Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA
DOI: 10.46254/BA08.20250102
Published: December 20, 2025

Hybrid Predictive Modelling on U.S. Recession Forecasting: A Comparison of Machine Learning and Response Surface Methodology

Miftaur Rahman Zisan and Anik Debnath
Department of Industrial and Production Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
miftaurr314@gmail.com, anikdbnath.mcc@gmail.com

Gourab Mondol
Institute of Business Administration
University of Dhaka, Dhaka-1000, Bangladesh
Iamgm651851@gmail.com

Abstract

Recessions forecasting is invaluable to policymakers, economists, and investors because it enables planning, decision-making, and risk mitigation. This paper suggests a hybrid approach that merges ensemble machine learning (ML), unsupervised clustering, and Response Surface Methodology (RSM) to expand predictive accuracy and interpretability. Employing U.S. quarterly data (1966–2025) of six macroeconomic variables, the methodology incorporates feature selection, SHAP importance analysis, Bayesian hyperparameter tuning, and RSM estimation. Unsupervised techniques (DBSCAN, Gaussian Mixture Models) revealed hidden structures consistent with supervised predictions, and XGBoost provided the greatest ROC-AUC. Though RSM resulted in lower accuracy, it produced explicit functional relations. Collectively, these techniques show that hybrid modeling provides predictive capability along with interpretability and that their integration is useful to make reliable recession forecasts.

Keywords

Recession forecasting, Machine learning, Response Surface Methodology (RSM), Hybrid models, Macroeconomic indicators, Interpretability.

1. Introduction

Recessions have devastating effects on employment, investment and social welfare, and thus the ability to anticipate them is critical for effective policy planning. Traditional econometric models, like the Probit and VAR, are stable dynamics models, but normally break down during turbulence. Machine learning (ML) techniques such as Random Forest, Gradient Boosting and XGBoost increase prediction accuracy by means of modeling nonlinearities but they are not interpretable. Although Response Surface Methodology (RSM) is not typically used in macroeconomics, it provides a way of being transparent about functional relationships. In this work, ML and regularised RSM models are compared using U.S. data (1966-2025) in a quest to achieve both statistical power and interpretability to help guide recession forecasting.

1.1 Research Objectives

1. How do ML models perform in U.S. recession forecasting (1966-2025) relative to RSM?
2. What is the trade-off between ML accuracy and RSM interpretability for policy application?
3. What macroeconomic triggers are the best predictors of recession, across models?
4. Can unsupervised machine learning reliably detect latent recessionary patterns in macroeconomic data without predefined labels?
5. Can the predictive strength of ML-RSM be balanced with interpretability in the decision-making context for policy?

2. Literature Review

The latest developments in computational algorithms have revolutionized macroeconomic prediction, and machine learning (ML) models have become an undeniable leap forward of the conventional econometric specifications. Empirical macroeconomics had over the decades been based on factor models, cointegration models, and vector autoregressions. Although these methods served well in capturing time-series dynamics and testing structural hypotheses, they did not do well with nonlinearities, high-dimensional predictors or structural breaks. They also made harsh assumptions of stationarity and linearity that are hardly true in practice when predicting recessions, inflation or exchange rates. Stock and Watson (2002) demonstrate that even factor augmented regressions, which were a significant innovation at one time, fail when the number of predictors grows excessively large or they get too noisy. These shortcomings have fostered the use of ML tools that have the capacity to process complex data in a manner that is flexible.

Supervised ML algorithms are among the most popular ones, and they enjoy predictive accuracy and the capability to capture nonlinear relationships. Rahman et al. (2022) provide evidence that Gradient Boosting, XGBoost and Random Forest are always superior to logistic regression in terms of predicting recession. These ensemble techniques minimize bias and variance by means of repeated subsampling and tree aggregation to make stable and reliable predictions. They have found their best application in cyclical recessions where financial; real and survey measures take centre stage. This opinion is supported by Chen et al. (2022), who demonstrate that ML models do not produce as many forecast errors as the econometric methodologies in high-dimensional conditions. Although neural networks and support vector machines are also discussed, they are more demanding in terms of their large datasets and are prone to overfitting, thus tree-based ensembles are more realistic in most macroeconomic settings.

Supervised ML has also been used effectively in real time prediction. Lahiri and Wang (2019) demonstrate that the integration of survey expectations on the basis of high-frequency data with the help of ML will contribute greatly to the accuracy of GDP nowcasting. ML models give more credible forecasts in the short term than the conventional regression-based models by incorporating financial indicators, commodity prices, and consumer sentiment. Similar findings may be traced to Medeiros et al. (2021), who state that machine learning techniques, specifically, boosting algorithms, are effective at making inflation forecasts in data intensive environments. The common theme of their work is that in the situation when numerous indicators are present, supervised ML algorithms can be quite efficient in filtering noise and producing predictive signals.

Despite these advantages, supervised ML does have issues limiting its direct use in the policy context. One of the problems is interpretability. Good predictions are required not only by policymakers but also a description of causal. drivers and mechanisms. Ensemble methods, however, tend to be black boxes, and do not need explicit functional. tables to predict (Varian, 2019). Attempts to achieve this non-transparency, incorporate post-hoc interpretability procedures. attaching SHAP values and LIME explanations. Even though the approaches have the capability of establishing the significance of variables. and local sensitivities, these are model-neutral approximations, which are not grounded in economic theory (Molnar, 2022). A weakness of the black-box predictions is that it is impossible to directly trace the policy shocks or structural relationships in the predictions. limit to institutional adoption.

Unsupervised ML techniques have also begun to enter the micro-economic analysis alongside the supervised methods. Unlike the supervised models, the results must not only be labeled such (e.g. no recession, recession, etc.) and also predetermined, unmonitored learning finds patterns in the background and structures without stating its objective. This fact makes it especially attractive in economic terms where the most attractive is. latent important variables of interest such as financial stability or business cycle phase, are not observable.

Clustering methods are one of the most popular unsupervised tools that are used in economics. Hoppmann et al. (2021) use clustering on composite indicators and can detect turning points on business cycle, which give. indications of recessions early. Unsupervised learning may also be useful in detecting regime shifts that cannot be easily detected. observed in raw data through grouping the observations that have similar dynamics. This technique is particularly helpful to policymakers, when transitions between cycles should be detected. The dimensionality reduction is the next crucial part of unsupervised learning.

The next important aspect of unsupervised learning is dimensionality reduction. Economic data is often full of hundreds of indicators, both financial series and survey outcomes. Since the times of principal component analysis (PCA) to compress information is not new to the econometrics field, more current methods like uniform manifold approximation and projection (UMAP) are becoming more flexible. Kaufmann and Schelling (2020) prove that UMAP can find meaningful low-dimensional structures in macroeconomics data that helps in visualization, as well as in further modeling. These methods enable analysts to store important data and also lessen the calculational loads.

The cross-country and sectoral analysis is also made easier with the help of clustering. Xu and Wunsch (2005) offer a thorough survey of clustering algorithms, which are shown to be useful in terms of categorizing economies in terms of their structural characteristics. This can be used in macroeconomic forecasting to peer-group compare and to improve the inter-regional transferability of models. Unsupervised approaches reveal new perspectives on systemic risk and heterogeneity by revealing latent typologies of the economies or sectors.

What is novel about unsupervised ML in macroeconomic forecasting is its complementary nature to the supervised ones. Whereas in the supervised ML, predictive accuracy is optimized to known outcomes, in the unsupervised learning, latent regimes and structures are discovered to deepen the explanatory foundation of the forecasts. Using clustering results to build supervised pipelines, researchers can build hybrid models that not only make accurate predictions but also provide new dimensions of interpretation. This two-sided view is a developing field of concern, and the direct juxtaposition of supervised and unsupervised and interpretable options in your study is directly related to this new stream of research.

Although ML can be used to provide predictive benefit, the interpretability question is urgent, particularly in policy-oriented systems. A possible solution is Response Surface Methodology (RSM). RSM, originally created in the field of engineering and optimization of processes, is based on the approximation of functional relations between a set of input and output variables by means of a set of equations of the form of polynomials. As observed by Montgomery (2019), RSM models usually have linear, quadratic, and interaction terms to enable them to represent curvature and interdependence of the variables in a transparent algebraic form.

RSM has potential based on applications to economics. Mendez et al. (2021) use RSM in the model of regional growth and demonstrate that it could offer reasonable functional equations, in which economic variables and growth performance are interrelated. The fact that RSM equations are explicitly formulated makes them well adapted to the situations where the decision-makers need explicit structural explanations. Furthermore, it is possible to use RSM in classification problems, which makes it applicable not only to continuous forecasting but to categorical forecasts in finance and industry.

The most important issue in RSM is multicollinearity, that is, when the predictors are very correlated, and it often happens with macroeconomic data. Myers et al. (2016) suggest regularization measures to address this problem, which will guarantee constant estimates of coefficients. This regularization does not affect interpretability but enables RSM to generalise to more high-dimensional data settings. The ongoing development of RSM algorithms is an example of how traditional statistical models may develop to become useful in the data-intensive world.

Recent studies indicate that integration can be used to overcome this trade-off. According to Yu et al. (2022), symbolic regression, which combines a stronger predictive power than ML with the transparency of algebraic models would be proposed in the middle. The complementary framework might utilize supervised ML to determine influential predictors, unsupervised learning to discover hidden structures and RSM to formalize results as explicit equations. These kinds of designs would give accurate and interpretable forecasts to the policymakers.

This study is a direct contribution to that agenda because it makes a comparison between supervised ML, unsupervised ML, and RSM in predicting recessions. In contrast to the previous work, the analysis does not just assess predictive performance, but also interpretability, which is a new contribution to methodological evaluation. In this way, it tackles one of the challenges in the field of modern economics: how to produce forecasts accurate enough to make a decision and transparent enough to hold policy accountable. This comparative framework, with input on United States recession data 1966-2025, illustrates that predictability needs to be balanced with interpretability of predictions to have any policy value.

3.Methods

3.1 Methodological Overview

The proposed framework (Figure 1) brings together machine-learning classifiers and Response Surface Methodology (RSM). It involves a step-by-step process including data acquisition, exploration data analysis, model training with cross-validation and hyperparameters tuning, performance assessment, feature interpretation and finally, RSM model building that extends the machine learning results with added interpretability.

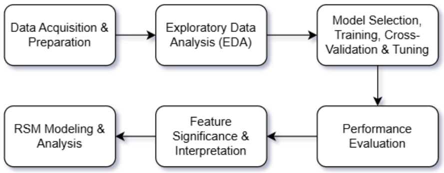

Figure 1. Workflow of the proposed ML-RSM approach

3.2 Data Acquisition, Preprocessing and Optimal Feature Selection

The dataset has been constructed by combining several macroeconomic data sources into a continuous series from 1966-2025. All these separate datasets were collected from Federal Reserve Bank of ST. lOUIS. A set of six variables were chosen as predictors (Unemployment Rate, Real GDP Growth, Debt-to-GDP Ratio, CPI, 10-Year Treasury Yield, and 3-Month Treasury Yield) and the dependent variable was a binary recession indicator based on NBER recession classifications. In the case of unsupervised algorithms, we have excluded two input parameters named Unemployment Rate and 10Y Treasury Rate, and the remaining feature has gone under standard scaling.

3.3 Exploratory Data Analysis

The macroeconomic indicators were assessed using Exploratory Data Analysis. Descriptive statistics showed (Table 1 and Table 2) that some key trends existed. Unemployment rates have not been decreasing significantly, whereas the growth of GDP became positive, which signified economic recovery. Debt to GDP ratios went high indicating growing fiscal burdens. There was high variation and inflation in CPI values. There was a general reduction in the levels of treasury rates indicating a relaxed monetary environment. These trends give the first hints about the changing economic environment.

Table 1. Statistics of Recession period

Feature	Mean	Median	Std
Unemployment Rate	5.91	5.7	1.76
Real GDP Growth	3.43	3.2	4.10
Debt to GDP Ratio	65.49	60.05	29.12
CPI	156.79	157.3	79.99
10Y Treasury Rate	5.73	5.62	2.82
3M Treasury Rate	4.21	4.67	3.01

Table 2. Statistics of Non-Recession period

Features	Mean	Median	Std
Unemployment Rate	5.99	5.48	1.71
Real GDP Growth	-1.24	-1.5	3.52
Debt to GDP Ratio	46.73	35.13	19.38
CPI	120.36	95.28	72.22
10Y Treasury Rate	7.67	7.415	3.82
3M Treasury Rate	6.67	7.28	4.12

Unemployment rates have not been decreasing significantly, whereas the growth of GDP became positive, which signified economic recovery. Debt to GDP ratios went high indicating growing fiscal burdens. There was high variation and inflation in CPI values. There was a general reduction in the levels of treasury rates indicating a relaxed monetary environment. These trends give the first hints about the changing economic environment.

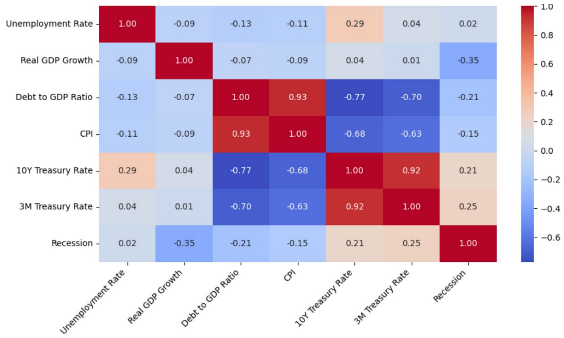

Figure 2. Pearson Correlation

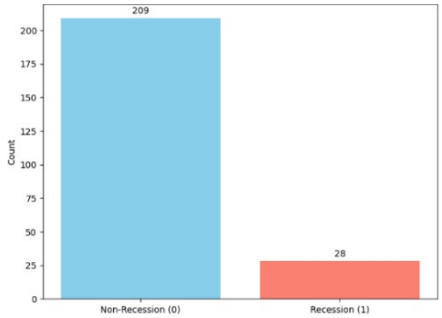

Figure 3. Count of Non-Recession vs Recession Quarters

The EDA also involved the distribution of the quarterly recessions and non-recessions (Figure 3) periods that revealed the class imbalance which is prevalent in recession prediction activities. Correlation analysis between the key variables was estimated between Real GDP Growth, Debt-to-GDP Ratio, Treasury rates, Consumer Price Index and unemployment to identify possible multicollinearity and linearity relationship. A Pearson correlation matrix (Figure 2) based on visualization indicated which indicators fluctuated in tandem in the various stages of the economic cycle. The use of summary statistics, as well as plots demonstrated the existence of clear differences in the behavior of features between recessionary and stable quarters. These descriptive observations did not only verify the usefulness of the selected indicators but also helped to apply supervised and unsupervised models by explaining the importance of features and the structure of the dataset.

3.3 Model selection, Training, Tuning, and Clustering

this paper, supervised and unsupervised learning was used to assess recession forecasting. In order to analyze it supervised, five classifiers were chosen, including Decision Tree (DTC), Extra Trees (ETC), Random Forest (RFC), AdaBoost (ABC), and Gradient Boosting (GBC), based on the fact that each could model nonlinear relationships and also could use ensemble learning to enhance performance. Repeated Stratified K-Fold Cross-Validation was used to train and validate the models to ensure even splits and discourage overfitting. The Bayesian optimization in Optuna was used to optimize hyperparameters to present an effective exploration of parameter spaces.

The unsupervised approaches were used to find the latent structures in the data. K-Means, Gaussian Mixture Models (GMM), DBSCAN, HDBSCAN and Spherical clustering were run, and performance evaluated based on the Silhouette score and Bayesian Information Criterion (BIC). Clusters were not labeled thus Hungarian algorithm was applied to match them with recession and non-recession clusters to allow a direct comparison with supervised methods.

3.4 Supervised ML Algorithms

Table 3 presents the training and validation accuracy and ROC-AUC values of the six supervised classifiers. XGBoost (XGBC) and Gradient Boosting (GBC) produced best overall performance, validation accuracy of 0.9203 and 0.9161 respectively, and a ROC-AUC of greater than 0.90 which suggests good predictive ability as well as generalization. Extra Trees (ETC) also-classified strongly well-balanced results (accuracy of 0.9132) and (ROC-AUC of 0.9215) whereas(RandomForest) (RFC) remained consistent once again with validation accuracy of 0.9051 and ROC-AUC of 0.8325. Decision Tree (DTC) and (AdaBoost) (ABC) although fairly less strong remained competitive strong performance (validation accuracy of 0.9110) and (ROC-AUC of near 0.75). Typically ensemble methods (RFC, ETC, GBC, and XGBC) dominate baseline DTC model consolidating strength of ensemble learning again to learn strong decision boundaries that tightly improve model robustness.

Table 3. Accuracy and ROC-AUC Score for Different Supervised Classification Model

Models	Accuracy		ROC-AUC	
	Train	Val	Train	Val
DTC	0.9208 ± 0.0091	0.9110 ± 0.0397	0.7719 ± 0.0268	0.7500 ± 0.1020
ETC	0.9398 ± 0.0085	0.9132 ± 0.0408	0.9132 ± 0.0408	0.9215 ± 0.0517
RFC	0.9149 ± 0.0100	0.9051 ± 0.0425	0.9257 ± 0.0129	0.8325 ± 0.1060
ABC	0.9208 ± 0.0091	0.9110 ± 0.0397	0.7722 ± 0.0265	0.7500 ± 0.1019
GBC	0.9590 ± 0.0098	0.9161 ± 0.0328	0.9902 ± 0.0036	0.9055 ± 0.0685
XGBC	0.9545 ± 0.0086	0.9203 ± 0.0353	0.9895 ± 0.0031	0.9077 ± 0.0571

Table 4 is a summary of the hyperparameters used for all classifiers. The decision-tree (DTC) classifier was limited to shallow depth, pruning, but the extra-trees (ETC) and random-forest (RFC) classifiers used gigantic numbers of estimators, utilized balanced sampling and were allowed to make deeper splits. Adaptive boosting (ABC), gradient-boosted (GBC) and extreme gradient-boosted (XGBC) boosting models were highly optimized using low learning rates, limiting tree depth and regularization. Of course, ensemble approaches had more estimators and deeper trees, and boosting algorithms focused on decreasing the learning rate to maximize stability.

Table 4. Optimized Hyperparameters of the Models

Model	Optimized Hyperparameters
DTC	criterion=gini, splitter=best, max_depth=1.0, min_samples_split=6.0, min_samples_leaf=8.0, max_features=None, class_weight=None, ccp_alpha=0.0006
ETC	n_estimators=883.0, criterion=log_loss, max_depth=6.0, min_samples_split=12.0, min_samples_leaf=1.0, max_features=None, bootstrap=True, class_weight=balanced, max_samples=0.5173
RFC	n_estimators=543.0, max_depth=27.0, min_samples_split=40.0, min_samples_leaf=19.0, max_features=None, bootstrap=False, class_weight=None, criterion=entropy
ABC	n_estimators=12.0, learning_rate=0.0126, algorithm=SAMME.R, max_depth=1.0, min_samples_leaf=8.0, tree_class_weight=None
GBC	loss=log_loss, n_estimators=31.0, learning_rate=0.1236, subsample=0.63, max_depth=8.0, min_samples_split=8.0, min_samples_leaf=14.0, max_features=None, max_leaf_nodes=32.0
XGBC	n_estimators=94.0, learning_rate=0.0847, max_depth=9.0, min_child_weight=0.7966, subsample=0.5661, colsample_bytree=0.9058, gamma=0.9289, reg_alpha=0.0089, reg_lambda=1.3959, grow_policy=depthwise

3.5 Feature Importance and Interpretation from Supervised Models

Table 5 demonstrate the following insights: Real GDP Growth (RGG) remains the most significant predictor in all the classifiers with the 3-Month Treasury Rate (3MTR) always coming in second or third place. The importance of CPI and Debt-to-GDP Ratio (DTGR) is moderate, and Unemployment Rate is always the lowest which proves its lagging indicator. On the whole, recession forecasts are driven by growth and short-term interest rates and a minor predictive value on unemployment is added.

Figure 4 shows the XGBoost feature-importance ranking and it can be seen that the most powerful predictors are the Real GDP growth followed by the 3-month Treasury rate and Debt-to-GDP ratio. By comparison, the contribution of Unemployment and the 10-Year Treasury Rate is minor, which highlights the preeminent contribution of growth and short-term rates to detecting recession in the early stages.

Table 5. Feature importance Ranking of Input Parameters across

Model	Feature					
	R.G.G.	D.G.R.	3M	10Y	CPI	U.R.
DTC	1	2	3	5	4	6
ETC	1	2	3	5	4	6
RFC	1	5	2	6	3	4
ABC	1	2	3	5	4	6
GBC	1	3	2	6	5	4
XGB	1	3	2	5	4	6

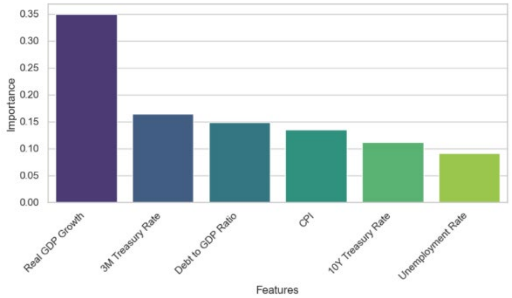

Figure 4. Feature Importance for XGBoost Classifier

3.6 Model Interpretability through SHAP Analysis

Table 6. SHAP Importance Ranking of Input Parameters across Supervised Classification Models

Model	Feature					
	R.G.G.	D.G.R.	3M	10Y	CPI	U.R.
DTC	1	4	2	5	6	3
ETC	1	2	3	4	5	6
RFC	1	5	2	6	3	4
ABC	1	4	2	5	6	3
GBC	1	4	2	5	6	3
XGBC	1	4	2	5	6	3

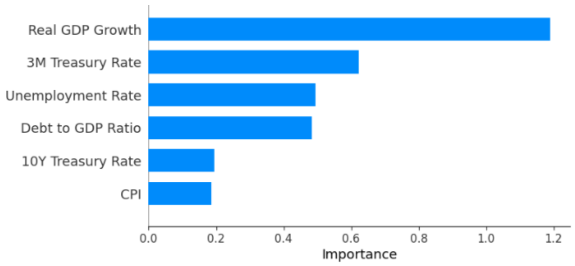

Figure 5. SHAP Importance for XGBoost Classifier

Table 6 and Figure 5 present the SHAP importance ranking of input parameters against supervised classifiers and supportive evidence that Real GDP Growth is the most important determinant of recession risk. The magnitude of its dominance demonstrates that macroeconomic growth variations have the highest explanatory power for classification outcomes. The 3M Treasury Rate and Unemployment Rate then become the next most important features, capturing the importance of short-run interest rates and labor market health. Debt-to-GDP ratio contributes modestly, capturing the fiscal nature of recession forecasting. The CPI and 10Y Treasury Rate, however, always exhibit smaller SHAP values, indicating minor incremental contribution compared to other predictors. Together, these results strengthen the prominence of growth and near-term financial metrics, with inflation and long-run rates playing more ancillary roles in model interpretability.

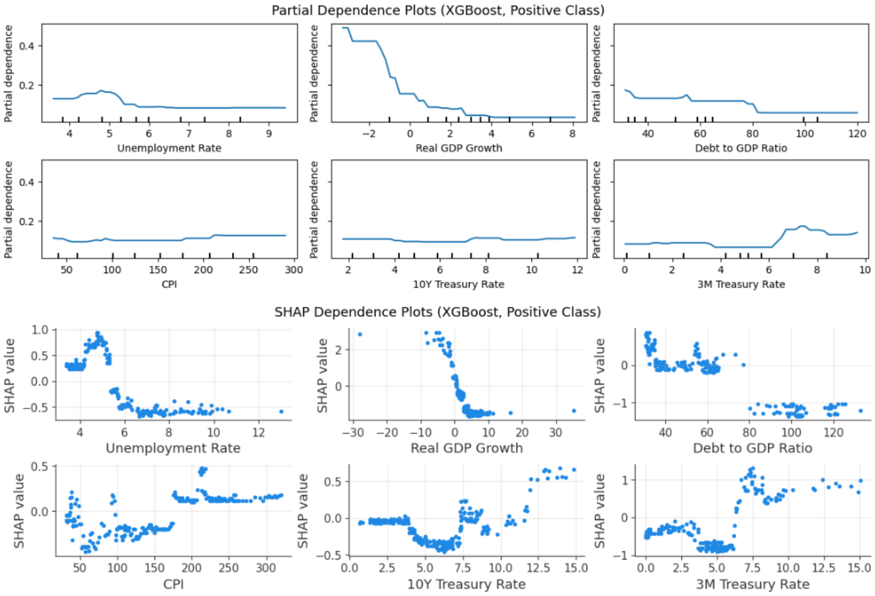

Figure 6. Partial and SHAP Dependence Plot (XGBoost, Recession)

Figure 6 shows the Partial Dependence Plots (PDPs) and SHAP dependence plots obtained using the XGBoost classifier with the input being recession prediction (positive class). The PDPs show the marginal effect of all macroeconomic variables, and it is apparent that high unemployment rates and negative GDP growth significantly raise the probability of the model to fall into recession. The debt-to-GDP, and CPI display non linear patterns, with specific threshold values initiating clearly distinguishable shifts in the output; treasury rates, although exhibiting a relatively linear gradient, nevertheless have statistically significant impact.

The SHAP dependence plots also explain the contribution by features at the instance level, both in magnitude and direction. Regularly, declining GDP and high unemployment rates lead to a strong positive SHAP value hence proving their significance in predictive outcomes. CPI and treasury rates on the other hand show non-homogeneous contributions depending on their range discussed as signs of interactions with other predictor variables.

3.7 Unsupervised ML Algorithms

Table 7 explains the clustering results of the unsupervised methods after Hungarian matching. KMeans and Spherical clustering generated two clusters after alignment, obtaining moderate accuracies of 0.7342 and 0.7426, respectively. Density-based methods, however, demonstrated improved performance, where DBSCAN obtained the best matched accuracy of 0.8945 and was followed by GMM of 0.8861. HDBSCAN, even as it shrunk to two clusters after matching, only reached a relatively lower accuracy of 0.6371. Bayesian GMM revealed a performance comparable to that of KMeans with an accuracy of 0.7342. Generally, density-based and mixture models (DBSCAN, GMM) of clustering methods indicated a higher effectiveness than centroid-based and spherical methods, which again exemplify their suitability in capturing the underlying structure of a dataset.

Table 7. Summary of Results from Unsupervised Models after Matching

Model	Number of Clusters Before Matching	Number of Clusters after Matching	Accuracy (Matched)
KMeans	3	2	0.7342
DBSCAN	7	2	0.8945
HDBSCAN	2	2	0.6371
GMM	8	2	0.8861
Bayesian GMM	3	2	0.7342
Spherical	3	2	0.7426

Unsupervised clustering by DBSCAN was assessed by diagnostic visualizations including confusion matrix (Figure 7) and UMAP plots (Figure 8). The confusion matrix gave a direct comparison between the labeling of the clusters predicted and the real labeling of recession classes after Hungarian matching. In the DBSCAN model which obtained one of the highest accuracy, the majority of non-recession quarters were correctly classified, and lesser portion of recession periods were wrongly classified. This is an indication of the challenge of representing brief declines and recoveries as opposed to steady periods. A two-dimensional UMAP projection was also used to analyze cluster separability. This dimensionality reduction procedure saved local and global structures in addition to allowing visual evaluation of how clusters were overlaid on the recession and non-recession states. In the DBSCAN UMAP plot, it was clear that recession quarters were a distinct grouping. The combination of the confusion matrix and the UMAP visualization proved that the clustering performance was satisfactory and that the dynamics of features distinguishing between crises and the times of the stable growth occur.

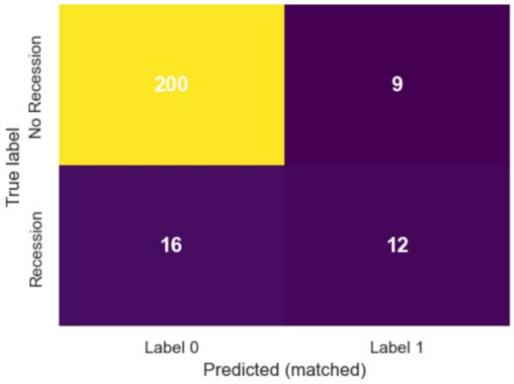

Figure 7. Confusion Matrix of DBSCAN with Hungarian Matching

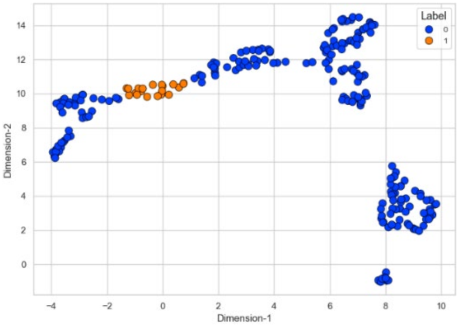

Figure 8. UMAP Visualization of DBSCAN Clusters after Hungarian Matching

3.8 Response Surface Methodology

During implementation of the Response Surface Methodology (RSM) application, macro variables were selected as per their ranking of importance as revealed in Table 8, thus conserving a focus on most influential recession predictors. Polynomial models including linear terms, squared terms, and interaction terms were formed in terms of between three to six such top-ranked inputs. Considering that the output variable of interest was binary, distinguishing recession from its absence, this problem was framed as a classification problem.The issue of multicollinearity was alleviated by use of ridge regularization, while model validation used optimized thresholds within an expanding-window time series cross-validation framework. Results, expressed as mean accuracy and ROC-AUC (std deviation) demonstrate that models utilizing three variables achieved optimum levels of performance (accuracy: 0.856 ± 0.165, ROC-AUC: 0.780 ± 0.213). As additional variables were included in set utilized by models, a degradation of performance occurred that serves to illustrate potential for overfitting and reinforces predictive role of a small set of quintessential macroeconomic variables.

Table 8. RSM Performance Across Variable

Number of Variables	**Accuracy**	**ROC-AUC**
3	0.856 ± 0.165	0.780 ± 0.213
4	0.790 ± 0.252	0.748 ± 0.202
5	0.769 ± 0.286	0.556 ± 0.381
6	0.779 ± 0.227	0.560 ± 0.362

4. Results

Comparative analysis (Figure 9) of models clearly shows that there is a difference in forecasting actions. XGBoost (XGBC) was the best performing supervised classifier with a validation accuracy of 0.920 and then Gradient Boosting (GBC) was second with a 0.916. It was confirmed that these ensemble methods performed better than the unsupervised and regression-based models and thus showed the capacity of modeling nonlinear macroeconomic dynamics. In unsupervised techniques, DBSCAN had competitive accuracy of 0.8945, marginally higher than Gaussian Mixture Models (GMM) of 0.8861, and in finding recessionary regimes without assigned labels, it had potential. On the contrary, Response Surface Methodology (RSM) model, although interpretable in its explicit functional equations, was slow in predictability with a lag of 0.856.

The feature-level analysis provided further insight into the economic determinants of classification effectiveness. Percentage changes in mean and median values were computed between categories defined by DBSCAN clusters, as tabulated in Table 9. Real GDP Growth (R.G.G) indicated the largest positive change as it registered a mean-value increment of 16.07 and a median increment of 66.67, which made it central to recession indicators. The Debt-to-GDP Ratio (D.G.R) and Consumer Price Index (CPI) also registered sizeable increments during periods of recession, indicating the role of fiscal imbalances and inflationary pressure. By contrast, 3-Month Treasury Rate (3M) registered a sharp decline, which indicated a tightening of financial policies.

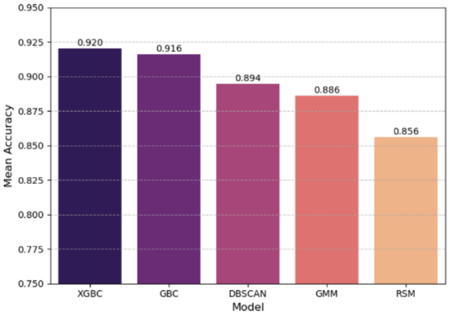

Figure 9. Accuracy Comparison between Different Models

Table 9. Percentage Change in Feature Statistics between Categories and DBSCAN Clusters

Feature	% Change in Mean		% Change in Median	
	N.R.- Label 0	R.- Label 1	N.R.- Label 0	R.- Label 1
R.G.G.	1.5	16.07	0	66.67
D.G.R.	-2.67	30.54	-1.25	9.59
CPI	-4.28	44.42	-4.53	27.37
3M	7.18	-37.29	4.39	-14.29

4.1 Quadratic Classification Function

The study produced a three-term quadratic Response Surface Methodology (RSM) classification equation using Real GDP Growth (RGG), Debt-to-GDP Ratio (DTGR) and the 3-Month Treasury Rate (M3TR). The choice of these variables was decided by feature importance rankings obtained from ensemble machine learning models. The quadratic model included linear, squared and interaction terms, yielding a polynomial decision surface to distinguish recessionary from non-recessionary periods. Ridge regularisation was applied to address multicollinearity and prevent coefficient unstable results.

$$\mathit{Score} = 0.103 - 0.023 \times \mathit{RGG} - 0.011 \times \mathit{DTGR} + 0.012 \times \mathit{3MTR} + 0.0026 \times \mathit{RGG}^2 + 0.0079 \times (\mathit{RGG} \times \mathit{DTGR}) - 0.0073 \times (\mathit{RGG} \times \mathit{3MTR}) - 0.0053 \times \mathit{DTGR}^2 - 0.0047 \times (\mathit{DTGR} \times \mathit{3MTR}) + 0.0151 \times \mathit{3MTR}^2$$

Classification Rule:

Recession (1): If the calculated score is ≥ 1.5940, then the model predicts a recession.

Non-recession (0): If the score is < 1.5940, then the model predicts no recession.

The resulting score function contained an explicit classification rule: if the score is greater than t = 1.594, then prediction is made that there is a recession. As the RSM equation provides interpretable functional relationships between variables and the probability of recession and complements the higher predictive but less interpretable ML models, it has a value for policy-relevant applications.

5. Limitations and Future Scope

The proposed study has been applied to the prediction of recession by the application of ensemble machine learning (ML) classifiers, unsupervised models, and Response Surface Methodology (RSM), however, its broader use is limited by several constraints. The assessment is based only on the U.S. quarterly data between 1966-2025. Although this long sample can be used to capture macroeconomic dynamics, it restricts generalization to other economies that have dissimilar fiscal institutions, institutional structure or monetary regime. Not all recession trends can be similar in countries, which lessens the transferability of policy. This was also the case with RSM where predictors rose in sensitivity to high-dimensional multicollinearity and overfitting. Its application to data-rich environments is not so strong without dimensionality reduction or regularization. Lastly, the binary yes/no treatment of recessions serves as a simplistic approach to reality because it disregards depth, duration and recovery. To develop specific policies, policymakers tend to require differences between gentle decelerations and great slumps.

These limitations can be dealt with in future studies. The addition of the cross-country data to the analysis would enable the comparison of the dynamics of recession in different institutional settings. Timeliness and responsiveness might be enhanced with higher-frequency indicators, e.g. monthly financial or labor market data. Further feature selection can also be used to promote stability and interpretability as SHAP-based scores, genetic algorithms, or evolutionary optimization. Combining the predictive capabilities of ML with the transparency of RSM, hybrid ML-RSM models have a very high potential in providing both policy-relevant and accurate predictions.

6. Conclusion

This paper will give a comparative analysis of ensemble machine learning (ML) classifiers, unsupervised learning systems, and Response Surface Methodology (RSM) in the forecasting of U.S. recessions between 1966-2025. The findings prove the higher predictive power of ensemble ML models, especially Gradient Boosting and XGBoost, which can represent nonlinearities and high-order macroeconomic interactions. Simultaneously, the meaning non-supervised methods such as DBSCAN and Gaussian Mixture Models suggested that latent. data block structures can be exposed well and give more information about economic cycles even when there is no. labeled results. By contrast, RSM, which was less accurate, could provide interpretable functional relationships in the type of explicit quadratic equations between Real GDP Growth, the Debt-to-GDP Ratio, and the 3-Month Treasury. Nominal ratio to the probability of recession. The following results show a primary trade-off between predictive performance and. interpretability, which is a challenge that lies at the core of economic forecasting that is of interest to policymakers.

The comparison analysis reveals the fact that the twin desires of precision and are not covered in any single way. disclosure comprehensively. Supervised ML is well-predicted but poorly interpretable; unsupervised ML is poorly predicted. are the implicit structures but not directly accountable; and RSM describes the things explicitly. with sacrificed accuracy. The strategy that promises is a combination of the said strategies. With the help of clustering insights, the hybrid ML-RSM models can merge predictive capability and transparency to make forecasts. more practical to policy makers. Macroeconomic conditions are becoming more complex and information based since this makes it more complex, intensive, one must work out methods that are precise and explainable. This research paper has a role to play. play in that agenda, and it demonstrates how ML, unsupervised learning and RSM can be complementary to one another, and it also outlines the means of building forecasting systems that will be statistically reliable as well as policy wise.

References

Stock, J. H. and Watson, M. W., Forecasting using principal components from a large number of predictors, Journal of the American Statistical Association, vol. 97, no. 460, pp. 1167–1179, 2002.

Rahman, S. A., Hossain, M. S. and Rahman, M. M., Predicting economic recessions using machine learning algorithms, Journal of Economic Studies, vol. 49, no. 5, pp. 835–855, 2022.

Chen, Z., Zhang, Y. and Yang, L., Comparing statistical and machine learning methods for macroeconomic forecasting, Entropy, vol. 24, no. 12, p. 1698, 2022.

Lahiri, K. and Wang, F., Combining surveys and high-frequency data for GDP forecasting, Econometrics, vol. 7, no. 1, pp. 1–21, 2019.

Medeiros, J., Vasconcelos, G., Veiga, A. and Zilberman, B., Forecasting inflation in a data-rich environment: The benefits of machine learning methods, Journal of Business & Economic Statistics, vol. 39, no. 1, pp. 98–119, 2021.

Varian, H., Machine learning and econometrics, Journal of Economic Perspectives, vol. 33, no. 2, pp. 3–22, 2019..

Lundberg, S. and Lee, S.-I., A unified approach to interpreting model predictions, Advances in Neural Information Processing Systems, vol. 30, pp. 4765–4774, 2017.

Ribeiro, M. T., Singh, S. and Guestrin, C., Why should I trust you? Explaining the predictions of any classifier, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 1135–1144.

Hoppmann, J., Schmitt, T. and Hoffmann, V., Unsupervised learning for business cycle detection: A clustering approach, Journal of Economic Dynamics and Control, vol. 132, p. 104222, 2021.

Molnar, C., Interpretable Machine Learning, 2nd ed., 2022.

Kaufmann, D. and Schelling, A., Dimensionality reduction for macroeconomic indicators: A UMAP approach, Computational Economics, vol. 56, no. 3, pp. 623–645, 2020.

Xu, R. and Wunsch, D., Survey of clustering algorithms, IEEE Transactions on Neural Networks, vol. 16, no. 3, pp. 645–678, 2005.

Montgomery, D. C., Design and Analysis of Experiments, 8th ed., Hoboken, NJ: Wiley, 2019.

Méndez, J. A., Maqueda, E. and Martínez, M., Application of RSM in economic modeling: A case study in regional growth, Applied Economics, vol. 53, no. 18, pp. 2062–2074, 2021.

Myers, R., Montgomery, D. and Anderson-Cook, C., Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 4th ed., Hoboken, NJ: Wiley, 2016.

Biographies

Miftaur Rahman Zisan is an undergraduate in Industrial and Production Engineering at the Bangladesh University of Engineering and Technology (BUET). He is keenly interested in operations research, improvement in processes and computational modelling and in system analysis and optimisation techniques in addition to data-oriented solutions to intricate problems in engineering. He has experience in the application of machine learning approaches, especially with predictive analytics and decision support based upon data. He is driven by the interest in transferring theoretical paradigms to applied problems and thrives in interdisciplinary contexts that interweave analytics, engineering, and strategic design.

Anik Debnath is in his fourth year of undergraduate studies in Industrial and Production Engineering from the Bangladesh University of Engineering and Technology (BUET). His specializations include artificial intelligence, data science, and statistical forecasting techniques, as well as his field of research interest in operation research, composite materials, and environmental sustainability. He has been involved in research that combine machine learning techniques and optimization frameworks to address challenging industrial and environmental problems. His activities reflect a strong will to create data-informed, sustainable, and innovative solutions that combine analytical accuracy and real-world relevance.

Gourab Mondal is an undergraduate student at the Institute of Business Administration, University of Dhaka, studying Finance with a minor in Economics. His areas of current interest for research include quantitative finance, behavioral economics, corporate governance, and regulatory policy. He has worked on projects at Bangladesh's national financial regulator on applying machine learning to detect fraud and equity research and collaborated with the Dhaka Stock Exchange on studying large-scale data on corporate governance. With growing experience in data analysis, econometrics, and financial modeling, he hopes to bring quantitative sophistication to economics and governance perspectives with which to approach finance's challenging problems.