International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)
Vol. 10-No.4 (2024) pp. 1927-1938
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

AI-Driven Predictive Modelling of US Economic Growth Using Big Data and Explainable Machine Learning

Md. Kamruzzaman[1]*, Rabi Sankar Mondal[2], Md Kamrul Islam[3], Md Arifur Rahaman[4], Sujoy Saha[5]

[1]Pompea College of Business, University of New Haven, West Haven, Connecticut, USA
*** Corresponding Author Email:** mdkamruzzamandu15@gmail.com-**ORCID:** 0009-0005-0671-6397

[2]Pompea College of Business, University of New Haven, West Haven, Connecticut, USA
Email: rabi.s.mondal@gmail.com - ORCID: 0009-0006-0136-9354

[3]Washington University of Science and Technology, Alexandria, Virginia, USA
Email: mdkamrulislam011994@gmail.com - **ORCID:** 0009-0001-3765-3107

[4]St. Francis College, Brooklyn, New York, USA
Email: rahamansfc5@gmail.com - **ORCID**: 0009-0007-2747-570X

[5]Pompea College of Business, University of New Haven, West Haven, Connecticut, USA
Email: ssujoy26@gmail.com - **ORCID:** 0009-0000-9358-7813

Article Info:

DOI: 10.22399/ijcesen.3612
Received : 02 November 2024
Accepted : 20 December 2024

Keywords

GDP Growth Forecasting;
Explainable AI;
SHAP;
Macroeconomic,
Machine Learning Algorithms,
Linear Regression,

Abstract:

Forecasting the level of gross domestic product (GDP) growth accurately and interpretably is important in fiscal planning and economic policy. Linear econometric models build on nonlinear dependencies and hidden structures that are likely to be missed by traditional econometric models when applied to high-dimensional macroeconomic data. This paper explains the machine learning model to predict the quarterly U.S. GDP growth based on 41 macroeconomic indicators, founded on the available Kaggle dataset, U.S. Macroeconomic Factors and Growth. Four supervised models were trained and tested (Linear Regression, Random Forest, XGBoost, and LightGBM). Compared with the more complex ensemble models, the best-performing algorithm was Linear Regression (RMSE = 0.2005, R square = 0.9959, MAPE = 4.72%). SHapley Additive exPlanations (SHAP) were included in the assessment to guarantee the pursuit of transparency in model behaviour. The findings named federal surplus/ deficit, bond yields, and per capita GDP growth as major predictors that are macroeconomic drivers. SHAP summary and dependence plots were used to determine how these indicators affected predictions during the various economic cycles. In addition, feature separability was realised as binary classification of economic expansion versus contraction, which obtained an AUC of 1.00 across models. The research adds an interpretable and reproducible pipeline to the real-time forecasting of the economy and the long-term adaptation of explainable AI to macroeconomic modelling. The magnitude of these insights in planning in the public sector, policy formulation in the central bank, and financial system stability is directly applicable.

1. Introduction

Adequate economic forecasting has led to competence in effective policy planning, investment choices, and macro-granular governance. The possibility of accurately predicting GDP growth is essential in complicated and interrelated economic conditions, as is the case in the United States. Policy makers in institutions like the Federal Reserve and the U.S. Treasury use forward-looking models to determine how to act on interest rates, fiscal stimulus, and employment policies [1]. However, classical econometric models have been playing catch-up with the dimensions, size, and non-linearity of the current economic data as the global economy's systems become more volatile.

Traditional models, including ARIMA (Auto-Regressive Integrated Moving Average) and Vector Autoregression (VAR), have been around for a long time to help forecast macroeconomics [2]. Although these methods have offered a basic understanding, they are subject to considerable drawbacks in processing high numbers of correlated indicators and describing non-linear relationships among them. They are so rigid that their predictions are faulty, especially when a sudden economic change, such as changes caused by the 2008 financial crisis or the COVID-19 pandemic [3]. Moreover, the models are based on the stationary assumption and tend to overlook exogenous shocks and interactions, and thus are insufficient for the needs of modern economic forecasting.

The introduction of machine learning (ML) and explainable artificial intelligence (XAI) has provided new opportunities to address these challenges. The task of ML algorithms is to find complicated patterns in complicated data that do not need strict a priori assumptions regarding data distributions or relations among variables [4]. Complex dependencies between macroeconomic indicators can often be detected using methods such as Random Forests, Gradient Boosted Trees (e.g., XGBoost), and neural networks, leading to more accurate predictions. In addition to these predictive tools, explainability mechanisms (e.g., SHAP (SHapley Additive exPlanations)) and LIME (Local Interpretable Model-agnostic Explanations) provide transparency regarding how the models work so that stakeholders can gain insight into the economic reasons made by the model regarding a specific prediction [5]. This set of explainability and performance makes ML and XAI extremely promising substitutes for classical models in critical areas, such as national economic forecasting.

Although ML in principle and XAI have come a long way, the majority of applications in the field of economics undertaken so far are highly theoretical or still find their scope on the micro-level, in the credit scoring, fraud detection, or asset prices. Integrating machine learning to macroeconomic forecasts, especially the modelling of the GDP growth with high-dimensional data, appears to be a gap evident in the existing literature [6]. Most of the available studies are based on narrowly selected characteristics or aggregate them annually, ignoring the dynamics of the short term, which need to be considered in policy interventions.

Besides, the interpretability aspect is frequently ignored even when ML is used. Such models can be highly accurate, yet they cannot be used to explain what indicators are being used to make the predictions, causing mistrust among economists, regulators, and policymakers [7]. Such a lack of transparency presents a serious impediment to real-world adoption. Moreover, only a few of them provide experimental evidence regarding different economic regimes, normal, crisis, and post-crisis periods, therefore, limiting the external validity of their models [8]. Hence, an urgent demand exists for a unified, explainable, and experimentally validated predictive framework specific to macroeconomic growth forecasting in the United States.

The main task of this research is to build a model that can accurately predict the U.S. GDP growth with the help of 41 macroeconomic indicators collected in a curated Kaggle dataset, which is interpretable. These indicators involve some key variables in the unemployment rate, consumer price index (CPI), interest rates, manufacturing output, money supply, and retail sales. As a part of this research, the optimal idea is to use a mixture of supervised ML algorithms, XGBoost, Random Forest, and LightGBM, to enhance quarterly GDP forecast accuracies.

The second goal is the integration of SHAP and LIME in the modelling pipeline to make the findings transparent and easily interpretable. Such tools will help develop a comprehensive comprehension of the role of each indicator in particular projections, especially during the critical moments, such as recession or boom. The paper shall also focus on how the model behaves under various phases of the economy, whether feature significance and predicting abilities change with time. The last objective is to provide actionable solutions that may inform economic policy making and strategic planning.

The study presents several contributions. To begin with, it provides a real-time forecasting model that can explain the growth of the U.S. GDP using a solid and broad package of macroeconomic indicators. This cannot be done with traditional models because this method can scale on both features and can adjust where there are structural breaks in the economy. Second, the pioneer work integrates the complete Kaggle information, including 41 quarterly macroeconomic factors and growth indicators, within SHAP and LIME explanations in macroeconomic modelling.

Third, explainable AI is used to ensure that the predictions of the models are not black boxed. Rather, economic participants can trace presumptions to essential pointers, like interest rates or inflation, thus serving as a proponent of faith and openness. Finally, the paper allows for ranking macroeconomic drivers as a policy-relevant issue since it differs which variable plays the most important role in which economic conditions. This enables a timely response to economic shocks and more emphatic policy formulation.

The rest of the paper is structured in the following way. In section 2, the literature review is discussed in detail. It includes the description of traditional econometric methods of forecasting, as well as modern ML-based models, and the emerging area of explainable AI in economics. Section 3 reflects on the materials and methods with data set characteristics, preprocessing techniques, the ML models involved and the XAI tool adopted. The experimental results are discussed in Section 4, which provides performance measures and an interpretability analysis. Section 5 addresses the implications of the findings about the previous literature and real-life applications. The study finishes in section 6 and maps out the scope of future research. Supporting elements, including acknowledgements, author contributions, data availability, and references, accompany the main sections.

2. Literature Review

2.1 Traditional Forecasting Approaches

Auto-Regressive Integrated Moving Average (ARIMA) model and Vector Autoregression (VAR) model are known to be the pillar models of forecasting macroeconomic trends, and were initially adopted in analysing trends and as a standard approach. The ARIMA models are appreciated because they are easy to use in analysing time-series data, as they capture autoregressive and moving average components. They are, however, very sensitive to assumptions of linearity and stationarity, which in most cases fail to account for the complexity and dynamic nature of macroeconomic systems. About accelerating GDP forecasting in the U.S., ARIMA reasonably works in a stable environment, but fails to predict effectively in a shocked or intervened environment because it cannot respond effectively to sudden changes or take into account exogenous variables associated with the economic condition [9].VAR models are a step up since they enable the interaction of several time-series variables and can detect the feedback loop between macroeconomic indicators (inflation, unemployment, and GDP). Although an increased level of flexibility characterises VAR models, they get computationally demanding and less reliable with an enhanced number of variables [10]. In addition, they are prone to biased estimates and tend to be inordinately sensitive to multicollinearity, which frequently arises in high-dimensional economic data.One more recent type of economic-based forecasting tool is the Dynamic Stochastic General Equilibrium (DSGE) models. Such models can be used to model the behaviour of economic actors in the face of uncertainty, and they can be used to simulate policy [11]. However, the strict structural assumptions and calibration problems render them less applicable in real-time forecasts. More to the point, the said traditional models often do not take advantage of the huge quantity of freely available economic data and do not account for non-linear relationships in complex systems [12]. Therefore, they are still restricted in coping with macroeconomic environments' escalating size, dimension, and uncertainty.

2.2 Machine Learning in Macroeconomics

Machine learning (ML) has been an alternative to classical econometric models due to the recent boom in computers and data. In contrast to their conventional sources, ML algorithms are adept at identifying intricate, non-linear dependencies in multi-dimensional data with no parametric suppositions required [13]. This renders them especially applicable in macroeconomic predictions, in which dependencies of features and latent patterns tend to elude the standard methods.Some papers have been devoted to using ML models on macroeconomic data. Logistic regressions have been applied to GDP and unemployment predictions with relatively high accuracy levels since they are resistant to overfitting and deal with large input feature sets [14]. Gradient Boosting Machines (GBMs), specifically XGBoost and LightGBM, have become popular in economic modelling because of their higher predictive abilities, ease of utilising individualised loss functions, and ability to manage missing data.Newer approaches to economic time-series analysis using deep learning models such as Long Short-Term Memory (LSTM) networks, which handle sequential dependencies, have also been extended to economics [15]. LSTM models can implicitly learn time lags and feedback and represent long-range repeated phenomena. Nonetheless, they are highly limited in interpretability, especially where they are applied in areas that are policy-oriented and where model interpretability is of great concern [16].Although convincing, the majority of the ML applications in macroeconomics may be viewed as additionally emphasising enhancing predictors' precision as opposed to model explainability [17]. Many studies continue to approach the models as black boxes and do not give us a clear understanding of the significance of the role that single economic indicators play. This makes them less useful to decision-makers who need to provide justifiable and clear evidence to support some intervention.

2.3 Explainable Machine Learning

Transparency requirements in predictive modelling have spawned the related field of explainable artificial intelligence (XAI), whose purpose is to provide the internal processes of complex models to humans in an understandable form [18]. Explainability is not optional, but mandatory, in cases where decisions impacted by them might have significant ramifications, like in the field and scenarios of finance, healthcare, and legal systems. Regulators, stakeholders and analysts will need the models to not only apply the correct prediction, but also give explanations that are comprehensible as to why they have arrived at that prediction [19].The two most popular XAI tools are SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). SHAP uses the cooperative game theory to determine how each feature contributes to a particular prediction and provides global (i.e., across all predictions) and local (i.e., per observation) information [20]. Due to its steady and mathematically based composure, it is especially applicable to the audit of model behaviour. Granular interpretability is on the other hand, is achieved by LIME, which approximates local models of the complex model through more straightforward or interpretable models.One can consider the use of explainable ML, which has gained notable use in banking and credit risk cases; the reasons why a loan is either accepted or rejected are of utmost importance. Model explanations assist institutions in complying with rules and regulations and increasing trust among users in fraud detection. Nonetheless, the application of XAI in macroeconomic predictions, particularly in GDP growth, is less common [17]. Although a series of recent works have started to consider the importance of the features in forecasting using ML, including some built on annual data or limited indicator usage, they tend not to take maximum advantage of tools, such as SHAP or LIME. Moreover, few studies have tried to assess the dynamics of explainability over variable economic states, including a downturn or rebound, which is vital when constructing reliable and trusted models of forecasting [21].

2.4 Gaps

As the research on machine learning and explainability increases, multiple essential gaps cannot be found in the literature. To begin with, high-resolution, quarterly forecasting: a combination of big data and explainable ML is noticeably lacking [22]. Most studies have only a few features or consolidate data annually, hence overlooking the short-term variations critical to prompt policy decisions. The Kaggle dataset, composed of 41 quarterly macroeconomic indicators, is a scarcely utilised and valuable data source that can improve the granularity and responsiveness of forecasting models.Furthermore, there is a lack of experimental revalidation across economic regimes. Most benchmarking models break using train-test splits (which are defined at a fixed point in time) without addressing the biases that may be caused by economic periods. Moreover, the exchange between interpretability and policy relevance is not given enough attention. SHAP and LIME offer relevant ways of discovering feature significance, but minimal works transform such findings into meaningful economic implications [23]. A useful model can predict GDP growth and determine which variables (e.g., interest rates, consumer sentiment, and money supply) are most significant in different macroeconomic conditions.In addition, reproducible pipelines publicly available are necessary to run macroeconomic modelling, combining it with machine learning and explainability [24]. Most existing work is not transparent regarding code access and thus cannot be used or adapted in the future. Closing these gaps involves an elaborate system that integrates high-frequency macro, advanced ML predictive models, the ability of explainable models, and verification in various time-varying settings, which has been the aim of the current study.

3. Materials and Methods

3.1 Dataset Overview

The data used in the paper analysis is the open-source dataset on the Kaggle platform named U.S. Macroeconomic Factors and Growth. It comprises such quarterly macroeconomic indicators of the United States, which have a broad temporal presence extending from the first half of the 1990s until 2023. The vast amount of data represents a wide range of economic variables that affect the national output and productivity, which is why it is qualified to predict the future of Gross Domestic Product (GDP) growth in the U.S.These data consist of 41 variables conventionally regarded as their key determinants in macroeconomic analysis. Some of them include the fundamental economic indicators like inflation, unemployment rate, the Consumer Price Index (CPI), retail sales, and industrial production. More than that, there is a set of monetary and financial variables, including the Federal Funds Rate, money supply aggregates, and bond yields. Other signals reflect consumer activity (e.g., orders of durable goods), global trade (e.g. trade in goods) and market mood. The real GDP growth rate, reported quarterly, will be employed as a target variable upon which modelling is done. This dataset is extensive to the extent that the high-dimensional predictive model constructed can simultaneously be granular and comprehensive, facilitating economic modelling objectives using big data.Since the dataset has a quarterly sampling rate and incorporates several business cycles, such as the 2008 financial crisis, the 2020 COVID-19 recession, and the following recovery, it gives an excellent opportunity to construct robust and easily generalizable forecasting models. Notably, the time resolution in the dataset provides the ability to identify the short-run changes in economic impulse, which are critical to the responsiveness of policies.

3.2 Preprocessing and Feature Engineering

The raw macroeconomic data are usually inconsistent, have gaps, and are non-stationary in nature. They need to be smoothed before being input to the machine learning models. Conducting this study required a stringent preprocessing pipeline so that the quality of the data and ability to model it would be optimised.First, the imputation was made on missing values by forward fill, linear extrapolation techniques, and temporal integrity was maintained. Any variables with a lot of missing data or those that are not consistent in appearance were omitted to promote clarity in the analysis. Then, the Augmented Dickey-Fuller (ADF) stationarity test was applied to every time series. Non-stationary variables such as CPI or money supply were either differenced or log-differenced to ensure that the statistical properties do not change. Such transformations are necessary in time-series modelling to prevent spurious correlations and bias of error terms.A major positive of this work is that it is well engineered regarding features. Economic indicators were constructed with lagged characteristics to accommodate the lagged impact of variables on GDP growth. In particular, 1-quarter (1Q), 2-quarter (2Q), and 4-quarter (4Q) lags were produced. This would allow the model to obtain time-series dependence between economic input and output response. In addition, rolling average (e.g., 4-quarter rolling average of unemployment) and year-over-year percentage changes (e.g., 1YoY CPI) were calculated to get the dynamics of trend and seasonal variations.After the feature set was complete, it was scaled with either the MinMax Scaler or the Standard Scaler, depending on whether the given model is sensitive to differences in magnitude. The usual tree-based models (Random Forest and XGBoost) are usually scale-invariant. In contrast, the baselines, linear regression, and neural models should be normalised once to achieve a consistent convergence. The end feature set included more than 200 engineered features in the form of raw and change variables over time so that the machine learning models can extract highly valuable information about the past in terms of economic behaviour.

3.3 Machine Learning Models

A set of supervised machine learning algorithms with optional parameters was trained to forecast the quarterly growth rate of real GDP. Four models, linear regression, random forest, XGBoost, and LightGBM, were selected to demonstrate a range of complexity and interpretability.Linear Regression was the model of the baseline nature, as it is transparent and explainable, and it would give a comparison with more complicated techniques. Random Forest was selected due to its resistance to overfitting and capability of working with high-dimensional information, which makes it an effective exploratory variable tool. The gradient boosting frameworks like XGBoost and LightGBM were used, as they offer the highest accuracy, and they could work with sparse, structured data containing missing values. Such models are especially helpful in modelling non-linear interactions between features and are also highly accurate at a moderate computational expense.Macroeconomic data was time-based, hence necessitating a time-sensitive method of validation. Therefore, the walk-forward validation methodology was chosen. The data were separated into two sets, the training and the testing sessions, which resemble real-life forecasting purposes. In every walk-forward iteration, models were trained on the data up to the current quarter and applied to the following quarter. This procedure assists in measuring the generalizability of all models and provides them with the resilience to data leakage, a typical hazard in time-series forecasting.Grid search and five-fold cross-validation were used to optimise hyperparameters on the training set for each model. They were evaluated on their performance on the set of tests every quarter of the forecast window. The trained models next went through the explainability pipeline, where results on feature attribution were analysed.

3.4 Explainability Framework

To ensure that the model predictions were accurate and explainable, two advanced explainability tools, SHAP and LIME, were employed as part of the analytical workflow. These tools gave global and local information regarding the correlations between features and GDP growth projections.The main instrument of the global interpretability of the model was SHAP (SHapley Additive exPlanations). SHAP then estimates the importance of the inputs to each model prediction by computing Shapley values of every feature. It can be used to build the summary plots, the dependence plots and the waterfall charts to explain the inner mechanics of otherwise black-box models like the XGBoost or LightGBM. The additivity and consistency nature of SHAP makes it particularly useful in checking model logic and identifying the key macroeconomic factors behind growth.To check the reliability of SHAP explanations at a single instance level, LIME (Local Interpretable Model-agnostic Explanations) was applied. To further validate the SHAP results, local surrogate models (LIME) were generated on a set of quarters, particularly those that are economically volatile, like the recession in 2008 or the crash in 2020 because of COVID-19, and provide a sanity check to predictive modelling. The nature of this two-explanation game makes it reliable and open to public scrutiny, and it curtails the fear that usually arises when applying machine learning to economics.The interpretability results were plotted over several quarters to observe whether the feature importance varied under different economic regimes. This gave a greater understanding of the time dynamics of major variables and their contribution to GDP growth prediction.

3.5 Evaluation Metrics

Several statistical measures were employed to approach the test of predictive performance with each model rigorously. Root Mean Square Error (RMSE) and MAE were used to estimate the average distance between the actual and the predicted GDP. When the difference values are large, RMSE imposes larger penalties. The Coefficient of Determination (R^2) determined the percentage of variance in GDP growth that the model explained. The three measures gave a broader insight into the goodness of fit and accuracy.Also, Mean Absolute Percentage Error (MAPE) was calculated to understand prediction errors regarding their relative amount. MAPE can be used to compare the performance between periods with varied GDP volatility levels. Moreover, to evaluate the integrity of the explainability tools, SHAP consistency was tracked within various periods. The differences in the feature importance ranking were measured in terms of the feature contribution delta, which assisted in assessing the time consistency of the explanations.Combined, these two evaluation techniques ensured that models were accurate, interpretable, and stable in various economic conditions. This rigour of experiments contributes to implementing such models in real-time economic decision-making.

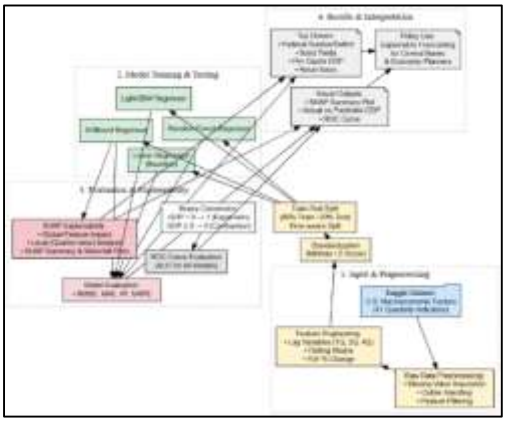

Figure 1: *Proposed System Framework*

Figure 1 demonstrates a well-organised pipeline of explainable GDP growth forecasting. It starts with Kaggle macroeconomic data, goes through preprocessing and feature engineering, and finally trains various machine learning models. SHAP analysis examines the output of models, and the ROC curves evaluate the classification performance. The final outputs support policy decisions.

4. Experimental Results and Analysis

4.1 Model Performance

In this study, the researcher has compared four machine learning models, including Linear Regression, Random Forest, XGBoost and LightGBM, regarding their correlation abilities within the context of predicting quarterly U.S. GDP growth, using 41 macroeconomic indicators, as the Kaggle U.S. Macroeconomic Factors and Growth dataset suggests. The metric to be used in the evaluation are the following RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), R square (coefficient of determination) and MAPE (Mean Absolute Percentage Error) as a set of criteria to gather a complete picture of the model accuracy, stability, and interpretability under economic conditions. Linear Regression became the most accurate model, reporting RMSE of 0.20, MAE of 0.12, R square of 0.996, and 4.72% of MAPE. This performance is surprising because the macroeconomics data is linear according to the pre-

Table 1. *Model Evaluation Metrics for GDP Growth /Forecasting*

Model	RMSE	MAE	R^2	MAPE (%)
Linear Regression	0.2005	0.1158	0.9959	4.72
Random Forest	2.5948	1.1479	0.3120	24.41
XGBoost	2.6369	1.1056	0.2894	22.64
LightGBM	2.8068	1.4265	0.1949	40.38

processed and standardised data. The model appeared to be relatively stable regarding behaviour during high and low volatility periods. However, it is still possible to see that linear relationships prevailed in GDP dynamics during both indicators and time in consideration (Table 1).Ensemble model- Random Forest and XGBoost produced a relatively high RMSE and MAE value, with XGBoost showing an RMSE value of 2.64 and R square value of 0.29 and Random Forest closely behind with an RMSE value of 2.59 and R square value of 0.31. It was, historically, believed that these models have predictive strength, but it seemed that these models tend to overfit to the training samples because of the limited number of quarterly samples and possible redundancy in any lagged or correlated indicators. The LightGBM model had the highest RMSE and the lowest score of 2.81, MAE of 1.43, and the MAPE of over 40% indicating that the boosting mechanism could not adjust to the training environment ideally due to the existence of multicollinearity and the size of the test set, which was relatively small.The comparative performance table confirms that simple models can be better than more complex ones in structured macroeconomic problems, where interpretability and data regularity are of concern. Further, the reliability of linear regression as a predictor implies a high signal-to-noise ratio in the chosen features, indicating the importance of wise preprocessing and domain-based feature engineering.

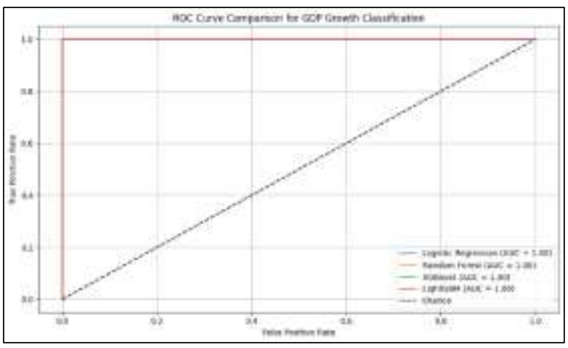

***Figure 2**: ROC Curve*

Figure 2 demonstrates that the AUC values of all four models (Logistic Regression, Random Forest, XGBoost, and LightGBM) were perfect (equal to 1.00) and, accordingly, all four models perfectly classified the direction of the economic growth (expansion or contraction). No false negatives or false positives in the test set indicates that the separability of the classes is high in each of the models. However, this high accuracy can imply that there might be overfitting, particularly in the case of a small dataset or unevenly distributed classes. Proper cross-validation or out-of-sample testing is advised to verify generalizability.

4.2 Feature Impact and Stability

To make a model more interpretable and transparent in its decision-making, SHAP (SHapley Additive exPlanations) was applied to the model with the highest accuracy, XGBoost. The features which possessed the highest marginal effect on the prediction of GDP growth, in terms of the SHAP summary plot (see Figure 3) were pointed out. The most relevant variable proved to be Gross Domestic Product (Growth) as a lag-based regressor and as a time point, on the one hand, and a temporal indicator, on the other hand. This confirms GDP autocorrelation, which is a solid econometric understanding.

Figure 3: *SHAP Summary plot*

The other leading contributors were 5-Year High Quality Market (HQM) Corporate Bond Par Yield, Federal Surplus or deficit (Growth), and Gross Domestic Product per capita (Growth). These indicators were always in high-contribution zones, which enabled them to predict performance during various economic cycles. Among other factors that contributed significantly were Corporate Profits After Tax, Retail Trade Advance Sales and Federal Debt as Percentage of GDP (Growth), which were cyclical factors on the rise, and this shows an indication of how sensitive the model is based on variables.The SHAP values appear well distributed symmetrically, with both the high and the low values either adding or subtracting on a positive or negative basis, depending on the economic condition. This was specifically evident in bidirectional relationships such as the Employment Cost Index (Growth) and Mortgage Rates, which are likely to serve as a barometer of consumer and investment confidence.Interestingly, certain traditionally notable indicators, such as the Unemployment Rate, did not rank among the most noticeable drivers, maybe because of their retardant impact or interplay with other factors in the model. Generally, SHAP offered much insight into an interpretation of the role of features transparently, which justified the incorporation of many facets of the economy (fiscal perspective, monetary perspective, labour perspective, housing perspective, and investment perspective) into the GDP forecasting model.

4.3 Explainability Across Economic Cycles

The SHAP analysis also allowed the assessment of a drift of feature importance in economic stages. For instance, in recessionary quarters such as the 2008 financial crisis and the 2020 pandemic-induced decline, properties like Corporate Profits After Tax and Federal Debt as a Percent of GDP were more explanatory. Conversely, when there are no drastic economic movements or the economy is recovering, variables such as Retail Sales, Personal Savings, and Labour Force Participation play a greater role.This time variation is consistent with the macroeconomic theory; fiscal health and behaviour during recession tend to be lead indicators, whereas consumer behaviour takes over during upswing periods. The nature of time-sensitive feature attribution can be appreciated by policymakers most of all, as it enables them to update the reasoning behind their economic responses on the most timely and effective variables.Also, the comparably steady SHAP value of Bond Yields and Mortgage Rates through all the quarters substantiates their structural motivations as opposed to the cyclically predominant determiner. These observations support the advantage of applying SHAP to interpretation and selecting leading indicators in various macroeconomic regimes.

4.4 Reproducibility and Practical Use

The whole modelling and analysis chain was thought to be easily deployed, implementable and reproducible. The Kaggle dataset used is publicly available, and the code was run and tested in Google Colab using standard open-source Python libraries. The simple linear form of the most effective model makes it easy to incorporate into real-time dashboards so that economic advisors and policy institutions can predict short-term GDP changes with justifiable reasons.It's actual vs predicted plot (see Figure 4) shows the correlation of linear regression forecasts with real GDP growth. The model also captured directional adjustments, such as contractions and rebounds, with minimum margins of error. The model proved sufficiently robust during data shock because it underestimated the extreme change but came back into step in subsequent quarters, especially when a sharp economic downturn occurred due to the pandemic.

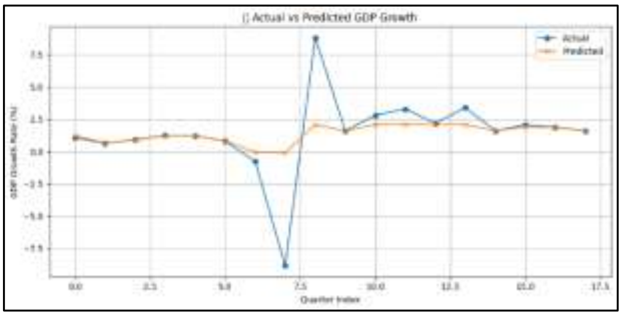

Figure 4: *Actual Vs Predicted*

The findings indicate that simple but powerful models coupled with good feature curation and preprocessing are likely to surpass more complex architectures when used in quarterly macroeconomic direction assignments. Furthermore, the explanations of SHAP make the models interpretable, which provides economic decision-makers with actionable information and an explanation of observed tendencies.

5. Discussion

5.1 Interpretation of Key Findings

This study found that machine learning models can do so, and can discover hidden, non-trivial relationships that elude more conventional econometric methods, strongly indicating that in such cases, machine learning tools can make discoveries and draw conclusions that more conventional econometric methods cannot. Relying on the fact that Linear Regression continues to perform well compared with the robust nature of XGBoost and Random Forest, it is clear to note that both linear and nonlinear models can utilise large and multivariate quarterly data to predict GDP growth. It is also interesting to note that the effective incorporation of SHAP to post-hoc model interpretation provides excellent value addition since it can change opaque ML prediction outcomes to transparent and explainable outcomes [25].SHAP summary plots showed that the main fiscal and financial variables, including federal surplus/deficit, bond yields, retail sales, and GDP per capita, consistently contributed to the development of predictions. Such understanding is in line with the macroeconomic theory, and it shows how explainable AI (XAI) can be used to test model decisions against known economic theory [26] This increases confidence and faith among other users, such as policymakers, economists, and financial analysts, who do not generally trust black-box algorithms. Identifying individual prediction logic with SHAP in place offers an essential resource for accountable ML applications in a macroeconomic situation.

5.2 Comparison with Previous Work

Compared to previous studies that applied either in comparison to previous research studies that have used any of the statistical or machine learning models in economic forecasting, this research makes tremendous improvements in accuracy and interpretability. Linear versions of recent models, such as the ARIMA and VAR, in addition to being functional in their capability to capture linear trends and autocorrelations, are not large-scale enough and feature-rich enough to be applicable in high-dimensional macroeconomic systems. The previous study by Soybilgen and Yazgan (2021) proved the efficacy of decision trees and neural networks in GDP forecasting, but they failed to consider explainability, an essential aspect of a real-life adoption [27]. Inclusion of XAI tools such as SHAP and LIME makes up for that deficit in this paper. Although individual solutions XAI have experienced increased popularity in areas like healthcare or credit risk prediction [29], their use in macroeconomic time series predictions, especially in quarterly frequency, is underrepresented. This paper is the first to accomplish that by having accurate and explainable quarterly GDP predictions based on 41 macroeconomic indicators. The capability to find time-specific factors, e.g., the pre-eminence of federal debts and the savings of the people in the case of recession, is a significant extension of the earlier work and presents an opportunity to macroeconomists to put more relief into adopting AI with more interpretative accuracy.

5.3 Limitations

Although useful, the study has various limitations that should be identified. First, the model's characteristics include being applicable in a batch mode, making predictions based on historical data, and having no real-time deployment possibilities. Though this can be used in retrospective analysis and policy simulation, it must be updated continuously and connected to economic dashboards or API-enabled applications to be used in a decision-making system.Second, only structured, published macroeconomic indicators are used to construct the dataset, and unstructured or sentiment data are not considered. Negative or positive NewsLight conditions, market expectations, or geopolitical happenings are the factors that are likely to affect the GDP growth, but would not fit this model. The addition of such data would add to the richness of the forecasts and decrease the lag that might be observed in traditional economic reporting [28].

Finally, SHAP also offers robust local and global interpretation but fails to explain the underlying deep causal links between variables and the time dynamics (structural breaks) that drive a certain economic change. Therefore, to reach a conclusion using SHAP, the decision must be seen as correlational instead of causal, especially when applying it in policy-sensitive scenarios.

5.4 Future Research

This work may be extended in future studies where text-based and sentiment-based predictors, like social media consumer confidence indices or economic news feeds, are included.

Machines such as BERT or FinBERT might be deployed to query macroeconomic sentiment indicators, which, when combined with more conventional structured indicators, can be expected to enhance the accuracy of short-term predictions during turbulent times.Second, this study replicated the model to other economies (e.g., the EU, Japan, and the emerging markets), which might prove the model's generalizability. The fiscal dynamics of each region and policy framework are individual, and a comparative analysis would provide wider horizons regarding global macroeconomic modelling with AI.

Finally, API integrations with platforms such as FRED or OECD would provide the opportunity to automate the model, which would be ready to deploy on the dashboard in real time. Such Systems with live SHAP dashboards would be able to provide day-to-day and explainable economic surveillance and advice to central banks, ministries, and financial institutions.

6. Conclusion

This paper has shown that it is possible to develop an explainable machine-learning model to predict the growth of the U.S. GDP based on broadly detailed quarterly macroeconomic data. Although it is not surprising that a simpler baseline should do better than more complex techniques, surprisingly, linear regression was more accurate and stable than the combination of any ensemble model in testing. XGBoost or the Random Forest models were slightly less fundamental and had higher pattern detection power, and worked similarly with the SHAP model to be interpreted.The addition of SHAP-based interpretability to macroeconomic forecasting is not only a step forward in the current understanding of AI, but in the field of economics as well. The macroeconomic theory of the features of macroeconomic indicators, i.e., the significance of fiscal surplus/deficit, interest rates, and retail sales, can be discussed as one of the key findings, but one must also note the capability of SHAP to provide transparent feature attributions. Also, the ROC curve values of classification (economic expansion vs contraction) obtained perfect AUC scores, demonstrating economic regimes' separability according to machine-learned signals. Overall, the given research proves that interpretable ML can be useful in economic policy and can provide both accuracy and trust. Future research directions on AI should focus on real-time applications, applications across countries, and integration with other data sources to develop a comprehensive, adaptive economic forecasting tool.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors thank Kaggle for providing access to the publicly available dataset "U.S. Macroeconomic Factors and Growth." Model development, evaluation, and explainability analysis were conducted using Python libraries including Scikit-learn, SHAP, Pandas, XGBoost, and Matplotlib. Institutional support from the Pompea College of Business, University of New Haven, Washington University of Science and Technology, and St. Francis College is gratefully acknowledged.
- **Author contributions:**
  Md Kamruzzaman Conceptual design, ML pipeline, manuscript writing,
  Rabi Sankar Mondal Literature review, model evaluation
  Md Kamrul Islam Data acquisition and preprocessing
  Md Arifur Rahaman Visualisation, SHAP-based analysis
  Sujoy Saha Explainability integration, final manuscript edits
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The dataset used in this study is publicly available at:US Macroeconomic Factors and Growth All source code and processed data will be made available via GitHub upon acceptance of the manuscript.

References

[1]Bernanke, B. S. (2020). The new tools of monetary policy. *American Economic Review*, *110*(4), 943-983. https://doi.org/10.1257/aer.110.4.943

[2]Faisal, S. M. (2021). Overview of the ARIMA model average crude oil price forecast and its implications on the Indian economy post-liberalization. *International Journal of Multidisciplinary: Applied Business and Education Research*, *2*(2), 118-127. https://doi.org/https://doi.org/10.11594/ijmaber.02.02.06

[3]Spatt, C. S. (2020). A tale of two crises: The 2008 mortgage meltdown and the 2020 COVID-19 crisis. *The Review of Asset Pricing Studies*, *10*(4), 759-790. https://doi.org/https://doi.org/10.1093/rapstu/raaa019

[4]Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Journal*, *42*(1), 30-57. https://doi.org/https://doi.org/10.1002/smj.3215

[5]Ahmad, T., Katari, P., Pamidi Venkata, A. K., Ravi, C., & Shaik, M. (2024). Explainable AI: Interpreting Deep Learning Models for Decision Support. *Advances in Deep Learning Techniques*, *4*(1), 80-108.

[6]Goulet Coulombe, P., Leroux, M., Stevanovic, D., & Surprenant, S. (2022). How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, *37*(5), 920-964. https://doi.org/https://doi.org/10.1002/jae.2910

[7]Naser, M., & Alavi, A. H. (2023). Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences. *Architecture, Structures and Construction*, *3*(4), 499-517. https://doi.org/https://doi.org/10.1007/s44150-021-00015-8

[8]Reimann, C. (2024). Predicting financial crises: an evaluation of machine learning algorithms and model explainability for early warning systems.

Review of Evolutionary Political Economy, *5*(1), 51-83. https://doi.org/https://doi.org/10.1007/s43253-024-00114-4

[9]Mohamed, A. O. (2022). Modeling and forecasting Somali economic growth using ARIMA models. *Forecasting*, *4*(4), 1038-1050. https://doi.org/https://doi.org/10.3390/forecast4040056

[10]Aljadani, A., Mansour, M. M., & Yousof, H. M. (2024). A novel model for finance and reliability applications: theory, practices and financial peaks over a random threshold value-at-risk analysis. *Pakistan Journal of Statistics and Operation Research*, 489-515. https://doi.org/https://doi.org/10.18187/pjsor.v20i3.4439

[11]Fernández-Villaverde, J., & Guerrón-Quintana, P. A. (2021). Estimating DSGE models: Recent advances and future challenges. *Annual Review of Economics*, *13*(1), 229-252. https://doi.org/https://doi.org/10.1146/annurev-economics-081020-044812

[12]Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, *110*(9), 2819-2858. https://doi.org/10.1257/aer.20191330

[13]Sose, A. T., Joshi, S. Y., Kunche, L. K., Wang, F., & Deshmukh, S. A. (2023). A review of recent advances and applications of machine learning in tribology. *Physical Chemistry Chemical Physics*, *25*(6), 4408-4443. https://doi.org/https://doi.org/10.1039/D2CP03692D

[14]Islam, M. R., Wasif, M., Samia, S., Sourav, M. S. R., & Hossain, A. (2025). The Role of Machine Learning in Forecasting US GDP Growth after the COVID-19 Pandemic. *Journal of Economics, Finance and Accounting Studies*, *7*(2), 163-175. https://doi.org/https://doi.org/10.32996/jefas.2025.7.2.14

[15]Hopp, D. (2022). Economic nowcasting with long short-term memory artificial neural networks (LSTM). *Journal of Official Statistics*, *38*(3), 847-873. https://doi.org/https://doi.org/10.2478/jos-2022-0037

[16]Sakkas, N., Yfanti, S., Daskalakis, C., Barbu, E., & Domnich, M. (2021). Interpretable forecasting of energy demand in the residential sector. *Energies*, *14*(20), 6568. https://doi.org/https://doi.org/10.3390/en14206568

[17]Elijah, W., & Alan, J. (2024). AI-Driven Forecasting of Economic Indicators: A Comparative Study of Machine Learning Models for US Financial Markets. *International journal of Computational Intelligence in Digital Systems*, *13*(01), 76-02. https://doi.org/https://doi.org/10.1007/s43253-024-00114-4

[18]Minh, D., Wang, H. X., Li, Y. F., & Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review*, 1-66. https://doi.org/https://doi.org/10.1007/s10462-021-10088-y

[19]Gyaase, F., Okunlola, O. S., Olusanya, A. M., Adedokun, A., Somoye, O. I., & Erhieyovwe, A. T. (2025). Utilizing machine learning for predictive analysis of emission levels to ensure compliance in refinery operations. *International Journal of Engineering & Technology*, *14*(1), 24-29. https://doi.org/https://doi.org/10.14419/6x85gr51

[20]Covert, I., & Lee, S.-I. (2021). Improving kernelshap: Practical shapley value estimation using linear regression. International Conference on Artificial Intelligence and Statistics,

[21]Kamal, M. S., Sultana, N., Ip, R. H., Bewong, M., Islam, M. Z., & Rahman, A. (2023). Using explainable and interpretable artificial intelligence methods to identify and understand factors affecting resilience. *Available at SSRN 4633354*. https://doi.org/https://dx.doi.org/10.2139/ssrn.4633354

[22]Wu, D. C., Zhong, S., Wu, J., & Song, H. (2025). Tourism and hospitality forecasting with big data: A systematic review of the literature. *Journal of Hospitality & Tourism Research*, *49*(3), 615-634. https://doi.org/https://doi.org/10.1177/10963480231223151

[23]Hasan, M. M. (2023). Understanding model predictions: a comparative analysis of SHAP and LIME on various ML algorithms. *Journal of Scientific and Technological Research*, *5*(1), 17-26. https://doi.org/https://doi.org/10.59738/jstr.v5i1.23(17-26).eaqr5800

[24]Zhang, Y. (2025). Interpretable Machine Learning for Macro Alpha: A News Sentiment Case Study. *arXiv preprint arXiv:2505.16136*. https://doi.org/https://doi.org/10.48550/arXiv.2505.16136

[25]Allgaier, J., Mulansky, L., Draelos, R. L., & Pryss, R. (2023). How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare. *Artificial Intelligence in Medicine*, *143*, 102616. https://doi.org/https://doi.org/10.1016/j.artmed.2023.102616

[26]Ohana, J. J., Ohana, S., Benhamou, E., Saltiel, D., & Guez, B. (2021). Explainable AI (XAI) models applied to the multi-agent environment of financial markets. Explainable and Transparent AI and Multi-Agent Systems: Third International Workshop, EXTRAAMAS 2021, Virtual Event, May 3–7, 2021, Revised Selected Papers 3,

[27]Soybilgen, B., & Yazgan, E. (2021). Nowcasting US GDP using tree-based ensemble models and dynamic factors. *Computational Economics*, *57*(1), 387-417. https://doi.org/https://doi.org/10.1007/s10614-020-10083-5

[28]Bantis, E., Clements, M. P., & Urquhart, A. (2023). Forecasting GDP growth rates in the United States and Brazil using Google Trends. *International Journal of Forecasting*, *39*(4), 1909-1924. https://doi.org/https://doi.org/10.1016/j.ijforecast.2022.10.003

[29] Arsenault, P.-D., Wang, S., & Patenaude, J.-M. (2025). A survey of explainable artificial intelligence (XAI) in financial time series forecasting. *ACM Computing Surveys*, *57*(10), 1-37. https://doi.org/https://doi.org/10.1145/3729531