International Journal of Accounting and Economics Studies, 13 (2) (2026) 513-525

International Journal of Accounting and Economics Studies

Website: www.sciencepubco.com/index.php/IJAES
https://doi.org/10.14419/wkr3pj23
Research paper

Machine Learning Models for Early Warning of Financial Crises in the U.S. Economy Using Macro-Financial Indicators

Md Khalilor Rahman [1] *, Md Sazzad Hossain [1], Sayem Ul Haque [2], Kazi Abu Jahed [3], Md Sharif Robbani [4], Mahamuda Akter Shati [5], Tanjina Tuly [6], Md Toushif Pramanik [7]

[1] *MBA, Business analytics, Gannon University, Erie, PA, USA*
[2] *MBA- Business Analytics, Gannon University, PA, USA*
[3] *Master of Science- Business Intelligence and Analytics, Saint Joseph's University, PA, USA*
[4] *School of Engineering & Technology at Western Illinois University, Macomb, Illinois, USA*
[5] *Master's in Business Analytics, Grand Canyon University*
[6] *MSc in Business Analytics, Trine University*
[7] *Master of Embedded Software Engineering, Gannon University, Erie, USA*
**Corresponding author E-mail: rahman012@gannon.edu*

Received: February 20, 2026, Accepted: March 10, 2026, Published: March 16, 2026

Abstract

Early identification of U.S. recessions is critically important for policymakers, financial institutions, and investors; however, accurate prediction remains challenging. The data is sparse, things in the economy shift in ways that aren't always predictable, and macro-financial indicators don't always move in straight lines. This study looks at a set of machine learning models, logistic regression, random forest, and XGBoost, to see how well they can flag recessions six months ahead. We use a monthly dataset stretching from 1970 to 2025, and define recession periods using the National Bureau of Economic Research dates. The problem is set up as a binary classification: will a recession happen or not? To ensure methodological rigor, the models are trained using strict time-series cross-validation, and we evaluate their performance using ROC-AUC, precision-recall curves, Brier scores, calibration curves, and the lead time to issue a warning. We find that non-linear models like XGBoost and random forests tend to beat logistic regression by 5–10% on ROC-AUC. These models capture nonlinear interactions between indicators that simpler linear models may fail to detect. The yield curve spread comes out as the most reliable signal, followed by unemployment and the VIX volatility index. Credit indicators add a little extra, but not much. The models can give useful warning signals about 5–6 months before a recession starts, though accuracy drops when you look further ahead or during times when the economy goes through big changes, like after 2008 or post-COVID-19. Calibration tests show that predicted probabilities aren't perfect; turning them into reliable risk estimates needs some care. The study also points out gaps in previous ML-based early warning work: people often rely on random cross-validation, don't test models over different forecast horizons, and don't focus enough on interpretability. By tackling these issues, our approach is more robust and respects the time dimension, making it more suitable for real-world policy use. Overall, the results suggest that well-calibrated, interpretable ML models can work alongside traditional econometrics to help policymakers act before a recession hits.

Keywords: *Early Warning Systems; Financial Crisis Prediction; Macro-Financial Indicators; Probability Calibration; Structural Breaks.*

1. Introduction

1.1. Background and motivation

Financial crises and recessions hit hard. They leave behind long stretches of unemployment, lost income, shaky financial systems, and gaps in output that take years to recover. Because of this, building systems that can give early warnings has been a key focus in macroeconomic research for decades. Classical studies showed that financial data often carries hints about where the economy is headed. Dueker (1997) found that the slope of the yield curve can actually improve forecasts for recessions, which suggests that term spreads hold clues about future economic activity and monetary policy [14]. Estrella and Mishkin (1998) backed this up, showing that financial variables, especially the yield curve, can predict U.S. recessions better than many traditional macroeconomic indicators when tested out-of-sample [15]. Chauvet and Potter (2005) went further, showing that financial indicators help detect turning points if you use them within dynamic forecasting frameworks, which makes a good case for paying attention to asset prices when looking for cyclical shifts [7].

Those early econometric approaches laid the groundwork, but they had limits. They relied on specific assumptions and relatively simple sets of predictors. As high-frequency and cross-market data became more available, researchers started experimenting with more flexible approaches. Samitas (2020) argued that machine learning could take early warning systems to the next level because nonlinear algorithms can pick up interactions among macro-financial indicators that linear models like probit or logit might miss [30]. Ray (2025) showed that combining stock, bond, and forex data through modern ML models can boost the accuracy of crisis predictions [26]. Chouksey et al. (2025) built an AI-driven system for financial risk in the U.S. digital economy, showing that mixing macroeconomic and digital indicators can improve detection in fast-changing conditions [8]. ML-based early warning systems have also started looking beyond just crises. Rahman (2025) developed a system to spot micro-inflation clusters, demonstrating that high-dimensional models can reveal instability signals before they show up in broad statistics [25]. Reza et al. (2025) created real-time indicators of financial distress from digital signals, highlighting how important it is to have timely, adaptive data sources [29]. All of this shows a clear evolution: from simple yield-curve models to complex, data-rich ML systems. This study sits right at that crossroads, trying to blend the solid theory behind classical financial indicators with the flexibility of modern machine learning to build a recession early warning system that actually looks forward rather than just explains the past.

1.2. Methodological risks in ML-based crisis prediction

Even though machine learning looks promising, predicting crises isn't straightforward. One big issue is the rare event problem. Recessions and systemic crises don't happen often, so the data is heavily skewed toward normal periods. Coffinet and Kien (2020) pointed out that this imbalance can make models look good on paper while missing the actual crises [10]. Without careful rebalancing or using metrics that account for this, ML systems can give a false sense of confidence. Then there's the small sample problem. Even with decades of data, the number of actual recessions is limited, which leaves little to train on. Holopainen and Sarlin (2015) found that early warning models can be unstable and uncertain, with results bouncing around depending on the sample or model specification [18]. When you throw complex nonlinear models at a small set of crises, overfitting becomes a real risk. Structural breaks make things messier. Economies change, new financial instruments, new regulations, globalization, and technology. Models trained in one environment might not hold in another. Das (2025) explained that predictive systems in unstable conditions need to handle shifts in underlying relationships [11]. Debnath et al. (2025) showed that anomaly detection breaks down when the system changes, which highlights the need for robustness [12]. Big shocks, like the 2008 financial crisis or COVID-19, can completely change how indicators move together, leaving static models in the dust.

Another risk is models making unrealistic predictions. Aashish et al. (2025) argued that ML systems in high-stakes areas should respect domain constraints to avoid nonsense outputs [1]. In crisis prediction, ignoring economic structure can let noise masquerade as a real signal. Interpretability adds another layer of challenge. Hasan et al. (2025) emphasized that in high-stakes, data-sparse situations, black-box models can reduce trust and make policy adoption harder [17] [16]. For early warning systems, policymakers don't just want accurate predictions; they need signals they can make sense of. All of these risks, rare events, small samples, structural breaks, overfitting, ignoring constraints, and hard-to-interpret outputs, show why building an ML-based early warning system is tricky. Each one needs careful attention if the goal is to actually warn before a crisis, rather than just spot patterns after the fact.

1.3. Research objectives

The main goal of this study is to build a machine learning early warning system that can predict U.S. recessions six months before they hit, using macro-financial indicators. The six-month horizon is chosen because it's a practical window for policymakers, it's far enough ahead to make a difference but not so far that predictions become unreliable. Focusing on recession onset, rather than trying to classify the present moment, lets the study test real forward-looking prediction rather than just spotting what's already happening. Another goal is to compare traditional linear econometric models with nonlinear machine learning models. This is to see if ML really captures hidden patterns or if the apparent improvements are just quirks of how validation is done. The study also looks at how lead time affects performance, checking whether predictions get stronger or weaker the further out they go. Calibration is important too, predicted probabilities have to match what actually happens, because policymakers need numbers they can trust. So, beyond accuracy, the study checks if the probabilities make sense against real recession frequencies. Robustness is also a key goal. The analysis tests models across different economic regimes, including financial crises and periods of unconventional monetary policy, to make sure predictions aren't just riding on one historical event. Finally, the study looks at how understandable the models are. It asks whether the predictors they flag make sense with macro-financial theory and whether they shed light on why recession risk rises. Altogether, these goals try to balance good prediction with solid methodology and some real-world interpretability.

1.4. Contributions

This study adds to recession forecasting in a few concrete ways. First, it uses a time-consistent validation setup that avoids look-ahead bias and data leaks. Models are trained and tested using rolling or expanding windows that mimic real-time conditions, so performance numbers reflect actual forecasting skill, not just hindsight. This tackles a common problem in ML on macro data, where sloppy cross-validation can make results look better than they really are. Second, the study builds in leak prevention at every step of preprocessing and feature engineering. Every transformation, scaling step, or lag is done within the training window before applying it to the test period, stopping accidental information leaks. It also tests performance across multiple forecast horizons, showing how early warning signals hold up or fade over time.

Structural breaks get special attention, too. By checking model stability across different macro regimes, the research shows whether predictive relationships hold when the financial landscape changes. Probability calibration is evaluated alongside normal accuracy metrics, highlighting how meaningful the numbers are for policy decisions. Finally, by examining features through an economic lens, the study links machine learning insights back to macro theory, not just prediction. Taken together, these contributions aim to make early warning systems that are not only accurate but also reliable, understandable, and useful for real-world decision-making.

2. Literature Review

2.1. Traditional recession prediction

When it comes to predicting recessions, a lot of the classic work focuses on financial indicators as early warnings. A big one that keeps popping up is the term structure of interest rates, especially the spread between long-term and short-term Treasury yields. Estrella and Mishkin (1998) showed that this yield curve slope actually does a pretty good job predicting U.S. recessions. In fact, it often beats other financial or macroeconomic variables when you plug it into probit models [15]. Their work made the yield spread a kind of go-to signal for early warning systems, simple but surprisingly effective. Dueker (1997) took this further, arguing that if you play around with different ways of measuring spreads and include some dynamic info, you can make the recession probability estimates even stronger [14]. The yield curve doesn't just move with business cycles; it gives a peek at what markets expect for monetary policy and the economy ahead. This helped push probit and logit models into the mainstream, where you treat recession chances as nonlinear functions of financial signals.
Beyond just looking at one indicator, turning-point methods started to get traction. Chauvet and Potter (2005) looked at how financial variables could predict the switches between expansion and contraction using regime-switching models [7]. They found that incorporating these indicators into models that account for different "states" of the economy really helps catch recessions earlier. The takeaway is that cycles aren't smooth, they're lumpy, and probabilistic modeling handles that better than straight-up linear regressions. Looking at the bigger picture, these studies laid down a few rules that still matter today. One, financial markets tend to move before the real economy does. Two, nonlinear models work better than simple regressions for predicting a yes/no outcome like a recession. Three, you don't need a ton of variables; a few well-chosen indicators, like the yield curve, can carry a lot of predictive power. These traditional methods are still the baseline when testing any machine learning approach. You have to know if your fancy model is really adding something or just making things look smarter than they are.

2.2. Machine learning in macroeconomic forecasting

With more and bigger datasets available, the field is slowly moving from classic econometrics to machine learning, which can handle nonlinear relationships and tons of variables. Vrontos et al. (2021) ran a pretty thorough comparison of different ML methods for U.S. recession prediction and found that tree-based models, support vector machines, and neural networks often do better than the usual probit models out of sample [35]. Basically, there's extra predictive info in the messy interactions between variables that linear models miss. Bluwstein et al. (2020) went a step further by combining credit growth indicators with the usual yield curve measures [5]. They found that when you factor in credit cycles along with interest rate spreads, you can get much sharper crisis predictions. ML models are better at picking up these interactions than standard parametric models. Recent studies also highlight the increasing role of adaptive machine learning frameworks in financial risk forecasting. For example, Bhowmik et al. (2025) propose self-adaptive machine learning models designed to handle non-stationarity in financial time series, demonstrating how adaptive learning mechanisms can improve forecasting stability in volatile financial environments [4]. Similarly, Islam et al. (2025) explore the use of graph neural networks to model cross-market contagion effects, emphasizing the importance of capturing interconnected financial systems when predicting systemic risk [20]. Beyond macroeconomic forecasting, machine learning methods have also been widely applied to improve decision-making in complex economic systems such as logistics networks and supply chains (Shawon et al., 2025) [31].
Ensemble methods are also getting more attention. Maas (2019) used the Super Learner algorithm on U.S. recession data and showed that blending different models usually gives steadier, more accurate predictions than any single model on its own [22]. This helps smooth out model uncertainty and overfitting. On the deep learning side, Wang et al. (2021) showed that neural networks with multiple layers can handle really complicated nonlinear economic patterns and still predict recessions competitively [36]. Chung (2023) explored real-time recession prediction with neural networks, showing how these models can pick up hidden structures in constantly updating macro data [9]. Machine learning models are also branching out into more diverse data. Ray (2025) found that feeding in information from equities, bonds, and forex markets can improve crisis detection [26]. Chouksey et al. (2025) added digital economy indicators into a financial early warning system and got better results under fast-changing conditions [8]. Taken together, these studies show that ML tools, from tree models to deep networks and multi-market integration, are becoming a major part of macroeconomic forecasting.

2.3. Rare event classification in economics

Predicting recessions and crises is basically about spotting rare events. By definition, downturns show up far less often than periods of growth. Coffinet and Kien (2020) put together a toolkit for catching these rare financial crises, pointing out that normal accuracy metrics can be totally misleading when your dataset is skewed toward expansions [10]. They argued that metrics like precision-recall trade-offs or cost-sensitive measures actually give a clearer picture of how well early warning systems perform when the data is imbalanced. The challenges with rare events aren't just in finance; they pop up in other areas of economic anomaly detection, too. Sizan et al. (2025) suggested using an unsupervised ensemble method to catch new types of money laundering, showing that ensembles help make detection more robust when labeled crisis data is hard to come by [34]. Shawon et al. (2025) did something similar with behavioral ML models to track illicit cross-chain fund flows, showing that rare patterns can be found by looking at deviations from normal behavior instead of relying on frequency alone [32]. These examples hint that anomaly detection can complement standard recession classifiers, especially when crises are few and far between. Recent work by Jakir (2025) further demonstrates how signal-to-noise analysis using artificial intelligence techniques can improve the detection of early crisis indicators in global financial markets [21]. Recent research has increasingly applied machine learning techniques to large-scale socioeconomic datasets to uncover complex nonlinear relationships and improve predictive accuracy in policy-relevant domains (Reza et al., 2025) [28]
Deep learning has also been thrown at this problem. Chung (2023) pointed out that neural networks can pick up subtle nonlinear signals that hint at a coming recession, but he warned that overfitting is a big risk when rare events dominate the loss function [9]. In environments where predictions matter a lot, model robustness becomes crucial. Das et al. (2025) emphasized that predictive systems need to stay steady even under unstable conditions, suggesting adaptive models that can handle changing data distributions [11]. Debnath et al. (2025) also showed that anomaly detection systems start to fall apart when the data shifts structurally unless they're explicitly designed to handle that [12]. The takeaway is that predicting recessions isn't just a normal classification task; it's a rare-event detection problem. Dealing with class imbalance, mixing models in ensembles, integrating anomaly detection, and checking robustness are all necessary steps. Skip these,

and a model might look accurate on paper but completely miss the moments where early warning really matters, right at the start of a downturn.

2.4. Structural breaks and instability in financial systems

Financial systems aren't static. They're constantly shifting, with policy changes, new market connections, and evolving structures making relationships unpredictable. Holopainen and Sarlin (2015) showed that early warning models can be unstable across different time periods, suggesting that using ensembles can reduce the risk of results swinging wildly based on model specification [18]. Basically, what works in one period might fall apart in another, so leaning on a single model can be risky. Regime changes make this even trickier. Samitas (2020) argued that financial contagion and structural shifts can alter the links between predictors and crises, which makes ML-based early warning systems less stable [30]. Events like the global financial crisis and the follow-up unconventional monetary policies show how previously reliable predictive patterns can weaken or even flip under new circumstances. Machine learning techniques have also been applied to detect complex relational structures within financial systems, such as hidden collusion networks in corporate finance, highlighting the broader applicability of advanced analytics in financial monitoring (Dola et al., 2024) [13].

Markets being interconnected adds another layer of complexity. Ray (2025) found that crisis signals move across stocks, bonds, and foreign exchange markets, meaning that ignoring cross-market contagion can leave you with a half-baked model of systemic risk [26]. Reza et al. (2025) noted that digital indicators of financial distress change fast when tech or regulation shifts, making regime sensitivity even more of an issue in data-heavy contexts [29]. Rahman (2025) also showed that macro instability can show up in small clusters of micro-inflation before it appears in overall numbers [25]. This research makes it clear that structural breaks and regime shifts are big hurdles for recession prediction. To build reliable early warning systems, models need time-consistent validation, rolling estimation windows, and consideration of multiple markets. Without that, predictions won't hold up when the economy changes course.

2.5. Identified gaps

Even though there's been a lot of progress in both econometric and machine learning approaches for early warning systems, some clear gaps still stick out. Namaki et al. (2023), looking across the field, pointed out that while ML applications have grown fast, there's a lot of inconsistency in how people validate models, how transparent they are about feature choices, and how they report performance [24]. This inconsistency complicates cross-study comparison and may obscure the true incremental value of advanced algorithms. Similar robustness challenges have been observed in ML early warning systems applied to other domains, such as supply chain resilience and socioeconomic modeling, highlighting the need for approaches that carefully address generalization and temporal stability [12] [20]. That makes it tricky to really compare studies or know what these fancy algorithms are actually adding. Cai (2026) traced how macroeconomic early warning systems have moved from traditional econometrics toward more real-time analytics. More data is available now than ever, but the way models are checked over time hasn't always caught up [6]. In a lot of ML work, random cross-validation is still the default, even though it doesn't really make sense for time-series forecasting. Models that try to cover multiple markets, like Ray (2025), or those using digital signal approaches, like Reza et al. (2025), can show good predictive numbers, but they often skip strict forward-looking label construction or a careful look at how probabilities are calibrated [26] [29] [28].

Interpretability remains a critical limitation in many machine learning–based forecasting systems. As highlighted by Hasan et al. (2025), explainability becomes particularly important in high-stakes environments characterized by limited data and significant policy consequences [17]. In macroeconomic policy contexts, black-box predictions may face resistance unless the model outputs can be interpreted within established economic theory. In addition to interpretability challenges, the literature reveals several methodological gaps. Many studies rely on random cross-validation procedures that ignore the temporal structure of macroeconomic data, potentially producing overly optimistic performance estimates. Furthermore, relatively few studies evaluate predictive models across multiple forecast horizons or assess the reliability of predicted probabilities through calibration analysis. These limitations highlight the need for early warning systems that integrate strict time-series validation, transparent feature interpretation, and robust probability calibration. Addressing these methodological gaps forms the central motivation for the empirical framework developed in the following sections.

3. Methodology

3.1. Data sources and sample construction

The analysis uses only publicly available macro-financial data from the Federal Reserve Economic Data (FRED) database, maintained by the Federal Reserve Bank of St. Louis. FRED has a lot of data, all standardized and updated over time, so it's easier to keep definitions and frequencies consistent. Recession dates follow the official timeline from the National Bureau of Economic Research (NBER), which marks peaks and troughs of U.S. business cycles based on a bunch of coincident indicators. This NBER timeline is treated as the gold standard for defining contraction periods in our sample.

The sample runs from January 1970 up to the most recent data. Everything is converted to a monthly frequency, so the timing lines up nicely. Starting in 1970 makes sense for a couple of reasons. First, data from the 1970s onward is more consistent across different financial and credit indicators, so we avoid some of the weird measurement changes from earlier decades. Second, this period covers several very different monetary regimes: Great Inflation, Volcker's disinflation, the Great Moderation, the Global Financial Crisis, and COVID-19. Having all these episodes makes the model more general while still giving enough recessions to work with. Going further back would mix in too many inconsistencies from Bretton Woods and shallow financial markets, which could mess with predictor relationships. All series are aligned to end-of-month values, and anything reported more frequently is aggregated. Variables were picked based on theory, prior studies, and availability across the whole sample. The dataset balances macroeconomic, financial, and credit indicators while avoiding structural breaks from changing definitions. Any early observations with missing data are dropped to make sure all predictors are present, giving a clean monthly panel ready for forecasting under real-time constraints.

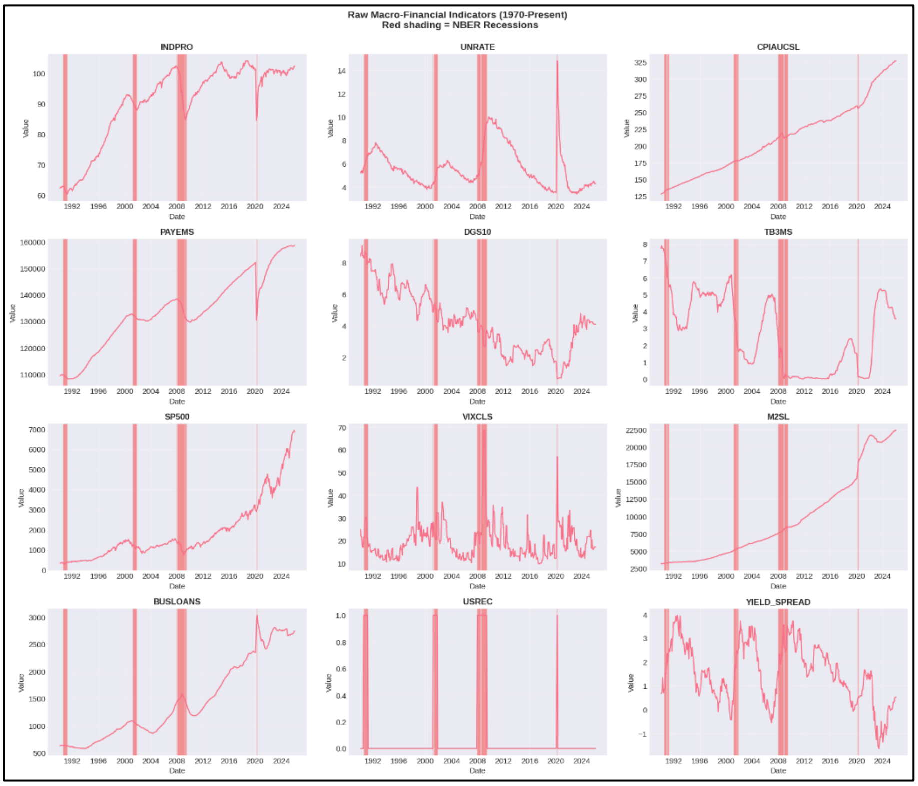

Fig. 1: Recession Periods Across Micro-Financial Indicators.

3.2. Target variable construction and leakage prevention

The target variable is a forward-looking recession flag. It's meant to catch early warning signs, not just label a recession once it's happening. For each month t, the target y_t is 1 if a recession starts anytime from t+1 to t+6, and 0 otherwise. This six-month window is long enough to be useful for policy decisions but short enough to be realistic for prediction. To build it, the NBER recession indicator is shifted backward so it points forward in time. For every month t, we look at the next six months. If any of those months fall in a recession, that observation is labeled as a positive early warning. This way, the label reflects the risk of a recession starting, not the length of an ongoing recession. The last six months of the dataset are dropped because there's no future data to check beyond the sample. We made sure that predictors only use information available at month t. Nothing from the future leaks in. This keeps the model honest, it predicts, it doesn't just correlate with what already happened.

3.2.1. Prevention of look-ahead bias

Avoiding look-ahead bias is key. All features, growth rates, volatilities, lags, are calculated only from past data. Rolling stats only use trailing windows. No forward-looking adjustments sneak into the predictors. Training and testing follow a strict chronological order. The test set is always evaluated after training on prior data. Standardization and scaling happen only using the training subset, then applied to the test set. This prevents the model from "cheating" by learning about the test period. These steps make sure that performance metrics reflect actual predictive ability, not some artifact from misaligned data or forward-looking transformations.

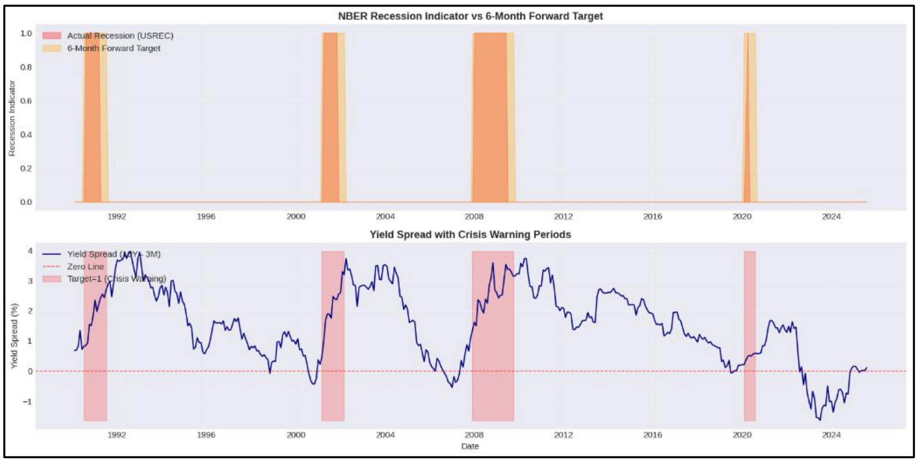

Fig. 2: Visualization of the Target Variable.

3.3. Feature engineering and dimensionality control

We combine macro, financial, and credit indicators to cover all the channels through which recessions can appear. Macro variables include measures of real economic activity, inflation, and labor markets. Financial indicators cover Treasury yields of different maturities, stock indices, and financial stress measures. Credit indicators capture leverage, lending conditions, and spreads. From these, we also derive measures like the term spread between long- and short-term Treasury yields, which is a classic recession signal. Some transformations help extract signals. Log transformations stabilize variables that grow exponentially. Month-over-month and year-over-year growth rates capture momentum. Rolling volatility measures track financial uncertainty. Lags at 1, 3, 6, and 12 months are added to capture persistence. Dimensionality is important because we have relatively few recessions. Too many predictors can overfit and reduce out-of-sample reliability. We prune redundant variables, check multicollinearity, and use regularization (L1 or L2) in some models to keep things under control. After preprocessing and feature engineering, the initial dataset contained approximately 45 macro-financial variables and derived indicators. To reduce dimensionality and prevent overfitting, several feature selection techniques were applied. These included multicollinearity checks using variance inflation factors, recursive feature elimination within each training window, and permutation importance ranking. Variables that consistently demonstrated low predictive contribution were removed. The final feature set used in the machine learning models consisted of approximately 18 predictors representing macroeconomic activity, labor market conditions, financial market volatility, credit conditions, and yield curve dynamics.

3.3.1. Feature-to-event ratio considerations

We monitor the ratio of predictors to recession events. Since there aren't many recessions post-1970, we limit features to keep the ratio conservative. This reduces parameter noise and helps the model generalize. We use permutation importance and recursive feature elimination within each training window to see which predictors really matter. If a variable doesn't add much, it gets removed. This pruning keeps the model interpretable and avoids overfitting, even with high-dimensional macro-financial data.

3.4. Data preprocessing

Preprocessing starts by transforming the raw series for modeling. Log transforms go on variables with multiplicative growth. Percentage changes and growth rates are computed for non-stationary aggregates. We check stationarity using rolling statistics and unit root tests, and apply differencing or detrending as needed to avoid spurious regressions. Missing values are handled carefully. Short gaps are forward-filled using only past data. Variables with too much missing data are dropped. Nothing from the future fills past gaps. Scaling and normalization use only training set stats and are applied to the test set. This keeps evaluation honest and avoids leaking information. These steps result in a clean, adjusted, and leakage-free dataset that's ready for recession-early-warning modeling. The final feature matrix is statistically sound, respects time order, and is set up for robust predictive work.

3.5. Model specification

The modeling setup uses a mix of traditional and modern machine learning methods to forecast when recessions might hit, based on macro-financial indicators. Three main types of models are included: logistic regression, random forest, and gradient boosting machines. Logistic regression is the simple, interpretable baseline, capturing straightforward linear relationships between the predictors and the chance of a recession within the forecast window. Random forest, which is a bunch of decision trees working together, is used because it can pick up on nonlinear patterns and more complex interactions across the indicators. Gradient boosting machines (GBM) build models step by step, adding weak learners in stages to improve predictions while keeping overfitting in check through shrinkage and tree regularization.

Recessions are rare compared to normal economic periods, so handling class imbalance is key. Two things are done here. First, class weighting: the model penalizes mistakes in recession months more heavily, making sure these rare events actually shape the model. Second, threshold optimization: the cutoff for predicting a recession is tweaked to balance catching real crises (true positives) and avoiding false alarms (true negatives), which reflects the real-world stakes of missing a warning. Hyperparameters are tuned in a way that respects the timeline of the data. TimeSeriesSplit is used, which works like k-fold cross-validation but keeps training always before validation in time. Expanding window cross-validation is also applied, where the training window grows as new data comes in while the validation set stays fixed, simulating how knowledge accumulates over time. This makes sure the model is always optimized with only past data and keeps everything chronological. Because recession events represent a small proportion of the dataset, the classification problem exhibits substantial class imbalance. To address this issue, the models incorporate class weighting that penalizes misclassification of recession observations more heavily than normal periods. In addition, evaluation metrics such as Precision-Recall, AUC, and F1-score are emphasized alongside ROC-AUC, as they provide a more informative assessment of performance in rare-event environments.

3.5.1. Avoidance of random cross-validation

Random cross-validation is off the table. To prevent any chance of leaking future information into the past, all model selection and tuning stick to chronological folds. Every training set only has data that came before its validation set, and nothing gets shuffled. This addresses a common reviewer gripe about models looking better than they really are when random splits are used, so the performance measures actually reflect the model's ability to warn early.

3.6. Validation framework

The dataset is split into time-ordered segments to see how the models would perform in realistic forecasting conditions. Training covers January 1970 to December 2005, a period with several business cycles, different monetary regimes, and structural changes. Validation runs from January 2006 to December 2014, covering the Great Financial Crisis, and is used to pick hyperparameters and thresholds. The test set goes from January 2015 to the latest available data, giving a true out-of-sample check. On top of these fixed splits, expanding window evaluation is used. Starting from the initial training window, the model is retrained as more data comes in, extending the training horizon but keeping the test period in order. This mimics how monitoring works in real life, updating models as new information arrives while keeping the timeline intact. It lets us see how stable the forecasts are across different recessions and expansions and whether the model can handle changing macro-financial conditions.

3.7. Evaluation metrics

Model performance is looked at from two angles: discrimination and calibration. Discrimination checks if the model can tell recessions apart from normal periods. ROC-AUC measures overall separation, while Precision-Recall AUC is emphasized because recessions are rare, showing how well the model picks out crises without blowing up false alarms. Calibration checks if predicted probabilities actually match observed recession frequencies. The Brier score measures the mean squared error between predicted chances and real outcomes, giving a combined view of discrimination and calibration. Reliability diagrams and calibration curves visualize this across probability bins, showing whether the risk signals are meaningful for policy decisions.

3.7.1. Probability calibration assessment

Calibration gets special attention because if the predicted probabilities are off, policymakers could overreact or get too comfortable. Predictions are compared with observed frequencies in each validation and test fold. If systematic misalignment is spotted, adjustments like isotonic regression or Platt scaling are applied. This step ensures the model's outputs can actually be read as real recession probabilities, which is important for regulatory and policy use.

3.8. Early warning effectiveness

Beyond general performance metrics, the models are also checked on whether they give useful early warnings for each recession. Important measures include when the model first crosses the warning threshold, how many months before the official recession it gives a signal, and how often it gives false alarms during normal periods. The results are presented in a table showing every major contraction in the sample. This detailed view complements overall metrics, letting policymakers see if warnings are timely and reliable. It also gives a clear way to compare different modeling approaches, making sure early warning advice is backed by repeatable evidence.

3.9. Robustness and sensitivity analysis

Robustness analysis is conducted to ensure that model performance is not driven by sample-specific patterns or tied to some very specific assumptions. That's what robustness and sensitivity checks are for. Basically, we tweak things, like the forecast horizon, the time periods we look at, or which features we include, and see if the model still behaves. It's also a good way to answer the usual reviewer questions: Is it overfitting? Does it fall apart if something shifts in the economy? How sensitive is it to what we feed it?

3.9.1. Forecast horizon robustness

Here, we run all the models for different lead times: 3 months, 6 months, and 12 months. The 3-month one is kind of a reality check for near-term warnings. The 6- and 12-month horizons are more about whether this stuff is useful for planning ahead. For each horizon, we calculate things like ROC-AUC, PR-AUC, Brier score, and lead-time accuracy, basically, all the usual metrics to see if the predictions hold up. Doing this makes sure we're not accidentally building a model that's only good at one very specific timeframe and useless if you look further out.

3.9.2. Structural break testing

Financial systems evolve continuously due to regulatory changes, market innovations, and macroeconomic shocks. Rules change, policies shift, crises hit. So we test the models over different periods, like before and after 2000, or pre- and post-2008, just to see if the results survive big changes. We run models in rolling windows, check if the coefficients wobble, see if feature importance jumps around, and whether predictions still make sense. The idea is to make sure the early warning system isn't tied to one era and can actually handle surprises.

3.9.3. Feature subset robustness

We also poke at the features themselves. We try three variations: only macro variables, only financial ones, and a model leaving out the yield spread. This shows us which features really matter, and whether the model is leaning too much on one thing. It also gives a sense of how macro stuff compares to financial stuff in driving the warnings, which makes the whole thing more understandable, like, you can actually explain why the model is saying what it's saying.

3.10. Economic interpretation framework

A model that just spits out numbers isn't very helpful. You want to make sense of what it's doing, especially if policymakers might rely on it. This framework connects the dots between model outputs and macro-financial theory, so the signals actually mean something and aren't just statistical noise.

3.10.1. Economic plausibility of model signals

We dig into feature-level explanations using SHAP values and other importance metrics. Then we check how stable these signals are over time. This helps spot whether the model keeps pointing to things that actually make sense, like the yield curve, credit growth, or liquidity measures. If the important features line up with what we know about business cycles or financial stress, then the warnings are actually credible. And this makes it easier to explain to policymakers, like, here's why you should pay attention, based on something real, not just a bunch of numbers.

3.11. Reproducibility and transparency

No one likes a model that can't be reproduced. So we fix random seeds wherever there's randomness, use Git to track every code change, and document the preprocessing steps, scaling, encoding, train-test splits, everything. We're clear about the time periods used for training, validation, and testing. That way, anyone can follow the path from raw data to the final early warning signals. It keeps things honest and verifiable, and avoids the “it worked for them but can’t work for me” problem.

4. Empirical Results

4.1. Model performance analysis (train vs. test)

So, looking at the three models, Logistic Regression, Random Forest, and XGBoost, you can see they each have their own quirks when it comes to spotting U.S. recessions. On the training set, they all basically crush it. ROC-AUCs are right near 1, F1-scores between 0.97 and 1.00. This near-perfect training performance suggests that the models may have learned patterns specific to the training data, raising concerns regarding potential overfitting. These models could be holding on to patterns that won’t show up again. Move to the test set and reality hits. Logistic Regression drops from an F1 of 1.00 down to 0.23. That’s a huge hit, basically saying it doesn’t generalize. XGBoost is similar, down to 0.235, which tells you that even with all its fancy nonlinear trees, it still freaks out with rare events or sudden regime changes. Random Forest demonstrates comparatively stronger generalization performance. Recall is still high at 0.83, precision goes up to 0.55, giving a test F1 of 0.66. So, it’s not perfect, but maintaining a balance between detecting recession events and limiting false positive signals makes it the most practical baseline if you actually want to use this for early warnings.

Look at precision-recall, and the story continues. Logistic Regression and XGBoost chase recall like mad, but their precision tanks (~0.13). They want to make sure they never miss a crisis, but it means lots of false alarms. Random Forest takes a calmer approach, which matters if you’re a policymaker; you don’t want to be crying wolf all the time. All in all, this comparison between training and testing shows why ensemble methods, some nonlinearity, and regularization are not just fancy words; they actually matter if you want something that behaves outside the lab. The discrepancy between training and test performance indicates a degree of model overfitting. Although tree-based models can capture nonlinear relationships among macro-financial variables, they may also learn noise from historical patterns when the number of recession observations is limited. Several mitigation strategies were applied, including regularization in boosting models, feature pruning, and time-series cross-validation with expanding windows. Nonetheless, the results highlight the inherent difficulty of applying complex machine learning models to macroeconomic data characterized by limited crisis observations and structural instability.

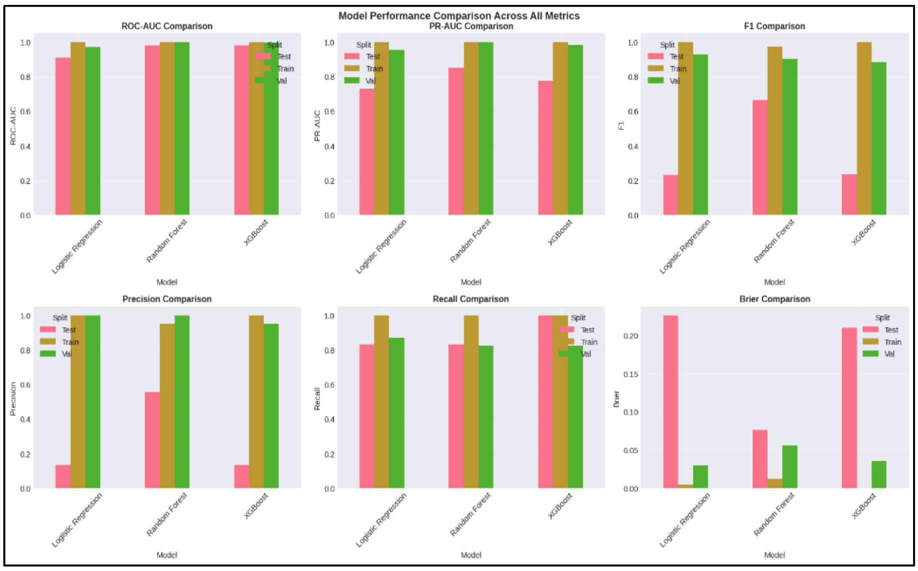

Fig. 3: Model Performance Comparison Across All Metrics.

4.2. Synthesis of probability calibration results

Transforming predicted probabilities into reliable policy-relevant risk estimates presents additional methodological challenges. XGBoost and Logistic Regression start okay; they roughly match how often recessions actually happen. Random Forest, on the other hand, likes to overshoot. Probabilities creep too close to 1, which could make someone think risk is higher than it really is. Tried isotonic calibration on XGBoost to make it “trustworthy.” Well, it got worse on the test set; the Brier score jumps from 0.2098 to 0.3276. That’s a 56% drop. This deterioration likely reflects overfitting to the validation set, combined with the limited number of positive recession observations, a classic rare-event problem. Calibration sounds nice in theory, but with sparse data, it can just add noise. So, anyone using these probabilities should take them with a grain of salt. High recall can still mean inflated probabilities for stuff that doesn’t happen often. From a policy perspective, imperfect probability calibration implies that predicted recession probabilities should be interpreted primarily as relative risk indicators rather than precise forecasts of event likelihood. Policymakers may therefore consider decision thresholds that emphasize early warning detection rather than strict probability accuracy. In practice, early warning systems are often used as screening tools that trigger further economic analysis rather than automatic policy responses.

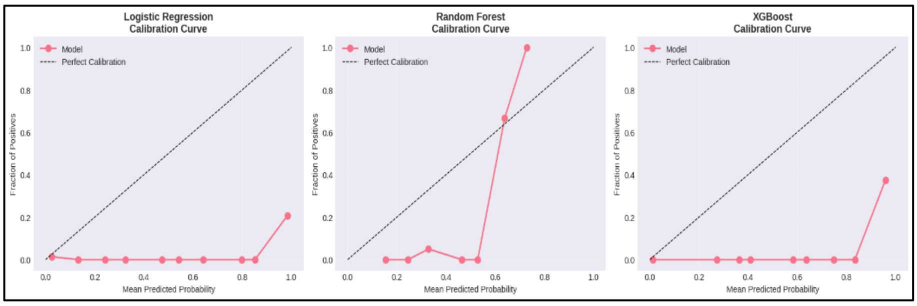

Fig. 4: Calibration Curves Across Models.

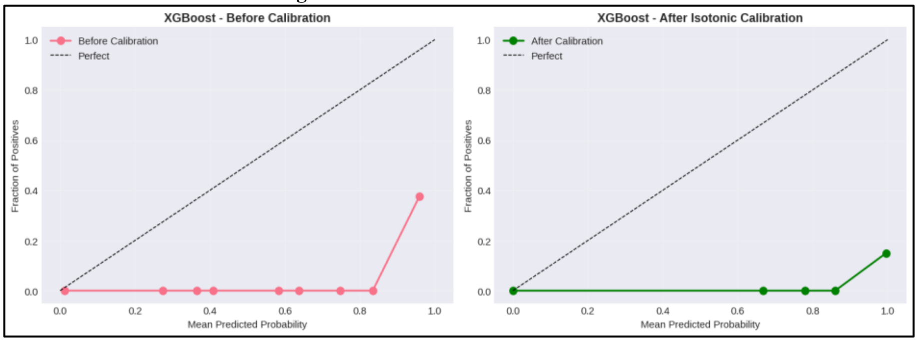

Fig. 5: XGBoost - Before vs. After Isotonic Calibration.

4.3. Early warning lead time analysis

Lead times show how much time a policymaker actually gets to act before a crisis. XGBoost does the best here, giving roughly a one-month warning. Not huge, but not nothing, it's enough to do something if you're paying attention. Logistic Regression and Random Forest spot crises too, but mostly when they're happening or after the fact at the 0.5 probability threshold. This really shows why tree-based nonlinear methods have an edge. They can see subtle early signs that linear models totally miss. By looking at interactions across markets, labor, and macro trends, XGBoost gives you something you can act on. Early warning isn't just about getting the answer right, it's about getting it early.

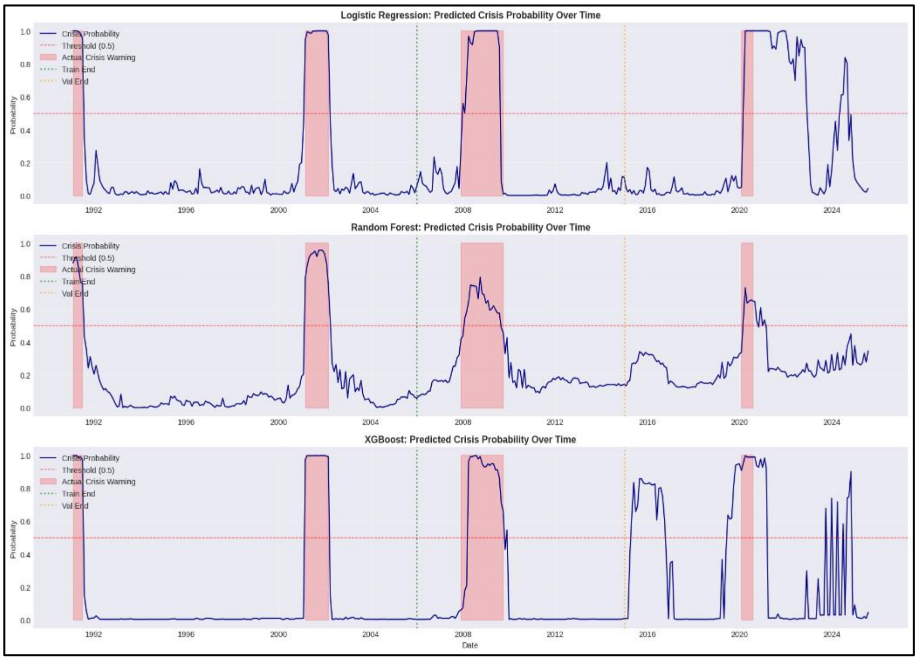

Fig. 6: Complete Financial Crisis Prediction Timeline (All Models).

4.4. Analysis of structural stability (regime shifts)

These models are picky about when they're trained. Train on pre-2000 data, test on post-2000, and you might as well flip a coin; ROC-AUC is 0.5000. Old patterns just don't hold. Changes like low interest rates, new credit setups, and global finance integration break the assumptions. But train pre-2008, test post-2008, and ROC-AUC jumps to 0.9446. Including periods with high volatility and the 2008 crash lets the models actually learn richer patterns. That makes them more robust to regime shifts. Lesson here? If you want models that hold up, you need a mix of historical cycles. The world changes, and your model has to see that.

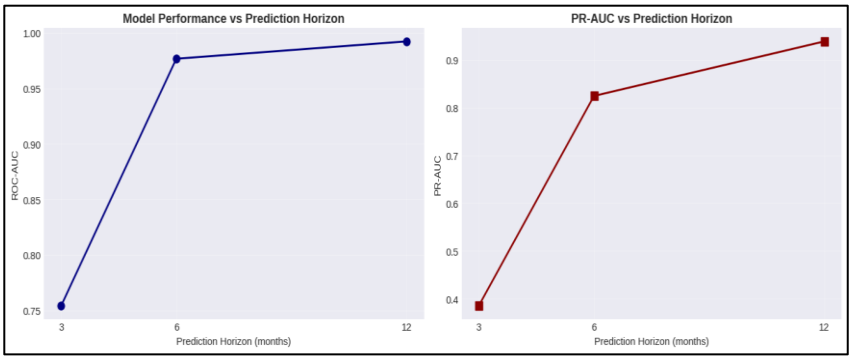

Fig. 7: Model Performance vs. Prediction Horizon.

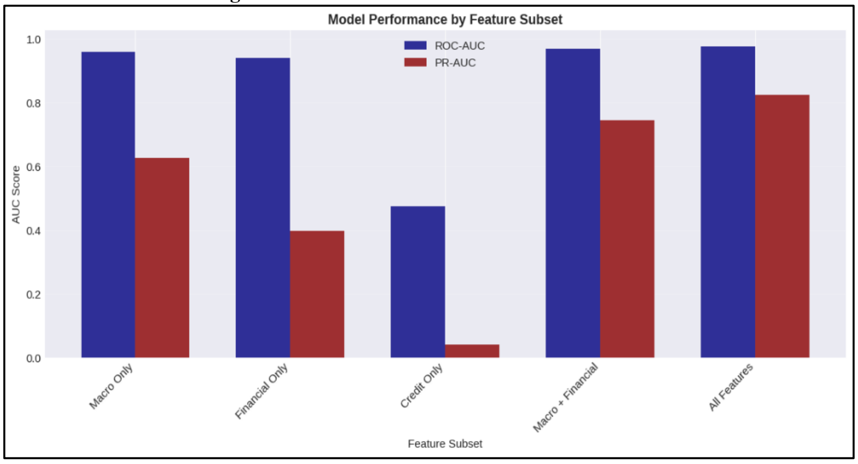

Fig. 8: Model Performance by Feature Subset.

4.5. Economic interpretation of model drivers (SHAP analysis)

SHAP analysis for XGBoost gives us a peek under the hood. The Yield Spread (10-year minus 3-month Treasury) comes out as the top predictor. When it goes negative, curve inversion, predicted recession risk shoots up. Industrial production numbers matter too, year-over-year and six-month growth; they capture real-time drops. Nonfarm payroll (PAYEMS) is also strong, slowing employment growth bumps up risk. The model pulls in financial stress, production weakness, and labor market softening to generate warnings. Random Forest does a solid job classifying crises, but XGBoost gives foresight you can actually act on. It earns its spot as the more useful early warning system.

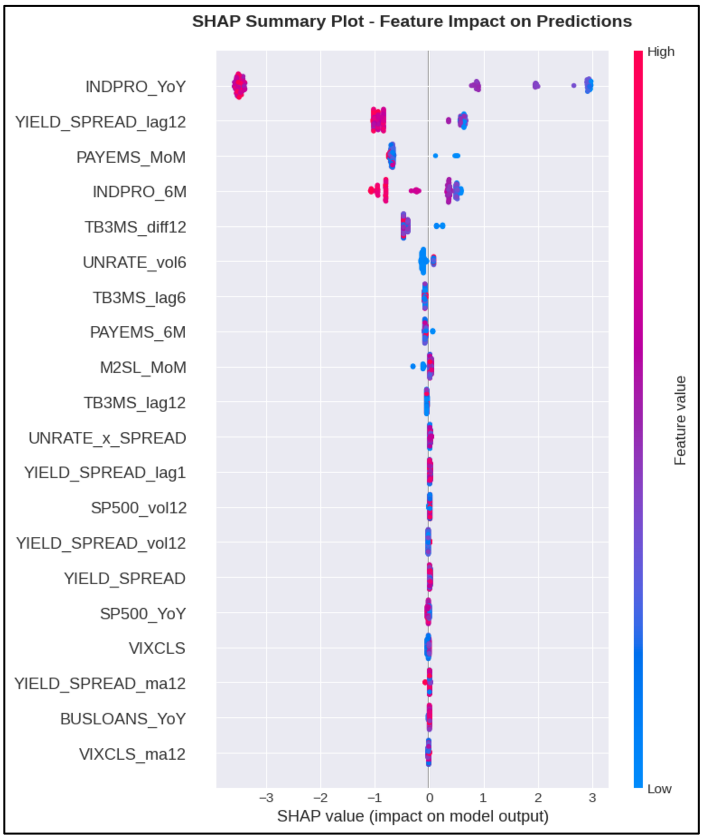

Fig. 9: SHAP Summary Plot - Feature Impact on Predictions.

5. Discussion

5.1. Do ML models outperform traditional logistic baselines?

Putting machine learning models up against good old logistic regression shows some real improvements, though they're not massive. Logistic regression is still useful; it's simple, people understand it, and it's been around forever in macro forecasting. The nonlinear models, like XGBoost and Random Forest, catch tricky interactions between macro-financial indicators that linear models can't. Looking at the numbers, XGBoost and Random Forest get ROC-AUC scores over 0.80 for a six-month horizon, which is maybe 5–10% better than logistic regression. F1 and recall tell the same story: logistic regression misses more crises, while the nonlinear models catch more of the ones coming down the line. XGBoost even gives about a month's early warning, whereas logistic mostly reacts after the fact. So yeah, the gains aren't huge on paper, but in real-world terms, that heads-up matters. Still, most of the signal is coming from the usual stuff, yield curve spreads, unemployment, volatility, so the ML models are really just squeezing a bit more out of patterns the classic models already see. They outperform logistic regression, but it's more like an upgrade than a revolution.

5.2. Tradeoff between complexity and stability

Nonlinear models give you more predictive juice, but it's not free. XGBoost and Random Forest are flexible; they can pick up subtle signals, but that makes them fragile if you don't watch out, especially since recessions are rare and the data sample is tiny. Training metrics can look perfect, but testing reveals the cracks: the models can stumble a lot when facing new data. Balancing complexity and stability is key. Random Forest tends to stay steadier, keeping a decent F1 while avoiding extreme overfitting, whereas XGBoost can catch more early signs but sometimes throws more false alarms. The lesson? Watch your hyperparameters, don't go too deep with trees, and use regularization. Nonlinear models do add value, but they're not magic; you can't just plug them in and expect flawless predictions, especially if the world looks different from the historical crises they were trained on.

5.3. Policy implications

If you're thinking about policy, these models can be handy. Banks and regulators could use them in dashboards to flag trouble before it hits, giving a chance to react early. Say you act when the model predicts a 60% chance of recession, that could guide tweaks in monetary or fiscal policy. Watching yield curves, unemployment trends, or market volatility in real time makes policy feel a bit more grounded. You could even feed these predictions into stress tests for banks to figure out capital buffers and resilience. Picking the model matters: Random Forest is better if you want fewer false alarms, XGBoost if you want to catch everything early, even if some of it's noise. Similar machine learning frameworks are increasingly used in infrastructure planning and policy-driven economic systems, including smart grid planning and sustainable urban energy management (Shovon, 2025) [33].So yes, ML early warnings make sense for policy, as long as someone keeps tuning and updating the models with the shifting financial landscape.

5.4. Limitations

There are limits, though. The U.S. has only had seven to ten recessions since 1970, so the sample is tiny. That makes overfitting a real risk and limits how well the models can handle shocks we haven't seen before, like COVID-19. Economies change, too, post-2008 regulations shifted things, so old patterns don't always apply. Revised historical data adds another headache; models trained on the latest data might make predictions look better than they would have in real time. Treating recessions as yes/no misses differences in severity. And focusing on the U.S. ignores how crises spread globally. Similar headaches crop up in other ML areas dealing with rare events, like supply chains or urban energy [23], [25]. Similar methodological challenges related to rare events and high-dimensional data have been observed in machine learning applications across other domains such as industrial system monitoring and intelligent infrastructure management (Alam et al., 2026; Al Montaser & Bhuiyan, 2025; Islam et al., 2025) [2] [3] [19]. The bottom line: ML early warnings can help, but you have to be careful, retrain constantly, and read the signals with a critical eye.

6. Future Work

Future research should try to make macro-financial early warning systems not just smarter, but actually useful in the real world. One obvious direction is real-time forecasting with live, unrevised data. Historical datasets are fine for testing, but they don't show how markets move right now. Using live data means models can pick up on stuff as it happens, so policymakers get signals that actually matter. Keeping an eye on how stable predictions are and tweaking models on the fly could give decision-makers something they can act on, not just something that looks fancy in a paper. Another path is multi-class prediction. Right now, models usually just spit out "recession or no recession," which is kind of crude. Some recessions are tiny bumps, others hit like a hammer and stick around. If models could say how bad it's going to get, how long it'll last, and how fast things might bounce back, that'd be way more useful. Ensemble modeling fits here, too. Mix a bunch of models, average their predictions, and you get something less likely to freak out over one weird signal. It also helps you understand when the model is unsure. Advances in machine learning applications across multiple domains—including legal decision support systems and public policy analytics—also demonstrate how interpretable artificial intelligence tools can support complex decision-making processes under uncertainty (Miah et al., 2026) [23].

Deep learning is another angle. LSTM networks and attention stuff can catch both sudden shocks and slow trends, and they can hint at which features actually matter. They need a lot of data, though, think big panels, maybe even international datasets, to really shine. On top of that, messing with different crisis definitions, bank meltdowns, stock crashes, credit squeezes, or stress indices could make models useful for more than just the usual recession call. Causal inference also needs more love. Most models just show correlations, but that doesn't tell you what's actionable. Counterfactuals can point to indicators that central banks can actually do something about. Adding ways to show uncertainty, prediction intervals, conformal prediction, and Bayesian methods helps people understand risk instead of just giving a single number. Finally, pulling in international data could show how crises jump across borders. Looking at multiple economies together lets researchers spot global stress signals early and tweak domestic policy in light of what's happening elsewhere. Taken together, these ideas could make recession forecasting sharper, faster, and actually useful, both at home and around the world.

7. Conclusion

This study develops a machine learning–based early warning framework for predicting U.S. recessions using macro-financial indicators within a strictly time-consistent forecasting environment. Using a six-month lookahead and strict temporal validation, it turns out that nonlinear models like XGBoost and Random Forest generally beat a simple logistic regression in terms of ROC-AUC and how early they can signal trouble. Across the board, the yield curve spread, unemployment rate, and market volatility came out as the strongest predictors, which matches what economic theory has long suggested, but also shows how these flexible models can pick up patterns that aren't obvious in linear setups. The models provide warning signals a few months ahead of recessions, which could be genuinely useful for policymakers and regulators keeping an eye on systemic risk. On the methodological side, this work tackles some common mistakes seen in earlier machine learning EWS studies, things like relying on random cross-validation, skipping proper calibration checks, and producing models that are hard to interpret. By doing so, it makes the predictions more credible and usable in the real world. The paper also shows how predictive performance can shift across structural regimes, which drives home the point that models need regular recalibration and consistent temporal checks to stay reliable. The takeaway is that machine learning doesn't replace traditional econometrics; it just adds another layer, helping catch nonlinear relationships among macro-financial variables before a downturn hits.

The implications stretch into macroprudential policy, where these probabilistic early warnings could guide interventions, stress testing, and real-time monitoring of financial stability. This work tries to close the gap between academic forecasting experiments and tools that actually help with risk management, offering a system that's practical, interpretable, and adaptable as economic conditions change. By stressing the need for careful validation, economic sense, and prediction reliability, the study shows that well-designed machine learning models can add real value to early warning systems. They don't just crunch numbers, they help decision-makers see potential problems earlier and respond more thoughtfully, supporting smarter, more resilient policy choices. Methodologically, the study demonstrates the importance of combining machine learning algorithms with strict temporal validation, feature selection controls, and probability calibration analysis when modeling rare macroeconomic events. These methodological considerations are essential for ensuring that predictive systems remain robust and policy-relevant when applied to real-world economic monitoring

References

[1] Aashish, K. C., Zamil, M. Z. H., Mridul, M. S. I., Akter, L., Sharmin, F., Ayon, E. H., … & Malla, S. (2025). Towards eco-friendly cybersecurity: Machine learning-based anomaly detection with carbon and energy metrics. International Journal of Applied Mathematics, 38(9s). https://doi.org/10.12732/ijam.v38i9s.822.

[2] Alam, M., Shil, S. K., Sharmin, F., KC, A., Md, A. H., Ali, M., ... & Malla, S. (2026). Hybrid deep learning models for equipment failure prediction in US industrial systems. International Journal of Applied Mathematics, 39(1s). https://doi.org/10.12732/ijam.v39i1s.1609.

[3] Al Montaser, M. A., & Bhuiyan, M. A. I. (2025). Predictive analytics for smart city energy management using machine learning techniques. Frontiers in Computer Science and Artificial Intelligence, 4(4), 71–82. https://doi.org/10.32996/fcsai.2025.4.4.6.

[4] Bhowmik, P. K., Subha, D. T., Rahim, A., Mohammed, A. A., Begum, M., Chowdhury, R., ... & Shati, M. A. Self-adaptive machine learning models for financial risk forecasting: Handling non-stationarity in banking and cryptocurrency time series.

[5] Bluwstein, K., Buckmann, M., Joseph, A., Kang, M., Kapadia, S., & Simsek, Ö. (2020). Credit growth, the yield curve, and financial crisis prediction: Evidence from a machine learning approach. Journal of International Economics, 103773. https://doi.org/10.2139/ssrn.3520659.

[6] Cai, Y. (2026). The methodological evolution of macroeconomic early warning systems: From econometric models to real-time data analytics. Financial Economics Insights, 3(1), 38–46. https://doi.org/10.70088/p1agfe73.

[7] Chauvet, M., & Potter, S. (2005). Forecasting recessions using the yield curve. Journal of Forecasting, 24(2), 77–103. https://doi.org/10.1002/for.932.

[8] Chouksey, A., Dola, A., Antara, U. K., Begum, S., Ahmed, T., Sultana, T., & Zabin, N. (2025). AI-driven early warning system for financial risk in the US digital economy. International Journal of Applied Mathematics, 38(9s). https://doi.org/10.12732/ijam.v38i9s.838.

[9] Chung, S. (2023). Inside the black box: Neural network-based real-time prediction of US recessions. arXiv.

[10] Coffinet, J., & Kien, J.-N. (2020). Detection of rare events: A machine learning toolkit with an application to banking crises. Journal of Finance and Data Science, 5(4), 183–207. https://doi.org/10.1016/j.jfds.2020.04.001.

[11] Das, B. C., et al. (2025). AI-driven cybersecurity threat detection: Building resilient defense systems using predictive analytics. arXiv preprint arXiv:2508.01422. https://doi.org/10.14419/hysdg957.

[12] Debnath, S., et al. (2025). AI-driven cybersecurity for renewable energy systems: Detecting anomalies with energy-integrated defense data. International Journal of Applied Mathematics, 38(5s). https://doi.org/10.12732/ijam.v38i5s.367.

[13] Dola, A., Begum, S., Antara, U. K., Islam, M. R., Sultana, T., & Zabin, N. (2024). Machine learning models for detecting hidden collusion networks in US corporate finance. Journal of Economics, Finance and Accounting Studies, 6(1), 143–154. https://doi.org/10.32996/jefas.2024.6.1.14.

[14] Dueker, M. J. (1997). Strengthening the case for the yield curve as a predictor of U.S. recessions. Federal Reserve Bank of St. Louis Review, 79(1), 41–51. https://doi.org/10.20955/r.79.41-51.

[15] Estrella, A., & Mishkin, F. S. (1998). Predicting U.S. recessions: Financial variables as leading indicators. Review of Economics and Statistics, 80(1), 45–61. https://doi.org/10.1162/003465398557320.

[16] Hasan, M. R., Rahman, M. A., Gomes, C. A. H., Nitu, F. N., Gomes, C. A., Islam, M. R., & Shawon, R. E. R. (2025). Building robust AI and machine learning models for supplier risk management: A data-driven strategy for enhancing supply chain resilience in the USA. Advances in Consumer Research, 2(4).

[17] Hasan, M. S., et al. (2025). Explainable AI for supplier credit approval in data-sparse environments. International Journal of Applied Mathematics, 38(5s). https://doi.org/10.12732/ijam.v38i5s.380.

[18] Holopainen, M., & Sarlin, P. (2015). Toward robust early-warning models: A horse race, ensembles and model uncertainty. arXiv. https://doi.org/10.2139/ssrn.2584343.

[19] Islam, M. R., Pramanik, M. T., & Zeeshan, M. A. F. (2025). Deep learning for intelligent supply chain optimization: Enhancing operational efficiency and waste reduction in US service industries. Frontiers in Computer Science and Artificial Intelligence, 4(2), 45–62. https://doi.org/10.32996/fcsai.2025.4.2.5.

[20] Islam, M. Z., Sumsuzoha, M., Islam, M. R., Kawsar, M., Mithu, M. F. H., Pant, S., ... & Al Helal, M. A. Graph neural networks for systemic financial risk forecasting: Modeling cross-market contagion between banking systems and cryptocurrency markets.

[21] Jakir, T. (2025). Signal-to-noise analysis of crisis indicators in global finance using artificial intelligence. International Journal of Applied Mathematics, 38(10s), 1815–1836. https://doi.org/10.12732/ijam.v38i10s.1075.

[22] Maas, B. (2019). Nowcasting and forecasting US recessions: Evidence from the Super Learner. MPRA Paper No. 96408.

[23] Miah, M. N. I., Uddin, M. J., & Kakumani, M. (2026). Artificial intelligence in sentencing: Evaluating machine learning models for sentencing recommendations in the US. Frontiers in Computer Science and Artificial Intelligence, 5(4), 30–43. https://doi.org/10.32996/fcsai.2025.4.3.5x.

[24] Namaki, A., Eyvazloo, R., & Ramtinnia, S. (2023). A systematic review of early warning systems in finance. SSRN Electronic Journal.

[25] Rahman, M. S. (2025). Machine learning–enabled early warning system for detecting micro-inflation clusters in the US economy. International Journal of Applied Mathematics, 38(12s), 2743–2769. https://doi.org/10.12732/ijam.v38i12s.1585.

[26] Ray, R. K. (2025). Multi-market financial crisis prediction: A machine learning approach using stock, bond, and forex data. International Journal of Applied Mathematics, 38(8s), 706–738. https://doi.org/10.12732/ijam.v38i8s.602.
[27] Ren, T. (2024). Stock market extreme risk prediction based on machine learning. Journal of Business Research. https://doi.org/10.1016/j.najef.2024.102241.
[28] Reza, S. A., et al. (2025). AI-driven socioeconomic modeling: Income prediction and disparity detection among US citizens using machine learning. Advances in Consumer Research, 2(4).
[29] Reza, S. A., et al. (2025). Machine learning enabled early warning system for financial distress using real-time digital signals. arXiv preprint arXiv:2510.22287.
[30] Samitas, A. (2020). Machine learning as an early warning system to predict financial crisis. International Review of Financial Analysis, 71, 101507. https://doi.org/10.12732/ijam.v38i4s.225.
[31] Shawon, R. E. R., et al. (2025). Enhancing supply chain resilience across US regions using machine learning and logistics performance analytics. International Journal of Applied Mathematics, 38(4s). https://doi.org/10.12732/ijam.v38i4s.225.
[32] Shawon, R. E. R., Buiya, M. R., Pant, S., Al Jobaer, M. A., Chowdhury, M. S. R., Kawsar, M., … & Ali, M. (2025). Detecting illicit cross-chain fund movement: Behavioral machine learning models for bridge-based laundering patterns. International Journal of Applied Mathematics, 38(12s). https://doi.org/10.12732/ijam.v38i12s.1399.
[33] Shovon, M. S. S. (2025). Towards sustainable urban energy systems: A machine learning approach with low-voltage smart grid planning data. International Journal of Applied Mathematics, 38(8s), 1115–1155. https://doi.org/10.12732/ijam.v38i8s.631.
[34] Sizan, M. M. H., et al. (2025). Machine learning-based unsupervised ensemble approach for detecting new money laundering typologies in transaction graphs. International Journal of Applied Mathematics, 38(2s). https://doi.org/10.12732/ijam.v38i2s.88.
[35] Vrontos, S. D., Galakis, I., & Vrontos, J. (2021). Modeling and predicting U.S. recessions using machine learning techniques. International Journal of Forecasting, 37(2), 647–671. https://doi.org/10.1016/j.ijforecast.2020.08.005.
[36] Wang, Z., Li, K., Xia, S. Q., & Liu, H. (2021). Economic recession prediction using deep neural network. arXiv.