Data Science in Finance and Economics

DSFE, 4(4): 469–514.
DOI: 10.3934/DSFE.2024020
Received: 02 May 2024
Revised: 25 August 2024
Accepted: 04 September 2024
Published: 11 October 2024

https://www.aimspress.com/journal/dsfe

Research article

Development and application of machine learning models in US consumer price index forecasting: Analysis of a hybrid approach

Yunus Emre Gur*

Department of Management Information Systems, Fırat University, Elazig, Centre, 23100, Turkey

* **Correspondence:** Email: yegur@firat.edu.tr.

Abstract: This study aims to apply advanced machine-learning models and hybrid approaches to improve the forecasting accuracy of the US Consumer Price Index (CPI). The study examined the performance of LSTM, MARS, XGBoost, LSTM-MARS, and LSTM-XGBoost models using a large time-series data from January 1974 to October 2023. The data were combined with key economic indicators of the US, and the hyperparameters of the forecasting models were optimized using genetic algorithm and Bayesian optimization methods. According to the VAR model results, variables such as past values of CPI, oil prices (OP), and gross domestic product (GDP) have strong and significant effects on CPI. In particular, the LSTM-XGBoost model provided superior accuracy in CPI forecasts compared with other models and was found to perform the best by establishing strong relationships with variables such as the federal funds rate (FFER) and GDP. These results suggest that hybrid approaches can significantly improve economic forecasts and provide valuable insights for policymakers, investors, and market analysts.

Keywords: Consumer Price Index (CPI); hyperparameter optimization; hybrid models; machine learning; macroeconomic indicators

JEL Codes: C10, C22, C52, C53

1. Introduction

The Consumer Price Index (CPI) is one of the most critical indicators for measuring inflation and tracking economic trends, as it reflects changes in the cost of a fixed basket of goods and services over time. Inflation erodes the purchasing power of money, significantly influencing consumer behavior, investment decisions, and monetary policy. Thus, accurate CPI forecasts are essential for informed economic assessments, effective policymaking, and governmental initiatives (Reed 2014; Hajdini et al. 2024; Subhani 2009; Arnone and Romelli 2013; Sharma 2016). The importance of CPI in guiding fiscal and monetary policies is well-documented, with central banks and institutions heavily relying on CPI data to gauge inflationary pressures and adjust interest rates accordingly.

However, the complexity of inflation dynamics and the scarcity of high-quality data pose significant challenges in producing reliable CPI forecasts. Traditional econometric models often struggle to capture the nonlinear and nonstationary nature of economic data, leading to suboptimal predictions. Despite advances in machine learning (ML) technology, which offers powerful tools for handling complex and large datasets, CPI forecasting remains a difficult task due to the intricate interplay of multiple economic factors (Yang and Guo 2021; Zhou et al. 2022; Cui et al. 2023). This study investigates the behavior of machine-learning methods for forecasting the US Consumer Price Index from January 1974 to October 2023. Ensemble methods, regression models, and recurrent neural networks are examined. In addition, hybrid models such as LSTM-MARS and LSTM-XGboost are used. Accuracy measures such as MAPE, R-squared, RMSE, and MAE were used to evaluate the models. Figure 1 shows how each model handles the CPI data, including data preprocessing, statistical testing, and training and testing steps. How the behavior of LSTM with respect to the data is realized, how XGBoost works in complex data situations (Dhamo et al., 2022; Zhang and Yang, 2024), and how deep learning and hybrid models can help improve economic forecasts are central areas of observational study. By demonstrating the accuracy of the LSTM-XGBoost model and showing that advanced machine-learning strategies can improve CPI forecasting, this research is intended to benefit multi-asset portfolio managers and policymakers in the field of monetary policy and investment tactics. The paper consists of introduction, methodology, findings, discussion, and literature sections; the initial experimental phases of the study are in Figure 1. The publication also includes Table 1, a list of essential terminology used in the research. According to the study, these models, especially LSTM-XGBoost, are the most suitable for predicting the US CPI because they yield highly precise CPI forecasts.

Table 1. Abbreviations meanings.

Abbreviations	Meaning
LSTM	Long short-term memory
MARS	Multivariate adaptive regression splines
XGBoost	Extreme gradient boosting
CPI	Consumer price index
FFER	Federal fund effective rate
GP	World gold price
OP	US crude oil first purchase price
GDP	US gross domestic product
UR	US unemployment rate
HPI	US house price index
ADF	Augmented Dickey-Fuller
PP	Phillips-Perron
KW	Kruskal–Wallis
SD	Standard deviation
RMSE	Root mean square error
MAPE	Mean absolute percentage error
MAE	Mean absolute error
R^2	R-squared

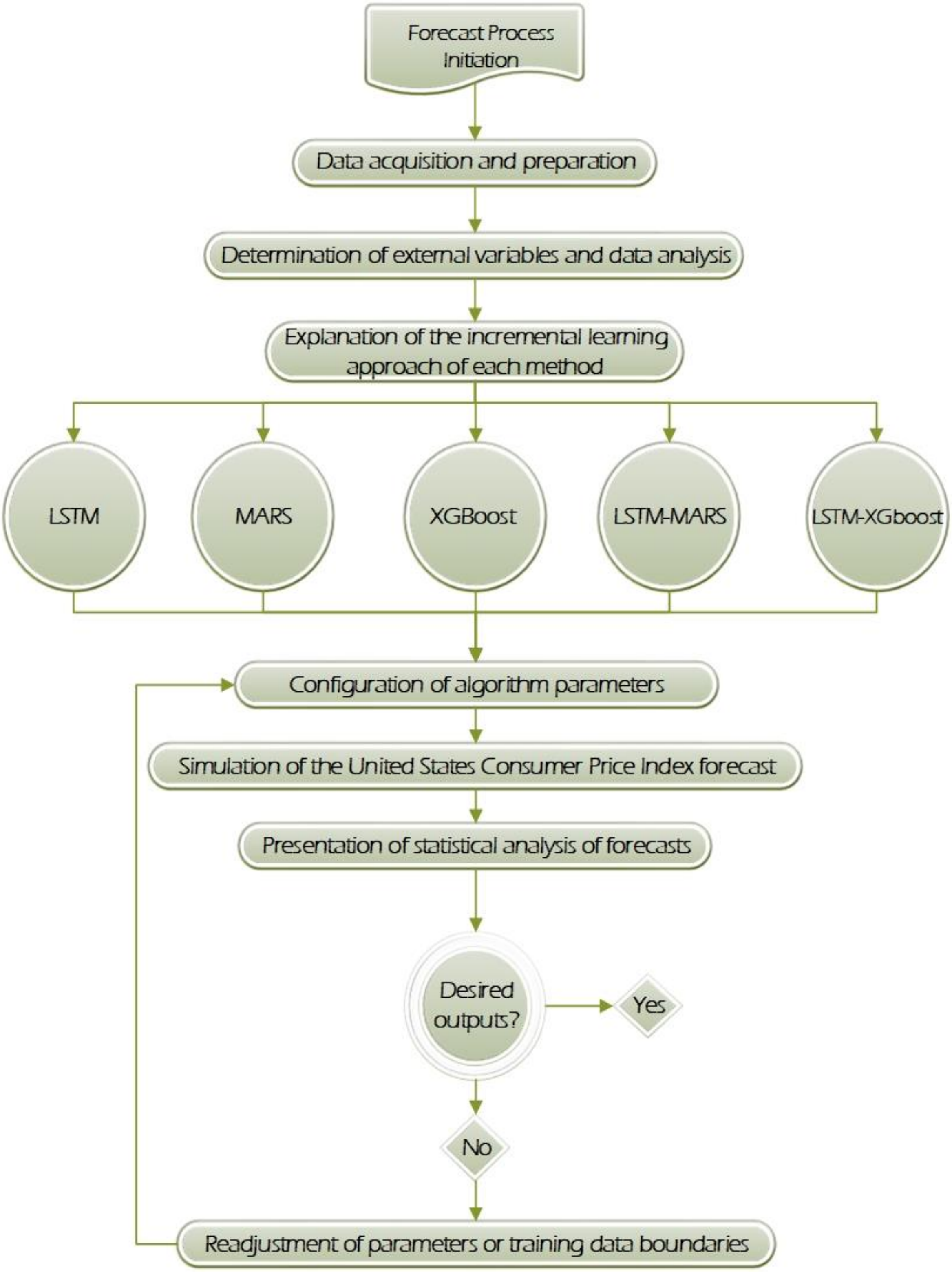

Figure. 1. Initial experimental phases of the study.

There is a growing need to enhance the precision and effectiveness of CPI forecasting methods, as accurate inflation predictions are crucial for economic stability and growth. Traditional econometric models, while widely used, often fall short of capturing the complexity of modern economic systems.

Machine-learning models, with their ability to learn from large datasets and adapt to changing patterns, offer a promising alternative. The evolution of economic modeling has witnessed a significant shift from traditional econometric approaches to the integration of machine learning (ML) techniques, which offer enhanced capabilities in handling complex economic systems. Traditional econometric models, often rooted in linear assumptions and limited data structures, frequently struggle to encapsulate the intricate dynamics of modern economies. This limitation stems from their reliance on predefined functional forms and the assumption of linear relationships among variables, which can lead to oversimplified interpretations of economic phenomena (Imbens and Athey, 2021; Liu, 2023; Wang and Zhao, 2022). In contrast, machine-learning models are designed to learn from vast datasets, allowing them to identify nonlinear relationships and adapt to evolving patterns in economic data, thereby providing a more nuanced understanding of economic behavior (Alhendawy et al., 2023; Chen, 2023).

The integration of machine learning into economic analysis has been propelled by the increasing availability of large-scale datasets, often referred to as big data. This phenomenon has transformed the landscape of economic research, enabling economists to extract valuable insights that were previously unattainable through conventional methods. For instance, machine-learning techniques can effectively manage high-dimensional data, uncovering complex relationships that traditional econometric models might overlook (Liu, 2023; Wang and Zhao, 2022). The challenges posed by big data require innovative analytical tools that can meet the unique characteristics of such datasets, including their size and complexity (Varian, 2014). (Varian, 2014). Consequently, the application of machine learning in economics has emerged as a powerful alternative, capable of enhancing predictive accuracy and providing deeper insights into economic trends (Imbens and Athey, 2021; Wang & Zhao, 2022; Harding and Lamarche, 2021). However, existing literature has not fully explored the potential of hybrid models that combine the strengths of different machine-learning techniques. This research fills that gap by demonstrating how hybrid models like LSTM-XGBoost can outperform traditional methods and standalone machine-learning models, thereby providing more accurate and actionable insights for policymakers and financial analysts.

The motivation for this study stems from the limitations observed in previous research, where models either lacked the necessary data or failed to incorporate the full range of economic indicators affecting CPI. By addressing these gaps, this study not only improves the accuracy of CPI forecasts but also contributes to the broader understanding of how machine learning can be effectively applied in economic forecasting.

2. Related works

Literature reviews of the Consumer Price Index mainly focus on the inflation-related factors within the index. The interaction between gold and oil prices, the federal funds rate, and the US Consumer Price Index (CPI) is an important area of economic research, particularly in understanding inflation dynamics. Recent studies provide valuable insights into how these factors affect the CPI, which reflects broader economic conditions. The fact that gold prices are often seen as a hedge against inflation and that their fluctuations can have a significant impact on the CPI suggests that gold prices are negatively affected by the CPI, the US dollar exchange rate, interest rates, and crude oil prices (Radev et al., 2023). This relationship suggests that as inflation rises, the demand for gold as a safe haven may increase, thus affecting its price and hence the CPI. In addition, Iqbal et al. emphasized the interconnectedness of gold prices, exchange rates, interest rates, and oil prices at the local level, although it focuses primarily on the Pakistani context and does not directly apply to the US (Iqbal et al., 2021).

Oil prices also play an important role in shaping inflation and CPI. Kilian and Zhou analyzed the impact of rising oil prices on US inflation and inflation expectations and revealed that oil price shocks can significantly affect consumer prices (Kilian and Zhou, 2021). This finding is in line with the broader understanding that increases in oil prices generally lead to higher transportation and production costs, which are then passed on to consumers and thus affect the CPI. Moreover, another study investigating the pass-through effects of oil price shocks on various US prices, including the CPI, suggests that oil price fluctuations can have direct effects on inflation (Yilmazkuday, 2024).

The federal funds rate is another critical factor affecting the CPI. Adhikari and Stevens highlighted the direct effects of the federal funds rate on both the CPI and the Producer Price Index (PPI) and showed that changes in monetary policy can lead to significant adjustments in consumer prices (Adhikari and Stevens, 2024). This relationship underlines the importance of the federal funds rate as a tool to manage inflation and stabilize the economy. This idea is also supported by evidence that consumer inflation expectations are closely tied to changes in the federal funds rate, which in turn can affect spending and pricing behavior (Knotek et al., 2024). Moreover, the combined effects of the federal funds rate as well as gold and oil prices on the CPI are crucial for understanding inflation dynamics. The study investigating the impact of global supply chain pressures and crude oil prices on inflation rates reveals that these factors collectively affect the CPI in both advanced economies and emerging markets (Ye et al., 2023). This research highlights the multifaceted nature of inflation, where various economic indicators interact to shape consumer price movements.

The Consumer Price Index (CPI) is influenced by a variety of broad economic indicators such as GDP, unemployment, inflation, and interest rates, while other factors such as currency transactions and real estate prices tend to have more indirect effects. The relationship between GDP and CPI is well documented in the economic literature. Shiferaw emphasized the interaction between GDP, unemployment, and inflation, showing that changes in GDP significantly affect inflation rates, which in turn affect the CPI (Shiferaw, 2023). Unemployment is another critical factor affecting CPI. Studies emphasizing the relationship between economic growth and inflation suggest that higher unemployment rates may lead to lower inflation, which in turn will affect the CPI (Upadhyaya and Kharel, 2022). Consumer spending is another vital component that directly affects the CPI. Analyses show that macroeconomic factors such as consumer spending, inflation, and GDP are interlinked and consumer spending is the primary driver of inflation (Islam et al., 2024). Government spending also affects the CPI, especially through fiscal policies that affect aggregate demand. The study examining the impact of government spending on economic growth and inflation suggests that increased government spending can lead to higher inflation rates, which are then reflected in the CPI (Correa, 2023). In contrast, factors such as foreign exchange transactions and real estate prices tend to affect the CPI indirectly. For example, fluctuations in exchange rates can affect import prices, which in turn can affect consumer prices. However, this effect is usually mediated through broader economic indicators such as inflation and GDP. Similarly, real estate prices can affect the cost of housing, an important component of the CPI, but their impact is often indirect, operating through changes in consumer spending and general economic conditions (Mohan et al., 2019). As a result, it is extremely important to recognize these factors in order to accurately predict the US CPI.

Recent studies have applied various econometric models to forecast the CPI, reflecting the complexity and multifaceted nature of inflation dynamics. A notable study was conducted by Njenga, who used the Holt-Winters model to analyze and forecast the CPI in Kenya and South Africa. The findings revealed that the model effectively captures seasonal patterns and trends in CPI data, demonstrating its applicability in different economic contexts (Njenga, 2024). Another important contribution comes from Imron et al., who focused on outlier detection in CPI forecasting using the ARIMA model. While their research emphasized the effectiveness of the ARIMA model in forecasting CPI, it also addressed the challenges posed by outliers in the data (Imron et al., 2022). This approach is particularly important as it highlights the need for robust modeling techniques that can deal with irregularities in economic data. In another comparative analysis, traditional SARIMA models as well as machine-learning techniques were examined for modeling CPI data in Pakistan. The study found that while SARIMA models provide a solid foundation for forecasting, machine-learning models can improve forecast accuracy by capturing complex nonlinear relationships in the data (Qureshi et al., 2023). This finding is in line with the growing trend of integrating machine-learning methods into traditional econometric frameworks to improve forecasting performance. Zhang and Yang investigated the impact of unemployment and construction spending on housing value indicators using vector autoregression (VAR) models. The study emphasized the interdependence of various economic indicators, including the housing consumer price index (HCPI), and how these relationships can inform the CPI forecast (Zhang and Yang, 2024). This research highlights the importance of considering multiple economic variables when modeling the CPI as they can significantly affect consumer prices. There has been extensive research in various geographical areas on whether machine-learning algorithms are superior to traditional econometric techniques in predicting inflation. Ivașcu stated that in data-sparse environments, simpler autoregressive models outperformed more sophisticated machine-learning methods (Ivașcu, 2023). Meanwhile, in Brazil, random forest and XGBoost outperformed traditional methods, benefiting from a wider range of data for forecasting, such as breakeven inflation rates and survey expectations (Araujo and Gaglianone, 2023). Nguyen et al. used various machine learning and classical econometric models to forecast the US CPI. In the study, multivariate linear regression (MLR), support vector regression (SVR), autoregressive distributed lag (ARDL), and multivariate adaptive regression splines (MARS) models were applied using monthly data from January 2017 to February 2022. The performance of the models was evaluated using metrics such as mean absolute percentage error (MAPE), mean absolute error (MAE), root mean square error (RMSE), and R-squared (R^2). The findings showed that the MARS model has the highest accuracy in the testing phase compared to the other models and that the impact of crude oil price (OP), world gold price (GP), and federal funds effective rate (FFER) on the US CPI is particularly strong. It has been suggested that the MARS model is an important tool for forecasting various aspects of the US economy and evaluating the effectiveness of economic policies (Nguyen et al., 2023). Sarangi et al. used various machine-learning techniques to model the CPI in South Africa. The study highlighted the increasing scope of machine-learning tools for forecasting and showed that these models can capture complex patterns in the data that traditional methods may miss. The findings showed that machine-learning approaches can provide lower forecast errors compared to traditional methods, suggesting a shift toward more data-driven forecasting techniques (Sarangi et al., 2022). Sibai investigated the accuracy of six different machine-learning methods in forecasting the Saudi Arabian CPI. The study used a comparative framework to evaluate the performance of these models against traditional econometric approaches. The results showed that machine-learning models, especially those using ensemble techniques, significantly outperformed traditional models in terms of forecasting accuracy, strengthening the potential of machine learning in economic forecasting (Sibai, 2024). Medeiros et al.

examined the benefits of machine-learning methods in a data-rich environment for inflation forecasting. Their findings show that machine-learning models, especially those with a large number of covariates, systematically outperform traditional measures and that the integration of machine learning in economic forecasting can lead to more accurate forecasts (Medeiros et al., 2019). In addition to these studies, Anderson and Ulrych utilized deep neural networks (DNN) to accelerate the pricing of American options. The research was conducted in order to provide faster and sufficiently accurate pricing compared to traditional methods, especially in environments where fast and accurate pricing is critical, such as market making. In the study, a DNN was trained based on the Heston stochastic volatility model and this network learned option prices from the data set obtained from the traditional methods used during training. The performance of the model showed a significant improvement in both speed and accuracy, especially when compared to traditional methods such as Monte Carlo and partial differential equations (PDE). The findings showed that the proposed DNN model significantly improves the balance between speed and accuracy in American-type option pricing and these methods can find a wide application in financial markets. By offering significant advantages in terms of speed and explainability, the model can be considered a reliable and transparent pricing tool for regulators and investors. According to Akbulut (2022), both linear and nonlinear machine-learning models were more accurate in predicting Turkey's inflation as data and computing power increased. In Russia, boosting and random forest techniques were found to be as accurate as random walk and autoregression in inflation rate forecasting (Baybuza, 2018). In the US, random forest data-rich environments were invaluable (Medeiros et al., 2019), but in Sri Lanka, the best inflation rate prediction model was support vector machine regression (Bandara and De Mel, 2024). Mohammed et al. (2023) used deep learning and hybrid ensemble techniques—LSTM, BiLSTM, and ANN-LSTM-AdaBoost—in CPI forecasting in India. Hybrid models generated time-efficient and GPU-efficient results. Meanwhile, hierarchical recurrent neural network was developed by Barkan et al. (2023) to analyze different segments of CPI inflation in more detail. Feng (2024) also mixed the LSTM and the traditional econometric model, the VAR model, for CPI forecast. In this case, the LSTM model was used less erroneously, and VAR was used to analyze Granger causality, a significant association. The result of combining the two techniques was still better.

While the current models present advancements in CPI prediction, there is a clear need to improve the rate of precision and effectiveness. Among the most significant advancements are sophisticated machine-learning approaches, such as ensemble methods, as they blend several predictive models to increase predictive accuracy (Choi and Lee, 2018; Hao et al., 2023). The method combines the predictions of different sources and, in CPI prediction, results in higher precision (Muruganandam and Arumugam, 2023). Ensemble machine-learning methods greatly benefit multivariate time-series analysis, especially when applied to complex problems. Such models merge multiple neural networks with autoregressive models and other computational approaches to create models that combine the best from each track and are generally highly adaptive forecasting tools (Widiputra et al., 2021). The integration of MARS or XGBoost with LSTM is currently a very promising tool for predictive tasks such as forecasting for macroeconomic indicators, e.g., for the US CPI (Vlachas et al., 2018).

The hybrid models based on the strengths of both approaches demonstrate increased predictive accuracy and usefulness (Wei et al., 2017). For instance, the results of the GARCH-ATT-LSTM model become not only more easily interpretable but also more accurate (Gao and Kuruoglu, 2023). It happens because the attention mechanisms identify the most critical data characteristics, enabling the LSTM to operate more accurately while, at the same time, the GARCH is used as a strong basis for volatility prediction. Other hybrid models, such as LSTM-GPR and LSTM-BO-XGBoost, also demonstrate satisfactory results by generating tighter prediction intervals than LSTM while offering hardly any counterbalance on the quality side. Thus, the LSTM-BO-XGBoost hybrid model performs better than LSTM in stock market prediction, and the LSTM-GPR hybrid is more precise in terms of COVID-19 prediction (Wang et al., 2021; Tian et al., 2021; Reddy et al., 2022). Lastly, coupling LSTM with XGBoost and MARS results in more intensive data analysis, thus improving feature extraction and analysis of the nature of multivariate time-series trended data (Temür and Yildiz, 2021). Overall, the impact is prominent in the capability to handle the nuances of complex data, significantly improving predictive accuracy.

To sum up, this paper argues in favor of the implementation of the combination of a rich dataset with macroeconomic takeaways and the excellent use of LSTM (RNN), MARS regression technique, XGBoost ensemble method, and hybrid approaches, such as LSTM-GBoost or LSTM-MARS. Seeing how the misconceptions persist in the literature concerning the subject matter, the choice of the proposed machine-learning tools to reach for an increase in the accuracy of the CPI forecasts, the explanations of the correlations within the CPI data, and the quality of the CPI forecasts is reasonable.

3. Methodology

Data collection was aimed at forecasting US Consumer Price Index for every month from January 1992 to December 2023, which gives 382 monthly released data. During the process, approximately 80% of the data were randomly defined based on the number of points for training, while the rest were dedicated for validation purposes. The models LSTM, MARS, and XGBoost, as well as the hybrid models LSTM-XGBoost and LSTM-MARS were selected to prepare the sample forecasts. They included variables such as OP, GP, FFER, US GDP, UR, and house price index (HPI). The step was finalized with a comparison of training and validation outcomes based on MAPE, MAE, R^2, and RMSE, as presented in Figure 2.

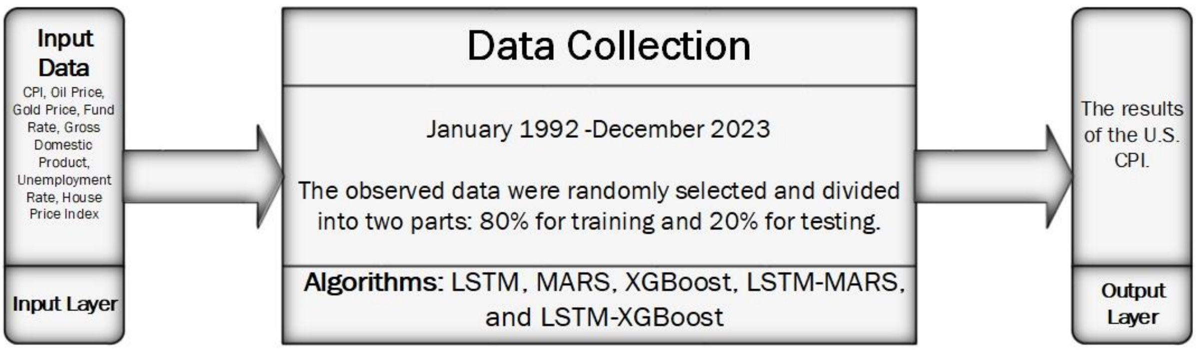

Figure. 2. Sequential experimental steps of the study.

3.1. Statistical description

The dataset was collected from credible sources such as the US Energy Information Administration and Statista for crude oil prices and gold prices. Meanwhile, the Federal Reserve Economic Data provided broader data concerning CPI, FFER, GDP, UR, and HPI. Table 2 clearly demonstrates the results of the statistical analysis, showing normal distribution for the variables, which is evidenced by a standard deviation and discrepancy among the results as the variables report little values of kurtosis and skewness estimated at near 0 as evidence for normal distribution (Groeneveld and Meeden, 1984; Cain et al., 2017).

Table 2. Statistical Analysis.

	GP	OP	FFER	GDP	UR	HPI	CPI
Standard Deviation	563,98	29,84	2,17	5,54	0,02	74,99	43,24
1st Quartile	359,17	19,96	0,18	9,98	0,04	135,81	168,93
Median	782,03	44,66	1,97	14,43	0,05	190,41	211,12
Mean	893,44	48,43	2,5	14,55	0,06	197,93	207,81
3rd Quartile	1.323,51	70,72	4,8	18,46	0,07	223,85	237,75
Min	256,08	8,03	0,05	6,24	0,03	102,24	138,3
Max	1.999,77	128,08	6,54	27,82	0,15	415,97	307,62
Kurtosis	−1,34	−0,91	−1,45	−0,64	1,89	0,84	−0,8
Skewness	0,39	0,48	0,31	0,44	1,25	1,03	0,25
Coefficient of Variation	0,63	0,62	0,87	0,38	0,31	0,38	0,21

Analysis of correlation was possible using Pearson's coefficient within the ranges of −1 to +1, summing to the highest positive correlation between the CPI and GDP, HPI, and GP and negative correlation with FFER; however, the changes in the UR variable had little to no effect on CPI. As presented in Figure 3, the correlation matrix allowed the in-depth understanding of relations existing between the presented variables, determining possible economic models that could help predict the behavior of the CPI.

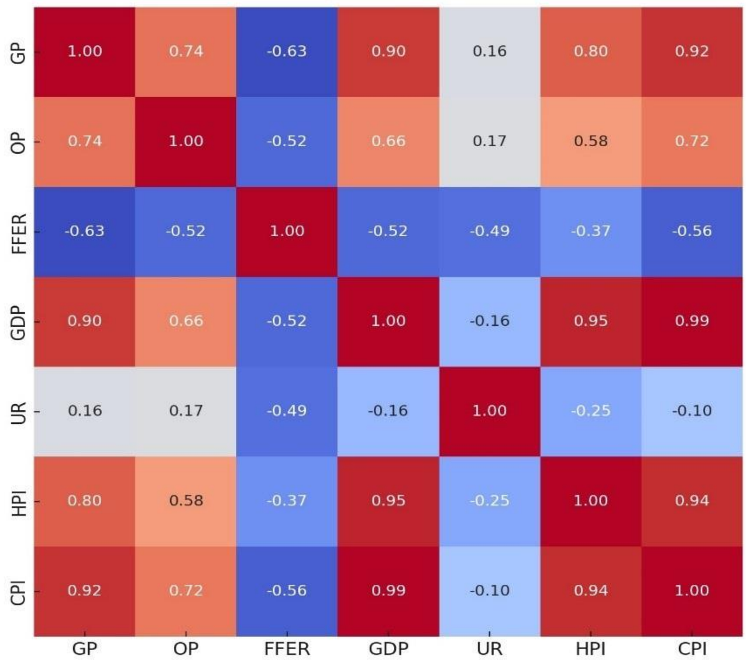

Figure 3. Pearson's coefficient correlation matrix.

Since the CPI is thought to be the outcome variable, the statistically significant correlations extending toward the remaining indicators serve a predictive purpose. Table 3 incorporates the aspects of the analytical framework to provide a clear definition and summary of all variables in concern.

When observing machine-learning applications, it was stressed to ensure dataset integrity to harvest the most obvious model; however, to do so, accurate economic data must be sought for easy collection.

Table 3. Inventory of the independent variables analyzed in the data set and the dependent variable that is the focus of the research.

Kind	Variable	Definition	Data source
Independent variables	x_1	World gold price	Statista
	x_2	US crude oil first purchase price	US Energy Information Administration
	x_3	Federal fund effective rate	Federal Reserve Economic Data
	x_4	US gross domestic product	Federal Reserve Economic Data
	x_5	US unemployment rate	Federal Reserve Economic Data
	x_6	US house price index	Federal Reserve Economic Data
Dependent variable	y	US consumer price index	Federal Reserve Economic Data

3.2. Unit root test

Unit root tests, such as the Augmented Dickey-Fuller test (Dickey and Fuller, 1981) and Phillips-Perron tests (Phillips and Perron, 1988), represent a feasible approach to determine whether one's data series is stationary and address the problem of autocorrelation since it incorporates lagged "formulations" of the data (DeJong et al., 1992; Paparoditis and Politis, 2018).

$$\Delta Y_t = C_0 + \beta_t + \gamma Y_{t-1} + \sum_{i=1}^{z} \delta_i \Delta Y_{t-i} + \epsilon_t \quad (1)$$

$$\Delta Y_t = C_0 + \beta_t + \gamma Y_{t-1} + \epsilon_t \quad (2)$$

The time-series variable is denoted as Y, and t refers to a linear time trend. Δ indicates the operation to calculate the first difference of the series. C_0 represents the constant part of the equation, z denotes the optimal lag length for the Augmented Dickey-Fuller test equation, and Ε ultimately stands for the stochastic error term within Equations (1) and (2). The primary hypotheses for the ADF and PP unit root tests are as follows:

H0. The series has a unit root (i.e., is nonstationary). In this case, the hypothesis $\gamma = 0$ is tested.

H1. There is no unit root in the series (i.e., it is stationary). The alternative hypothesis is $\gamma < 0$.

The null hypothesis (H0) is statistically rejected if the p-values of the observation series are low at specified significance levels, such as 1%, 5%, or 10%.

3.3. The Johansen cointegration test

Although unit root tests at least show that time-series data become stationary after one differencing, they do not imply that behaviors are consistent up to time nor imply a pattern is stable. We use the Johansen cointegration test to determine long-run relationships between two nonstationary time series. The test also tells us the number and types of cointegration links that exist within time series (Poh and Tan, 1997; Toraman and Basarir, 2014; Naidu et al., 2017). The trace test tests the null hypothesis that *r* variable is cointegrated against the alternative that *w* variable is cointegrated, where *w* is the number of endogenous variables. *r* ranges from 0 up to w-1 to specify the rank of the data matrix (Johansen, 2009; Hossain and Mitra, 2017; Raheem Ahmed et al., 2017).

3.4. Data normalization

One of the most critical issues concerns the data normalization process. Sometimes, different attributes in a dataset have their own scales, and this process is applied to prevent any one attribute from being dominant due to its scale. Likewise, preserving the relationship between different attributes maintains the level of precision of the analysis performed later. Accordingly, some techniques, such as Min-Max normalization, are applied, and the data are linearly adjusted to fit within a specified range, which keeps their relationship unchanged and makes them easily comparable (Zyprych-Walczak et al., 2015; Wang et al., 2021; Mulenga et al., 2021; Ampomah et al., 2021; Henderi, 2021).

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3}$$

The representation of this normalization in mathematics is shown in Equation (3). In this equation, x_{scaled} indicates the adjusted data value after normalization, and x_{min} and x_{max} are the minimum and maximum values of the data set (Jadiya et al., 2020). This adjustment can proportionally shift the data to the specified range, which can maintain the distribution of the data and make it possible to compare the data (Djordjevic et al., 2022).

3.5. Long short-term memory (LSTM)

Long short-term memory (LSTM) networks have gained prominence due to their proficiency in identifying long-term dependencies within sequential data. This characteristic has facilitated their widespread adoption across various fields, including natural language processing (Zazo et al., 2016; Setyanto et al., 2022), financial management (Budiharto, 2021; Park et al., 2022; Jakubik et al., 2023), environmental science (Lees et al., 2021; Liu et al., 2023), and healthcare (Kumar and Subha, 2019; Amin et al., 2020). LSTM is designed to identify and scan for these patterns, especially when analyzing the elements within a given window). It is important to note that their adoption is largely driven by modern models that ensure high standards in both data accuracy and the development of analysis tools (Sagheer and Kotb, 2019; Wan et al., 2019; Saputra et al., 2022; Cahyono et al., 2023; Dinh et al., 2023; Zhou et al., 2023).

LSTM is yet another architecture in the family of RNNs' variants. As a distinctive variant, this type is capable of handling more refined long-term dependencies in time-series data (Bhanja and Das, 2021). Prediction of future values with its help is more suitable, as the structure of the described neural network accommodates the long sequence of the learning of multivariate observation (Khodabakhsh et al., 2020).

The overall architecture of the LSTM network is based on three major figures—the forget gate, as explained in Equation (4), the input gate, detailed by Equations (5) and (6), and the output gate, which can be seen from the first LSTM block Equations (7) and (8). By modulating updates to the cell state, the network becomes capable of retaining and removing pieces of information, optimizing memory use in a successful and productive way (Siami-Namini et al., 2019).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f \tag{4}$$

Where W_f is the weight matrix for the forget gate, h_{t-1} is the previous cell output, x_t is the current input, b_f is the bias vector for the forget gate, and σ is the sigmoid activation function. This gate allows the cell to forget old information.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{5}$$

$$g_t = \tanh(W_g \cdot [h_{t-1}, x_t] + b_g) \tag{6}$$

i_t is the activation of the input gate, and g_t is the new candidate that should be added to the cell state. Also, W_i and W_g are weight matrices, and b_i and b_g are bias vectors.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t * \tanh(c_t) \tag{8}$$

o_t is the activation of the output gate, and h_t is the output of the existing cell. Finally, W_o is the weight matrix of the output gate, b_o is the bias vector, and c_t is the cell net after computing all. For instance, o_t is the output value, and h_t can be a value between -1 and 1.

3.6. Multivariate adaptive regression splines (MARS)

MARS is a time series toolkit that applies backward partitioning to turn the source data with a particular output function attached to it, which makes more pieces much longer but reduces slope; distance measurements after these parts have been migrated and now come as spline fittings. MARS solves the inefficiencies associated with linear modeling that arise when the true output has nonlinear behavior at higher dimensional points and great flexibility (Lewis and Stevens, 1991).

However, nowadays the most common use of MARS are various data science applications, environmental studies, and engineering. It provides a computationally efficient way to uncover hidden patterns in data sources such as predicting workforce requirements and job trends (Sharda et al., 2008; Balshi et al., 2009; Arthur et al., 2020; Shahbaz et al., 2020; Hasanah, 2021; Murat, 2023).

The effectiveness of MARS lies in its capability for modeling a variety of time years and thus the nonlinear nature of data. It can make consistent predictions over short periods or into the future. In addition, MARS has an elegant design that provides a flexible framework tailored to both classification and prediction tasks that go hand in hand with this method. With MARS, the modeling of the dynamics of complex data configurations is effective and efficient (Porcher and Thomas, 2003; Rezaie-Balf et al., 2017). Equation (9) highlights the central application of MARS in modeling nonlinear dynamics in data (Koc et al., 2019; Yildiz and Ozdemir, 2022).

$$GCV = \frac{\frac{1}{N}\sum_{i=1}^{n}\left[y_i - \hat{f}(x_i)\right]^2}{\left[1 - \frac{C(B)}{N}\right]^2} \tag{9}$$

Where N is the total number of observations in the dataset, and $C(B)$ is a complexity penalty that grows linearly with the number of subspaces in the model and has the expression shown in Equation (10) (Friedman, 1991; Friedman et al., 2010).

$$C(B) = (B + 1) + d(B) \tag{10}$$

3.7. Extreme gradient boosting (XGBoost)

Praise for XGBoost, one of the most highly rendered predictors in different industries (including healthcare, insights markets, and finance), commonly includes the following properties (Chen and Guestrin, 2016; Tan et al., 2021; Bhati et al., 2020; Nguyen et al., 2020; Papíková and Papík, 2022;

Simsek, 2024). XGBoost is recognizable for its fast processability and high prediction accuracy; the algorithm is always considered self-explanatory, enabling a successful combination with natural languages, like medical diagnostic reports or the novelty of predicting trends in the financial market (Guo et al., 2022; Yuan et al., 2020; Pan et al., 2023; Brzan et al., 2017; Elgui et al., 2020; Cao et al., 2020; Jung et al., 2023).

Ranked as one of the highest waterfronts in the multivariate time-series prediction, XGBoost has higher flexibility and outcome rates compared with traditional models, such as ARIMA and random forest (Wang et al., 2017; Jin et al., 2019; Alim et al., 2020; Lv et al., 2021; Noorunnahar et al., 2023; Li and Zhang, 2018; Wang and Ye, 2020; Sukarsa et al., 2021). An intelligent gradient-boosting library develops decision trees in iterations to eliminate prediction inaccuracy and has established itself as a dynamic tool in multivariate regression analysis (Zhai et al., 2020; Xu et al., 2022).

$$L(\emptyset) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{11}$$

$L(\emptyset)$ denotes the error function of the model and l marks the loss function that we apply to our metric after the computation. In the above example, y_i displays the real value for the ith instance, and $\hat{y}_i$ stands for the calculated value of the model for the ith instance. Finally, the total number of trees in the model is defined by K and f_k marks the k-th tree within the whole collection. Ω is the complexity of these trees.

Equation (12) describes the complexity measure of the trees with a function that integrates the number of leaves divided by regularizing the score in the leaves.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{J=1}^{T} w_j^2 \tag{12}$$

Where T is the number of leaves, w_j is the score of the j'th leaf, and γ and λ are regularization terms.

3.8. LSTM-MARS model

The specificity of the LSTM-MARS framework is that it combines the power of LSTM on sequential data with MARS's ability to manage nonlinear relationships to boost multivariate time-series prediction. In this case, Anagnostis et al. argued that the presentation offers LSTM the possibility to spot temporal ideals, while MARS can explore nonlinear actions between distinct variables (Anagnostis et al., 2022). The model consists of two stages: LSTM identifies features associated with the time factor and MARS utilizes this information to prepare the final predictions.

In this case, the LSTM model is responsible for acquiring the rhythmic features and dependencies within the dataset. For instance, LSTM for each input vector X_t in time t produces the output H_t which is the projection of what the model assigned for X_t in time t:

$$H_t = f_{LSTM}(X_t, H_{t-1}) \tag{13}$$

Where f_{LSTM} is the function of the LSTM model. X_t is the input vector at time t, and H_{t-1} is the latent state obtained in the previous step.

The outputs (or feature representations) from the LSTM model are then given as inputs to the MARS model. The produced features are then used to predict the final prediction $\hat{Y}$, as shown in Equation (14).

$$\hat{Y} = f_{MARS}(H) \tag{14}$$

where f_{MARS} is the prediction function for the MARS model, and H denotes the outputs or feature representations of the LSTM model produced in all time steps. Additionally, H consists of an output array $H_1, H_2, \dots, H_T$,, where T is the full length of the time series. The overall estimation procedure of the LSTM-MARS model using this dual-step methodology is given in Equation (15):

$$\hat{Y} = f_{MARS}(f_{LSTM}(X)) \tag{15}$$

Here, X reflects the full input dataset throughout the indicated time range, whereas $\hat{Y}$ shows the depicted output prediction as predicted by this model. The growth of the above algorithm therefore outlines the fundamental and multi-stage process through which the combined LSTM and MARS methodology generates forecasts. Initially, the proposed LSTM segment seeks to encode dataset features into the model through multiple embedded layers. The grown seeds are in turn fed into the MARS segment in order to explicitly identify high-order relationships and generate the final predictions. The above algorithm therefore aims to forecast both the temporal properties and its complex correlational qualities in a more mainstream way, benchmarking the native strengths of its component modeling methods.

3.9. Proposed model (hybrid LSTM-XGBoost)

The LSTM-XGBoost model is based on the strengths of both LSTM and traditional decision trees to enhance prediction accuracy in complex data environments. The LSTM component has the power to observe time-sensitive pattern detection. XGBoost deals well with intricate data interactions. According to Li et al. (2022), the modeling process is depicted graphically in two stages. At first, the LSTM model is built and trained to predict the CPI. To work on predictors during this training phase, then you train with feature data. XGBoost follows the LSTM outputs, forming a two-tiered strategy that develops both fine timing patterns and complex interactions for improved accuracy of predictions. This way includes multivariate time-series data input to the LSTM; it produces features for the XGBoost, and detailed expansions of LSTM outputs at each time step are given by Equation (16).

$$H_t = f_{LSTM}(X_t, H_{t-1}) \tag{16}$$

Where X is the input vector at time t, H_{t-1} is the latent state from the previous timestep, and f_{LSTM} is the function used for the LSTM model. Finally, the features produced by the LSTM are fed into the XGBoost model, which is trained on these features to produce the final outcomes. The following Equation (17) summarizes the algorithm's steps:

$$\hat{Y} = f_{XGBoost}(H) \tag{17}$$

In this case, H is the union of all of the feature representations generated by the LSTM model at each timestep while $f_{XGBoost}$ is the predictive function of the XGBoost model used.

The hybrid model makes accurate predictions via LSTM's ability to learn temporal structure and XGBoost's ability to perform nonlinear regression. Consequently, this approach captures well both the large volume, long-term dependences within time series and the complicated inherent interaction configuration and dynamics existing within the data. The formula for the forecasting method is shown in Equation (18).

$$\hat{Y} = f_{XGBoost}(f_{LSTM}(X)) \tag{18}$$

Where X is the multivariate input dataset, $f_{LSTM}(X)$ is the feature representations generated by the LSTM model, and $f_{XGBoost}$ is the predictions made over these features.

The research introduces a hybrid forecasting model combining LSTM and XGBoost, specifically tailored for predicting the US CPI. This model leverages LSTM's capability to detect long-term sequences in time-series data with XGBoost's ensemble learning approach to manage complex data interactions, thereby enhancing forecast accuracy. Initially, the model uses LSTM to process CPI-influencing variables and generate feature representations (as depicted in Figure 4).

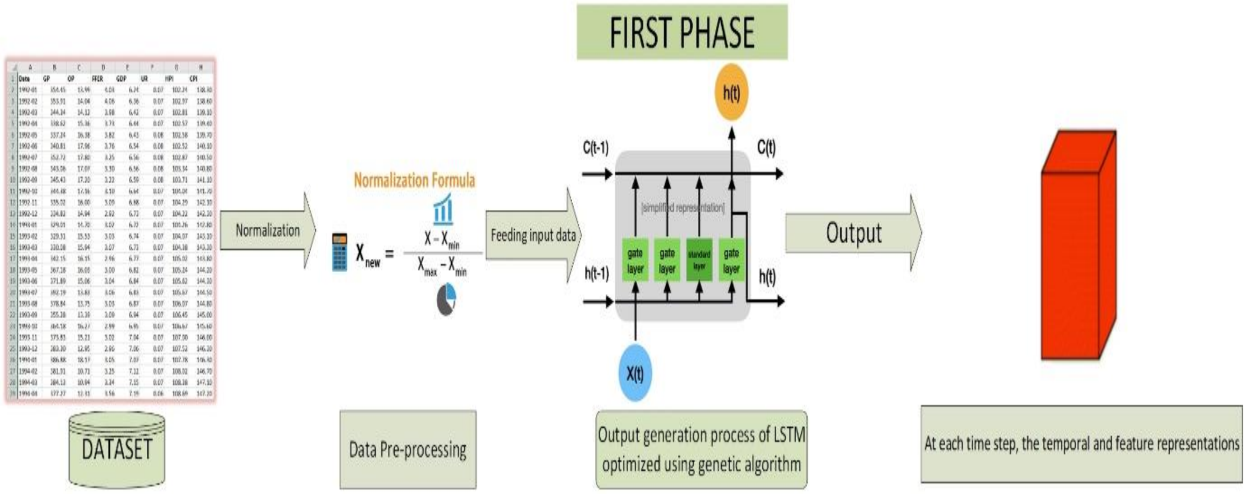

Figure 4. At each time step, the first stage where features extracted from the dataset produce a set of latent states or feature representations.

These LSTM-derived features are then inputted into XGBoost in a second phase (illustrated in Figure 5) to finalize the forecasts. This dual-phase method maximizes the strengths of both models, optimizing the hyperparameters through a genetic algorithm to improve the precision and adaptability of forecasts. This strategy is also applied to the LSTM-MARS model, exploring its potential for CPI forecasting within the study.

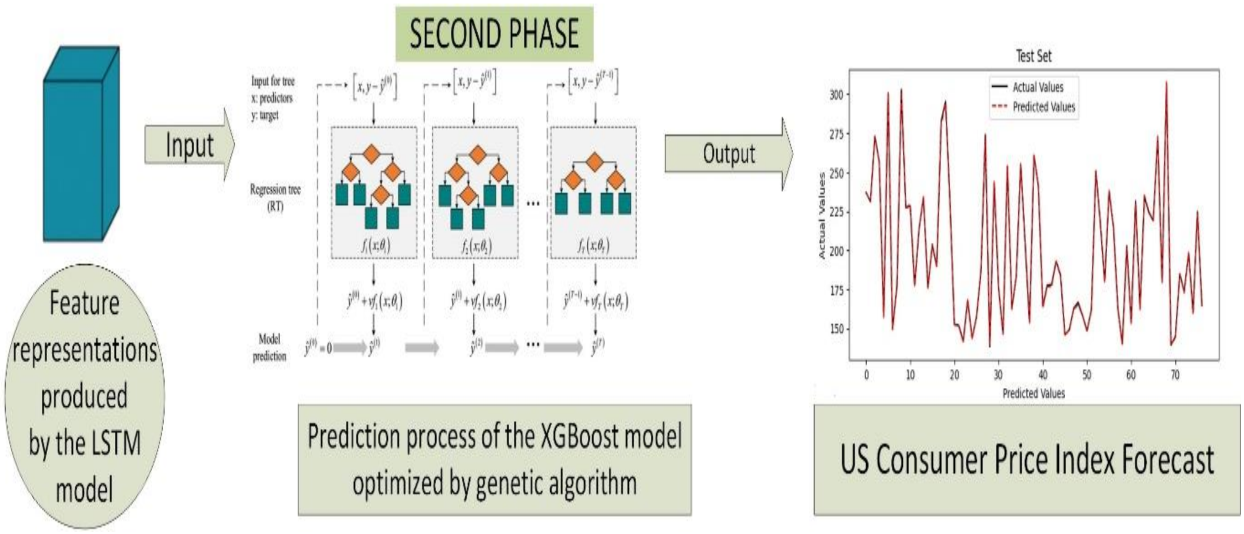

Figure 5. Final prediction process of the LSTM-XGBoost model.

4. Results

In this study, the PyCharm IDE was utilized for the preprocessing, analysis, development, evaluation, and deployment of predictive models, specifically focusing on US CPI forecasts using Python. PyCharm supports code writing, debugging, and project management, streamlining the development process for complex tasks like machine learning. The models were developed and evaluated using PyCharm and Python, supported by 16 GB of RAM, a 3.20 GHz Intel® Xeon® CPU, and a 4 GB GPU, providing the necessary computational power for high-performance tasks. Python's libraries and PyCharm's features greatly assist in building sophisticated machine-learning models. The effectiveness of these models is measured using four performance indicators: MAPE (Farsi et al., 2021), MAE (Li and Yang, 2022), RMSE (Li et al., 2020), and R^2 (Sun and Tian, 2022). Each of these metrics is articulated through Equations (19), (20), (21), and (22), correspondingly.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \tag{19}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{20}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{21}$$

$$R^2 = 1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{\bar{y}})^2} \tag{22}$$

Here, n denotes the number of aggregate values, and y_i, $\hat{y}_i$ and $\bar{\bar{y}}$ denote the true value at time i, the value predicted by the model, and the arithmetic mean of the predicted values at time i, respectively.

4.1. Stationarity test results

Findings shown in Table 4 reveal the time series as nonstationary at the chosen level, given that the p-values for both the Augmented Dickey-Fuller and Phillips-Perron tests exceed 0.01 across all lag intervals. Nonetheless, upon first differencing, all variables under observation achieve stationarity, suggesting that the time series for these variables can be modeled with relative ease (Amalu et al., 2021).

Table 4. Unit root test results.

Variable	ADF statistic			PP statistic		
At level						
	None	Intercept	Trend and intercept	None	Intercept	Trend and intercept
CPI	2.834 (0.999)	1.378 (0.997)	−1.587 (0.797)	7.549 (1.000)	2.232 (0.998)	0.764 (1.000)
GP	1.782 (0.982)	0.262 (0.975)	−1.975 (0.614)	1.765 (0.981)	0.235 (0.974)	−2.023 (0.588)
OP	−0.995 (0.289)	−2.523 (0.109)	−3.303 (0.065)	−0.528 (0.484)	−1.926 (0.319)	−2.524 (0.315)
FFER	−1.480 (0.129)	−2.727 (0.069)	−3.432 (0.047)	−1.141 (0.230)	−1.915 (0.324)	−1.690 (0.754)
GDP	7.066 (1.000)	3.284 (1.000)	1.252 (1.000)	8.187 (1.000)	3.116 (1.000)	0.897 (1.000)

Continued on next page

Variable	ADF statistic			PP statistic		
UR	−1.149 (0.228)	−3.174 (0.021)	−3.187 (0.086)	−1.115 (0.240)	−3.078 (0.028)	−3.097 (0.106)
HPI	1.343 (0.954)	0.108 (0.966)	−2.059 (0.568)	4.714 (1.000)	2.912 (1.000)	0.817 (1.000)
At first difference						
	None	Intercept	Trend and intercept	None	Intercept	Trend and intercept
CPI	−1.678* (0.088)	−3.032** (0.031)	−3.409** (0.049)	−10.544*** (0.000)	−12.170***(0.000)	−12.144*** (0.000)
GP	−16.246*** (0.000)	−16.421 (0.000)	−16.472*** (0.000)	−16.577*** (0.000)	−16.518*** (0.000)	−16.504*** (0.000)
OP	−12.925*** (0.000)	−12.916*** (0.000)	−12.900*** (0.000)	−12.015*** (0.000)	−11.991*** (0.000)	−11.970*** (0.000)
FFER	−4.252*** (0.000)	−4.254*** (0.000)	−4.241*** (0.000)	−11.193***(0.000)	−11.183***(0.000)	−11.245***(0.000)
GDP	−3.183*** (0.001)	−15.928*** (0.000)	−11.703*** (0.000)	−16.115***(0.000)	−16.497***(0.000)	−16.382***(0.000)
UR	−11.616*** (0.000)	−11.611*** (0.000)	−11.596*** (0.000)	−19.761***(0.000)	−19.755***(0.000)	−19.722***(0.000)
HPI	−1.595 (0.104)	−2.421 (0.135)	−2.784*** (0.007)	−4,936***(0.000)	−6,270*** (0.000)	−7.143***(0.000)

Table 4 presents the t-statistics and p-values in parentheses.

***/**/* represent significance levels at 1%, 5%, and 10%.

The terms “none”, “intercept”, and “trend and intercept” denote unit root test scenarios.

4.2. Johansen cointegration test results

Specifically, if the relationship between variables is nonstationary, then accepting the null hypothesis implies no cointegration or long-term equilibrium (Engle and Granger, 1987). If the null hypothesis is rejected, on the other hand, then its acceptance suggests cointegration and a stable long-term relationship (Hassouneh, 2018). As evidenced in Table 5, there are at least three cointegrating relationships between the variables, meaning that one rejects the null hypothesis of “At most 3 (r ≤ 3)” and concedes the alternative, also confirming long-term equilibrium and cointegration. This serves as a basis for further econometric analysis of variable connections.

Table 5. Johansen cointegration test results using CPI, GDP, GP, OP, FFER, UR, and HPI.

Hypothesized no. of CE(s)	Eigenvalue	Trace statistics	0.05 critical value
None (r = 0)	0.345	288.989	111.779
At most 1 (r ≤ 1)	0.131	127.148	83.938
At most 2 (r ≤ 2)	0.126	73.144	60.062
At most 3 (r ≤ 3)	0.035	21.498	40.174

Note: r represents the degree of the data matrix.

4.3. Time-series decomposition technique

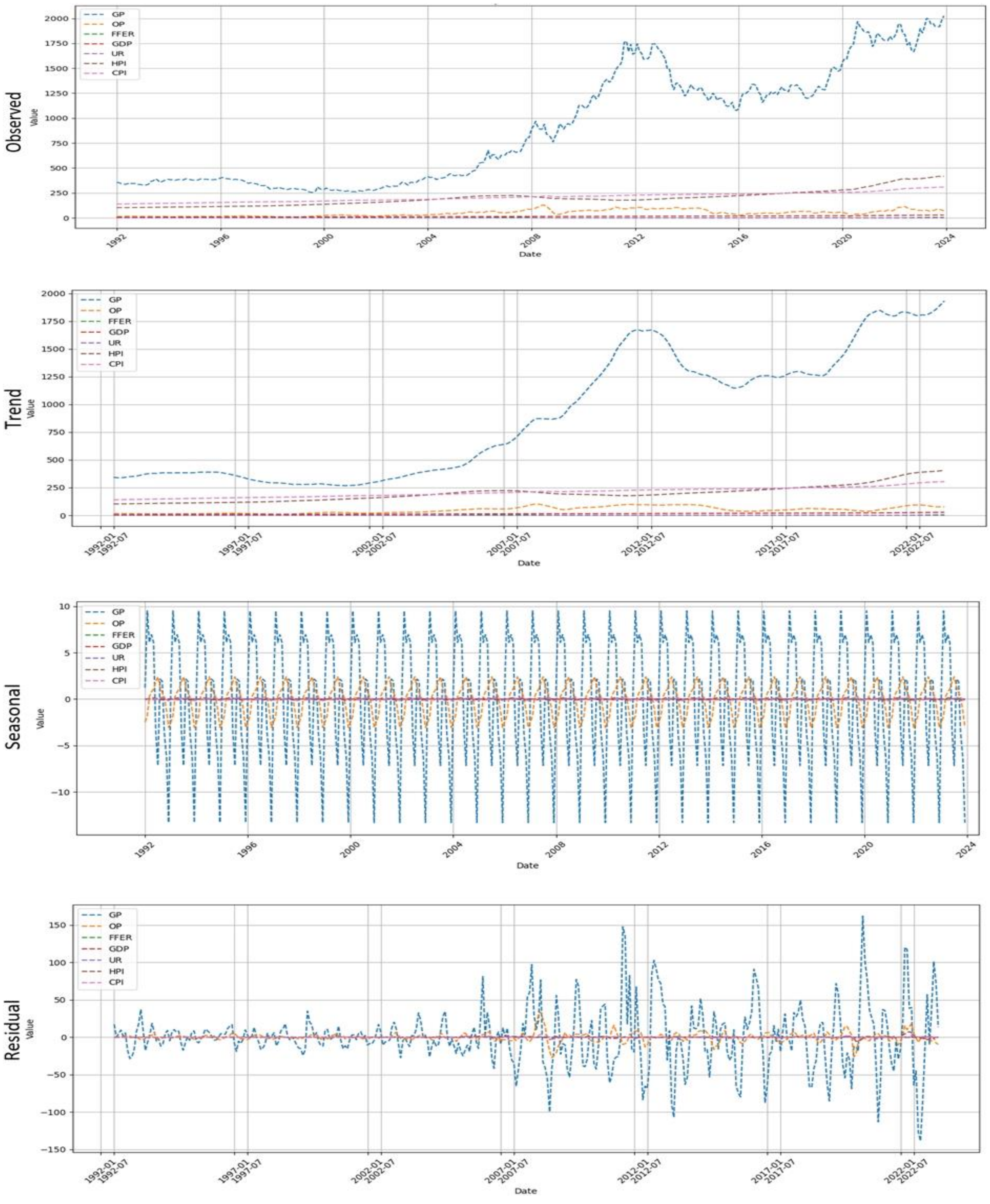

Figure 6. Calibrated vector decomposition of seven variables.

Decomposition is critical in improving time-series analysis to increase understanding and predictive accuracy by disassembling data into an additive or multiplicative-based trend, seasonality, and residuals' component (Delage et al., 2022). Decomposition methods were used to analyze and forecast a technically complex, nonlinear, nonstationary time series, which helped to reduce data complexity to enhance forecast accuracy (Jaber et al., 2014; Bandara et al., 2021). Empirical mode (EMD) and seasonal-trend decomposition techniques are the most common advanced decomposition methods used to solve diverse analytical challenges. They have applicability across multiple disciplines, including atmospheric science, hydrology, and finance (Balocchi et al., 2004; Choudhary et al., 2019). EMD is a particularly valuable decomposition for a nonstationary time series since it produces interpretable, orthogonal modal characteristics. Decomposition methods generate underlying patterns, which then enable more accurate predictions and hence more informed, data-based judgments. Figure 6 illustrates a time series that incorporates level, trend, seasonality, and noise, indicating the trends and seasonality of GP, OP, and FFER from January 1992 to December 2023. GDP, UR, HPI, and CPI display lower seasonality, although the analysis captured most of the variance and contained random errors within the margin of errors. Thus, decomposition methods are effective tools in time-series forecasting.

4.4. Granger causality test and VAR results

For the Granger causality test used in this study, the maximum lag value is set to 12. Since the data are collected monthly, the choice of 12 lags makes it possible to examine how the data affect each other over time, covering a full one-year period. This allows for short-term and seasonal effects between variables as well as annual cycles to be taken into account. Hence, through this analysis, it is possible to better understand how dynamics occurring over annual periods shape causal relationships.

Table 6. Granger causality test results of all variables against CPI changes.

Variable	P-value (Max Lag 12)	Significance status
GP	0.0003	High significant
OP	0.0002	High significant
FFER	0.0447	Significant
GDP	0.0001	High significant
HPI	0.0544	Meaningful at the border
UR	0.0521	Meaningful at the border

Granger causality test results from Table 6 show the relationship between CPI changes and other economic indicators. The p-values and their significance are important for assessing the impact of these indicators on the CPI. The p-values for GP and OP variables (0.0003 and 0.0002, respectively) are quite low, indicating that their effects on the CPI are *highly significant*. This finding emphasizes the strong impact of gold and oil prices on the CPI and hence their role in inflation dynamics. Higher oil prices lead to inflationary pressures by increasing energy costs and hence overall production costs. Gold, on the other hand, is often considered a proxy for inflation; higher gold prices can be associated with higher inflation expectations. The p-value of the FFE) is considered *significant* at 0.0447. The impact of interest rates on CPI is related to its direct impact on economic activity and borrowing costs. An increase in the federal funds rate can alleviate demand-driven inflationary pressures, thus having a significant impact on the CPI. The p-value of 0.0001 for GDP indicates that the impact of GDP on CPI is *highly significant*. Economic growth can often lead to inflationary pressures by increasing consumption demand. This result emphasizes the important role of GDP on CPI. The HPI and UR variables also show borderline significance (p-values of 0.0544 and 0.0521, respectively). These findings suggest that the HPI and UR could potentially have significant effects on the CPI, but these effects are significant at a lower confidence level. HPI can affect inflation through housing market dynamics and rental prices, while it can trigger inflation through labor market pressures and wage increases when the unemployment rate is low. In addition, a vector autoregression (VAR) model is used to examine the dynamic effects of the macroeconomic variables used in the study on the US CPI. The VAR model allows us to assess the effects of these variables over time, taking into account the interactions between variables (Thapa, 2023). The optimal lag length of the VAR model was determined using criteria such as Akaike information criterion (AIC), Bayesian information criterion (BIC), and Hannan-Quinn information criterion (HQIC). The optimal lag length for the model was chosen as 5 because this lag length provided the lowest information criterion values. The result of this selection is shown in Table 7.

Table 7. VAR order selection (* highlights the minimums).

Order	AIC	BIC	FPE	HQIC
0	−5.922	−5.848	0.002679	−5.893
1	−11.31	−10.72	1.22e-05	−11.08
2	−11.64	−10.53	8.815e-06	−11.2
3	−11.78	−10.16	7.655e-06	−11.14
4	−11.8	−9.662	7.486e-06	−10.95
5	−11.84*	−9.176	7.269e-06	−10.78
6	−11.77	−8.595	7.771e-06	−10.51
7	−11.68	−7.987	8.548e-06	−10.21
8	−11.64	−7.431	8.935e-06	−9.97
9	−11.53	−6.802	1.006e-05	−9.653
10	−11.53	−6.287	1.014e-05	−9.449
11	−11.56	−5.792	1.002e-05	−9.267
12	−11.56	−5.274	1.016e-05	−9.061

The summary of the VAR model is shown in Table 8. The AIC, HQIC, and BIC criteria are used to measure the appropriateness of the model; lower values indicate that the model is better. AIC (−11.9075), BIC (−9.28422), and HQIC (−10.8664) values indicate that the model is appropriate. Moreover, the results for CPI are shown in Table 9.

Table 8. Summary of results of VAR model.

Parameter	Value
No. of equations	7.00000
BIC	−9.28422
Nobs	378.000
HQIC	−10.8664
Log likelihood	−1252.00
FPE	6.76713e-06
AIC	−11.9075
Det (Omega_mle)	3.57969e-06

Table 9 presents the results of the VAR model examining the impact of various macroeconomic variables on the CPI across different lag periods. The analysis reveals that CPI is significantly influenced by its own past values, with the first and fourth lags showing positive and statistically significant coefficients (0.198 and 0.134, respectively, with $p < 0.05$). This suggests that past CPI values are crucial predictors of current CPI levels, highlighting the persistence of inflationary trends.

Oil prices (OP) also demonstrate a notable impact on CPI, particularly at the first lag, where the coefficient is 0.042 with high statistical significance ($p < 0.01$). This finding underscores the direct and immediate effect of oil price fluctuations on consumer prices, likely due to the role of energy costs in overall production and transportation. Conversely, the impact of gold prices (GP) on CPI is consistently insignificant across all lags, indicating that gold, often considered a hedge against inflation, does not have a direct or substantial influence on the CPI in the short term. The federal funds rate (FFER) exhibits a significant relationship with CPI at the first and third lags. The first lag shows a positive coefficient (0.384, $p < 0.05$), while the third lag presents a negative coefficient (−0.520, $p < 0.05$). This suggests that while initial changes in the federal funds rate may lead to an increase in CPI, the effect reverses over time, possibly reflecting the delayed impact of monetary policy on inflation. GDP consistently shows a strong positive effect on CPI at multiple lags, particularly at the first, second, and third lags (coefficients of 1.096, 0.890, and 1.373, respectively, with $p < 0.01$ for L1 and $p < 0.05$ for L2 and L3). This indicates that economic growth is a significant driver of consumer prices, likely due to increased demand and production costs associated with higher GDP. The house price index (HPI) only shows significance at the fourth lag, with a positive coefficient of 0.135 ($p < 0.05$). This delayed effect suggests that changes in housing prices may eventually influence consumer prices, possibly through wealth effects and changes in consumer spending patterns. Finally, the unemployment rate (UR) has a significant impact at the first and fourth lags, with a positive coefficient at L1 (36.117, $p < 0.05$) and a negative coefficient at L4 (−24.333, $p < 0.05$). This mixed result indicates that while higher unemployment initially correlates with higher inflation, potentially due to reduced labor supply and increased wage pressures, the long-term effect may dampen inflationary pressures as overall economic activity declines.

Table 9. VAR model results for CPI and macroeconomic variables.

Variables	Lag 1 coefficient	Lag 2 coefficient	Lag 3 coefficient	Lag 4 coefficient	Lag 5 coefficient	Statistical significance
CPI	0.198*	−0.175*	0.047	0.134*	0.050	Significant at various lags
OP	0.042**	0.004	0.008	−0.011	−0.011	Highly significant
GP	0.000	0.001	0.001	0.000	0.000	Not significant
FFER	0.384*	−0.033	−0.520*	−0.176	0.404	Significant at L1 and L3
GDP	1.096**	0.890*	1.373*	−0.409	0.169	Highly significant
HPI	0.001	−0.040	−0.024	0.135*	0.011	Significant at L4
UR	36.117*	−9.364	−13.653	−24.333*	12.600	Significant at L1 and L4

** $p < 0.05$, ** $p < 0.01$*

Note: The table displays coefficients for each variable at different lags, indicating the effect of each variable on CPI.

Impulse response function (IRF) analyses performed as a result of VAR analysis show the effects of various macroeconomic variables in the model on the CPI. Each graph shown in Figure 7 analyzes the impact of a shock in the relevant variable on the CPI over a certain period of time (12 periods).

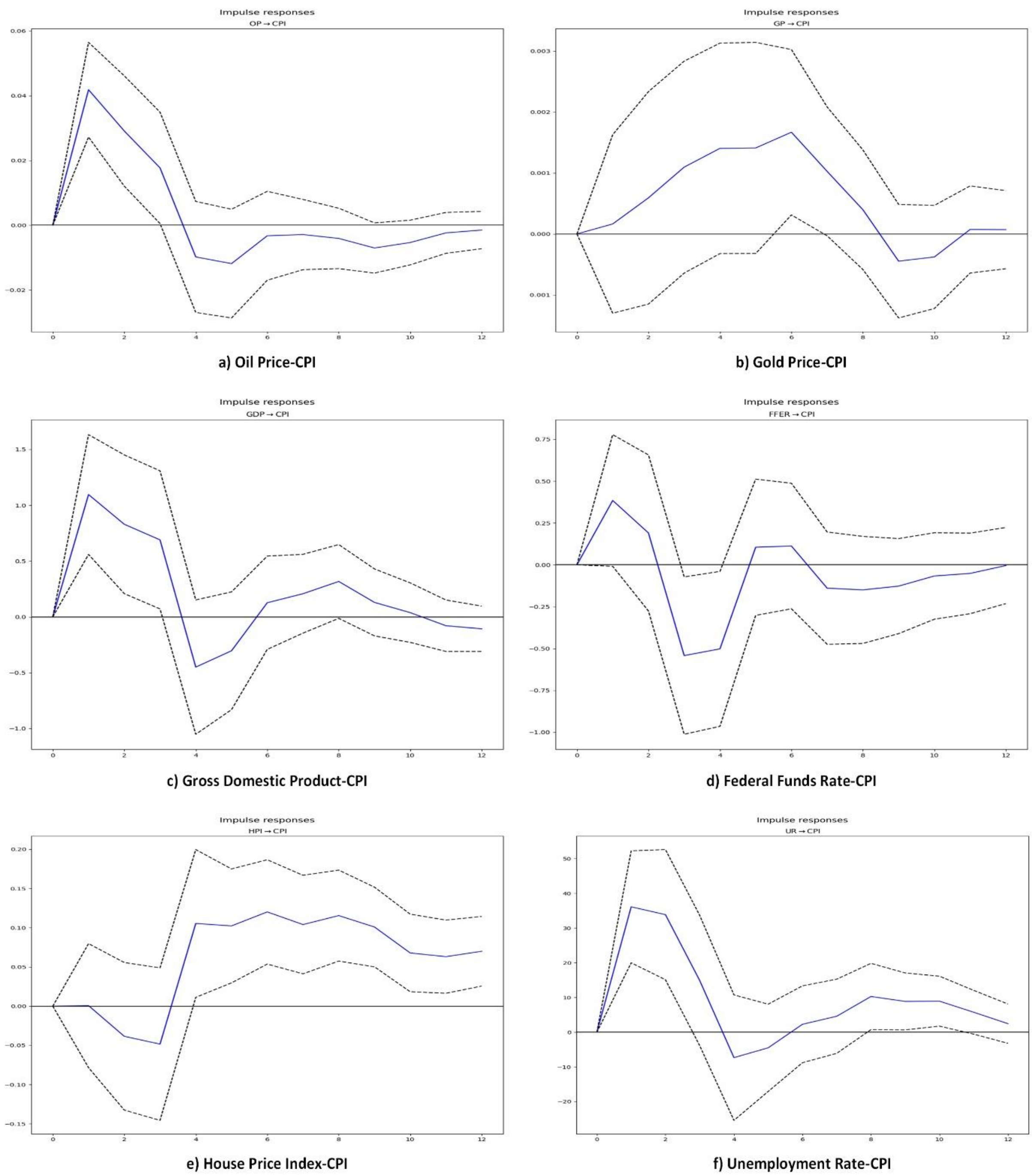

Figure 7. Impulse response function (IRF) analyses for CPI.

An increase in oil prices has a positive effect on the CPI in the first period, and this effect declines rapidly and shows a negative trend. This suggests that oil price increases initially push consumer prices up, but this effect diminishes over time and becomes negative. The shock effect reaches a level close to zero after about four cycles, after which a slight recovery trend is observed. In contrast, a shock to gold prices has a limited impact on the CPI, which is generally positive. However, the magnitude of this effect is low and fluctuates over time. Although the impact of changes in gold prices on the CPI initially increases, it is generally weak and limited. On the other hand, a shock in GDP has a strong and significant impact on the CPI. This effect shows a positive trend in the initial periods and then declines rapidly. Increases in GDP initially boost consumer prices by stimulating economic activity, but this effect diminishes over time and becomes negative. On the other hand, a shock to the federal funds rate initially has a negative effect on the CPI, which turns positive over time. Increases in interest rates initially depress consumer prices, but this effect reverses in the long run and leads to an increase in prices. In addition, changes in the house price index have a limited impact on the CPI. Although the magnitude of this effect is low, it is generally positive. Increases in house prices have a limited impact on consumer prices. Finally, changes in the unemployment rate have a strong negative impact on the CPI, which turns positive over time. Increases in the unemployment rate initially reduce consumer prices, but then lead to an increase in prices in later periods.

4.5. Hyperparameter settings of algorithms used in forecasting

Hyperparameter optimization is a key feature of time-series forecasting to improve the machine learning model's performance (Zhang et al., 2021). The decision randomly regulates the state variables to achieve optimal consistency of the high-dimensional relationships in large multivariate corpora (Vasco-Carofilis et al., 2020; Bouktif et al., 2020). Hyperparameters are not linear and enhance foretelling, notably when the relationship between different results and data is chaotic. Various rebel approaches have also been developed to explain deep learning, especially those related to the LSTM framework, which is perfect for presenting chaotic data series (Ribeiro et al., 2020; Ibrahim et al., 2021; Sumita et al., 2023). Other sorts of factors, such as extreme gradient boost, can also be prepared, notably through the CNN and ensemble methods (Zhou et al., 2019; Gastinger et al., 2021; Qinghe et al., 2022; Utama et al., 2022; Alizadeh, 2023).

Hybrid algorithms can help to optimize your Bayesian optimization as they can be used among various hyperparameter search methods, called hyperparameter optimization techniques. Hyperparameter optimization, particularly by employing adaptive genetic algorithms, significantly boosts LSTM models over standard methods (He et al., 2022). Bayesian optimization and swarm algorithms, a kind of metaheuristic approach, also effectively refine the parameters of LSTM networks, which in turn yields better results on tasks than do applications with simply default values for these settings or no fine-tuned hardware setup at all (Alibabaei et al., 2021; Alshahrani et al., 2023). By making use of these methods together with Coyote Optimization and genetic algorithms, when applied to XGBoost models, there must be a high guarantee that the results will be both stable as well as predictable (Ribeiro et al., 2021; Shi et al., 2022). Bayesian optimization is recognized as an effective method for tuning hyperparameters over a variety of machine-learning models, improving precision and performance in this way (Feurer et al., 2015; Yang and Shami, 2020; Xiao et al., 2023; Aghaabbasi et al., 2023). This technique is very useful since it is possible to adjust parameters simultaneously when it comes to time-series forecasting and, thus, to adjust optimal parameters into models. When combining Bayesian optimization with genetic algorithms, numerous experts have discovered that they can solve by themselves their own previously infeasible empirical problems and generate satisfying results (Wang et al., 2013; Jurado et al., 2015; Khandelwal et al., 2015; Adeloye, 2016).

Table 10. Details on hyperparameter settings of forecasting algorithms.

Forecasting algorithms	Hyperparameters	Hyperparameter learning algorithm
LSTM	Number of Layers: 2	Bayesian optimization
	Number of Units: 23	
	Batch Size: 50	
	Number of Epochs: 100	
	Learning Rate: 0.003	
	Activation Function: sigmoid	
	Dropout Rate: 0.2	
	Loss Function: mean squared error	
	Optimizer: Adam	
MARS	Max Terms: 68	Bayesian optimization
	Max Degree of Interactions: 6	
	Penalty Parameter: 1.2988	
XGBoost	Hyperparameters for genetic algorithm	Genetic algorithm
	Population size: 50	
	Crossover rate: 0.5	
	Mutation rate: 0.2	
	Number of generations: 40	
	Hyperparameters for XGBoost	
	n_estimators: 230	
	max_depth: 5	
	learning_rate (eta): 0.07758	
	subsample: 0.6244	
	colsample_bytree: 0.9841	
	gamma: 0.0	
	reg_lambda: 4.3522	
	reg_alpha: 0.0	
LSTM-MARS	Hyperparameters for LSTM	Bayesian optimization
	Number of Layers: 2	
	Number of Units: 24	
	Batch Size: 47	
	Number of Epochs: 100	
	Learning Rate: 0.001	
	Activation Function: sigmoid	
	Dropout Rate: 0.2	
	Loss Function: mean squared error	
	Optimizer: Adam	Bayesian optimization
	Hyperparameters for MARS	
	Max Terms: 65	
	Max Degree of Interactions: 5	
	Penalty Parameter: 1.2005	
LSTM-XGBoost	Hyperparameters for LSTM	Genetic algorithm
	Number of Layers: 2	
	Number of Units: 20	
	Batch Size: 45	
	Number of Epochs: 100	
	Learning Rate: 0.002	
	Activation Function: sigmoid	
	Dropout Rate: 0.2	
	Loss Function: mean squared error	
	Optimizer: Adam	Genetic algorithm
	Hyperparameters for XGBoost	
	n_estimators: 200	
	max_depth: 10	
	learning_rate (eta): 0.04890	
	subsample: 0.4371	
	colsample_bytree: 0.6793	
	gamma: 0.0	
	reg_lambda: 4.7170	
	reg_alpha: 0.0	

In this study, a combination of Bayesian optimization and genetic algorithms is used to improve the performance of each hybrid model component by fine-tuning its hyperparameters. The use of a unified and optimized platform enables automatic tuning of the entire hybrid model, thus improving model performance. The hyperparameter settings presented in Table 10 were determined using various hyperparameter optimization algorithms to optimize the performance of each model. The hyperparameter optimization for the LSTM, MARS, and LSTM-MARS hybrid models was performed using Bayesian optimization. This method is preferred due to its high computational cost and effective results in multidimensional search spaces. Bayesian optimization searches for the best hyperparameter combination by selecting new points based on the available information at each iteration. Especially in complex models such as LSTM and MARS, the use of this method has played a critical role in improving the accuracy of the model. In addition, genetic algorithms were used for the XGBoost model and the LSTM-XGBoost hybrid model. Genetic algorithms evaluate diverse populations of hyperparameter combinations by mimicking natural selection and genetic processes. This method works effectively in large search spaces and is suitable for improving the overall performance of the model. By optimizing the hyperparameters of the XGBoost model, the genetic algorithm has enabled the model to make more accurate predictions. These hyperparameter optimization methods were carefully chosen to maximize the performance of each model on the data set and improve the prediction accuracy. Optimizing the hyperparameters is critical to improve the reliability and generalizability of the model's results.

4.6. Training and test results of models

The scatter plots in Figure 8 show that US CPI predictions by LSTM, MARS, XGBoost, LSTM-MARS, and LSTM-XGBoost models are consistent with the actual data, showing their high level of performance in prediction. It is clear from the plots that hybrid modeling has a high level of accuracy, with LSTM-MARS and LSTM-XGBoost models having a slightly higher accuracy than traditional ones in both training and testing. These scatterplot results imply that the hybrid models have a high level of model flexibility to predict time-series data and are thus able to accurately track and predict the dynamics of the dataset.

The following analysis indicates a close alignment to reality, seen in training and testing data for the LSTM-XGBoost and LSTM-MARS, indicating high accuracy in US CPI forecasting. Superior performance metrics are demonstrated in Tables 11 and 12. These measures include RMSE, MAPE, MAE, and R^2. During training, the LSTM-XGBoost model indicates the best results with extremely low RMSE at 0.00037, MAPE of 0.00423, MAE of 0.00031, and near-perfect R^2 of 0.99964. The LSTM-MARS model also has a superior RMSE of 0.00054 and the highest R^2 of 0.99908. In testing, the following LSTM model demonstrates best-predicted values, with the lowest RMSE of 0.00090, MAPE of 0.0075, MAE of 0.00036, and high R^2 of 0.99995, showing a nearly perfect forecast. MARS, XGBoost, and the previous biggest LSTM model have drastically higher error metrics, which underline the high level of these hybrid models in capturing time series complexity and generalizing properly.

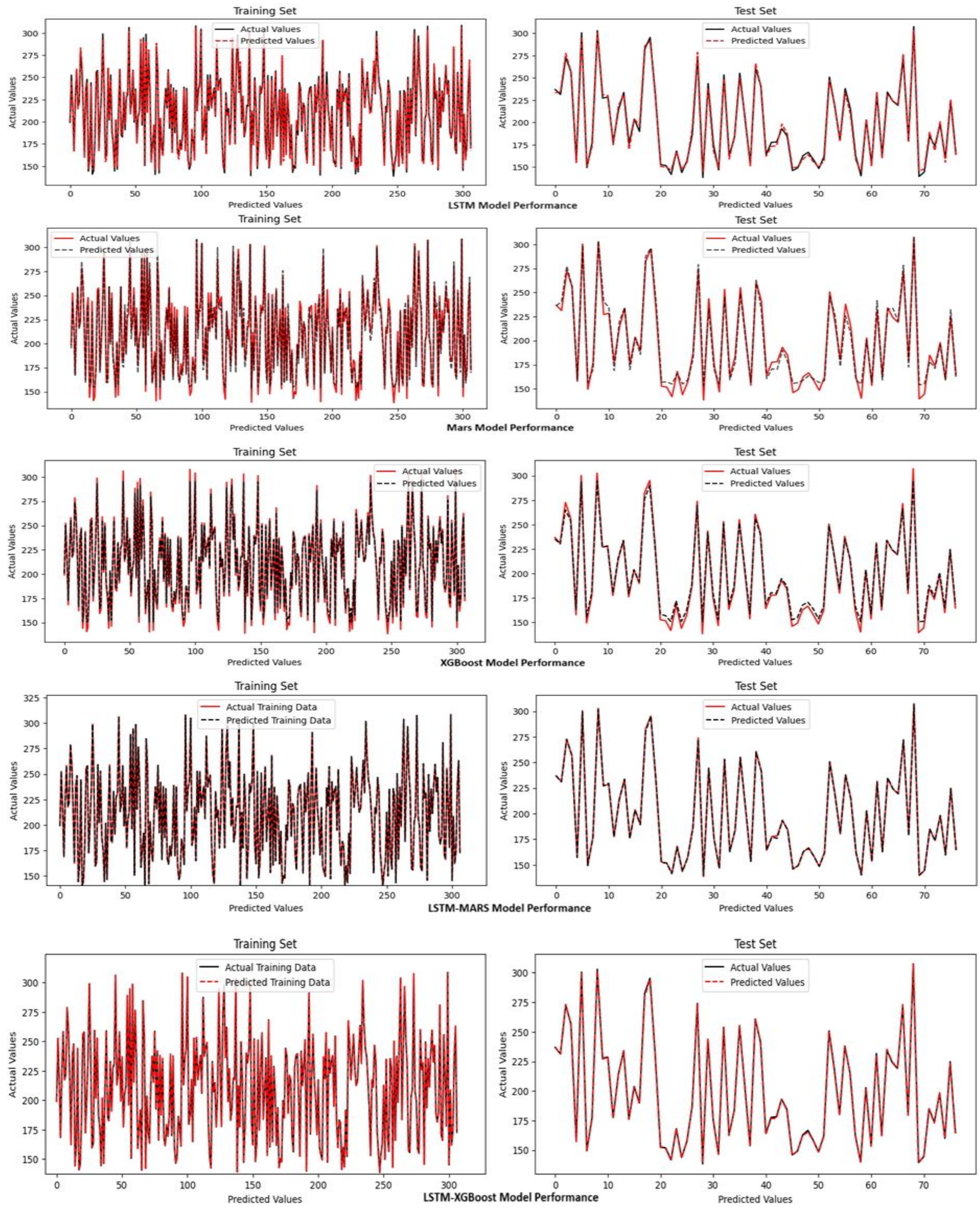

Figure 8. Actual and predicted CPI values based on LSTM, MARS, XGBoost, LSTM-MARS, and LSTM-XGBoost algorithms for the training set and test set.

Table 11. Accuracy metrics for the training process in the context of US CPI forecasting.

Metrics	LSTM	MARS	XGBoost	LSTM-MARS	LSTM-XGBoost
RMSE	0.00072	0.00158	0.00120	0.00054	0.00037
MAPE	0.00930	0.01120	0.01050	0.00610	0.00423
MAE	0.01610	0.05460	0.03550	0.00092	0.00031
R^2	0.99380	0.97750	0.97990	0.99908	0.99964

Table 12. Accuracy metrics for the test process in the context of US CPI forecasting.

Metrics	LSTM	MARS	XGBoost	LSTM-MARS	LSTM-XGBoost
RMSE	0.00135	0.00186	0.00158	0.00110	0.00090
MAPE	0.00910	0.01774	0.01200	0.00810	0.00375
MAE	0.01978	0.08745	0.03780	0.00180	0.00036
R^2	0.99888	0.97490	0.97748	0.99915	0.99995

The Kruskal–Wallis test is used to check if there are significant statistical differences among groups by analyzing the medians of the dataset from different groups, whereby the null hypothesis assumes no difference in medians, meaning that the models perform the same (Fan et al., 2010). The test is non-parametric and is preferred over ANOVA when using non-normal distributions (Fan et al., 2010). A p-value under 0.05 rejects the null hypothesis, which means there is a significant difference according to Greenland et al. (2016). According to both Tables 13 and 14, the Kruskal–Wallis test results for the LSTM, MARS, XGBoost, LSTM-MARS, and LSTM-XGBoost models when processing the training dataset and testing data had a p-value above 0.05. This supports the null hypothesis and suggests there is not a significant difference between these models. As such, the LSTM-XGBoost model also had the highest p-values for both the training data and testing data, equal to 0.998 and 0.984. It can, therefore, be said that the forecast of the LSTM-XGboost model is statistically similar to the other models. In conclusion, the acceptance of the null hypothesis can maintain its validity on both these datasets.

Table 13. Results of the KW test conducted at 0.05 significance level for the training phase.

Model	P-value	H_a^0
LSTM	0.866	accept
MARS	0.834	accept
XGBoost	0.849	accept
LSTM-MARS	0.979	accept
LSTM-XGBoost	0.998	accept

Table 14. Results of the KW test conducted at 0.05 significance level for the test phase.

Model	P-value	H_a^0
LSTM	0.978	accept
MARS	0.813	accept
XGBoost	0.835	accept
LSTM-MARS	0.961	accept
LSTM-XGBoost	0.984	accept

4.7. Shapley additive explanations (SHAP) results

The Shapley additive explanations (SHAP) method interprets complex model outcomes by attributing predictive value to every input feature, where it measures each feature's effect on specific prediction (Rodríguez-Pérez and Bajorath, 2020; Khan, 2024). This is especially effective for multivariate time-series forecasting since it demonstrates to what extent every variable has contributed to predicting specific values and allows for interpreting the entire model on a local and global level (Li et al., 2021; Lee et al., 2022). SHAP is a universal post-model construction interpreter, ranking features based on their predictive benefits and highlighting their significance with SHAP values (Rodríguez-Pérez and Bajorath, 2020; Kitani and Iwata, 2023; Nguyen et al., 2023). In Figure 9, SHAP values for different models in CPI prediction of consumers are demonstrated, reflecting the feature impact distribution in shades as trained and tested during training and testing.

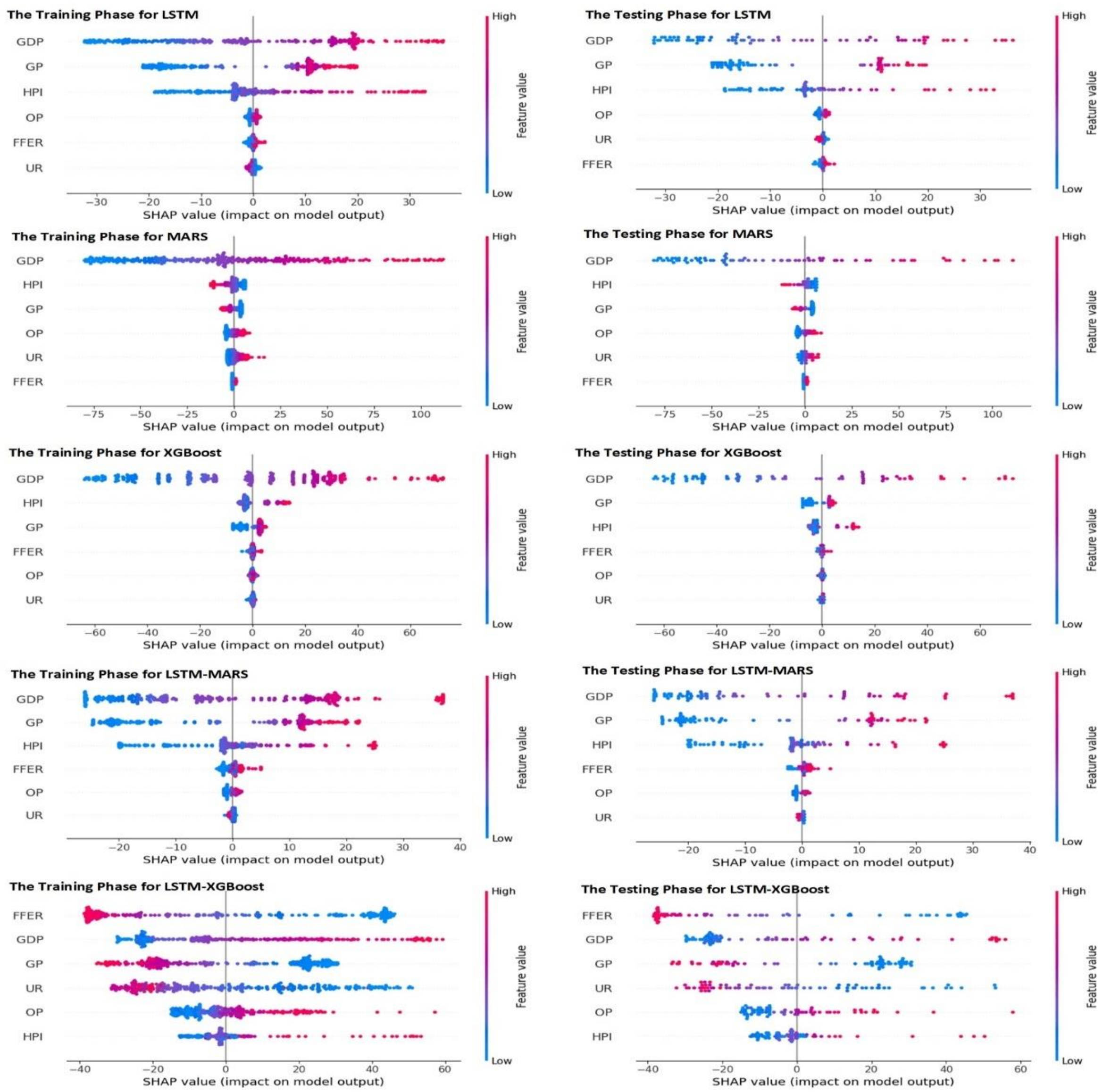

Figure 9. Importance of the features identified for all models in the training and testing processes.

Figure 9 shows the distribution of SHAP values for the LSTM, MARS, XGBoost, LSTM-MARS, and LSTM-XGBoost models in both the training and testing phases. SHAP values are used to determine the contribution of each feature to the model output and this figure evaluates the impact of various macroeconomic variables (GDP, GP, HPI, OP, FFER, UR) on the forecasts for each model. In the figure, the distribution of the SHAP value of each characteristic shows its impact on model output and whether this impact is positive or negative. For example, GDP has high SHAP values and has a strong positive impact on model forecasts, which is particularly pronounced in models such as LSTM and XGBoost. In contrast, variables such as GP and HPI have a lower impact on model outputs, showing low or unstable SHAP values in some models. As for the LSTM-XGBoost model, in both the training and testing phases of this model, the impact of variables such as FFER and GDP on forecasts is much more pronounced. FFER made the largest contribution to the model output, especially in the test phase, showing both positive and negative effects. This reflects the complex effects of FFER on inflation dynamics over time, while GDP also plays a decisive role in model forecasts. Another aspect of the LSTM-XGBoost model that differentiates it from other models is that the effects of all variables are more evenly distributed and therefore the model is generally more flexible and capable of capturing complex interactions. These findings emphasize the ability of the LSTM-XGBoost model to accurately reflect both short-term and long-term economic dynamics and explain why this model has a superior performance in economic forecasting.

5. Discussion

The evaluation of several predictive models, such as LSTM, MARS, XGBoost, and their hybrids, was conducted in this research, focusing on forecasting the US Consumer Price Index. The choice of machine-learning methods used in this study was largely driven by the characteristics of the dataset and the phenomena encountered in this dataset. Time-series data, such as the US CPI, can be influenced by economic and social factors, creating complex patterns. In particular, we use a large dataset spanning from 1974 to 2023, during which there have been several extraordinary economic events. The dataset includes various economic crises, oil shocks, major financial collapses (e.g., the 2008 global financial crisis), the COVID-19 pandemic, and the subsequent recovery. Such extraordinary events led to sudden changes in data patterns and extreme volatility. For example, dramatic increases in oil prices caused sharp fluctuations in the CPI, directly affecting both production costs and consumer prices. Similarly, the economic recession during the COVID-19 pandemic and the subsequent expansionary monetary policies had sudden and strong effects on the CPI. These extraordinary circumstances have increased the complexity of the dataset and revealed complex and nonlinear relationships for which traditional econometric models may be inadequate. Therefore, the selection of machine-learning methods focused on models that could best handle this complexity and the extreme cases in the dataset.

Since the dataset contains time series with long-term dependencies and seasonal patterns, LSTM models were chosen. The ability of LSTM to capture long-term dependencies and complex patterns in time-series data makes it particularly suitable for modeling the effects of economic crises and sudden shocks. LSTM has also been used to model recovery processes after extraordinary events such as pandemics. Moreover, since the dataset has a large number of attributes and complex relationships, a model with a powerful attribute selection and modeling capability such as XGBoost was used. XGBoost is characterized by its fast processing of large datasets and its capacity to identify interactions between attributes. This has been particularly useful for modeling the effects of extreme events such as oil shocks and financial crises. On the other hand, MARS has been used to model nonlinear relationships in the dataset and the changing effects of specific attributes. In cases where economic indicators may show different effects during certain periods, the adaptive nature of MARS has played an important role in capturing these changes. For example, in times of crisis, some economic variables may react very differently than usual, which requires the flexibility of MARS. Finally, the choice of hybrid models (LSTM-XGBoost, LSTM-MARS) was made in order to best model the complexity of the dataset and the effects of extreme situations. The hybrid models, combining the time series strengths of LSTM with the feature selection capability of XGBoost, provided higher forecasting accuracy, especially during large economic shocks. Likewise, the combination of LSTM and MARS offered flexibility and accuracy in changing economic conditions.

On the other hand, each model has its own strengths as well as certain limitations. These limitations can affect the model results and the accuracy of forecasts. While the LSTM is highly effective in capturing long-term dependencies in time-series data, it is time-consuming and computationally costly to train due to the complexity of the model. It can also run the risk of overfitting, which can lead to the model underperforming on new data. Although XGBoost has a strong modeling capacity, it is highly sensitive to hyperparameter settings. Incorrectly set hyperparameters can cause the model to overfit or underperform. Furthermore, the interpretability of the model is often low, which can make the results difficult to interpret. Although MARS offers flexibility in modeling nonlinear relationships, it can tend to overfit within the data set. This can negatively affect the generalization ability of the model and lead to poor performance on new data. Furthermore, the results of MARS can be more difficult to interpret due to the complexity of the model. Although hybrid models improve prediction accuracy by combining the strengths of various methods, these models are more complex to configure and train. Furthermore, when the potential limitations of multiple methods used in hybrid models are combined, there is a risk that the model becomes overly complex, and the results are less explainable. These limitations can affect model performance and the accuracy of results. When faced with such complex and extreme situations, researchers should make model choices based on the characteristics of their data sets. The complexity and extreme cases present in the data set played a decisive role in model selection. Ultimately, it is critical to consider the dynamics of the data set, the extraordinary circumstances, and the research questions in the selection of forecasting methods.

Several evaluation criteria were used in the study to assess the model performance. These metrics were chosen to analyze the model's prediction accuracy, errors, and generalization capacity. MAE calculates the average magnitude of the model's forecast errors. It was chosen to directly measure the magnitude of errors in CPI forecasts. Since MAE is less sensitive to large errors, it is an effective measure for understanding the average performance of the model (Li and Yang, 2022). RMSE is a more intuitive measure of the model's prediction performance and is sensitive to large errors (Li et al., 2020). Therefore, it is preferred in economic forecasts where large deviations are important. R^2 shows the ratio of the variance that the model can explain to the total variance. It is used to assess the overall performance of the model and how well the independent variables explain the CPI. R^2 is a critical metric to measure the generalization capacity of the model and to understand the effects of independent variables (Sun and Tian, 2022). MAPE expresses the proportion of errors as a percentage and normalizes it to predicted values (Farsi et al., 2021). MAPE makes data at different scales comparable. It is particularly useful for analyzing changes in variables such as the CPI over time. These metrics are carefully chosen to understand both the overall performance of the model and its behavior under specific economic conditions. Accuracy, minimization of large errors, and the generalizability of the model are important in economic forecasting; therefore, the metrics used were chosen to cover these factors.

The LSTM-XGBoost hybrid model has demonstrated the lowest mean absolute percentage error of 0.00375 throughout the training, suggesting its highest precision among the other models, such as the MARS model with the highest MAPE. Also, this model has shown the best performance throughout the testing beyond any doubt. Coulibaly and Baldwin (2005), Alizamir et al. (2023), and Reddy et al. (2022) found that the models mentioned are efficient in forecasting applied to different industries and greatly enhance accuracy. The research by Zahara and Ilmiddaviq (2020) and Ali and Mohamed (2022) also tested ARIMAX and LSTM models, discovering that LSTM is more accurate in terms of predicting the detail feature of CPI; besides, the latter can indeed become even more accurate with optimization methods, such as Adam. The research of Nguyen et al. (2023) has also confirmed that MARS is quite efficient when predicting CPI. Also, the same work indicates the fact that the LSTM-XGBoost model is much better than MARS in these regards. Therefore, it is possible to state that using these hybrid and deep learning models when working with complex datasets is suitable as they provide the possibility of achieving precision. However, it is crucial to add that it is recommended to choose the advanced option, in all, the last one for economic and financial forecasting.

In their 2024 study, Bandara and De Mel reported that support vector regression was an effective forecasting tool in predicting inflation rates in Sri Lanka, with a MAPE of 0.12%. However, this analysis indicates that the LSTM-XGBoost model also outperformed SVR in forecasting US CPI, which shows the increased effectiveness of hybrid models in working with complex data patterns. Thus, it is important to compare various forecasting techniques to identify the most efficient. Over the period from the 1990s to the 2020s, with dramatic changes in economies and policies, LSTM, MARS, and XGBoost and their hybrid models, such as LSTM-MARS and the model under analysis, showed better accuracy in predicting CPI than many traditional models. Overall, the development of models with high complexity, such as LSTM, MARS, XGBoost, and their combinations, has improved the field of economic forecasting. MARS is still valuable for certain predictions, but older models are no longer effective in processing complex and sequential data patterns. Newer methodologies demonstrate great potential, but hybrid models improve data processing by identifying complex correlations, which significantly increases forecast accuracy and reliability. In fact, when it comes to global CPI and the macroeconomy, modeler analysts have a wide choice of tools to use.

Based on this information, LSTM can detect long-term dependencies and complex patterns in time series, while XGBoost can effectively model nonlinear relationships using these features. As a result, the combination of these two powerful methods improved forecasting performance, leading to higher accuracy. This finding is of great importance to policymakers. Increased accuracy in CPI forecasts allows for more effective implementation of disinflationary policies. More accurate forecasts enable central banks and other financial authorities to identify inflationary pressures in advance and take timely measures to counter them. This is critical for maintaining economic stability and ensuring sustainable growth. Compared to previous research, the innovative hybrid model approach presented in this paper goes beyond traditional econometric models and stand-alone machine-learning techniques. Traditional models are often based on linear assumptions and therefore may not accurately reflect complex economic dynamics. However, hybrid models such as LSTM-XGBoost can both capture nonlinear relationships and analyze complex patterns over time. The findings of this study demonstrate the superiority of hybrid models with lower error rates and higher R2 values compared to previous studies in the literature. For example, while traditional time-series models used in previous studies can be effective in certain situations, their accuracy is often limited. Likewise, studies using only machine-learning techniques have not been able to take full advantage of the advantages of hybrid models, although they have been able to achieve a certain level of accuracy. In this context, the hybrid models presented in this paper make an important contribution to the literature and provide a roadmap for more accurate and reliable economic forecasts.

The correlation matrix shown in Figure 3 indicates strong relationships among the macroeconomic indicators used. These relationships show how the factors that play a decisive role in the forecasting process interact with each other. In particular, indicators such as gold prices, oil prices, and the federal funds rate are directly related to the CPI, and changes in these indicators have a significant impact on the CPI. Gold prices are often seen as a safe haven against inflation, and an increase in demand for gold in the event of higher inflation may increase gold prices. This may indirectly affect the CPI. Oil prices, on the other hand, affect consumer prices by directly affecting production and transportation costs. Increases in oil prices may lead to a rise in the general price level and thus to an increase in the CPI. The federal funds rate also plays a decisive role in inflation by influencing consumer spending and investment decisions. A rise in interest rates can reduce consumer demand, which in turn depresses the price level, leading to a decline in the CPI. The rationale behind the selection of these indicators is underpinned by such mechanisms explaining their impact on the CPI. The relationship of each indicator with the CPI has been carefully analyzed in order to improve the forecasting performance of the model. In particular, the use of such a wide range of macroeconomic indicators has improved the forecasting power of the model, enabling it to more accurately predict inflation dynamics.

According to the VAR model results, the effect of oil prices on CPI is quite high and statistically significant. However, gold prices do not have a direct or significant effect on the CPI in the short run. This finding contradicts the significant effect of gold prices on CPI reported by Radev et al. (2023) and Iqbal et al. (2021). The strong effect of oil prices is consistent with the findings of Yilmazkuday (2024). However, it shows that the federal funds rate has a positive effect on the CPI in the first lag period and a negative effect in the third lag period. This reflects the effects of monetary policy changes on consumer prices as reported in Adhikari and Stevens (2024). While there is an initial upward effect on prices, this effect is reversed over time. It also shows that GDP has a strong and positive effect on CPI and that this effect is significant in the first, second, and third lag periods. This finding is fully consistent with the effect of GDP on inflation in Shiferaw (2023). The unemployment rate has a positive effect on CPI in the first lag period and a negative effect in the fourth lag period. As stated by Upadhyaya and Kharel (2022), the unemployment rate reduces inflation, but this effect diminishes and even reverses in the long run. These results are consistent with the findings in the literature. In conclusion, the VAR model findings are consistent with most of the studies in the related literature, especially on the effects of oil prices and GDP on CPI. However, the effect of gold prices on the CPI is not fully consistent with the findings in the literature. This suggests that the impact of gold prices on the CPI may be more long-run or indirect. Moreover, the complex effects of macroeconomic variables such as the federal funds rate and the unemployment rate on the CPI are revealed by the VAR model and are largely consistent with other findings in the literature.

6. Conclusions

In conclusion, the LSTM-XGBoost combination model performed better than individual models and outperformed LSTM, MARS, XGBoost, and the hybrid LSTM-MARS models in predicting the US CPI in both training and testing. This accurate prediction of the CPI has a massive implication. Accurate CPI prediction helps policymakers plan for measures to reduce the unfavorable effects of inflation and manipulate the financial market for investment purposes and the government in terms of the distribution of resources like social security and budgeting. It is also critical in central bank monetary decisions such as inflation targeting to assure stability of the economy and maintain economic growth. This research also gives investors insights into the direction the CPI means in investment, thus an informed investment decision. High-level machine-learning models such as LSTM-MARS and LSTM-XGBoost based on historical data can capture real-life trends that ordinary methods can hardly capture, assisting in predicting market behaviors and the economy through the ages. The models also help policymakers determine the effectiveness of a policy before making corrections for economic stability. Hybrid machine-learning models have become a great tool in economic forecasting and policy-based decision-making, presenting a great improvement approach to analyzing economic data.

In light of the VAR model analysis and impulse response function (IRF) findings, a number of recommendations for policymakers, investors, and economic analysts can be offered by considering the effects of various macroeconomic variables on the CPI. First, given the persistent impact of past values of the CPI on current inflation rates, policymakers should carefully monitor inflation trends and make timely interventions. This finding suggests that future inflation rates can be accurately predicted based on past levels of inflation and thus a more proactive monetary policy can be pursued. The direct and significant impact of oil prices on the CPI emphasizes the critical role of energy costs on inflation. In this context, it is important to develop an energy policy that is sensitively adjusted to fluctuations in energy prices. Given that increases in energy prices may lead to inflationary pressures, it would be appropriate for governments to develop policies to increase energy efficiency and promote renewable energy sources. Moreover, effective utilization of strategic reserves against the volatility of oil prices may also contribute to containing inflation. While gold prices do not have a direct impact on the CPI in the short term, they have a limited impact on long-term inflation expectations. This shows that although gold investments are considered a protective measure against inflation, they do not create direct price pressure in the short run. Investors may continue to use gold and similar commodities in their portfolios as a long-term hedging instrument. The complex and time-varying impact of the federal funds rate (FFER) on inflation suggests that the effects of monetary policies should be carefully monitored over time. While interest rate hikes may alleviate inflationary pressures in the initial phase, the fact that this effect may reverse in the long run shows how critical the timing of monetary policies is. In this context, central banks are advised to adopt a balanced approach when setting interest rates, taking into account both short-term and long-term effects. The strong and persistent impact of GDP on inflation suggests that economic growth can exacerbate inflationary pressures. During periods of rapid economic growth, governments should review their macroeconomic stabilization strategies through tax policies or public spending to minimize demand-side inflation risks. Moreover, policies to increase production capacity and strengthen the supply side are also important to offset demand pressures. The lagged effect of the house price index (HPI) suggests that real estate prices can affect inflation over time, and hence regulations for the housing market should be carefully designed. In order to minimize the potential effects of price increases in the housing market on inflation, governments should develop policies to increase housing supply and facilitate access to housing finance. Finally, the complex effects of the unemployment rate on inflation suggest that low unemployment rates may increase inflation in the short run, but this effect may diminish in the long run. This calls for a flexible approach to designing labor market policies. Measures to increase labor supply and reduce labor market imbalances can help alleviate inflationary pressures. These findings highlight the need for macroeconomic policymakers to carefully monitor the key determinants of inflation and make strategic decisions taking into account both short- and long-term effects. At the same time, for investors and economic analysts, taking these variables into account in economic decision-making is critical for risk management and strategy development.

Most importantly, the present investigation contributes to US CPI forecasting research by rectifying several limitations of the literature. First, the use of advanced hyperparameter optimization approaches has not been applied in the context of LSTM and MARS in the previous studies. Second, introducing hybrids like LSTM-MARS and LSTM-XGBoost is not common in the literature but substantially increases forecasting performance. While precision is significantly higher in the considered models, they impose higher complexity on computations and costs but may cause a problem in low computing-resource situations. Future research may consider broader lists of machine learning algorithms and datasets expanded with the help of alternative data streams, such as social media data, which can add to the precision of the models. Better optimizing the models or testing the hybrids on more economic indicators can also bring a new horizon for their application in the economic domain. Studying possible implications on societal and moral issues might also be considered due to the potential harms of the broad implementation of ML in the economic domain.

Use of AI tools declaration

The author declares they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The author declares no conflicts of interest in this paper.

References

Adhikari DR, Stevens DP (2024) Effect of federal funds rate on cpi and ppi. *J Appl Bus Econ* 26. https://doi.org/10.33423/jabe.v26i1.6887

Aghaabbasi M, Ali M, Jasiński M, et al. (2023) On hyperparameter optimization of machine learning methods using a bayesian optimization algorithm to predict work travel mode choice. *IEEE Access* 11: 19762–19774. https://doi.org/10.1109/access.2023.3247448

Ahmed N, Assadi M, Zhang Q, et al. (2023) Assessing impact of borehole field data's input parameters on hybrid deep learning models for heating and cooling forecasting: a local and global explainable ai analysis. *IOP Conference Series: Materials Science and Engineering* 1294: 012056. https://doi.org/10.1088/1757-899x/1294/1/012056

Akbulut H (2022) Forecasting inflation in Turkey: A comparison of time-series and machine learning models. *Econ J Emerg Market* 14.

Alhendawy HAA, Abdallah Mostafa MG, Elgohari MI, et al. (2023) Determinants of renewable energy production in egypt new approach: machine learning algorithms. *Int J Energy Econ Policy* 13: 679–689. https://doi.org/10.32479/ijeep.14985

Ali M, Apriliana T, Fathonah AN (2023) The Effect of Money Supply and Bank Indonesia Rate on Consumer Price Index in Indonesia 2018–2022. *J Ekonomi Bisnis Entrep* 17: 488–497. https://doi.org/10.55208/jebe.v17i2.471

Alibabaei K, Gaspar PD, Lima TM (2021) Modeling soil water content and reference evapotranspiration from climate data using deep learning method. *Appl Sci* 11: 5029. https://doi.org/10.3390/app11115029

Alim M, Ye G, Guan P, et al. (2020) Comparison of arima model and xgboost model for prediction of human brucellosis in mainland China: a time-series study. *BMJ Open* 10: e039676. https://doi.org/10.1136/bmjopen-2020-039676

Alizadeh M, Beheshti MTH, Ramezani A, et al. (2023) An optimized hybrid methodology for short-term traffic forecasting in telecommunication networks. *T Emerg Telecommun T* 34: e4860. https://doi.org/10.1002/ett.4860

Alizamir M, Shiri J, Fard AF, et al. (2023) Improving the accuracy of daily solar radiation prediction by climatic data using an efficient hybrid deep learning model: Long short-term memory (LSTM) network coupled with wavelet transform. *Eng Appl Artif Intell* 123: 106199. https://doi.org/10.1016/j.engappai.2023.106199

Alshahrani SM, Alrayes FS, Alqahtani H, et al. (2023) Iot-cloud assisted botnet detection using rat swarm optimizer with deep learning. *Cmc-Comput Mater Con* 74: 3085–3100. https://doi.org/10.32604/cmc.2023.032972

Amalu HI, Agbasi LO, Olife LU, et al. (2021) Responsiveness of service sector growth to financial development in nigeria: evidence from 1981–2019. *J Adv Res Econ Adm Sci* 2: 1–12. https://doi.org/10.47631/jareas.v2i3.305

Amin J, Sharif M, Raza M, et al. (2020) Brain tumor detection: a long short-term memory (LSTM)-based learning model. *Neural Comput Appl* 32: 15965–15973. https://doi.org/10.1007/s00521-019-04650-7

Ampomah EK, Nyame G, Qin Z, et al. (2021) Stock market prediction with gaussian naïve bayes machine learning algorithm. *Informatica* 45. https://doi.org/10.31449/inf.v45i2.3407

Anagnostis A, Moustakidis S, Papageorgiou EI, et al. (2022) A hybrid bimodal lstm architecture for cascading thermal energy storage modelling. *Energies* 15: 1959. https://doi.org/10.3390/en15061959

Araujo GS, Gaglianone WP (2023) Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Lat Am J Cent Bank* 4: 100087. https://doi.org/10.1016/j.latcb.2023.100087

Arnone M, Romelli D (2013) Dynamic central bank independence indices and inflation rate: A new empirical exploration. *J Financ Stabil* 9: 385–398. https://doi.org/10.1016/j.jfs.2013.03.002

Arthur CK, Temeng VA, Ziggah YY (2020) Multivariate Adaptive Regression Splines (MARS) approach to blast-induced ground vibration prediction. *Int J Min Reclam Env* 34: 198–222. https://doi.org/10.1080/17480930.2019.1577940

Attoh-Okine NO, Cooger K, Mensah S (2009) Multivariate adaptive regression (MARS) and hinged hyperplanes (HHP) for doweled pavement performance modeling. *Constr Build Maters* 23: 3020–3023. https://doi.org/10.1016/j.conbuildmat.2009.04.010

Balocchi R, Menicucci D, Santarcangelo L, et al. (2004) Deriving the respiratory sinus arrhythmia from the heartbeat time series using empirical mode decomposition. *Chaos Soliton Fract* 20: 171–177. https://doi.org/10.1016/s0960-0779(03)00441-7

Balshi MS, McGuire AD, Duffy P, et al. (2009) Assessing the response of area burned to changing climate in western boreal North America using a Multivariate Adaptive Regression Splines (MARS) approach. *Global Change Biol* 15: 578–600. https://doi.org/10.1111/j.1365-2486.2008.01679.x

Bandara K, Hyndman R, Bergmeir C (2021) Mstl: a seasonal-trend decomposition algorithm for time series with multiple seasonal patterns. https://doi.org/10.48550/arxiv.2107.13462

Bandara WMS, De Mel WAR (2024) Evaluating the Efficacy of Supervised Machine Learning Models in Inflation Forecasting in Sri Lanka. *Am J Appl Stat Econ* 3: 51–60. https://doi.org/10.54536/ajase.v3i1.2385

Barkan O, Benchimol J, Caspi I, et al. (2023) Forecasting CPI inflation components with hierarchical recurrent neural networks. *Int J Forecast* 39: 1145–1162. https://doi.org/10.1016/j.ijforecast.2022.04.009

Baybuza I (2018) Inflation forecasting using machine learning methods. *Russ J Money Financ* 77: 42–59. https://doi.org/10.31477/rjmf.201804.42

Bhanja S, Das A (2021) Deep neural network for multivariate time-series forecasting. In: *Proceedings of International Conference on Frontiers in Computing and Systems*: COMSYS 2020 (267–277). Springer Singapore. https://doi.org/10.1007/978-981-15-7834-2_25

Bhati BS, Chugh G, Al-Turjman F, et al. (2020) An improved ensemble based intrusion detection technique using xgboost. *T Emerg Telecommun T* 32: e4076. https://doi.org/10.1002/ett.4076

Bouktif S, Fiaz A, Ouni A, et al. (2020). Multi-sequence lstm-rnn deep learning and metaheuristics for electric load forecasting. *Energies* 13: 391. https://doi.org/10.3390/en13020391

Brzan PP, Obradovic Z, Stiglic G (2017) Contribution of temporal data to predictive performance in 30-day readmission of morbidly obese patients. *Peer J* 5 e3230. https://doi.org/10.7287/peerj.3230v0.1/reviews/2

Budiharto W (2021) Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM). *J Big Data* 8: 1–9. https://doi.org/10.1186/s40537-021-00430-0

Cahyono ND, Sumpeno S, Setiiadi E (2023) Multivariate Time Series for Customs Revenue Forecasting Using LSTM Neural Networks. In*: 2023 International Conference on Information Technology and Computing (ICITCOM),* 357–362. https://doi.org/10.1109/ICITCOM60176.2023.10442562

Cain MK, Zhang Z, Yuan KH (2017) Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behav Res Methods* 49: 1716–1735. https://doi.org/10.3758/s13428-016-0814-1

Cao L, Li Y, Zhang J, et al. (2020) Electrical load prediction of healthcare buildings through single and ensemble learning. *Energy Rep* 6: 2751–2767. https://doi.org/10.1016/j.egyr.2020.10.005

Chen S (2023) Multiple stock prediction based on linear and non-linear machine learning regression methods. *Advances in Economics, Management and Political Sciences* 46: 225–232. https://doi.org/10.54254/2754-1169/46/20230343

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*. https://doi.org/10.48550/arxiv.1603.02754

Choi JY, Lee B (2018) Combining lstm network ensemble via adaptive weighting for improved time series forecasting. *Math Probl Eng* 2018: 1–8. https://doi.org/10.1155/2018/2470171

Choudhary K, Jha GK, Kumar RR, et al. (2019) Agricultural commodity price analysis using ensemble empirical mode decomposition: a case study of daily potato price series. *Indian J Agr Sci* 89: 882–886. https://doi.org/10.56093/ijas.v89i5.89682

Correa E (2023) Effect of unemployment, inflation and foreign direct investment on economic growth in sub-saharan africa. *J Dev Econ* 8: 297–315. https://doi.org/10.20473/jde.v8i2.47283

Coulibaly P, Baldwin CK (2005) Nonstationary hydrological time series forecasting using nonlinear dynamic methods. *J Hydrol* 307: 164–174.

Cui Q, Rong S, Zhang B (2023) Advancing the comprehension of consumer price index and influencing factors: insight into the mechanism based on prediction machine learning models. *Adv Econ Manage Res* 7: 125–125. https://doi.org/10.56028/aemr.7.1.125.2023

DeJong DN, Nankervis JC, Savin NE, et al. (1992). The power problems of unit root test in time series with autoregressive errors. *J Econometrics* 53: 323–343. https://doi.org/10.1016/0304-4076(92)90090-E

Delage O, Portafaix T, Benchérif H, et al. (2022). Empirical adaptive wavelet decomposition (eawd): an adaptive decomposition for the variability analysis of observation time series in atmospheric science. *Nonlinear Proc Geoph* 29: 265–277. https://doi.org/10.5194/npg-29-265-2022

Dhamo D, Dhamo X, Spahiu A, et al. (2022) PV production forecasting using machine learning and deep learning techniques: Albanian case study. *Adv Eng Days* 5: 68–70.

Dickey DA, Fuller WA (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: J Econom Soc* 1057–1072. https://doi.org/10.2307/1912517

Dinh TN, Thirunavukkarasu GS, Seyedmahmoudian M, et al. (2023). Predicting Commercial Building Energy Consumption Using a Multivariate Multilayered Long-Short Term Memory Time-Series Model. *Appl Sci* 13: 7775. https://doi.org/10.3390/app13137775

Djordjević K, Jordović-Pavlović MI, Ćojbašić Ž, et al. (2022) Influence of data scaling and normalization on overall neural network performances in photoacoustics. *Opt Quant Electron* 54. https://doi.org/10.1007/s11082-022-03799-1

Elgui K, Bianchi P, Portier F, et al. (2020) Learning methods for rssi-based geolocation: a comparative study. *Pervasive Mob Comput* 67: 101199. https://doi.org/10.1016/j.pmcj.2020.101199

Enke D, Mehdiyev N (2014) A hybrid neuro-fuzzy model to forecast inflation. *Proc Comput Sci* 36: 254–260. https://doi.org/10.1016/j.procs.2014.09.088.

Fan C, Zhang D, Zhang C (2010) On sample size of the kruskal-wallis test with application to a mouse peritoneal cavity study. *Biometrics* 67: 213–224. https://doi.org/10.1111/j.1541-0420.2010.01407.x

Farsi B, Amayri M, Bouguila N, et al. (2021) On short-term load forecasting using machine learning techniques and a novel parallel deep lstm-cnn approach. *IEEE Access* 9: 31191–31212. https://doi.org/10.1109/access.2021.3060290

Feng H (2024) Analysis and Forecast of CPI in China Based on LSTM and VAR Model. *Advances in Digital Economy and Data Analysis Technology The 2nd International Conference on Internet*

Finance and Digital Economy, Kuala Lumpur Malaysia, 339–357. https://doi.org/10.1142/9789811267505_0025

Feurer M, Springenberg JT, Hutter F (2015) Initializing bayesian hyperparameter optimization via meta-learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 29. https://doi.org/10.1609/aaai.v29i1.9354

Friedman JH (1991) Multivariate adaptive regression splines. *Annals Stat* 19: 1–67.

Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Software* 33: 1–22.

Gao Z, Kuruoglu EE (2023) Attention based hybrid parametric and neural network models for non-stationary time series prediction. *Expert Syst* 41. https://doi.org/10.1111/exsy.13419

Gastinger J, Nicolas S, Stepić D, et al. (2021) A study on ensemble learning for time series forecasting and the need for meta-learning. In: *2021 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.48550/arxiv.2104.11475

Gil-Cordero E, Rondán-Cataluña FJ, Sigüenza-Morales D (2020) Private label and macroeconomic indicators: Europe and USA. *Adm Sci* 10: 91. https://doi.org/10.3390/admsci10040091

Greenland S, Senn S, Rothman KJ, et al. (2016) Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 31: 337–350. https://doi.org/10.1007/s10654-016-0149-3

Groeneveld RA, Meeden G (1984) Measuring skewness and kurtosis. *J Roy Stat Soc Series D* 33: 391–399. https://doi.org/10.2307/2987742.

Guo Y, Strauss VY, Prieto-Alhambra D, et al. (2022) Use of machine learning for comparing disease risk scores and propensity scores under complex confounding and large sample size scenarios: a simulation study. *medRxiv* 1–12. https://doi.org/10.1101/2022.02.03.22270151

Hajdini I, Knotek II ES, Leer J, et al. (2024) Indirect consumer inflation expectations: Theory and evidence. *J Monetary Econ* 103568.

Hao J, Feng Q, Li J, et al. (2023) A bi-level ensemble learning approach to complex time series forecasting: taking exchange rates as an example. *J Forecasting* 42: 1385–1406. https://doi.org/10.1002/for.2971

Harding M, Lamarche C (2021) Small steps with big data: using machine learning in energy and environmental economics. *Annu Rev Resour Econ* 13: 469–488. https://doi.org/10.1146/annurev-resource-100920-034117

Hasanah SH (2021) Multivariate Adaptive Regression Splines (MARS) for Modeling The Student Status at Universitas Terbuka. *J Mat MANTIK* 7: 51–58.

Hauke J, Kossowski T (2011) Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaest Geogr* 30: 87–93.

He Y, Zeng X, Li H, et al. (2022) Application of lstm model optimized by individual-ordering-based adaptive genetic algorithm in stock forecasting. *Int J Intell Comput* 16: 277–294. https://doi.org/10.1108/ijicc-04-2022-0104

Henderi H (2021) Comparison of min-max normalization and z-score normalization in the k-nearest neighbor (knn) algorithm to test the accuracy of types of breast cancer. *IJIIS: Int J Inf Inf Syst* 4: 13–20. https://doi.org/10.47738/ijiis.v4i1.73

Hossain MS, Mitra R (2017) The determinants of price inflation in the United States: a multivariate dynamic cointegration and causal analysis. *J Dev Areas* 51: 153–175. https://www.jstor.org/stable/26415701

Ibrahim A, Mirjalili S, El-Said M, et al. (2021) Wind speed ensemble forecasting based on deep learning using adaptive dynamic optimization algorithm. *IEEE Access* 9: 125787–125804. https://doi.org/10.1109/access.2021.3111408

Imbens GW, Athey S (2021) Breiman's two cultures: a perspective from econometrics. *Obs Stud* 7: 127–133. https://doi.org/10.1353/obs.2021.0028

Imron M, Utami WD, Khaulasari H, et al. (2022) Arima model of outlier detection for forecasting consumer price index (cpi). *BAREKENG: J Ilmu Matematika Dan Terapan* 16: 1259–1270. https://doi.org/10.30598/barekengvol16iss4pp1259-1270

Islam H, Islam MS, Saha S, et al. (2024) Impact of macroeconomic factors on performance of banks in bangladesh. *J Ekon*. https://doi.org/10.58251/ekonomi.1467784

Iqbal Z, Akbar M, Amjad W (2021) Nexus of gold price-exchange rate-interest rate-oil price: lessons for monetary policy in pakistan. *Int J Bus Manag* 16: 1–16. https://doi.org/10.52015/nijbm.v16i1.50

Ivașcu C (2023) Can Machine Learning Models Predict Inflation? In: *Proceedings of the International Conference on Business Excellence*, 17: 1748–1756.

Jaber AM, Ismail MT, Altaher AM (2014) Empirical mode decomposition combined with local linear quantile regression for automatic boundary correction. *Abstr Appl Anal* 2014: 1–8. https://doi.org/10.1155/2014/731827

Jadiya AK, Chaudhary A, Thakur R (2020) Polymorphic sbd preprocessor: a preprocessing approach for social big data. *Indian J Comput Syst Sci Eng* 11: 953–961. https://doi.org/10.21817/indjcse/2020/v11i6/201106169

Jakubik J, Nazemi A, Geyer-Schulz A, et al. (2023) Incorporating financial news for forecasting Bitcoin prices based on long short-term memory networks. *Quant Financ* 23: 335–349. https://doi.org/10.1080/14697688.2022.2130085

Khan A, Kandel J, Tayara H, et al. (2024) Predicting the bandgap and efficiency of perovskite solar cells using machine learning methods. *Mol Inform* 43. https://doi.org/10.1002/minf.202300217

Knotek ES, Mitchell J, Pedemonte MO, et al. (2024) The effects of interest rate increases on consumers' inflation expectations: the roles of informedness and compliance. *Working Paper 24–01 Federal Reserve Bank of Cleveland*. https://doi.org/10.26509/frbc-wp-202401

Jin Q, Fan X, Liu J, et al. (2019) Using extreme gradient boosting to predict changes in tropical cyclone intensity over the western north pacific. *Atmosphere* 10: 341. https://doi.org/10.3390/atmos10060341

Johansen S (2009) Cointegration: Overview and development. *Handbook Financ Time Ser* 671–693. https://doi.org/10.1007/978-3-540-71297-8_29

Jung HS, Lee SH, Lee H, et al. (2023) Predicting bitcoin trends through machine learning using sentiment analysis with technical indicators. *Comput Syst Sci Eng* 46: 2231–2246. https://doi.org/10.32604/csse.2023.034466

Jurado S, Nebot À, Mugica F, et al. (2015) Hybrid methodologies for electricity load forecasting: entropy-based feature selection with machine learning and soft computing techniques. *Energy* 86: 276–291. https://doi.org/10.1016/j.energy.2015.04.039

Khandelwal I, Adhikari R, Verma G (2015) Time series forecasting using hybrid arima and ann models based on dwt decomposition. *Procedia Comput Sci* 48: 173–179. https://doi.org/10.1016/j.procs.2015.04.167

Khodabakhsh A, Ari I, Bakır M, et al. (2020) Forecasting multivariate time-series data using LSTM and mini-batches. In: *Data Science: From Research to Application*, 121–129. Springer. https://doi.org/10.1007/978-3-030-37309-2_10

Kilian L, Zhou X (2021) The impact of rising oil prices on u.s. inflation and inflation expectations in 2020-23. *Energy Econ* 113: 106228. https://doi.org/10.2139/ssrn.3980337

Kitani R, Iwata S (2023) Verification of interpretability of phase-resolved partial discharge using a cnn with shap. *IEEE Access* 11: 4752–4762. https://doi.org/10.1109/access.2023.3236315

Kumar SD, Subha DP (2019) Prediction of depression from EEG signal using long short term memory (LSTM). In: *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 1248–1253. https://doi.org/10.1109/ICOEI.2019.8862560

Kunstmann L, Pina D, Silva F, et al. (2021) Online deep learning hyperparameter tuning based on provenance analysis. *J Inf Data Manage* 12. https://doi.org/10.5753/jidm.2021.1924

Lee K, Ayyasamy MV, Ji Y, et al. (2022) A comparison of explainable artificial intelligence methods in the phase classification of multi-principal element alloys. *Sci Rep* 12: 11591. https://doi.org/10.1038/s41598-022-15618-4

Lees T, Buechel M, Anderson B, et al. (2021) Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. *Hydrol Earth Syst Sci* 25: 5517–5534. https://doi.org/10.5194/hess-25-5517-2021

Lewis PA, Ray BK (1997) Modeling long-range dependence, nonlinearity, and periodic phenomena in sea surface temperatures using TSMARS. *J Am Stat Assoc* 92: 881–893. https://doi.org/10.1080/01621459.1997.10474043

Lewis PA, Stevens JG (1991) Nonlinear modeling of time series using multivariate adaptive regression splines (MARS). *J Am Stat Assoc* 86: 864–877. https://doi.org/10.1080/01621459.1991.10475126

Li G, Yang N (2022) A hybrid sarima-lstm model for air temperature forecasting. *Adv Theor Simul* 6. https://doi.org/10.1002/adts.202200502

Li P, Zhang JS (2018) A new hybrid method for China's energy supply security forecasting based on ARIMA and XGBoost. *Energies* 11: 1687. https://doi.org/10.3390/en11071687

Li S, Huang H, Lu W (2021) A neural networks based method for multivariate time-series forecasting. *IEEE Access* 9: 63915–63924. https://doi.org/10.1109/access.2021.3075063

Li T, Hua M, Wu X (2020) A hybrid cnn-lstm model for forecasting particulate matter (pm2.5). *IEEE Access* 8: 26933–26940. https://doi.org/10.1109/access.2020.2971348

Li X, Huo H, Liu Z (2022) Analysis and prediction of pm2.5 concentration based on lstm-xgboost-svr model. https://doi.org/10.21203/rs.3.rs-2158285/v1

Liu Z (2023) Review on the influence of machine learning methods and data science on the economics. *Appl Comput Eng* 22: 137–141. https://doi.org/10.54254/2755-2721/22/20231208

Liu Y, Yang Y, Chin RJ, et al. (2023) Long Short-Term Memory (LSTM) Based Model for Flood Forecasting in Xiangjiang River. *J Civil Eng* 27: 5030–5040. https://doi.org/10.1007/s12205-023-2469-7

Lv C, An S, Qiao B, et al. (2021) Time series analysis of hemorrhagic fever with renal syndrome in mainland china by using an xgboost forecasting model. *Bmc Infect Dis* 21. https://doi.org/10.1186/s12879-021-06503-y

Medeiros MC, Vasconcelos GF, Veiga Á, et al. (2019) Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *J Bus Econ Stat* 39: 98–119. https://doi.org/10.1080/07350015.2019.1637745

Mitchell DJB (1999) Review of Getting Prices Right: The Debate over the Consumer Price Index, by D. Baker. *Ind Labor Relat Rev* 52: 317–318. https://doi.org/10.2307/2525170

Mohammed AA, Immanuel PJ, Roobini MS (2023) Forecasting Consumer Price Index (CPI) Using Deep Learning and Hybrid Ensemble Technique. *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Chennai, India, 1–8. https://doi.org/10.1109/ACCAI58221.2023.10200153

Mohan S, Hutson A, MacDonald I, et al. (2019) Impact of macroeconomic indicators on housing prices. *Int J Hous Mark Anal* 12: 1055–1071. https://doi.org/10.1108/IJHMA-09-2018-0070

Mulenga M, Kareem SA, Sabri AQM, et al. (2021) Stacking and chaining of normalization methods in deep learning-based classification of colorectal cancer using gut microbiome data. *IEEE Access* 9: 97296–97319. https://doi.org/10.1109/access.2021.3094529

Murat N (2023) Outlier detection in statistical modeling via multivariate adaptive regression splines. *Commun Stat-Simul C* 52: 3379–3390. https://doi.org/10.1080/03610918.2021.2007400

Muruganandam NS, Arumugam U (2023) Dynamic ensemble multivariate time series forecasting model for pm2.5. *Comput Syst Sci Eng* 44: 979–989. https://doi.org/10.32604/csse.2023.024943

Naidu S, Pandaram A, Chand A (2017) A Johansen cointegration test for the relationship between remittances and economic growth of Japan. *Mod Appl Sci* 11: 137–151. https://doi.org/10.5539/mas.v11n10p137

Naser AH, Badr AH, Henedy SN, et al. (2022) Application of Multivariate Adaptive Regression Splines (MARS) approach in prediction of compressive strength of eco-friendly concrete. *Case Stud Constr Mat* 17: e01262. https://doi.org/10.1016/j.cscm.2022.e01262

Nguyen LT, Chung HH, Tuliao KV, et al. (2020) Using xgboost and skip-gram model to predict online review popularity. *SAGE Open* 10: 215824402098331. https://doi.org/10.1177/2158244020983316

Nguyen TT, Nguyen HG, Lee JY, et al. (2023) The consumer price index prediction using machine learning approaches: Evidence from the United States. *Heliyon* 9.

Njenga JK (2024) Analysis and Forecasting of Consumer Price Index (CPI) in Kenya and South Africa using Holt Winter Model. *Asian J Econ Bus Account* 24: 322–331. https://doi.org/10.9734/ajeba/2024/v24i41283

Noorunnahar M, Chowdhury AH, Mila FA. (2023) A tree based extreme gradient boosting (xgboost) machine learning model to forecast the annual rice production in bangladesh. *Plos One* 18: e0283452. https://doi.org/10.1371/journal.pone.0283452

Pan J, Zhang Z, Peters S, et al. (2023) Cerebrovascular disease case identification in inpatient electronic medical record data using natural language processing. https://doi.org/10.21203/rs.3.rs-2640617/v1

Paparoditis E, Politis DN (2018) The asymptotic size and power of the augmented Dickey–Fuller test for a unit root. *Econ Rev* 37: 955–973.

Papíková L, Papík M (2022) Effects of classification, feature selection, and resampling methods on bankruptcy prediction of small and medium-sized enterprises. *Intell Syst Account Financ Manage* 29: 254–281. https://doi.org/10.1002/isaf.1521

Park HJ, Kim Y, Kim HY (2022) Stock market forecasting using a multi-task approach integrating long short-term memory and the random forest framework. *Appl Soft Comput* 114: 108106. https://doi.org/10.1016/j.asoc.2021.108106

Phillips PC, Perron P (1988) Testing for a unit root in time series regression. *Biometrika* 75: 335–346. https://doi.org/10.1093/biomet/75.2.335

Poh CW, Tan R (1997) Performance of Johansen's cointegration test. In: *East Asian Economic Issues: Volume III,* 402–414.

Porcher R, Thomas G (2003) Order determination in nonlinear time series by penalized least-squares. *Commun Stat-Simul C* 32: 1115–1129. https://doi.org/10.1081/SAC-120023881

Qinghe Z, Wen X, Huang B, et al. (2022) Optimised extreme gradient boosting model for short term electric load demand forecasting of regional grid system. *Sci Rep* 12: 19282. https://doi.org/10.1038/s41598-022-22024-3

Radev L, Golitsis P, Mitreva M. (2023) Economic and financial determinants of gold etf price volatility on the u.s. futures market (comex). *J Econ* 8: 12–26. https://doi.org/10.46763/joe2382012r

Raheem Ahmed R, Vveinhardt J, Štreimikienė D, et al. (2017) Estimation of long-run relationship of inflation (cpi & wpi), and oil prices with kse-100 index: Evidence from johansen multivariate cointegration approach. *Technol Econ Dev Econ* 23: 567–588. https://doi.org/10.3846/20294913.2017.1289422

Reddy S, Akashdeep S, Harshvardhan R, et al. (2022) Stacking Deep learning and Machine learning models for short-term energy consumption forecasting. *Adv Eng Inform* 52: 101542. https://doi.org/10.1016/j.aei.2022.101542

Reed SB (2014) One hundred years of price change: The Consumer Price Index and the American inflation experience. *Monthly Lab Rev* 137: 1.

Rezaie-Balf M, Zahmatkesh Z, Kim S (2017) Soft computing techniques for rainfall-runoff simulation: local non–parametric paradigm vs. model classification methods. *Water Resour Manag* 31: 3843–3865. https://doi.org/10.1007/s11269-017-1711-9

Ribeiro MHDM, Silva RG, Mariani VC, et al. (2021) Dengue cases forecasting based on extreme gradient boosting ensemble with coyote optimization. *Anais Do* 15. *Congresso Brasileiro De Inteligência Computacional*. https://doi.org/10.21528/cbic2021-36

Ribeiro MHDM, Stefenon SF, Lima JD, et al. (2020) Electricity price forecasting based on self-adaptive decomposition and heterogeneous ensemble learning. *Energies* 13: 5190. https://doi.org/10.3390/en13195190

Rippy D (2014) The first hundred years of the Consumer Price Index: a methodological and political history. *Monthly Lab Rev* 137: 1.

Rosado R, Abreu AJ, Arencibia JC, et al. (2021) Consumer price index forecasting based on univariate time series and a deep neural network. In: *International Workshop on Artificial Intelligence and Pattern Recognition*, 33–42. Cham: Springer. https://doi.org/10.1007/978-3-030-89691-1_4

Rodríguez-Pérez R, Bajorath J (2020) Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *J Comput Aid Mol Des* 34: 1013–1026. https://doi.org/10.1007/s10822-020-00314-0

Sagheer A, Kotb M (2019) Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems. *Sci Rep* 9: 19038. https://doi.org/10.1038/s41598-019-55320-6

Saputra AW, Wibawa AP, Pujianto U, et al. (2022) LSTM-based Multivariate Time-Series Analysis: A Case of Journal Visitors Forecasting. *ILKOM J Ilm* 14: 57–62.

Sarangi PK, Sahoo AK, Sinha S (2022) Modeling consumer price index: a machine learning approach. *Macromol Sym* 401. https://doi.org/10.1002/masy.202100349

Setyanto A, Laksito A, Alarfaj F, et al. (2022). Arabic language opinion mining based on long short-term memory (LSTM). *Appl Sci* 12: 4140. https://doi.org/10.3390/app12094140

Shahbaz M, Khraief N, Mahalik MK (2020) Investigating the environmental Kuznets's curve for Sweden: Evidence from multivariate adaptive regression splines (MARS). *Empir Econ* 59: 1883–1902. https://doi.org/10.1007/s00181-019-01698-1

Sharda VN, Prasher SO, Patel RM, et al. (2008) Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data/Performances de régressions par splines multiples et adaptives (MARS) pour la prévision d'écoulement au sein de micro-bassins versants Himalayens d'altitudes intermédiaires avec peu de données. *Hydrolog Sci J* 53: 1165–1175. https://doi.org/10.1623/hysj.53.6.1165

Sharma SS (2016) Can consumer price index predict gold price returns? *Econ Model* 55: 269–278. https://doi.org/10.1016/j.econmod.2016.02.014

Shi F, Lu S, Gu J, et al. (2022) Modeling and evaluation of the permeate flux in forward osmosis process with machine learning. *Ind Eng Chem Res* 61: 18045–18056. https://doi.org/10.1021/acs.iecr.2c03064

Shiferaw Y (2023) An understanding of how gdp, unemployment and inflation interact and change across time and frequency. *Economies* 11: 131. https://doi.org/10.3390/economies11050131

Siami-Namini S, Tavakoli N, Namin AS (2019) The performance of LSTM and BiLSTM in forecasting time series. In: *2019 IEEE International conference on big data (Big Data)*, 3285–3292, IEEE.

Sibai N, El-Moursy F, Sibai A (2024) Forecasting the consumer price index: a comparative study of machine learning methods. *Int J Comput Digit Syst* 15: 487–497. https://doi.org/10.12785/ijcds/150137

Simsek AI (2024) Improving the Performance of Stock Price Prediction: A Comparative Study of Random Forest, XGBoost, and Stacked Generalization Approaches. In: *Revolutionizing the Global Stock Market: Harnessing Blockchain for Enhanced Adaptability*, 83–99. IGI Global.

Subhani MI (2009) Relationship between Consumer Price Index (CPI) and government bonds. *S Asian J Manage Sci* 3: 11–17.

Sukarsa IM, Pinata NNP, Rusjayanthi NKD, et al. (2021) Estimation of gourami supplies using gradient boosting decision tree method of xgboost. *TEM J* 144–151. https://doi.org/10.18421/tem101-17

Sumita S, Nakagawa H, Tsuchiya T (2023) Xtune: an xai-based hyperparameter tuning method for time-series forecasting using deep learning. https://doi.org/10.21203/rs.3.rs-3008932/v1 Shimon

Sun Y, Tian L (2022) Research on stock prediction based on simulated annealing algorithm and ensemble neural learning. *Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)*. https://doi.org/10.1117/12.2663138

Tan KR, Seng JJB, Kwan YH, et al. (2021) Evaluation of machine learning methods developed for prediction of diabetes complications: a systematic review. *J Diabetes Sci Techn* 17: 474–489. https://doi.org/10.1177/19322968211056917

Temür AS, Yildiz Ş (2021) Comparison of forecasting performance of arima lstm and hybrid models for the sales volume budget of a manufacturing enterprise. *Istanb Bus Res* 50: 15–46. https://doi.org/10.26650/ibr.2021.51.0117

Thapa KB (2023) Macroeconomic determinants of the stock market in nepal: an empirical analysis. *NCC J* 8: 65–73. https://doi.org/10.3126/nccj.v8i1.63087

Tian L, Feng L, Sun Y, et al. (2021) Forecast of lstm-xgboost in stock price based on bayesian optimization. *Intell Autom Soft Comput* 29: 855–868. https://doi.org/10.32604/iasc.2021.016805

Toraman C, Basarir Ç (2014) The long run relationship between stock market capitalization rate and interest rate: Co-integration approach. *Procedia-Soc Behav Sci* 143: 1070–1073. https://doi.org/10.1016/j.sbspro.2014.07.557

Upadhyaya Y, Kharel K (2022) Inflation with gdp, unemployment and remittances: an outline of the joint effect on nepalese economy. *Interd J Manage Soc Sci* 3: 154–163. https://doi.org/10.3126/ijmss.v3i1.50244

Utama ABP, Wibawa AP, Muladi M, et al. (2022) Pso based hyperparameter tuning of cnn multivariate time- series analysis. *J Online Inform* 7: 193–202. https://doi.org/10.15575/join.v7i2.858

Varian HR (2014) Big data: New tricks for econometrics. *J Econ Perspect* 28: 3–28. https://doi.org/10.1257/jep.28.2.3

Vasco-Carofilis RA, Gutiérrez–Naranjo MA, Cárdenas-Montes M (2020) Pbil for optimizing hyperparameters of convolutional neural networks and stl decomposition. *Lect Notes Comput Sci* 147–159. https://doi.org/10.1007/978-3-030-61705-9_13

Vlachas PR, Byeon W, Wan Z, et al. (2018) Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *P Roy Soc A-Math Phy* 474: 20170844. https://doi.org/10.1098/rspa.2017.0844

Wan R, Mei S, Wang J, et al. (2019) Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting. *Electronics* 8: 876. https://doi.org/10.3390/electronics8080876

Wang L, Zhao L (2022) Digital economy meets artificial intelligence: forecasting economic conditions based on big data analytics. *Mob Inf Syst* 2022: 1–9. https://doi.org/10.1155/2022/7014874

Wang L, Haofei Z, Su J, et al. (2013) An arima-ann hybrid model for time series forecasting. *Syst Res Behav Sci* 30: 244–259. https://doi.org/10.1002/sres.2179

Wang W, Shi Y, Lyu G, et al. (2017) Electricity consumption prediction using xgboost based on discrete wavelet transform. *DEStech T Comput Sci Eng*. https://doi.org/10.12783/dtcse/aiea2017/15003

Wang Y, Ye G (2020) Forecasting method of stock market volatility in time series data based on mixed model of arima and xgboost. *China Commun* 17: 205–221. https://doi.org/10.23919/jcc.2020.03.017

Wang Y, Bao F, Hua Q, et al. (2021). Short-term solar power forecasting: a combined long short-term memory and gaussian process regression method. *Sustainability* 13: 3665. https://doi.org/10.3390/su13073665

Wei B, Yue J, Rao Y (2017) A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *Plos One* 12: e0180944. https://doi.org/10.1371/journal.pone.0180944

Weinzierl M (2014) Seesaws and social security benefits indexing(No. w20671). *National Bureau of Economic Research*, Cambridge.

Widiputra H, Mailangkay ABL, Gautama E (2021) Multivariate cnn-lstm model for multiple parallel financial time-series prediction. *Complexity* 2021: 1–14. https://doi.org/10.1155/2021/9903518

Qureshi M, Khan A, Daniyal M, et al. (2023) A comparative analysis of traditional sarima and machine learning models for cpi data modelling in pakistan. *Appl Comput Intell S* 2023: 1–10. https://doi.org/10.1155/2023/3236617

Xiao C, Wang Y, Wang S (2023) Machine learning to set hyperparameters for overlapping community detection algorithms. *J Eng* 2023. https://doi.org/10.1049/tje2.12292

Xu J, He J, Gu J, et al. (2022) Financial Time Series Prediction Based on XGBoost and Generative Adversarial Networks. *Int J Circ Syst Signal Process* 16: 637–645. https://doi.org/10.46300/9106.2022.16.79

Yang C, Guo S (2021) Inflation prediction method based on deep learning. *Comput Intel Neurosc* 2021. https://doi.org/10.1155/2021/1071145

Yang L, Shami A (2020) On hyperparameter optimization of machine learning algorithms: theory and practice. *Neurocomputing* 415: 295–316. https://doi.org/10.1016/j.neucom.2020.07.061

Ye M, Mohammed KS, Tiwari S, et al. (2023) The effect of the global supply chain and oil prices on the inflation rates in advanced economies and emerging markets. *Geo J* 58: 2805–2817. https://doi.org/10.1002/gj.4742

Yilmazkuday H (2024) Pass-through of shocks into different u.s. prices. *Rev Int Econ* 32: 1300–1315. https://doi.org/10.1111/roie.12726

Yildiz M, Ozdemir L (2022) Determination of the sensitivity of stock index to macroeconomic and psychological factors by MARS method, in: Insurance and Risk Management for Disruptions in Social, Economic and Environmental Systems: Decision and Control Allocations within New Domains of Risk. *Emerald Publishing Limited* 2022: 81–105. https://doi.org/10.1108/978-1-80117-139-720211005

Yuan M, Yang N, Qian Z, et al. (2020) What makes an online review more helpful: an interpretation framework using xgboost and shap values. *J Theor Appl El Comm* 16: 466–490. https://doi.org/10.3390/jtaer16030029

Zahara SS, Ilmiddaviq MB (2020) Consumer price index prediction using Long Short Term Memory (LSTM) based cloud computing. *J Phys* 1456: 1–8, IOP Publishing. https://doi.org/10.1088/1742-6596/1456/1/012022

Zazo R, Lozano-Diez A, Gonzalez-Dominguez J, et al. (2016) Language identification in short utterances using long short-term memory (LSTM) recurrent neural networks. *PloS One* 11: e0146917. https://doi.org/10.1371/journal.pone.0146917

Zhai N, Yao P, Zhou X (2020) Multivariate time series forecast in industrial process based on XGBoost and GRU. In: *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)* 9: 1397–1400.

Zhang J, Meng Y, Jin W (2021) A novel hybrid deep learning model for sugar price forecasting based on time series decomposition. *Math Probl Eng* 2021: 1–9. https://doi.org/10.1155/2021/6507688

Zhang J, Wen J, Yang Z (2022) China's GDP forecasting using Long Short Term Memory Recurrent Neural Network and Hidden Markov Model. *Plos One* 17: e0269529. https://doi.org/10.1371/journal.pone.0269529

Zhang X, Yang E (2024) Have housing value indicators changed during COVID? Housing value prediction based on unemployment, construction spending, and housing consumer price index. *Int J Hous Mark Anal* 17: 242–260. https://doi.org/10.1108/IJHMA-01-2023-0015

Zhou S, Zhou L, Mao M, et al. (2019) An optimized heterogeneous structure lstm network for electricity price forecasting. *IEEE Access* 7: 108161–108173. https://doi.org/10.1109/access.2019.2932999

Zhou X, Pranolo A, Mao Y (2023) AB-LSTM: Attention Bidirectional Long Short-Term Memory for Multivariate Time-Series Forecasting. In: *2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3)*, 1–6.

Zhou Z, Song Z, Ren T (2022) Predicting China's CPI by Scanner Big Data. *arXiv preprint* arXiv:2211.16641.

Zhu C, Ma X, Zhang C, et al. (2023) Information granules-based long-term forecasting of time series via BPNN under three-way decision framework. *Inf Sci* 634: 696–715. https://doi.org/10.1016/j.ins.2023.03.133

Zyprych-Walczak J, Szabelska A, Handschuh L, et al. (2015) The impact of normalization methods on rna-seq data analysis. *Biomed Res Int* 2015: 1–10. https://doi.org/10.1155/2015/621690