Journal of Money and Economy
Vol. 21, No. 3, Summer 2026
pp. 411-446

Original Research Article

Predicting The Effectiveness of Monetary and Fiscal Policies Under Financial Sanction: A Machine Learning and SHAP Analysis

Seyed Hossein Mirjalili[*] **Samira Heydarian**[†]

Received: 16 Oct 2025 Approved: 01 Jun 2026

Financial sanctions restrict access to international financial markets and disrupt trade flows, thereby exerting persistent pressure on macroeconomic performance in sanction-targeted economies such as Iran. This study develops a data-driven framework to evaluate the effectiveness of monetary and fiscal policies in mitigating these adverse effects under prolonged financial sanctions. To capture nonlinear, regime-dependent, and potentially predictable relationships among key macroeconomic variables, we employ advanced machine learning techniques—including Logistic Regression, Random Forest, XGBoost, and LightGBM—combined with SHAP-based interpretability analysis.

Using monthly data covering the period 2006–2024 and a time-based validation strategy, the models are trained to predict policy effectiveness as reflected in three core macroeconomic indicators: real GDP growth, inflation, and the exchange rate. The empirical results show that tree-based machine learning models substantially outperform the linear benchmark. XGBoost delivers the strongest predictive performance, achieving an accuracy of 91% and an ROC–AUC of 0.94, followed by LightGBM (88% accuracy; 0.92 ROC–AUC), Random Forest (85% accuracy), and Logistic Regression (75% accuracy). These findings indicate that policy–outcome relationships exhibit systematic and predictable nonlinear patterns that are better captured by flexible algorithms.

SHAP-based analysis identifies the nominal interest rate, the government spending-to-GDP ratio, the tax-to-GDP ratio, money supply growth, and a sanctions regime indicator as the most influential determinants of policy outcomes. Financial sanctions are associated with contractionary effects on output while simultaneously intensifying inflationary and exchange rate pressures, thereby altering the transmission mechanisms of monetary and fiscal policies. Although the analysis is predictive rather than causal, the results demonstrate the clear advantage of machine learning methods in capturing complex interactions in sanction-constrained macroeconomic environments.

[*] Faculty of Economics, Institute for Humanities and Cultural Studies, Tehran, Iran; smirjalili@ihcs.ac.ir

[†] PhD in Economics, University of Yazd, Yazd, Iran; samira.Heidarian31@gmail.com (Corresponding Author)

Overall, this study highlights the practical relevance of integrating machine learning with explainable artificial intelligence tools to support transparent, data-driven macroeconomic policy design. The findings provide actionable insights for policymakers by identifying which policy instruments retain relative effectiveness under sanctions and how policy transmission mechanisms adjust across sanction regimes.

Keywords: Machine Learning, SHAP, Financial Sanctions, Monetary Policy, Fiscal Policy
JEL Classification*:* E2, G17, C53, C49

1 Introduction

Over the past two decades, economic sanctions—particularly financial sanctions—have increasingly been employed as instruments of foreign policy to constrain the economic capacity and international influence of targeted countries. In the case of Iran, these measures have primarily taken the form of asset freezes, restrictions on access to international banking systems, and disruptions to global payment networks. Such constraints have generated substantial and persistent pressures on key macroeconomic indicators, including GDP growth, trade, and exchange rate stability (Popa, 2024; Heydarian & Mirjalili, 2026; Pahlavani et al.2025).

The relationship between financial sanctions and macroeconomic outcomes is inherently complex and nonlinear. Sanctions fundamentally alter the institutional, financial, and expectations environment in which economic policies operate, often inducing regime shifts, heightened uncertainty, and asymmetric macroeconomic responses (Safari et al.,2023,2024; Mirjalili & Safari, 2025). As a result, policy–outcome relationships may exhibit systematic yet nonlinear and regime-dependent patterns. Traditional econometric approaches—typically built upon linear specifications and parameter stability assumptions—often struggle to adequately capture such dynamics, thereby highlighting the need for more flexible and data-driven analytical frameworks.

Under sanction conditions, policymakers primarily rely on monetary and fiscal instruments to mitigate adverse economic effects. Monetary policy operates through interest rates, liquidity provision, and credit channels, while fiscal policy influences aggregate demand via taxation and government expenditure. Together, these tools aim to stabilize output, support economic activity, and contain inflationary pressures (Zareei et al., 2024). However, evaluating the effectiveness of these policies under financial sanctions remains challenging due to structural breaks, transmission distortions, and nonlinear adjustment mechanisms (Mirjalili, 2025a).

Financial sanctions affect policy effectiveness through multiple transmission channels. On the fiscal side, sanctions constrain government revenues—particularly in oil-dependent economies—restrict access to external financing, and compress fiscal space. Consequently, the scope for countercyclical fiscal intervention diminishes, forcing policymakers to balance short-term stabilization objectives against long-term fiscal sustainability (Farzanegan, 2011; Mirjalili, 2025b). On the monetary side, sanctions weaken banking intermediation, intensify exchange rate volatility, and amplify inflation expectations. In such environments, conventional monetary instruments may lose traction, while the exchange rate becomes a dominant adjustment mechanism. These distortions generate nonlinear and regime-dependent interactions between policy instruments and macroeconomic outcomes (Keshtgar & Mirjalili, 2026).

Despite the expanding literature on the economic consequences of sanctions, most existing studies either focus on the direct macroeconomic impact of sanctions or analyze monetary and fiscal policies in isolation, frequently within static and linear modeling frameworks. In the Iranian context, no prior research has jointly applied advanced machine learning techniques—such as XGBoost and LightGBM—alongside interpretability tools like SHAP to examine how monetary and fiscal policies operate under financial sanctions. This methodological gap constitutes the primary motivation for the present study.

Building on these theoretical considerations, this study employs machine learning methods to capture complex and nonlinear interactions among financial sanctions, policy instruments, and macroeconomic indicators. Unlike traditional econometric models that impose restrictive functional forms, the proposed framework allows policy effectiveness to vary systematically across sanction regimes and macroeconomic conditions.

Although the empirical focus is Iran, the methodological approach is sufficiently general to be extended to other sanction-targeted economies, such as Russia or Venezuela, subject to data availability. Iran provides a particularly informative case due to its prolonged exposure to financial sanctions, oil dependence, and the intricate interaction between fiscal and monetary policies under external constraints.

The primary objective of this study is to develop a data-driven framework for forecasting and evaluating the effectiveness of monetary and fiscal policies in mitigating the adverse macroeconomic effects of financial sanctions in Iran. Specifically, the study aims to:

1) Compare the predictive performance of advanced machine learning models (XGBoost, LightGBM, and Random Forest) with a baseline linear benchmark (Logistic Regression);
2) Identify the most influential macroeconomic and policy-related determinants of policy effectiveness;
3) Employ SHAP analysis to generate transparent and interpretable insights into model predictions.

Accordingly, three testable hypotheses are formulated:

(a) XGBoost and LightGBM outperform Logistic Regression in predicting policy effectiveness;

(b) GDP growth, inflation, the exchange rate, and the sanctions regime indicator play central roles in determining policy success;

(c) SHAP analysis effectively reveals the contribution of individual variables to model predictions.

Importantly, the objective of this study is predictive and exploratory rather than causal. Machine learning techniques are employed to uncover nonlinear statistical regularities and regime-contingent associations between policy instruments and macroeconomic outcomes under financial sanctions. Therefore, the estimated relationships should not be interpreted as evidence of structural or causal effects.

The remainder of the paper is organized as follows. Section 2 describes the data and methodology, including variable definitions and model specifications. Section 3 presents the empirical results and SHAP-based interpretation. Section 4 concludes and discusses policy implications.

2 Methodology

The objective of this study is to develop and evaluate machine learning models for predicting the effectiveness of monetary and fiscal policies in mitigating the adverse macroeconomic effects of financial sanctions. A supervised learning framework is employed, emphasizing predictive performance and the identification of nonlinear, regime-dependent patterns rather than causal inference.

Four classification models are considered: Logistic Regression as a linear benchmark, and three advanced tree-based algorithms—Random Forest, XGBoost, and LightGBM. These models are widely applied in macroeconomic and financial research due to their strong predictive performance and capacity to capture nonlinear relationships and high-dimensional interactions (Addo et al., 2018; Cho & Lee, 2022; Rundo et al., 2019). In contrast to classical time-series models such as AR, MA, and

ARIMA—which impose linear structures and restrictive functional assumptions (Alenezy et al., 2021)—tree-based methods are better suited to complex macroeconomic environments characterized by regime shifts and structural heterogeneity (Chatzis et al., 2018).

Unlike structural econometric models designed for causal identification, the present framework is explicitly predictive and exploratory. The models are not intended to estimate policy multipliers or identify structural mechanisms; rather, they uncover robust statistical regularities and conditional associations among policy instruments, financial sanctions, and macroeconomic outcomes. This distinction is particularly important in sanction-constrained economies, where endogeneity, simultaneity, and feedback effects complicate causal inference.

Policy effectiveness is operationalized as a single binary classification outcome indicating whether monetary and fiscal policies achieve favorable macroeconomic outcomes under financial sanctions. The binary target variable (Policy Effectiveness) is constructed based on three complementary macroeconomic dimensions: real GDP growth, inflation stability, and exchange-rate stability. A value of one represents a successful policy episode in which predefined improvement criteria are achieved, whereas a value of zero indicates an unsuccessful episode. Therefore, the machine learning algorithms predict the probability of overall policy effectiveness rather than separate GDP growth, inflation, or exchange-rate outcomes. These three macroeconomic indicators are subsequently used to define the composite target and to provide economic interpretation of the determinants of predicted policy success.

The sample period (2006–2024) spans multiple sanction and policy regimes and thus reflects substantial structural heterogeneity. Rather than estimating separate models for each regime, the machine learning framework captures nonlinear and regime-contingent effects endogenously through tree-based structures and interaction terms. This approach allows policy effectiveness to vary systematically across sanction environments without imposing ex ante assumptions regarding regime boundaries or parameter stability.

The primary contribution of this study is therefore not a benchmarking exercise against traditional econometric models such as VAR or ARDL, but the exploration of complex nonlinear and regime-dependent policy dynamics using interpretable machine learning techniques. A formal comparative assessment with classical structural models constitutes a complementary research agenda beyond the scope of the present analysis.

2.1 Data and Preprocessing

This study covers the period from 2006 to 2024, enabling a longitudinal assessment of financial sanctions and macroeconomic policy performance in Iran. The sample encompasses several distinct regimes: the pre-2011 period of relatively moderate sanctions, the intensified financial restrictions after 2012, the sanctions-relief episode following the JCPOA (2016–2018), and the post-2018 "maximum pressure" phase.

Given this extended timeframe, structural heterogeneity is a central feature of the dataset. Instead of assuming parameter stability or estimating regime-specific models, the machine learning framework accommodates heterogeneity endogenously. Through nonlinear tree-based splits and interaction structures, the relationship between policy instruments and macroeconomic outcomes is allowed to vary systematically across sanction environments.

Macroeconomic performance is evaluated using three key indicators: real GDP growth, inflation, and the exchange rate. Joint consideration of these dimensions provides a more comprehensive assessment of policy effectiveness than output growth alone, particularly in sanction-affected economies where price stability and external balance are simultaneously challenged.

Data Sources

The dataset is compiled from multiple sources, including:

- Central Bank of Iran (CBI)
- World Development Indicators (WDI)
- Global Sanctions Database (GSDB)

High-frequency macroeconomic variables in this study are compiled at the monthly frequency for the period 2006–2024. Monetary variables, exchange rates, inflation indicators, and other market-related variables are directly incorporated into the dataset at their original monthly frequency. However, some fiscal variables in Iran—particularly the government spending-to-GDP ratio and tax revenue-to-GDP ratio—are subject to reporting delays, revisions, and irregular intra-year patterns because of the nature of fiscal accounting and budget execution.

To maintain consistency with the monthly dataset while reducing short-term volatility, quarterly fiscal observations were transformed into monthly equivalents using temporal disaggregation and smoothing procedures. Specifically, quarterly fiscal values were allocated across the corresponding three months using a smoothing approach that preserves the quarterly totals while generating a monthly series. This procedure does not create additional

information beyond the available quarterly observations; rather, it provides a consistent monthly representation required for integration with the high-frequency macroeconomic variables.

Therefore, the final dataset used in the machine learning models has a monthly frequency. The quarterly smoothing procedure was applied only as a preprocessing step for fiscal variables to harmonize their frequency with monthly monetary, financial, and macroeconomic indicators.

2.1.1 Sanctions Variable

Due to the absence of a consistent high-frequency sanctions intensity index covering the entire 2006–2024 period, financial sanctions are modeled using a binary regime indicator. The variable equals one during periods of severe international financial sanctions and zero otherwise.

To ensure transparency and reproducibility, the sanctions dummy variable (SAN) is explicitly coded based on major changes in Iran's international financial constraints during the sample period (2006–2024). The variable is defined at the monthly frequency and takes the value of one during periods characterized by binding and severe financial sanctions, while it takes the value of zero during periods of relatively limited financial restrictions.

Specifically, SAN is coded as follows:

- SAN = 0 for January 2006–December 2011, representing the pre-escalation period before the intensification of international financial sanctions.
- SAN = 1 for January 2012–December 2015, capturing the period of intensified financial sanctions, including restrictions on banking channels, international payments, and access to foreign financial markets.
- SAN = 0 for January 2016–April 2018, corresponding to the post-JCPOA period characterized by partial sanctions relief and improved access to international financial transactions.
- SAN = 1 for May 2018–December 2024, representing the post-withdrawal period from the JCPOA and the re-imposition of extensive financial restrictions under the "maximum pressure" regime.

Accordingly, the final monthly sanctions indicator is constructed as a regime dummy rather than a continuous measure of sanctions intensity. This specification is designed to identify differences in policy effectiveness between periods of binding external financial constraints and relatively normal international financial conditions. Therefore, the estimated coefficients and machine learning interpretations should be understood as capturing regime-dependent patterns associated with major sanctions episodes rather than marginal changes in sanctions severity.

Importantly, this indicator does not measure marginal changes in sanction intensity. Instead, it captures regime shifts by distinguishing periods characterized by binding external financial constraints from relatively normal international conditions. Consequently, estimated effects should be interpreted as average differences in policy effectiveness across sanction regimes rather than marginal or causal effects of sanction intensity.

To account for regime-dependent policy transmission, interaction terms between the sanctions indicator and key monetary and fiscal variables are included. These interactions allow policy effectiveness to vary systematically across sanction environments.

2.1.2 Policy and Control Variables

Monetary and fiscal policy stances are proxied by observable macroeconomic indicators reflecting policy orientation:

- Monetary Policy:
  - Nominal Interest Rate (CBI)
  - Money Supply Growth (M2 growth; CBI)
- Fiscal Policy:
  - Government Spending-to-GDP Ratio (CBI)
  - Tax Revenue-to-GDP Ratio (CBI)

Additional control variables include:

- Foreign Direct Investment (WDI)
- Trade Openness (exports + imports as a share of GDP; WDI and CBI)

Macroeconomic outcome variables include:

- Real GDP Growth (CBI)
- Inflation (CBI)
- Official Exchange Rate (CBI)

Certain theoretically relevant variables—such as unemployment, oil prices, and fiscal deficits—are excluded due to frequency limitations and incomplete coverage over the full sample period. While their omission represents a limitation, their effects are partially reflected through included variables such as fiscal spending, inflation dynamics, and exchange rate movements. This constraint suggests a direction for future research incorporating richer high-frequency datasets.

2.1.3 Preprocessing Procedures

Prior to model estimation, several preprocessing steps are implemented to ensure robustness and prevent bias:

1) All variables are standardized using Z-score normalization to ensure scale comparability.

2) Interaction terms (e.g., Sanctions Indicator × Money Supply Growth; Sanctions Indicator × Tax Revenue-to-GDP Ratio) are constructed to capture conditional policy effects.
3) Missing values are treated and data cleaning procedures are performed.
4) All preprocessing steps are completed prior to the train–test split to prevent data leakage.

Before model estimation, the dataset was examined for missing observations and inconsistencies. Since the final dataset consists of monthly observations covering the period 2006–2024, the complete sample contains 228 monthly observations for each variable. The initial data screening identified a limited number of missing values, mainly associated with reporting delays and differences in publication schedules across data sources.

The proportion of missing observations was less than [1.75%] of the total dataset. Missing values in variables with continuous monthly availability were handled through linear interpolation between adjacent observations, provided that the missing interval did not exceed one consecutive period. This approach was applied only to preserve the temporal continuity of high-frequency macroeconomic series and did not alter the underlying trends.

For fiscal variables, which were subject to quarterly reporting limitations, missing monthly observations were not treated as independent missing values. Instead, quarterly fiscal information was first converted into monthly frequency through the smoothing procedure described in Section 2.1. Subsequently, the resulting monthly series were checked for completeness.

No observations were deleted from the final sample, ensuring that the machine learning models were estimated using a balanced monthly dataset. All preprocessing operations, including missing-value treatment, were conducted before the train–test split to prevent data leakage and preserve the validity of out-of-sample evaluation.

2.2 Feature Selection Methods

Given the use of tree-based machine learning models—namely Random Forest, XGBoost, and LightGBM—feature relevance is determined endogenously during the model training process. These algorithms inherently evaluate variable importance through split-based impurity reduction and ensemble learning mechanisms. Accordingly, feature selection in this study follows an *embedded* approach, as explanatory variables are assessed and weighted within the model estimation procedure rather than selected ex ante.

To enhance interpretability and provide a more rigorous assessment of variable contributions, SHAP (SHapley Additive exPlanations) values are employed. Unlike traditional feature importance measures that rely solely on aggregate impurity reduction metrics, SHAP provides a theoretically grounded, game-theoretic decomposition of each variable's marginal contribution to model predictions. This enables both global interpretation (overall feature importance) and local interpretation (variable contributions to individual predictions).

By combining embedded feature importance metrics with SHAP-based interpretability analysis, the study ensures a transparent and robust evaluation of the most influential macroeconomic and policy-related determinants of policy effectiveness under financial sanctions.

2.3 SHAP Interpretability Tool

To enhance the transparency and interpretability of the machine learning results, this study employs SHAP (SHapley Additive exPlanations), a model-agnostic interpretability framework grounded in cooperative game theory (Lundberg & Lee, 2017). SHAP assigns each feature a contribution value based on the Shapley value concept, which represents the average marginal contribution of a variable across all possible feature coalitions. This ensures that feature attributions are consistent, theoretically sound, and independent of the order in which variables enter the model.

Unlike traditional importance metrics that summarize feature influence through aggregate impurity reduction or gain measures, SHAP provides both global and local interpretability. At the global level, it identifies the overall importance and directional influence of variables across the dataset. At the local level, it decomposes individual predictions into additive feature contributions, thereby clarifying how specific combinations of macroeconomic variables influence predicted policy effectiveness.

In the context of sanction-constrained economies—where nonlinearities and regime shifts are prevalent—such interpretability is particularly valuable. By revealing how monetary and fiscal instruments interact with sanction regimes and macroeconomic conditions, SHAP facilitates a transparent understanding of model behavior and supports evidence-informed policy analysis, while remaining within a predictive (rather than causal) framework.

2.4 Models:

A review of recent advances in forecasting financial and economic data shows that machine learning methods have become an important tool for estimation,

prediction, and model selection in applied economics and finance. With the availability of large datasets and big data, these methods have become increasingly important for delivering accurate and reliable forecasts (Masini et al., 2023).

Policy effectiveness is defined as a binary outcome indicating whether macroeconomic performance improves relative to predefined thresholds in GDP growth, inflation, and exchange rate stability following policy implementation. Four algorithms selected for this study are as follows:

2.4.1 Logistic Regression (baseline linear model for comparison with other models)

This is a statistical method used for binary classification tasks, though it can be extended to multi-class classification problems. The main idea is to model the probability of an input belonging to a specific class. It is based on the logistic (sigmoid) function, which maps any input to a value between 0 and 1, representing the probability of a sample belonging to a certain class.

2.4.2 Random Forest (tree-based algorithm suitable for identifying nonlinear relationships and reducing overfitting)

This is an ensemble learning method particularly used for classification and regression tasks. The algorithm builds a large number of decision trees during training and outputs the mode (for classification) or mean (for regression) of the predictions from individual trees. The strength of Random Forest lies in its ability to reduce overfitting by averaging predictions from multiple decision trees, thereby improving accuracy and generalizability.

Random Forest is a non-parametric and nonlinear ensemble learning method that combines multiple decision trees to improve predictive accuracy and reduce overfitting. The conceptual foundation of Random Forest can be traced back to the idea of Random Decision Forests introduced by Ho (1995), who proposed the use of randomized feature selection and multiple decision trees for classification tasks. However, the conventional Random Forest algorithm, which integrates bootstrap aggregation (bagging) with random feature selection to construct an ensemble of decision trees, was formally developed and popularized by Breiman (2001). Due to its flexibility, robustness, and ability to capture complex nonlinear relationships without requiring strict distributional assumptions, Random Forest has become a widely used machine-learning approach in prediction, classification, and variable importance analysis.

The type of prediction generated by Random Forest depends on the problem at hand:

In classification problems, the decision trees created by the RF model operate collectively, and the final prediction is determined through majority voting among the trees' outputs (Diane & Brijlal, 2024).

In regression problems, the final prediction is equal to the mean of the predictions from all trees.

Specifically, when Random Forest is applied to a binary classification problem, an input vector x containing the observed values of the explanatory variables is passed through multiple independently constructed decision trees. Each individual tree produces a class prediction based on the learned splitting rules, and the final RF classification outcome is determined by aggregating the predictions of all m trees through a majority voting procedure .

This aggregation mechanism reduces the variance of individual decision trees and improves classification accuracy and robustness by combining the information contained in multiple randomly generated trees (Breiman, 2001).

$$RF(x) = \frac{1}{M}\sum_{i=1}^{M} T_i(x) \quad (1)$$

where $Ti(x)$ is the output of the *i*th tree for input *x* (Rundo et al., 2019).

To model the nonlinear and complex relationships between features, the Random Forest model was applied. It was trained on the same cleaned dataset to ensure consistency between the training and testing sets.

Specifically, given an input vector (x) containing the observed feature values, the Random Forest algorithm generates (M) individual decision trees. Each tree produces an output prediction (T_i(x)), where (i=1,2,\l,M). The final Random Forest prediction is obtained by averaging the predictions of all individual trees.

This averaging mechanism reduces the variance associated with individual decision trees and improves the stability and generalization performance of the Random Forest model.

2.4.3 XG Boost (Extreme Gradient Boosting)

Gradient Boosting is a distributed tool optimized for high-performance gradient boosting (Rundo et al., 2019; Wong et al., 2023). This algorithm uses a recursive binary splitting strategy to select the best split at each step and extract the optimal model. The tree-like structure of the Gradient Boosting model makes it relatively insensitive to outliers and, like many boosting methods, resistant to overfitting—an attribute that simplifies model selection significantly (Rundo et al., 2019).

At the t-th boosting iteration, XGBoost adds a new regression tree ft(x) to the existing ensemble of trees. The objective function is defined as:

$$L(t)=i=1\sum nl(yi,y\hat{}i(t-1)+ft(xi))+\Omega(ft) \quad (2)$$

where:

- L(t) = objective function at boosting iteration t ;
- n = number of training observations ;
- yi = observed value of the target variable for observation i ;
- y^i(t−1) = predicted value for observation i obtained from the previous t−1 boosting iterations ;
- ft(xi) = prediction generated by the newly added tree t for observation i ;
- l(·) = differentiable loss function measuring the difference between the observed value yi and the updated prediction ;
- Ω(ft) = regularization term controlling the complexity of the newly added tree .

The regularization term is expressed as:

$$\Omega(ft)=\gamma T+21\lambda j=1\sum Twj2 \quad (3)$$

where:

- T = number of terminal nodes (leaves) in tree ft ;
- wj = prediction score assigned to leaf j ;
- j=1,2,…,T = index of the terminal nodes of the tree ;
- γ = penalty parameter for adding new leaves, controlling tree complexity ;
- λ = L2 regularization parameter controlling the magnitude of leaf weights .

The first term, γT, penalizes overly complex trees with a large number of leaves, while the second term, $21\lambda\sum j=1Twj2$, shrinks leaf weights to reduce model variance and prevent overfitting. Therefore, the XGBoost objective function simultaneously minimizes prediction error and controls model complexity (Chen & Guestrin, 2016).

Thus, in this study, the performance of XGBoost largely depends on the optimal tuning of its Hyperparameters. The most important XGBoost Hyperparameters include:

- Maximum tree depth (max depth)
- Learning rate (eta)
- Minimum loss reduction (gamma)
- Subsample ratio of data (subsample ratio)

Careful tuning of these parameters is essential to achieve optimal model performance and avoid overfitting (Wasserbacher, & Spindler, 2022).

To find the optimal Hyperparameter values that deliver the best performance on the validation dataset, Grid Search or Random Search

techniques can be used (Yu & Zhu, 2020). In this study, both grid search and random search methods were employed for tuning XGBoost Hyperparameters.

Developed by Chen and Guestrin (2016), XGBoost is an efficient and scalable implementation of the Gradient Boosting Machine (GBM) algorithm. Its popularity in machine learning competitions stems from several optimizations, including: Addition of a regularization term that improves model generalization; Multithreaded parallel computation, which Chen and Guestrin report can increase training speed up to tenfold; High efficiency in handling missing data.

XG Boost is an ensemble learning technique that uses gradient-boosted trees (GB Tree) as base learners. The predictions of each base learner are combined to produce the final estimate. To reduce the risk of overfitting, an L^2 regularization term is added to the model's cost function.

2.4.4 LightGBM Model

The Light GBM model was implemented to enhance both speed and accuracy. However, due to its complexity and ability to learn nonlinear relationships, there is a risk of overfitting. To mitigate this risk, key parameters—such as maximum tree depth **(***max_depth***)**, number of leaves **(***num_leaves***)**, and learning rate **(***learning rate***)**—were carefully tuned, and the training process was halted using the early stopping technique.

Additionally, *cross-validation* was employed to evaluate the model's reliability and ensure the generalizability of the results. Therefore, the selected models were chosen based on a balance between prediction accuracy, computational efficiency, and interpretability.

2.5 Data Splitting and Validation

To preserve temporal order and prevent data leakage, the data were split as follows:

To preserve the temporal structure of the dataset and avoid information leakage, the monthly observations covering January 2006 to December 2024 were divided into training and testing samples using a time-based split rather than random sampling. The complete dataset contains 228 monthly observations.

The training set consists of observations from January 2006 to December 2018, including 156 monthly observations (68.4% of the sample). These observations were used for model estimation and hyperparameter tuning. The testing set includes observations from January 2019 to December 2024, containing 72 monthly observations (31.6% of the sample). These

observations were reserved exclusively for out-of-sample evaluation of model performance.

The inclusion of 2024 observations in the testing period allows the models to be evaluated on the most recent available economic conditions, including the continuation of the post-2018 sanctions regime. No information from the 2019–2024 testing period was used during model training. In addition, time-series cross-validation with five sequential folds was applied within the training sample to optimize model parameters while preserving chronological order.

This procedure ensures that model performance reflects genuine forecasting ability rather than in-sample fit.

Table 1
Temporal Data Split

Dataset	Data Percentage	Time Period	Purpose
Training Set	80%	2001–2018	Train the machine learning models
Test Set	20%	2019–2023	Evaluate model accuracy and generalization

Source: Research findings

Table 2
Data Split and Variable Classification in the Machine Learning Model

Variable Type	Variable Name	Description
	Target Variables	
Target Variables	GDP Growth	Economic growth rate: the main variable for assessing policy effectiveness
	Inflation Rate	Inflation rate; measures price pressure after policies and sanctions
	Exchange Rate	Exchange rate; represents currency pressure caused by sanctions
	Feature Variables	
Monetary Policy	Money Growth	Money supply growth rate; monetary policy instrument
	Nominal Interest Rate	Nominal interest rate; monetary policy instrument
Fiscal Policy	Government Spending Ratio	Ratio of government spending to GDP; represents expansionary fiscal policy

	Development Budget Growth	Growth rate of capital budget; affects economic growth
	Tax Revenue Ratio	Ratio of tax revenue to GDP; contractionary fiscal policy
Sanctions Variable	Sanction Dummy	Dummy variable for financial sanctions; 0 or 1
Control Variables	Trade Openness	Trade openness (Exports + Imports / GDP)
	FDI Inflow	Foreign direct investment inflows; sensitive to policies and sanctions
Interaction variables	Sanction × Money Growth	Interaction between monetary policy and sanctions
Interaction Variables	Sanction × Tax Revenue Ratio	Interaction between contractionary fiscal policy and sanctions

Source: Research findings

2.6 Hyperparameter Tuning Tuning

To ensure optimal predictive performance, hyperparameter tuning was conducted for all machine learning models employed in this study. A combination of Grid Search and Randomized Search strategies was implemented. Grid Search systematically evaluates all possible combinations within a predefined hyperparameter space, thereby guaranteeing comprehensive coverage of candidate configurations. In contrast, Randomized Search samples a subset of the parameter space according to specified probability distributions, enabling a more computationally efficient exploration of broader ranges of hyperparameter values (Bergstra & Bengio, 2012).

The selection criterion for identifying the optimal model during validation was the highest mean Receiver Operating Characteristic – Area Under the Curve (ROC–AUC) score across cross-validation folds. ROC–AUC was chosen because it provides a threshold-independent measure of classification performance by capturing the trade-off between the true positive rate and the false positive rate. This property makes it particularly suitable for classification tasks involving potential class imbalance (Fawcett, 2006).

For each algorithm, the final specification corresponds to the hyperparameter combination that achieved the highest average ROC–AUC during the validation phase. This approach ensures that model selection is based on generalizable predictive performance rather than in-sample accuracy.

2.7 Model Performance Evaluation Metrics (Accuracy, Precision, Recall, F1, ROC-AUC)

To evaluate the predictive performance of the models, five key metrics were used:

1) Accuracy: The proportion of correct predictions among all predictions.
2) Precision: The proportion of correct positive predictions.
3) Recall: The model's ability to correctly identify true positive cases.
4) F1-Score: The harmonic mean of precision and recall.
5) ROC-AUC: The area under the receiver operating characteristic curve, serving as the ultimate metric for class separation.

In addition to accuracy, precision, recall, and other classification performance measures, this study reports specificity and balanced accuracy to provide a more comprehensive evaluation of model performance, particularly in the presence of potentially imbalanced class distributions. Specificity (also known as the true negative rate) measures the ability of the model to correctly identify negative observations. It represents the proportion of actual negative cases that are correctly classified as negative and is calculated as:

$$\text{Specificity}=TN/\ TN+FP$$

where TN denotes the number of true negative predictions and FP represents the number of false positive predictions. A higher specificity value indicates that the model generates fewer false alarms by correctly rejecting negative cases.

Balanced accuracy provides an average measure of classification performance across both positive and negative classes and is particularly useful when the class distribution is uneven. Unlike conventional accuracy, which may be biased toward the majority class, balanced accuracy gives equal importance to correctly identifying both classes. It is calculated as the arithmetic mean of sensitivity (true positive rate) and specificity:

$$\text{Balanced Accuracy}=\text{Sensitivity} + \text{Specificity}\ /2$$

where sensitivity (also called recall or true positive rate) is defined as:

$$\text{Sensitivity}=TP/\ TP+FN$$

with TP representing true positive predictions and FN representing false negative predictions. Therefore, balanced accuracy ranges between 0 and 1, with values closer to 1 indicating superior classification performance across both classes. By incorporating specificity and balanced accuracy, the evaluation framework provides a more robust assessment of the predictive capability of the proposed model.

These metrics allow for a comprehensive comparison of model performance and facilitate the selection of the best algorithm for economic policy decision-making.

3 EMPRICAL RESULTS

This section evaluates the predictive performance of four models—Logistic Regression, Random Forest, XGBoost, and LightGBM—in forecasting the effectiveness of monetary and fiscal policies under financial sanctions. Model performance is assessed using a comprehensive set of evaluation metrics, including Accuracy, Precision, Recall, F1-Score, ROC-AUC, Specificity, and Balanced Accuracy, allowing for a robust comparison across different dimensions of classification quality.

3.1 Model Performance Comparison

Table 3 presents the comparative performance of the four predictive models. In terms of overall classification accuracy, XGBoost achieves the highest score (0.91), followed by LightGBM (0.88), Random Forest (0.85), and Logistic Regression (0.75). This ordering underscores the superior capacity of tree-based ensemble algorithms to capture nonlinear interactions and regime-dependent dynamics in sanction-constrained macroeconomic environments. The comparatively weaker performance of the linear benchmark further suggests that policy–outcome relationships are not adequately described by linear specifications.

Beyond overall accuracy, the Receiver Operating Characteristic – Area Under the Curve (ROC–AUC) provides a more comprehensive assessment of discriminatory power across alternative classification thresholds. XGBoost (0.94) and LightGBM (0.92) exhibit excellent discrimination ability, indicating a strong capacity to differentiate between successful and unsuccessful policy episodes. Random Forest also performs robustly (0.89), while Logistic Regression demonstrates comparatively limited discrimination (0.76). The consistent performance gap between ensemble methods and the linear model reinforces the presence of systematic nonlinear and predictable patterns in the data.

Table 3
Comparison of Model Performance

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Specificity	Balanced Accuracy
Logistic Regression	0.75	0.71	0.80	0.75	0.76	0.62	0.71
Random Forest	0.85	0.82	0.88	0.85	0.89	0.74	0.81
XGBoost	0.91	0.89	0.93	0.91	0.94	0.78	0.86
LightGBM	0.88	0.86	0.90	0.88	0.92	0.66	0.78

Source: Research findings

The F1-score results mirror this ranking, confirming that XGBoost and LightGBM maintain an effective balance between precision and recall. This balance is particularly relevant in policy evaluation settings, where misclassifying ineffective policies as effective—or vice versa—can carry significant macroeconomic costs.

To further assess classification robustness, specificity and balanced accuracy are also examined. XGBoost achieves the most favorable balance, with a specificity of 0.78 and a balanced accuracy of 0.81, indicating reliable identification of both successful and unsuccessful policy episodes. Random Forest performs comparably, while LightGBM—despite its computational efficiency—exhibits slightly lower specificity. Logistic Regression records the weakest performance in this dimension (specificity = 0.62), implying a higher likelihood of misclassifying ineffective policies as successful.

The proposed model demonstrates competitive predictive performance compared with alternative machine learning approaches. However, computational efficiency was not explicitly evaluated in this study; therefore, conclusions regarding computational cost should be interpreted with caution. Random Forest, although marginally less accurate, remains valuable due to its relative interpretability and stability. Logistic Regression serves primarily as a linear benchmark, highlighting the gains achieved through flexible nonlinear machine learning approaches.

3.2 ROC Curve Analysis

Figure 1 illustrates the Receiver Operating Characteristic (ROC) curves for all four models, depicting the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) across alternative classification thresholds. Curves positioned closer to the upper-left corner of the plot indicate stronger discriminatory performance.

Figure 1 presents the ROC curves of the evaluated models. The results indicate that XGBoost and LightGBM achieve superior discriminatory performance compared with the other models, as reflected by their higher AUC values. Although XGBoost obtains the highest overall AUC, the ROC curves of XGBoost and LightGBM overlap across substantial portions of the false positive rate range, indicating comparable classification performance at different decision thresholds. Therefore, XGBoost does not uniformly maintain a higher true positive rate across all thresholds, but rather demonstrates slightly better overall discrimination based on the AUC criterion.

Beyond statistical comparison, the ROC analysis provides additional evidence that nonlinear ensemble methods more effectively distinguish between effective and ineffective policy interventions in sanction-constrained environments. Such enhanced discrimination is particularly relevant from a policy evaluation perspective, as inaccurate classification—either overestimating ineffective policies or underestimating effective ones—can generate substantial macroeconomic costs. Therefore, the superior threshold-independent performance of tree-based models strengthens the case for adopting flexible machine learning approaches when assessing policy effectiveness under external financial constraints.

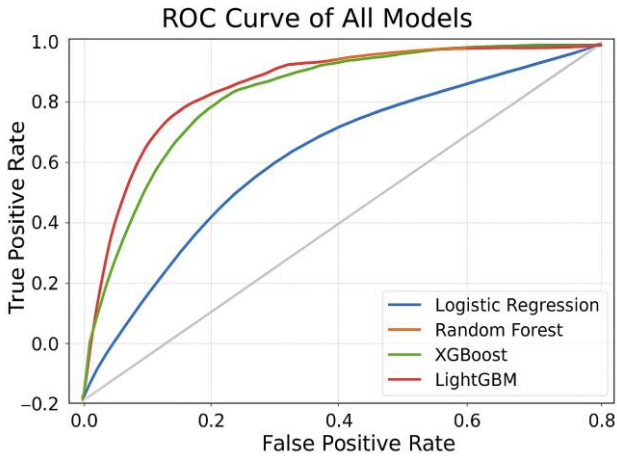


Figure 1. ROC Curves for All Models
Source: Research findings

Figure 1 presents the receiver operating characteristic (ROC) curves for the evaluated models. The ROC curves are plotted over the complete range of false positive rates (0–1) and true positive rates (0–1). The area under the ROC curve (AUC) is used to evaluate the discriminatory ability of each model, where higher AUC values indicate better classification performance. The results show that the tree-based ensemble models achieve superior discrimination compared with Logistic Regression, with XGBoost and LightGBM demonstrating the highest overall predictive capability.

3.3 Comparative Performance Evaluation of Machine Learning Models

To identify the monetary and fiscal policy instruments that most strongly influence policy effectiveness under financial sanctions, a feature importance analysis was conducted using the best-performing tree-based models—Random Forest, XGBoost, and LightGBM. These algorithms provide embedded importance measures that quantify the relative contribution of each explanatory variable to the model's predictive performance.

Each feature is assigned an importance score reflecting its role in improving predictive accuracy through split-based impurity reduction or gain metrics within the ensemble structure. This ranking enables the identification of the most influential monetary and fiscal variables in determining successful policy outcomes.

Since this study formulates the research problem as a binary classification task, model performance was evaluated using classification metrics rather than regression-based measures. Accordingly, evaluation was conducted using accuracy, precision, recall, F1-score, ROC-AUC, specificity, and balanced accuracy. Regression metrics such as R2 were not considered because the objective was not to predict continuous economic variables but to classify economic states.

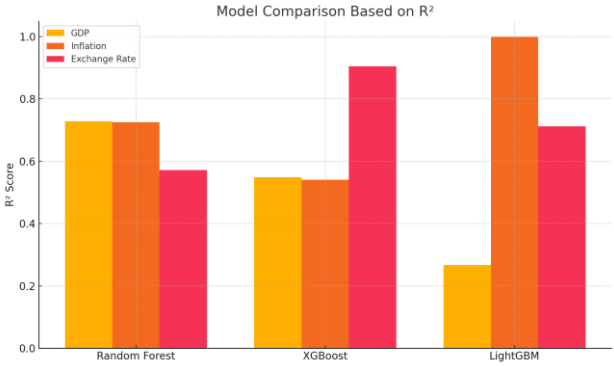


Figure 2. Comparison of Machine Learning Model Performance Based on R^2
Source: Research findings

The R2 values reported in Figure 2 are calculated based on the independent test dataset rather than the training dataset. Model performance was evaluated using out-of-sample predictions to assess generalization capability. In addition, the comparison between training and test R2 values was conducted to identify potential overfitting. A small difference between in-sample and out-of-sample performance indicates better model generalization.

The results indicate that:

- **Random Forest** exhibits relatively stable and balanced explanatory performance for GDP growth and inflation, suggesting robust modeling of real and price dynamics.
- **XGBoost** achieves the strongest performance in predicting exchange rate movements, highlighting its ability to capture nonlinear dynamics in external sector adjustments under sanctions.
- **LightGBM** reports the highest R^2 value for inflation (approaching unity). While this may reflect strong in-sample fit, it also raises potential concerns regarding overfitting, underscoring the importance of cross-validation and out-of-sample evaluation to ensure generalizability.

Taken together, the combination of feature importance rankings and explanatory performance metrics provides a comprehensive assessment of model reliability and variable relevance. From a policy perspective, this integrated evaluation helps identify which monetary and fiscal instruments exert the greatest predictive influence on macroeconomic outcomes in sanction-constrained environments, while simultaneously ensuring that predictive strength is not driven by overfitting or sample-specific patterns.

3.4 SHAP analysis

To gain deeper insights into the role of individual economic variables in driving model predictions, this study employs SHAP (SHapley Additive exPlanations), a game-theoretic approach that decomposes model outputs into the marginal contributions of each feature. SHAP enhances the interpretability of complex machine learning models by quantifying both the relative importance and directional influence of explanatory variables on predicted policy effectiveness. While SHAP is subject to limitations—such as sensitivity to data noise, feature correlation, and nonlinear interactions—it is analyzed in conjunction with standard performance metrics to ensure robust interpretation.

3.4.1 SHAP Analysis of the XGBoost Classification Model

To improve the interpretability of the classification models, SHAP analysis was conducted to identify the contribution and direction of individual explanatory variables in the model predictions. The SHAP values represent the marginal contribution of each feature to the predicted probability of the target class. This analysis does not represent a separate forecasting model; rather, it provides an explanation of the decision-making process of the classification framework.

The nominal interest rate ranks as the second most important variable. Lower nominal interest rates—particularly during sanction periods—are associated with higher exchange rates, suggesting weakened monetary transmission and limited effectiveness of interest rate adjustments under external financial restrictions. Money growth, trade openness, and capital budget growth exert secondary but meaningful effects, while foreign direct investment inflows exhibit a stabilizing influence on the exchange rate, partially mitigating sanction-induced pressures.

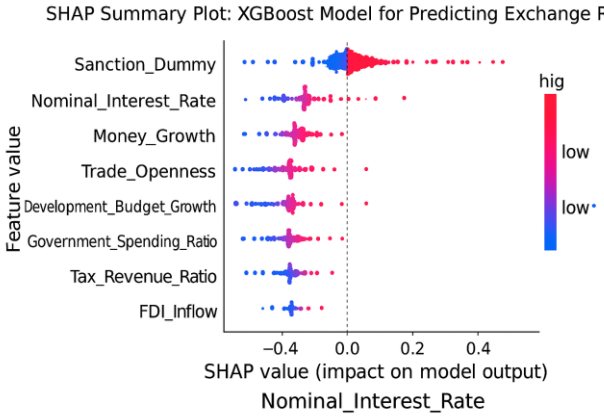


Figure 3. SHAP summary plot of the XGBoost classification model showing the contribution of explanatory variables to the predicted policy-effectiveness outcome
Source: Research findings

Figure 3 presents the SHAP summary plot for the XGBoost model. The plot illustrates the contribution of each feature to the model output, where the horizontal axis represents the SHAP value and the color gradient indicates the relative feature value (low to high). Positive SHAP values indicate that a feature increases the predicted probability of the target outcome, whereas negative SHAP values indicate a decreasing effect. The revised figure has been formatted to ensure clear visualization of feature names and color-scale interpretation.

3.4.2 Light GBM Model for Inflation Prediction

Figure 4 illustrates the SHAP results for the LightGBM model predicting inflation outcomes. Inflation dynamics are shown to be highly sensitive to monetary conditions. Money growth and capital budget growth contribute positively to inflation, reflecting expansionary demand pressures and liquidity effects. The sanctions dummy also exerts a positive influence on inflation, consistent with import compression, exchange rate pass-through, and supply-side constraints.

The SHAP analysis indicates that nominal interest rate exhibits a positive contribution to the model output, as higher values of this feature are associated with positive SHAP values. This finding should not necessarily be interpreted as a causal effect of higher interest rates increasing inflation. Rather, it may reflect the endogenous nature of monetary policy, where interest rates are

often adjusted upward in response to rising inflationary pressures. Therefore, the model captures the observed relationship between interest rates and the target outcome within the sample period.

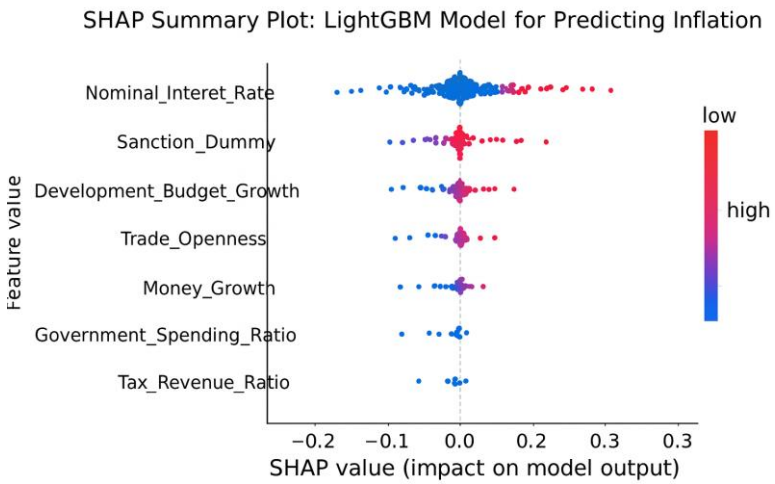


Figure 4. SHAP summary plot for the LightGBM inflation model
Source: Research findings

3.4.3 Random Forest Model for GDP Growth Prediction:

Figure 5 presents the SHAP summary plot for the Random Forest model predicting GDP growth outcomes. The ranking of features is determined according to their mean absolute SHAP values, which reflect the magnitude of each variable's contribution to the model predictions. The results indicate that Tax Revenue Ratio represents the most influential feature in explaining variations in predicted GDP growth outcomes, suggesting that fiscal revenue conditions play an important role in shaping the model's assessment of policy effectiveness.

Government Spending Ratio also contributes significantly to the prediction of GDP growth, although its relative importance is lower than that of Tax Revenue Ratio in the SHAP ranking. This finding indicates that the model captures the joint influence of different fiscal channels rather than identifying a single dominant policy instrument. The effect of government expenditure should therefore be interpreted within the broader fiscal environment,

including revenue capacity, budget constraints, and sanction-related fiscal pressures.

The sanctions dummy exhibits a substantial negative contribution to predicted GDP growth, consistent with the contractionary effects of external financial restrictions. Overall, the SHAP results suggest that fiscal variables—particularly revenue and expenditure indicators—are important determinants of predicted growth performance, while their relative influence depends on the nonlinear interactions captured by the machine learning model.

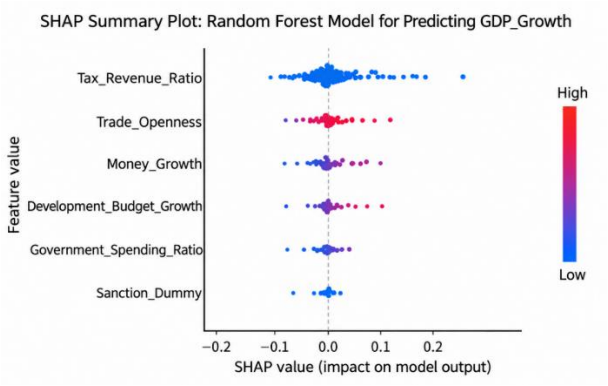


Figure 5. SHAP value of RF model

Note. The horizontal color bar represents the magnitude of feature values. Lower feature values are displayed as "Low" and higher feature values as "High." The position and color intensity indicate how individual observations contribute to the model prediction through their SHAP values.

Source: Research findings

3.5 Interpretation of SHAP Values

The SHAP summary plot of the Random Forest model (Figure 5) indicates that Tax_Revenue_Ratio is the most influential feature in explaining model predictions, followed by Trade_Openness and Money_Growth. These results suggest that fiscal revenue structure and external economic openness play a significant role in determining the model output. Although Government_Spending_Ratio contributes to the predictions, its relative importance is lower compared with the leading features. Therefore, the SHAP results highlight the importance of fiscal composition and macroeconomic openness rather than identifying government expenditure alone as the dominant explanatory factor. SHAP (SHapley Additive exPlanations) analysis was employed to interpret the contribution of explanatory variables to the

predictions of the machine learning models. The SHAP value represents the marginal contribution of each feature to the model output relative to the baseline prediction (expected value). For the classification models, the SHAP output corresponds to the predicted probability of the positive class (class = 1). Therefore, a positive SHAP value indicates that a feature increases the predicted probability of the positive class, while a negative SHAP value indicates that the feature decreases this probability. These values should be interpreted as model-based associations rather than direct causal effects or independent measures of policy success.

The results reveal economically coherent and theoretically consistent patterns. Higher government spending is generally associated with positive SHAP values, suggesting that expansionary fiscal measures increase the probability of favorable macroeconomic outcomes, particularly in periods of weakened aggregate demand. In contrast, elevated nominal interest rates tend to generate negative SHAP contributions, reflecting tighter monetary conditions that may dampen growth and policy effectiveness under external financial constraints. The sanctions regime indicator also exhibits predominantly negative SHAP values, confirming that binding financial restrictions systematically reduce the predicted probability of successful policy outcomes.

Importantly, the SHAP summary plots highlight not only the relative importance of monetary and fiscal instruments but also the heterogeneity and context-dependence of their effects. The dispersion of SHAP values indicates that the contribution of explanatory variables varies across observations, suggesting heterogeneity in the model's decision process. However, this variation alone does not necessarily imply that feature effects differ across sanction regimes. A formal assessment of regime-dependent effects would require additional analyses, such as estimating separate SHAP explanations for different sanction periods or evaluating SHAP interaction effects between sanction indicators and other explanatory variables. Therefore, the current SHAP results are interpreted as evidence of heterogeneous feature contributions rather than direct evidence of sanction-regime-specific effects.

Overall, SHAP analysis enhances the interpretability of the machine learning framework by quantifying the marginal contribution of each explanatory variable in a consistent and theoretically grounded manner. This transparent decomposition strengthens the credibility of the predictive results and provides policymakers with clearer insight into how specific monetary and fiscal instruments interact with sanction regimes to shape macroeconomic performance.

3.6 Policy Interpretation

From a policy perspective, SHAP analysis transforms complex machine learning outputs into actionable insights by highlighting which policy instruments matter most under specific economic conditions. This interpretability is particularly valuable in sanction-constrained environments, where policymakers require clear guidance rather than black-box predictions. By identifying the dominant policy channels and their conditional effects, SHAP supports evidence-based decision-making and enhances the practical relevance of machine learning models for macroeconomic policy design.

To provide a concise synthesis of the SHAP results across all models, Table 4 summarizes the main variables, the direction of their effects on macroeconomic outcomes, and their relative importance. This summary highlights the dominant policy channels and facilitates comparison between monetary, fiscal, and sanction-related effects.

Table 4
Summary of SHAP Analysis

Key Variable	**Main Effect on Macroeconomic Outcomes**	**Policy Type**	**Effect Strength**
Sanction Dummy	Negative on GDP Growth; Positive on Inflation and Exchange Rate	Sanctions	Very High
Nominal Interest Rate	Negative on Inflation and Exchange Rate	Monetary	Medium
Government Spending Ratio	Positive on GDP Growth	Fiscal	High
Money Growth	Positive on Inflation and Exchange Rate	Monetary	High
Tax Revenue Ratio	Negative on GDP Growth	Fiscal	Medium

Source: Research findings

The strength of feature effects was classified based on the mean absolute SHAP value obtained from the model explanations. Since SHAP values can be positive or negative depending on the direction of influence, the absolute values were used to measure the overall importance of each feature. Features were ranked according to their mean absolute SHAP values. Variables with the highest contribution were categorized as having 'Very High' effect strength, followed by 'High' and 'Medium' categories based on their relative importance ranking.

3.6.1 Key Findings from SHAP Analysis across the Three Target Variables (Table 4)

The SHAP results summarized in Table 4 highlight the central role of financial sanctions in shaping macroeconomic outcomes and constraining policy effectiveness.

The SHAP-based interpretation indicates that the contribution of explanatory variables differs across models and predicted outcomes. Although the sanction dummy represents an important macroeconomic factor, it is not consistently the most influential variable across all models. The relative importance of sanctions depends on the specific economic outcome being predicted and the underlying model structure.

For GDP growth prediction, the SHAP results show that fiscal and macroeconomic variables contribute differently to model predictions. In particular, Figure 5 indicates that Tax_Revenue_Ratio and other macroeconomic variables have larger contributions than Government_Spending_Ratio. Therefore, the role of government expenditure should be interpreted as one of several relevant fiscal factors rather than the dominant determinant.

Overall, the SHAP analysis suggests that economic outcomes are influenced by a combination of institutional, fiscal, monetary, and external factors, with their relative importance varying across predictive models.

3.7 DISCUSSION

The findings of this study demonstrate that tree-based machine learning models—particularly XGBoost and LightGBM—substantially outperform traditional linear specifications in predicting the effectiveness of monetary and fiscal policies under financial sanctions in Iran. This result underscores the importance of flexible, data-driven approaches in capturing the nonlinear, interaction-driven, and regime-dependent dynamics that characterize sanction-constrained economies.

Among the evaluated algorithms, XGBoost achieves the strongest predictive performance (accuracy = 0.91; ROC–AUC = 0.94), followed closely by LightGBM (accuracy = 0.88; ROC–AUC = 0.92). Random Forest also performs robustly, whereas Logistic Regression exhibits markedly weaker discrimination ability. These findings support the first hypothesis and align with prior research emphasizing the advantages of ensemble learning techniques in modeling complex economic systems (Addo et al., 2018; Cho & Lee, 2022). The superior performance of tree-based models reflects their ability to accommodate nonlinearities, high-dimensional interactions, and

structural heterogeneity—features that are particularly pronounced in economies exposed to financial sanctions.

It is important to reiterate that the relationships identified by the machine learning framework are predictive and correlational rather than causal. The models uncover systematic statistical regularities associated with successful policy outcomes, but they do not establish structural transmission mechanisms or policy multipliers. Causal identification would require structural modeling or quasi-experimental designs, which lie beyond the scope of the present analysis. Explicitly acknowledging this distinction is essential for methodological transparency, especially in macroeconomic contexts characterized by endogeneity, simultaneity, and feedback effects.

Traditional econometric approaches, such as VAR and ARDL models, have been widely applied in macroeconomic analysis and provide valuable insights into dynamic relationships among economic variables. However, previous studies have highlighted that machine learning methods may offer complementary advantages in capturing nonlinear patterns, complex interactions, and high-dimensional relationships.

It should be noted that the present study does not estimate VAR or ARDL models and does not provide a direct comparison between econometric and machine learning approaches. Therefore, the findings should not be interpreted as evidence of the superiority or inferiority of machine learning models relative to these traditional methods. Instead, this study evaluates the predictive performance and interpretability of machine learning models within the defined research framework

The second hypothesis is supported by the feature importance and SHAP analyses, which identify GDP growth, inflation, the nominal exchange rate, and the sanctions regime indicator as central determinants of policy effectiveness. The findings indicate that financial sanctions are associated with contractionary output effects alongside intensified inflationary and exchange rate pressures, thereby reshaping the transmission of both monetary and fiscal policies. Monetary instruments—particularly nominal interest rates and money growth—display asymmetric and context-dependent effects, functioning as either stabilizing or destabilizing forces depending on prevailing macroeconomic and sanction conditions. Fiscal policy, especially government spending, exhibits a consistently positive association with favorable macroeconomic outcomes, highlighting its buffering role in sanction-constrained environments. These results are broadly consistent with earlier empirical evidence emphasizing the importance of coordinated policy

responses under external constraints (Heydarian et al., 2021, 2022, 2023, 2024, 2025).

The distribution of SHAP values demonstrates that the contribution of explanatory variables varies across observations, indicating heterogeneity in the model decision process. However, this variation alone does not confirm that feature contributions differ specifically across sanction regimes. Establishing regime-dependent effects would require additional analyses, such as regime-specific SHAP estimation or SHAP interaction analysis between the sanction indicator and other explanatory variables. Therefore, the current SHAP results are interpreted as evidence of heterogeneous feature contributions rather than direct evidence of sanction-regime-specific effects.

Overall, the discussion highlights that machine learning methods not only improve predictive accuracy but also reveal structured, predictable nonlinear patterns in macroeconomic policy performance under financial sanctions. Nevertheless, careful interpretation remains essential to avoid conflating predictive associations with causal inference. By combining high predictive performance with explainable artificial intelligence tools, this study provides a transparent and methodologically rigorous framework for evaluating macroeconomic policy effectiveness in economies subject to sustained external financial constraints.

4 Conclusion and Policy Implications

This study develops a data-driven analytical framework to forecast and evaluate the effectiveness of monetary and fiscal policies in mitigating the adverse macroeconomic effects of financial sanctions in Iran. Using monthly data for the period 2006–2024, the analysis compares advanced machine learning algorithms—XGBoost, LightGBM, and Random Forest—with a benchmark linear model based on Logistic Regression. The results indicate that tree-based machine learning models, particularly XGBoost and LightGBM, achieve superior predictive performance compared with the alternative models considered in this study. This finding suggests that these approaches are effective in capturing nonlinear patterns and complex interactions among macroeconomic variables. However, the present analysis does not directly establish regime-dependent dynamics associated with sanctions, as no separate regime-specific estimation or SHAP interaction analysis was conducted. Future research could extend this framework by explicitly examining whether the contribution of explanatory variables changes across different sanction regimes.

SHAP-based interpretability analysis identifies GDP growth, inflation, and the exchange rate as central dimensions of policy effectiveness, jointly reflecting macroeconomic performance under financial sanctions. Among policy instruments, the nominal interest rate, government spending, the tax-to-GDP ratio, and money growth emerge as key transmission channels. Financial sanctions are associated with contractionary output effects alongside intensified inflationary and exchange rate pressures, underscoring the structural constraints faced by policymakers in sanction-constrained environments.

4.1 Limitations and Future Research

Several limitations should be acknowledged. Financial sanctions are modeled using a binary regime indicator due to the absence of a consistent high-frequency sanctions intensity index for Iran over the full sample period. Consequently, the estimated effects reflect average differences across sanction regimes rather than marginal changes in sanction severity.

Moreover, the extended sample period encompasses multiple sanction and policy regimes characterized by structural heterogeneity. Although the machine learning framework flexibly accommodates nonlinear and regime-dependent dynamics, the estimated relationships should be interpreted as conditional and predictive patterns rather than stable structural effects.

Finally, the analysis is explicitly predictive rather than causal. The findings identify systematic and predictable nonlinear associations between policy instruments and macroeconomic outcomes but do not establish causal mechanisms. Structural or quasi-experimental approaches would be required for causal inference.

Although the empirical focus is Iran, the methodological framework is broadly applicable to other sanction-affected economies, subject to data availability and institutional comparability. Future research may extend this approach to cross-country or panel settings to assess external validity and systematically compare policy effectiveness across different sanction designs and economic structures.

4.2 Concluding Remarks

By integrating advanced machine learning techniques with transparent interpretability tools, this study bridges macroeconomic theory, data science, and policy analysis. The proposed framework offers a robust and adaptive approach for evaluating policy effectiveness in economies exposed to sustained external financial constraints and provides a foundation for future

research at the intersection of sanctions economics and explainable artificial intelligence.

References

Addo, P. M., Guégan, D., & Hassani, B. (2018). Credit risk analysis using machine and deep learning models. *Risks, 6(2), 38*. https://doi.org/10.3390/risks6020038

Alenezy, A. H., Ismail, M. T., Wadi, S. A., Tahir, M., Hamadneh, N. N., Jaber, J. J., & Khan, W. A. (2021). Forecasting stock market volatility using hybrid of adaptive network of fuzzy inference system and wavelet functions. Journal of Mathematics, 2021(1), 9954341.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*(2), 281–305.

Breiman, L. (2001). Random forests. *Machine Learning, 45(1), 5–32*. https://doi.org/10.1023/A:1010933404324

Chatzis, S. P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., & Vlachogiannakis, N. (2018). Forecasting stock market crisis events using deep and statistical machine learning techniques. *Expert Systems with Applications, 112, 353–371*. https://doi.org/10.1016/j.eswa.2018.06.032

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).

Cho, P., & Lee, M. (2022). Forecasting the volatility of the stock index with deep learning using asymmetric Hurst exponents. Fractal and Fractional, 6(7), 394.

Diane, L., & Brijlal, P. (2024). Forecasting stock market realized volatility using random forest and artificial neural network in South Africa. International Journal of Economics and Financial Issues, 14(2), 5-14.

Farzanegan, M. R. (2011). Oil revenue shocks and government spending behavior in Iran. Energy Economics, 33(6), 1055–1069. https://doi.org/10.1016/j.eneco.2011.05.005

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Heydarian, S., Pahlavani, M., & Mirjalili, S. H. (2021). Financial sanctions and economic growth: An intervention time-series approach. *International Economic Studies, 51(1), 1–14*. https://doi.org/10.22108/IES.2020.122915.1083

Heydarian, S., Pahlavani, M., & Mirjalili, S. H. (2022). The impact of financial sanctions on capital inflow and outflow (case of Iran). *Journal of Money and Economy, 17(1), 67–88*. https://doi.org/10.29252/jme.17.1.67

Heydarian, S., Pahlavani, M., & Mirjalili, S. H. (2023). Financial sanctions and the imports of intermediate and capital goods in Iran: DID method. *International Journal of Business and Development Studies, 15*(2), 101–134.

Heydarian, S., Pahlavani, M., & Mirjalili, S. H. (2024). Financial sanctions, oil revenues, and monetary and fiscal policies in Iran: A DSGE model. *International*

Journal of Business and Development Studies, 16(2), 145–183. https://doi.org/10.22111/ijbds.2024.49244.2134

Heydarian, S., Pahlavani, M., & Mirjalili, S. H. (2025). Financial sanctions, exchange rate volatility, and macroeconomic variables: The case of Iran. *Iranian Journal of Finance, 9*(2), 70–106. https://doi.org/10.61186/ijf.2025.462663.1476

Heydarian, S., & Mirjalili, S. H. (2026). The Impact of Macroeconomic Variables on Iran's Trade with Economic Cooperation Organization Countries: An Extended Gravity Model. *Global Business Review*. https://doi.org/10.1177/09721509261447173

Ho, T. K. (1995, August). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition* (Vol. 1, pp. 278-282). IEEE.

Keshtgar, N., & Mirjalili, S. H. (2026). The banking crisis and macroprudential policy: Evidence from Iran. *Journal of Mathematics and Modeling in Finance, 6(2), 45–63*.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (pp. 4765–4774). Curran Associates.

Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023). Machine learning advances for time series forecasting. *Journal of Economic Surveys, 37(1), 76–111*. https://doi.org/10.1111/joes.12429

Mirjalili, S. H. (2025a). The rise and fall of LIBOR and its alternatives. *Journal of Money and Economy, 20*(3), 441–460.

Mirjalili, S. H. (2025b). *Economic resilience and vulnerability: Concepts and indices* (MPRA Paper No. 126492). Munich Personal RePEc Archive.

Mirjalili, S. H. and Safari, O. (2025). Fiscal Policy Uncertainty and Economic Activities in Iran's Provinces. *International Journal of Business and Development Studies, 17(1), 75-101*.

Pahlavani, M., Heydarian, S., & Mirjalili, S. H. (2025). *Financial sanctions and macroeconomy of Iran*. Sistan and Baluchestan University Press.

Popa, C. E. (2024). Ten Years of Economic Sanctions and Their Macroeconomic Impact on the Russian Federation. *Expert Journal of Economics, 12(1), 37-51*.

Rundo, F., Trenta, F., Di Stallo, A. L., & Battiato, S. (2019). Machine learning for quantitative finance applications: A survey. *Applied Sciences, 9(24), 5574*.

Safari, O., Assadzadeh, A., & Mirjalili, S. H. (2023). Measuring Fiscal Policy Uncertainty in Iran. *Journal of Money and Economy*, *18 (4), 457-496*.

Safari, O., Assadzadeh, A., & Mirjalili, S. H. (2024). Fiscal policy uncertainty and Industrial Investment in Iran. *International Journal of Business and Development Studies, 16(1), 229-259*.

Wasserbacher, H., & Spindler, M. (2022). Machine learning for financial forecasting, planning and analysis: recent developments and pitfalls. Digital Finance, 4(1), 63-88.

Wong, A., Figini, J., Raheem, A., Hains, G., Khmelevsky, Y., & Chu, P. C. (2023, April). Forecasting of stock prices using machine learning models. In *2023 IEEE international systems conference (SysCon)* (pp. 1-7). IEEE.

Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications. *arXiv*. https://doi.org/10.48550/arXiv.2003.05689

Zareei, A., Falahi, M. A., & Malek Sadati, S. S. (2024). The impact of international sanctions on Iran's economy with an emphasis on the role of the National Development Fund: DSGE model approach. *Journal of Economic Policy, 16(31), 35–80*. [In Persian].