fiscal monitoring frameworks. As a growing number of U.S. states—including New York, Michigan, Ohio, and Pennsylvania—implement or refine early warning systems to detect fiscal distress, this study demonstrates that ML-based tools can significantly improve the timeliness, precision, and adaptability of these systems. By incorporating such models, oversight agencies and policymakers can better identify at-risk municipalities, allocate resources more efficiently, and intervene proactively to prevent crises.

Despite these contributions, the study has several limitations. The focus on New York State—while justified by its rich fiscal oversight infrastructure—limits the generalizability of findings to other institutional and policy contexts. Additionally, although interpretability methods such as SHAP values and dependence plots were employed, some ML models—especially deep neural networks—remain inherently opaque. Future research should explore more advanced explainable AI techniques and evaluate how predictive tools can be institutionalized within routine public financial management processes.

Several promising directions for future inquiry emerge. Expanding the scope to include cross-state comparative analyses would provide broader generalizability and allow for the assessment of model transferability across institutional settings. Incorporating higher-frequency or real-time data could improve the timeliness of predictions. Finally, combining predictive analytics with causal inference approaches could help not only forecast fiscal distress but also uncover its underlying drivers—offering evidence to design more targeted and effective policy interventions.

In an era of growing fiscal uncertainty, public sector decision-makers need forecasting tools that are not only accurate but also adaptive, interpretable, and actionable. This study shows that machine learning provides a powerful, yet underutilized, resource for enhancing fiscal foresight. By embedding ML techniques into state-level fiscal monitoring systems, governments can take a critical step toward building smarter, more resilient, and data-informed frameworks for public financial oversight.

REFERENCES

Alm, James, Robert D. Buschman, and David L. Sjoquist. "Foreclosures and Local Government Revenues from the Property Tax: The Case of Georgia School Districts." *Regional Science and Urban Economics* 46 (May 2014): 1–11. https://doi.org/10.1016/j.regsciurbeco.2014.01.007.

Alm, James, Robert D. Buschman, and David L. Sjoquist. "Rethinking Local Government Reliance on the Property Tax." *Regional Science and Urban Economics* 41, no. 4 (2011): 320–331. https://doi.org/10.1016/j.regsciurbeco.2011.03.006.

Ashraf, S., G. S. Félix, and Z. Serrasqueiro. "Do Traditional Financial Distress Prediction Models Predict the Early Warning Signs of Financial Distress?" *Journal of Risk and Financial Management* 12, no. 2 (2019): 55. https://doi.org/10.3390/jrfm12020055.

Baret, Kevin, Anne Barbier-Gauchard, and Theofanis Papadimitriou. "Forecasting Stability and Growth Pact Compliance Using Machine Learning." *The World Economy* 47, no. 1 (2023): 188–216. https://doi.org/10.1111/twec.13518.

Belly, Guillaume, Lukas Boeckelmann, Carlos Mario Caicedo Graciano, Andrea Di Iorio, Kleopatra Istrefi, Vasilios Siakoulis, and Agnès Stalla-Bourdillon. "Forecasting Sovereign Risk in the Euro Area via Machine Learning." *Journal of Forecasting* 42, no. 4 (2023): 657–684. https://doi.org/10.1002/for.2938.

Breiman, Leo. "Random Forests." *Machine Learning* 45, no. 1 (2001): 5–32. https://doi.org/10.1023/A:1010933404324.

Chen, Mingjie, Matthew DeHaven, Herbert Kitschelt, Seung Jin Lee, and Michael J. Sicilian. "Identifying Financial Crises Using Machine Learning on Textual Data." *International Finance Discussion Papers* No. 1374, Board of Governors of the Federal Reserve System, 2023. https://doi.org/10.17016/IFDP.2023.1374.

De Marchi, Riccardo, and Andrea Moro. "Forecasting Fiscal Crises in Emerging Markets and Low-Income Countries with Machine Learning Models." *Bank of Italy Temi di Discussione (Working Papers)* No. 1405 (2023). https://doi.org/10.32057/0.TD.2022.1405.

Demyanyk, Yuliya, and Iftekhar Hasan. "Financial Crises and Bank Failures: A Review of Prediction Methods." *Federal Reserve Bank of Cleveland Working Paper* No. 09-04 (2009). https://doi.org/10.26509/frbc-wp-200904.

Edinak, E. A., and A. A. Shirov. "Assessment of the Relationship between the Qualification Structure of Employment and Household Consumption Using Input–Output Tables." *Studies on Russian Economic Development* 32, no. 6 (2021): 593–602. https://doi.org/10.1134/S1075700721060046.

Felix, Alison, and Kate Watkins. "The Impact of an Aging U.S. Population on State Tax Revenues." Accessed April 26, 2025. https://www.kansascityfed.org.

Fioramanti, Marco. "Predicting Sovereign Debt Crises Using Artificial Neural Networks: A Comparative Approach." *Journal of Financial Stability* 4, no. 2 (2008): 149–164. https://doi.org/10.1016/j.jfs.2008.01.001.

Friedman, Jerome H. "Greedy Function Approximation: A Gradient Boosting Machine." *The Annals of Statistics* 29, no. 5 (2001): 1189–1232. https://doi.org/10.1214/aos/1013203451.

Geurts, Pierre, Damien Ernst, and Louis Wehenkel. "Extremely Randomized Trees." *Machine Learning* 63, no. 1 (2006): 3–42. https://doi.org/10.1007/s10994-006-6226-1.

Gorina, Evgenia, Craig Maher, and Marc Joffe. "Local Fiscal Distress: Measurement and Prediction." *Public Budgeting & Finance* 38, no. 1 (2018): 72–94. https://doi.org/10.1111/pbaf.12165.

Hellwig, Kristina. "Predicting Fiscal Crises: A Machine Learning Approach." *IMF Working Paper* No. WP/21/150 (2021). https://doi.org/10.5089/9781513573843.001.

Hondroyiannis, George, and Evangelia Papapetrou. "Do Demographic Changes Affect Fiscal Developments?" *Public Finance Review* 28, no. 5 (2000): 468–488. https://doi.org/10.1177/109114210002800505.

Holopainen, Marko, and Peter Sarlin. "Toward Robust Early-Warning Models: A Horse Race, Ensembles, and Model Uncertainty." *Bank of Finland Research Discussion Papers* 6 (2015): 1–48.

Hosmer, David W., Stanley Lemeshow, and Rodney X. Sturdivant. *Applied Logistic Regression.* 3rd ed. Hoboken, NJ: Wiley, 2013. https://doi.org/10.1002/9781118548387.

Justice, Jonathan B., and Eric A. Scorsone. "Measuring and Predicting Local Government Fiscal Stress: Theory and Practice." In *Handbook of Local Government Fiscal Health*, edited by Helisse Levine, Jonathan B. Justice, and Eric A. Scorsone, 43–74. Burlington, MA: Jones & Bartlett, 2012.

Justice, Jonathan B., Melanie Fudge, Helisse Levine, D'Leslie Bird, and Muhammad Naveed Iftikhar. "Using Fiscal Indicator Systems to Predict Municipal Bankruptcies." In *The Palgrave Handbook of Government Budget Forecasting*, 275–302. Palgrave Macmillan, 2019.

Leiser, Stephanie, and Sarah Mills. "Local Government Fiscal Health: Comparing Self-Assessments to Conventional Measures." *Public Budgeting & Finance* 39, no. 3 (2019): 75–96. https://doi.org/10.1111/pbaf.12226.

Lipton, Zachary C. "The Mythos of Model Interpretability." *Communications of the ACM* 61, no. 10 (2018): 36–43. https://doi.org/10.1145/3233231.

Liu, Lanbiao, Chen Chen, and Bo Wang. "Predicting Financial Crises with Machine Learning Methods." *Journal of Forecasting* 41, no. 5 (2021): 871–910. https://doi.org/10.1002/for.2840.

Luitel, Hari Sharan, and Mehmet Serkan Tosun. "A Reexamination of State Fiscal Health and Amnesty Enactment." *International Tax and Public Finance* 21, no. 5 (2014): 874–893. https://doi.org/10.1007/s10797-013-9278-8.

Office of the New York State Comptroller. "Local Government | Office of the New York State Comptroller." Accessed March 4, 2025. https://www.osc.ny.gov/local-government.

"Open Book New York | Office of the New York State Comptroller." Accessed March 4, 2025. https://www.osc.ny.gov/open-book-new-york.

Piermarini, Davide, Antonio Maria Sudoso, and Vittorio Piccialli. "Predicting Municipalities in Financial Distress: A Machine Learning Approach Enhanced by Domain Expertise." *arXiv* (2023). https://arxiv.org/abs/2302.05780.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning Representations by Back-Propagating Errors." *Nature* 323, no. 6088 (1986): 533–536. https://doi.org/10.1038/323533a0.

Sarlin, Peter. "Neuro-Genetic Predictions of the Global Financial Crisis." *Neural Computing & Applications* 24, no. 3–4 (2014): 663–673. https://doi.org/10.1007/s00521-012-1281-y.

Sarlin, Peter. "On Policymakers' Loss Functions and the Evaluation of Early Warning Systems." *Economics Letters* 124, no. 3 (2014): 500–504. https://doi.org/10.1016/j.econlet.2014.07.017.

Skidmore, Mark, and Eric Scorsone. "Causes and Consequences of Fiscal Stress in Michigan Cities." *Regional Science and Urban Economics* 41, no. 4 (2011): 360–371. https://doi.org/10.1016/j.regsciurbeco.2011.02.007.

Sohl, Shannon, Michael T. Peddle, Kurt Thurmaier, Curtis H. Wood, and Gregory Kuhn. "Measuring the Financial Position of Municipalities: Numbers Do Not Speak for Themselves." *Public Budgeting & Finance* 29, no. 3 (2009): 74–96. https://doi.org/10.1111/j.1540-5850.2009.00937.x.

Stone, Samuel B., Akheil Singla, James Comeaux, and Charlotte Kirschner. "A Comparison of Financial Indicators: The Case of Detroit." *Public Budgeting & Finance* 35, no. 4 (2015): 90–111. https://doi.org/10.1111/pbaf.12079.

Tölö, Eero. "Predicting Systemic Financial Crises with Recurrent Neural Networks." *Journal of Financial Stability* 49 (2020): 100746. https://doi.org/10.1016/j.jfs.2020.100746.

UCLA Institute for Digital Research and Education. "Logit Regression | R Data Analysis Examples." Accessed April 16, 2025. https://stats.oarc.ucla.edu/r/dae/logit-regression/.

APPENDIX

APPENDIX I Definitions of Variables and Indicators, abbreviations, and Data Sources

Variable Category and Name	Abbreviation	Definition	Data Source
Financial Variables			Open Book New York
Total Revenues	TR	The total amount of income received by a municipality from all sources during a fiscal year	
Transfer Revenue	TransRev	Total amount of revenue received through transfers, including intergovernmental aid and financial grants	
Federal and State Aid Revenue	FSAidRev	Total amount of revenues derived from Federal and State Aid	
Sales of Assets Revenue	AssetSaleRev	Total amount of revenue generated from the sale of goods, services, or other taxable activities	
Sales and Use Tax Revenue	SalesUseTaxRev	Total amount of revenue derived from sales and use taxes of tangible personal property and/or the consumption of goods and/or services	
Property Tax Revenues	PropTaxRev	Total amount of revenue generated from real property taxes and assessments	
Total Expenditures	TE	The total amount of money spent by a municipality on all functions during a fiscal year	
Economic Development Expenditure	EconDevExp	Total amount of expenditures for economic development	
Disaster Expenditure	DisasterExp	Total amount of expenditures related to preparedness and response to natural disasters and emergency	
Police Expenditure	PolExp	Total amount of expenditures for police services, sheriff, traffic control, etc.	
Transportation Expenditure	TransportExp	Total amount of expenditures for highway and transportation services	
Personnel Expenditure	PerExp	Total amount of expenditure on personnel services and employee benefits	
Total Fund Balance	TotFundBal	The cumulative difference between the government's total assets and total liabilities at the end of a fiscal period	
Total Cash Balance	TotCash	Total amount of liquid assets available to a government	
Debt Service Expenditure	DebtServExp	Total amount of expenditures for debt service	
Bond Issued	IssuedBnd	The total value of bonds issued by the entity to raise funds for public projects or services	
Bond Paid	PaidBnd	The amount of outstanding bonds that have been repaid, reducing overall debt obligations.	

Bond Issued during the Current Year	BndIssuedCY	The total value of bonds issued within the current fiscal year	
Bond Paid during the Current Year	BndPaidCY	The amount of outstanding bonds that have been repaid within the current fiscal year	
Bond Anticipation Notes issued	BANIssued	The total value of short-term borrowing through bond anticipation notes	
Bond Anticipation Notes paid	BANPaid	The amount of bond anticipation notes that have been repaid	
Bond Anticipation Notes issued during the current year	BANIssuedCY	The total value of bond anticipation notes issued withing the current fiscal year	
Bond Anticipation Notes paid during the current year	BANPaidCY	The amount of bond anticipation notes that have been repaid within the current fiscal year	
Socioeconomic Variables			ACS Census
IND_GINI	IND_GINI	Gini Index	
Unemployment Rate	PCT_UNEMPLOYED	The percentage of the population that is unemployed within the jurisdiction	
Service Sector Employment rate	PCT_SERVICEJOB	The percentage of jobs that are in the service sector within the jurisdiction	
High-Skilled Occupation Employment Rate	PCT_MANAGEJOB	The percentage of the labor force employed in management, business, science, and arts-related occupations within the jurisdiction	
Public Sector Employment Rate	PCT_PUBLICJOB	The percentage of the labor force employed in public administration within the jurisdiction	
Median Household Income	MED_HHINCOME	The median household income (in U.S. dollars) within the jurisdiction	
Public Assistance Recipients	PCT_PUBLICASST	The percentage of households receiving income from public assistance programs	
Poverty Rate	PCT_POVERTY	The percentage of individuals whose household or personal income falls below the official poverty threshold	
Uninsured Population Rate	PCT_NOHEALTHINS	The percentage of the civilian noninstitutionalized population without health insurance coverage	
FEMA Disaster Declarations	FEMAcount	The total number of FEMA disaster declarations within the jurisdiction	
Demographic Variables			ACS Census

Sex Ratio	RATIO_SEX	The number of male residents per 100 female residents	
Total Population	TOT_POP	The total population estimated by the U.S. Census Bureau within the jurisdiction	
Adult Population (18+)	PCT_OVER18	The percentage of the population aged 18 years or older	
Senior Population (65+)	PCT_OVER65	The percentage of the population aged 65 years or older	
Households with Minors	PCT_HHMINOR	The percentage of households that have at least one resident under the age of 18	
Average Household Size	AVG_HHSIZE	The average number of individuals residing in a single household	
High School Graduation Rate ~~(25+)~~	PCT_HSGRAD	The percentage of individuals aged 25 and over who have completed high school or an equivalent qualification	
Veteran Population	PCT_VETERAN	The percentage of individuals aged 18 and older who have served in the U.S. military	
Disability Rate	PCT_DISABILITY	The percentage of the civilian non-institutionalized population with a disability	
Black Population Percentage	PCT_BLACK	The percentage of the population that identifies as Black or African American.	
Hispanic Population Percentage	PCT_HISPANIC	The percentage of the population that identifies as Hispanic or Latino	
Housing Variables			ACS Census
Vacancy Rate	PCT_VACANT	The percentage of housing units that are unoccupied	
Renter-Occupied Housing	PCT_RENTAL	The percentage of housing units that are occupied by renters	
Median Home Value	MED_HVALUE	The median value of owner-occupied housing units	
GRAPI	PCT_GRAPI	GRAPI (Gross Rent as a Percentage of Household Income)	
SMOCAPI	PCT_SMOCAPI	SMOCAPI (Selected Monthly Owner Costs as a Percentage of Household Income)	

APPENDIX II: Machine Learning Models for Fiscal Stress Prediction

A. Logistic Regression (Baseline Model)

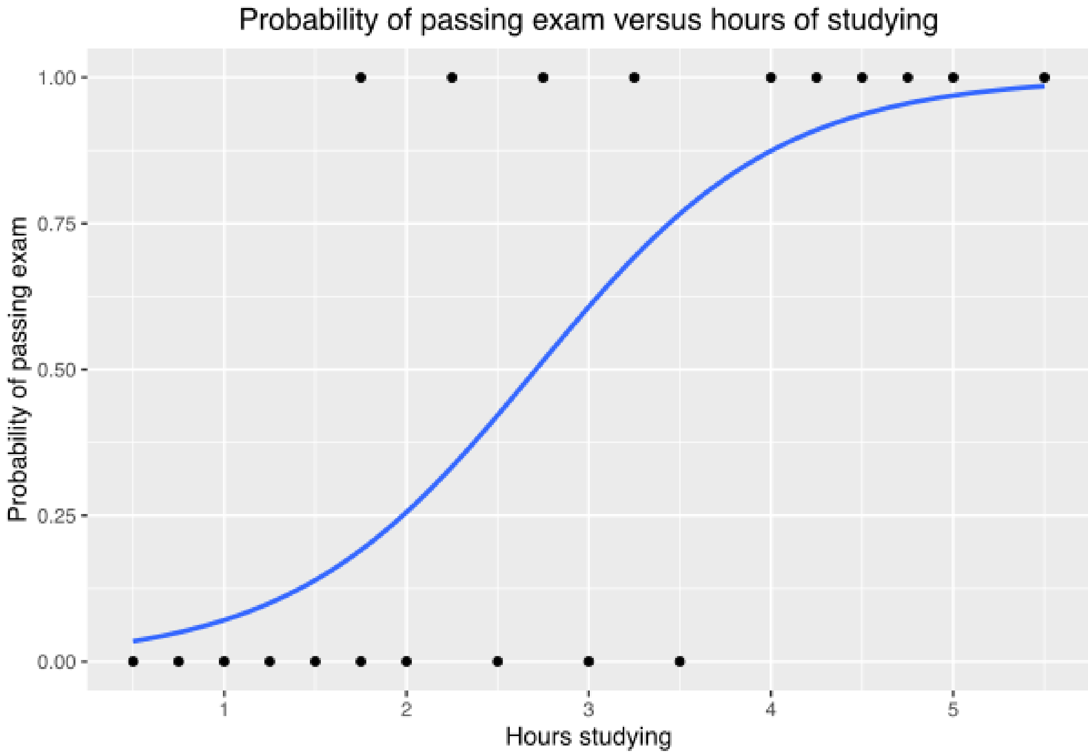


Figure A1: Example logistic regression curve fitted to data, showing how the probability of an outcome (e.g., passing an exam) increases with a predictor (hours studying) in an S-shaped manner.

Logistic regression is a generalized linear model commonly employed for binary classification tasks and is extensively utilized as a baseline method in econometric analyses of fiscal distress (Hosmer, Lemeshow, & Sturdivant, 2013). Specifically, logistic regression models the log-odds of the occurrence of fiscal stress (coded as 1) through a linear combination of predictor variables. The linear predictors are then transformed via the logistic function (sigmoid), mapping outputs into predicted probabilities ranging from 0 to 1. Model parameters are generally estimated using maximum likelihood estimation, ensuring that the predicted probabilities closely reflect observed binary outcomes. In the present study, logistic regression serves as an essential baseline model, leveraging lagged fiscal indicators to predict the likelihood of fiscal stress occurrence in subsequent periods. Despite its interpretability and simplicity, logistic regression may inadequately capture complex nonlinear relationships inherent in real-world data. This limitation often motivates the integration of more sophisticated and flexible machine learning algorithms to enhance predictive performance.

B. Gradient Boosting (Ensemble Tree Model)

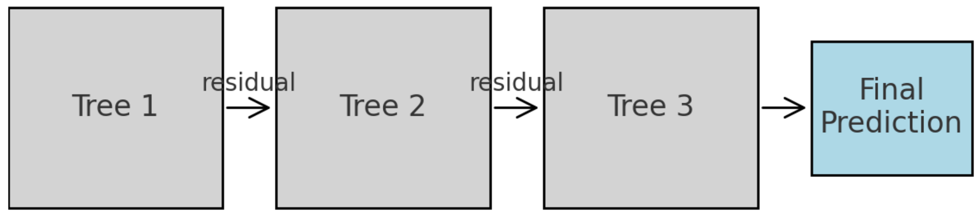


Figure A2: Schematic of gradient boosting, where decision trees (weak learners) are added sequentially. Each new tree is trained on the residuals (errors) of the previous ensemble, and the final prediction is the aggregated output of all trees.

Gradient boosting is an ensemble learning method that iteratively builds a predictive model by sequentially adding weak learners, typically shallow decision trees, aimed at correcting residual errors from previous iterations (Friedman, 2001). Unlike bagging methods, boosting focuses on bias reduction: each tree fits on the pseudo-residuals (the remaining prediction errors) of the current model, effectively performing a gradient descent optimization in function space. Over many iterations, the ensemble of trees converges to a strong predictive model that can capture complex nonlinear patterns. In our study, we employ gradient boosting to predict fiscal stress by training an ensemble of trees on lagged fiscal features, where each successive tree improves the prediction of stress by addressing the shortcomings (residual errors) of the prior trees. This method allows optimization of an arbitrary differentiable loss function (we use a classification loss for fiscal stress) and often achieves higher accuracy than single models or bagged trees.
Gradient boosting produces an ensemble model (commonly known as gradient boosted trees) by iteratively fitting new trees to the residuals of the model.

C. Artificial Neural Network (ANN)

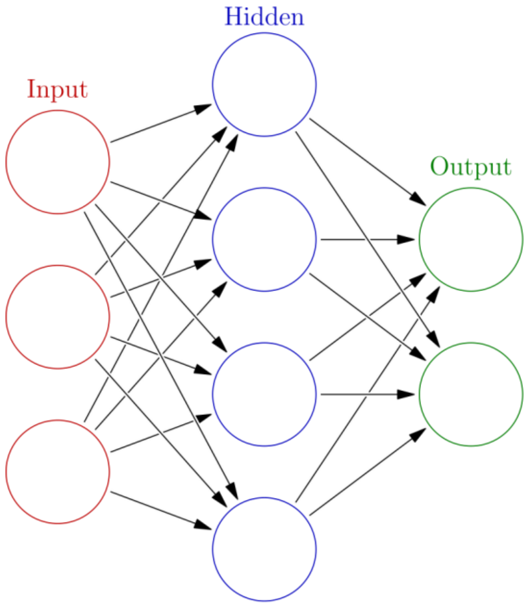


Figure A3: Topology of a simple Artificial Neural Network with an input layer, one hidden layer, and an output layer. Each circle represents a neuron, and connections (not shown) carry weighted signals forward.

An artificial neural network is a layered computational model inspired by biological neurons, capable of approximating complex nonlinear functions. The network consists of interconnected units called neurons arranged in an input layer, one or more hidden layers, and an output layer. Each neuron receives inputs (either raw features or outputs from previous layers), computes a weighted sum, applies a nonlinear activation function, and passes the result to the next layer. This layered structure enables ANNs to learn high-level feature representations. The learning principle of an ANN is error backpropagation: during training, the network's output is compared to the true outcome, and the error is propagated backward through the layers to adjust the connection weights via gradient-based optimization (Rumelhart, Hinton, and Williams 1986). Over many iterations (epochs), the network "learns" weight values that minimize the prediction error on the training data. In our study, we utilize a feed-forward ANN to predict fiscal stress, using past fiscal and socio-economic variables as inputs. The network's hidden layer allows it to capture nonlinear interactions between predictors (e.g., trends in revenues and expenditure) that might signal impending fiscal stress. We trained the ANN on historical data (with fiscal stress labels) and used a validation process to tune its complexity and prevent overfitting.

Artificial neural networks consist of layers of interconnected "neurons" that apply weighted sums and activation functions to learn complex mappings. They are trained via supervised learning, iteratively adjusting weights to minimize a loss function using techniques like backpropagation (Rumelhart, Hinton, and Williams 1986). This enables ANNs to model nonlinear relationships that simpler models might miss.

D. Random Forest (Ensemble Tree Model)

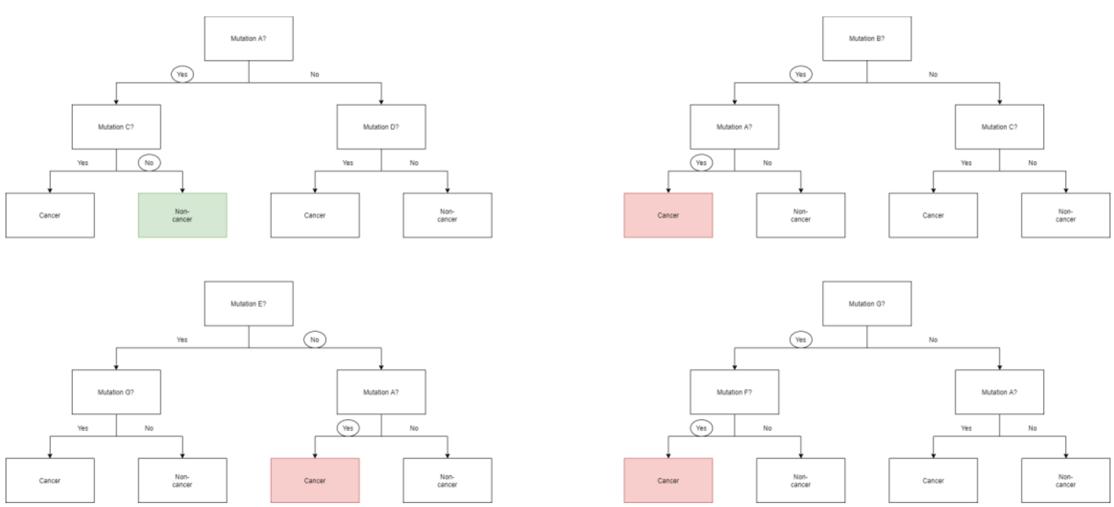


Figure A4: Example of a random forest classification process with four decision trees. Each tree votes on whether a sample is "Cancer" or "Non-cancer" (illustrative task), and the majority vote determines the final classification.

Random forests, introduced by Breiman (2001), an ensemble learning method that aggregates a large number of decision trees to improve predictive accuracy and robustness. Each individual decision tree in the forest is trained on a bootstrap sample of the data (random sampling with replacement) and typically using a random subset of features at each split. This randomization ensures the trees are de-correlated, so that their errors are largely independent. For classification tasks, a random forest outputs the class that is the majority vote among the predictions of all constituent trees; for regression, it outputs the average of their predictions. By averaging many deep trees, random forests achieve a strong balance of variance reduction (through averaging) while maintaining low bias, thus mitigating the overfitting that a single complex tree might exhibit. In our application, we train a random forest on lagged financial indicators (e.g., fund balances, debt ratios, revenue trends) to predict future fiscal stress. Each tree provides a "vote" (stress or no stress), and the ensemble's majority vote yields the final prediction. The random forest also allows us to measure variable importance, helping to identify which fiscal indicators are the strongest predictors of distress.

E. Extremely Randomized Trees (Extra Trees)

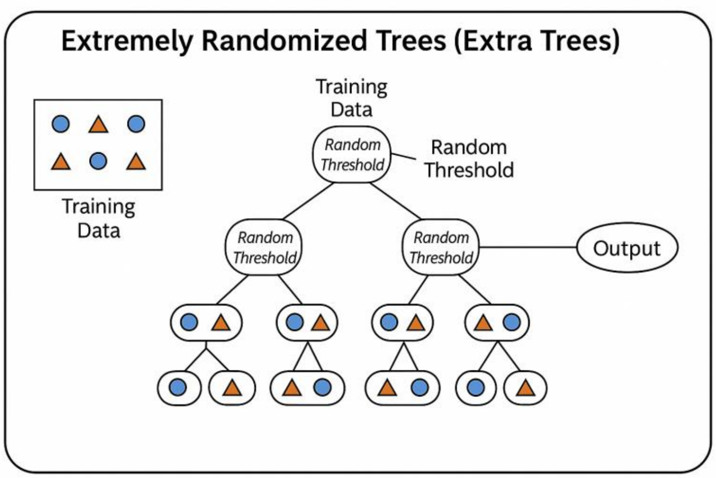


Figure A5: Illustration of an ensemble of decision trees (comparable to a random forest). Extra Trees uses the entire dataset for each tree and selects split points extremely randomly (e.g., random thresholds for "Mutation" features in this illustrative example) instead of the best splits.

Extremely randomized trees (Extra Trees), proposed by Geurts, Ernst, and Wehenkel (2006), extend the random forest methodology by incorporating additional randomization during the tree-building process. Like a random forest, an Extra Trees model constructs many decision trees and combines them by averaging or majority voting. However, Extra Trees differ in two keyways: (1) No Bootstrapping: each tree is trained on the full training sample rather than a bootstrap resample. (2) Random Splits: when growing a tree, split thresholds are chosen at random for each candidate feature (within that feature's range) instead of computing the optimal split point; among these random splits, the best one is then selected to split the node. These differences mean that Extra Trees typically produce even more diverse trees, reducing variance at the cost of a slight increase in bias. In the context of our study, we use the Extra Trees classifier on the same feature set (lagged fiscal data) to predict fiscal stress. The model trains multiple extremely randomized decision trees on all the data, where each tree's structure is highly random. The final prediction is obtained by aggregating the predictions of all these trees. We found that this additional randomness can improve generalization, and indeed Extra Trees achieved strong predictive performance in our fiscal stress experiments.