Leveraging machine learning for Australian macroeconomic forecasting

Boyuan Li[1] Zihui Deng[2] Gaurav Gupta[3]
Yingxi Chen[4]

(Received 2 March 2025; revised 18 October 2025)

Abstract

Macroeconomic forecasting has traditionally relied on regression and time series methods, which have often struggled to outperform basic benchmarks. In this study, we explore the use of machine learning (ML) and deep learning (DL) techniques to forecast key macroeconomic indicators for Australia, specifically focusing on Gross Domestic Product growth, Consumer Price Index inflation, and the Interbank Overnight Cash Rate. Using a comprehensive dataset spanning from 1985 to 2023, we incorporate 16 predictors identified from previous literature. Our findings indicate that ML methods, particularly ensemble approaches such as Random Forest and XGBoost, deliver superior predictive performance compared to traditional time series models and

DOI:10.21914/anziamproc.v66.19641 Published 2026-04-07, as part of the Proceedings of the 22nd Biennial Computational Techniques and Applications Conference. ISSN 1445-8810. (Print two pages per sheet of paper.)

DL methods. Although DL models such as Long Short-Term Memory (LSTM) and Convolutional Neural Network-LSTM were also tested, they did not achieve comparable accuracy in this context, highlighting the effectiveness of density-based ML approaches for macroeconomic forecasting.

Contents

1 Introduction

Macroeconomic forecasting is essential in macroeconomic policies development as it provides essential input for forward-looking policy deliberations [7].

However, macroeconomic data is often characterised by low serial correlation, high volatility, or structural changes, which make it hard to predict [5].

Forecasting models that have become widely used since the 1960s, such as econometric models [3, 10], input–output models [17, 28], and dynamic systems [15], were later shown to have incorrect economic predictions [14]. According to Fildes and Stekler [11], such forecasting models struggle with precision in critical areas. Notably, they tend to perform poorly in forecasting economic downturns and peaks, and they are often unable to assess inflation trends accurately. Diebold [9] explained that the failure of large-scale Keynesian frameworks to account for structural changes in the economy weakened their forecasting ability, and emphasized that the limitations posed by the assumption of stable relationships between economic variables often broke down during periods of rapid economic change. Unstructured methods such as Vector Autoregressive (VAR) Dynamic Stochastic General Equilibrium (DSGE), and Autoregressive Moving Average (ARMA) models have been proven to be more flexible and promising in capturing the dynamic characteristics of economic systems [9].

By enabling systems to learn from data and generate accurate predictions or decisions, Machine Learning (ML) models have become popular for economic forecasting tasks[2, 19, 12]. Giannone et al. [13] demonstrated that ML techniques, such as Bagging and Boosting, are effective for better predictions in high-dimensional data. Similarly, Medeiros et al. [20] pointed out that simple models, such as Phillips curve-based or time series models, often fail to predict inflation effectively. Moreover, Random Forest (RF) systematically outperforms benchmark models in experiments [20]. As Deep Learning (DL) models, based on natural networks, gained momentum in forecasting due to their ability to capture nonlinear patterns and model long-term dependencies in complex datasets, they were also put to use in numerous economic and financial forecasting [24, 27, 30]. Algorithms such as VAR DSGE and ARMA are based on linear equations, which cannot adapt to real-world economic data composed of non-linearities and structural breaks. ML and DL models, like tree-based models or neutral network like Long Short-Term Memory (LSTM),

can capture complex, non-linear relationships without stringent linearity assumptions.

For Australian macroeconomic forecasting, Panagiotelis et al. [23] applied models including the Dynamic Factor Model (DFM), Ridge Regression (RR), Least Angle Regression (LAR), and Bayesian Vector Autoregression (BVAR), and compared these to benchmark models such as the Autoregression Model and Random Walk (RW). They found that the accuracy of most methods is only slightly better than or equal to that of simple benchmark models when predicting Gross Domestic Product (GDP) growth, Consumer Price Index (CPI) inflation and the Interbank Overnight Cash Rate (IBR). Furthermore, large-scale datasets with more predictors and advanced models may not significantly improve prediction performance [23].

Based on the research of Panagiotel et al. [23] we incorporated 16 predictors, focusing on real GDP growth, CPI inflation, and IBR to demonstrate the predictive capability of macroeconomic models. We further apply advanced ML and DL models to evaluate their performance relative to the RW model benchmark, aiming to determine whether advanced models outperform traditional approaches. Additionally, we draw on Yoon's [29] findings that gradient boosting models and RF exhibit superior accuracy compared to benchmark models in forecasting GDP growth . Our objective is to assess whether Yoon's conclusions hold true when applied to the Australian dataset.

2 Method

In this study, we evaluated the performance of several forecasting models by comparing them to a benchmark model (RW) to test the predictive performance of macroeconomics. We focused on three key macroeconomic indicators: GDP, CPI, and IBR. The models analyzed include ARIMAX, RF, XGBoost, LSTM, and Convolutional LSTM Fully Connected Networks (CNN-LSTM). In addition, we explore an ensemble model that combines RF and XGBoost.

2.1 Dataset

We extracted the dataset from Wind financial terminal [6] and the Australian Bureau of Statistics [25]. The data set contains 19 key economic indicators, identified by a thorough review of the existing literature [23]. These indicators, spanning from June 1985 to September 2023, provide a quarterly snapshot of essential macroeconomic activities. The full list of variables is detailed in Table 1. In particular, our analysis emphasizes three critical factors: gross GDP growth, the CPI reflecting inflation trends, and the IBR, which serves as a primary measure of monetary policy adjustments in Australia.

The indicators within this dataset capture a broad spectrum of economic dimensions, such as international trade (exports and imports), financial liquidity (Credit-Total, Broad Money), price movements (commodity prices, terms of trade), labor market dynamics (unemployment rate, private dwelling approvals), and industrial activities (industrial production index). Each of these variables plays a significant role in understanding the economic health and fluctuations within the market over time.

2.2 Models

We apply a combination of ML and DL techniques to forecast key macroeconomic indicators for Australia. Model performance is evaluated against a basic benchmark, with the RW model serving as the baseline. We employ ARIMAX as the time series method, RF and XGBoost as ML methods, and an ensemble of these two models. For DL methods, we utilize LSTM, CNN-LSTM, and FNN. Model performance is assessed using Root Mean Squared Error (RMSE) and Mean Absolute Scaled Error (MASE).

2.2.1 Random walk

Random Walk (RW) is a classic time series model used to describe the evolution of a variable over time, assuming that the current value is simply a continuation of the previous value with the addition of a random error term.

Table 1: Macroeconomic variables of Australia.

Name	Description
RGDP	Real Gross Domestic Product, Chain Volume Measures (Inflation-adjusted GDP)
CPI	Consumer Price Index: All Groups
IBR	Interbank Overnight Cash Rate
Exports	Exports of Goods and Services
Imports	Imports of Goods and Services
COMMP	Commodity Price Index (All Items). In AUD, with the base period 2022/23 = 100.
Credit-Total	Total Credit Available
Mbase	Monetary Base
Termtrade	Terms of Trade Index (Ratio of export prices to import prices)
M1	Money Supply M1 (Narrow Money)
M3	Money Supply M3 (Broad Money)
BM	Broad Money (alternative measure)
10 year T-bond	10-Year Treasury Bond Yield
Hstarts-PDA	Private Dwelling Approvals (Housing Starts)
IP-Total	Industrial Production Index: Total Industrial Industries
SP ASX AllOrd	S&P ASX All Ordinaries Index Adjusted Closing Prices
AveComp	Average Compensation per Employee (Non-farm sector), Current Prices (AUD)
Urate	Unemployment Rate (Percentage of the labour force)
consfinal	Household Final Consumption Expenditure: Chain Volume Measures, AUD (Real household spending)

The fundamental formula for the RW model is

$$y_t = y_{t-1} + \epsilon_t \,, \tag{1}$$

where y_t is the value at time t, y_{t-1} is the value at the previous time step, and ϵ_t is an independent and identically distributed random error term, under the hypothesis of having a mean of 0 and constant variance. The model assumes no inherent trend or seasonality in the time series, which means the past value and an unpredictable random disturbance determine the future value. The RW model's simplicity makes it a useful baseline for evaluating more complex forecasting models [21, 22].

2.2.2 ARIMAX

The ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) model extends the ARIMA model by including external explanatory variables, which aims to enhance forecasting accuracy. The standard form of the ARIMAX model is

$$\phi_p(B)\nabla^d y_t = \mu + \nu(B)x_{mt} + \theta_q(B)\epsilon_t \,. \tag{2}$$

In this equation, y_t denotes the time series value at time t; $\phi_p(B)$ and $\theta_q(B)$ are polynomials that describe the autoregressive (AR) and moving average (MA) components [1]; and x_{mt} represents the external explanatory variables, with $\nu(B)$ as their associated polynomial. Lastly, ϵ_t stands for a white noise error term, characterized by a mean of 0 and constant variance. This model is particularly useful when external factors play a major role in influencing the target variable.

2.2.3 Random forest

Random Forest (RF) is an ensemble learning technique that creates multiple independent decision trees. Each tree is trained using a bootstrap sample from the original dataset, and at each node, a random subset of features, $\mathcal{F}_k$,

is selected to determine the best split. The optimal split point is chosen to maximize information gain:

$$\Delta I = I(D) - \sum_{v} \frac{|D_v|}{|D|} I(D_v), \tag{3}$$

where D represents the dataset at the current node, and D_v are the subsets formed after splitting. For classification tasks, RF uses a majority voting scheme across all trees to decide the final class label for a given input. During the construction of each tree, the data at each node is split based on the feature and split point that yields the greatest information gain, measured through metrics such as entropy or Gini impurity. This process is repeated recursively until the trees are fully grown.

For regression tasks, each tree $f_t(x)$ produces a numerical prediction and the final prediction of the model is the average of predictions from all trees:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} f_t(x), \tag{4}$$

where T is the total number of trees, and $f_t(x)$ is the prediction from tree number t [26].

2.2.4 eXtreme gradient boosting (XGBoost)

XGBoost is an optimized gradient-boosting algorithm, that constructs an ensemble of decision trees to predict continuous values in regression tasks. The core idea of XGBoost is to iteratively add new trees that correct the residuals of previous models, improving the overall accuracy of predictions. The model formally adds a new tree f_t to the ensemble at each iteration t. The prediction for the ith instance is

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i), \tag{5}$$

where $f_t(x_i)$ is the output of the new tree, η is the learning rate, which is a shrinkage parameter to avoid overfitting, and $\hat{y}_i^{(t-1)}$ is the prediction from the previous iteration.

For a regression task, the function f_t is learned by minimizing a regularized objective that combines a squared error loss with a penalty for model complexity:

$$\mathrm{Obj}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t)\,, \tag{6}$$

where $l(\cdot)$ is a differentiable convex loss function (e.g., squared error for regression), and $\Omega(f_t)$ is a regularization term that penalizes the complexity of tree f_t, defined as

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2\,. \tag{7}$$

Here, T is the number of leaves in the tree, w_j is the score on leaf j, γ is the complexity cost for adding a new leaf, and λ is the L2 regularization term on the leaf weights.

To make the optimization tractable, the objective is approximated using a second-order Taylor expansion. For a squared error loss $l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$, the gradient g_i and Hessian h_i are

$$g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) = -2(y_i - \hat{y}_i^{(t-1)})\,, \quad h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) = 2\,.$$

The optimal weight w_j^* for a leaf j is determined by leveraging the gradient and Hessian of the squared error loss function:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}\,. \tag{8}$$

This framework allows XGBoost to efficiently minimize the loss by leveraging both the gradient (first-order) and the Hessian (second-order) information, leading to robust and highly accurate predictive models [8].

2.2.5 Long short-term memory

Long Short-Term Memory (LSTM) networks capture long-term dependencies in sequential data by using a gating mechanism to regulate information flow,

addressing the vanishing gradient problem. Each LSTM cell contains three primary gates: the forget gate, input gate, and output gate, which collectively manage the cell state C_t and hidden state h_t. These gates remain identical across both classification and regression tasks. However, in regression tasks, adjustments are made to the output layer and loss function to accommodate continuous value prediction.

The forget gate determines the portion of the previous cell state C_{t-1} to retain

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)\,, \tag{9}$$

where σ is the sigmoid function, W_f is the weight matrix, h_{t-1} represents the prior hidden state, and x_t is the input at time t. As a bias term, b_f allows the model to adjust the output of the forget gate.

The input gate regulates new information added to the cell state, combining it with a candidate cell state $\tilde{C}_t$:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)\,, \quad \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)\,, \tag{10}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t\,. \tag{11}$$

The output gate computes the hidden state h_t for the current time step:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)\,, \quad h_t = o_t \cdot \tanh(C_t)\,. \tag{12}$$

In regression tasks, the main adjustment lies in the output layer, where the hidden state h_t is passed through a linear layer to produce the predicted value

$$\hat{y}_t = W_h \cdot h_t + b_h\,, \tag{13}$$

where W_h and b_h are the weight matrix and bias term of the linear layer, respectively. This linear activation ensures that the output is a continuous value, suitable for regression tasks. These modifications allow LSTM networks to capture continuous patterns and dependencies in sequential data [16].

2.2.6 Feedforward neural network

A Feedforward Neural Network (FNN) is built on interconnected layers, formed by neurons. The main objective of an FNN is to approximate a function $f : \mathbb{X} \to \mathbb{Y}$, with $\mathbb{X}$ as the input features and $\mathbb{Y}$ as the continuous target value. The neurons' outputs in each layer l is

$$a^{[l]} = \sigma(W^{[l]} a^{[l-1]} + b^{[l]}), \tag{14}$$

where $W^{[l]}$ is the weight matrix, $b^{[l]}$ is the bias vector, and σ are showed as the activation function. The output layer employs a linear activation function to generate the continuous predicted value

$$\hat{y} = W^{[L]} a^{[L-1]} + b^{[L]}. \tag{15}$$

Here $W^{[L]}$ and $b^{[L]}$ are the weight matrix and bias vector of the output layer [4].

2.2.7 CNN-LSTM

In the CNN-LSTM model, the CNN component extracts short-term patterns from the input sequence through 1D convolution operations. The convolution operation that defines the output at position i is

$$z_i = \sum_{j=0}^{k-1} W_j \cdot x_{i+j} + b, \tag{16}$$

where W_j represents the filter's weight at j, x_{i+j} is the input value at position $i+j$, and b stands for the bias term. This convolution operation extracts local features from the sequence. The filters move across the input data, with each filter learning to recognize specific patterns within the sequence. In the end, the resulting output is then passed to the LSTM component for further temporal processing [18].

2.2.8 Ensemble model

We build an ensemble model by combining RF and XGBoost, which are two models that showed best performance, averaging two minimised variances and bias to enhance performance and increase robustness. This ensemble model is

$$\hat{y}_{\text{ensemble}} = \frac{1}{2}\left(\hat{y}_{\text{RF}} + \hat{y}_{\text{XGBoost}}\right), \tag{17}$$

where $\hat{y}_{\text{RF}}$ and $\hat{y}_{\text{XGBoost}}$ are the predictions from the RF and XGBoost models, respectively. The final prediction $\hat{y}_{\text{ensemble}}$ is the average of these two predictions, giving equal weight to each model.

2.3 Evaluation indicators

We assess the performance of our models with Root Mean Squared Error (RMSE) and Mean Absolute Scaled Error (MASE). The RMSE measures the standard deviation of the residuals:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{18}$$

where n is the number of predictions, y_i represents true value at time i, and $\hat{y}_i$ is the predicted value at time i. RMSE gives higher weight to larger errors, making it particularly sensitive to outliers. A lower RMSE indicates better performance by the model.

The MASE compares the model's forecast accuracy against a naïve forecast:

$$\text{MASE} = \frac{\frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|}{\frac{1}{n-1}\sum_{t=2}^{n}|y_t - y_{t-1}|}, \tag{19}$$

where y_t is the actual value at time t, $\hat{y}_t$ is the predicted value, and $|y_t - y_{t-1}|$ represents the absolute error from the naïve forecast.

Table 2: Relative performance of forecasting models.

Indicator	ARIMAX	RF	XgBoost	LSTM	CNN-LSTM	FNN	Ensemble
CPI RMSE	3.943	**0.571**	1.114	3.714	12.343	11.686	**0.137**
CPI MASE	3.979	**0.707**	1.313	4.301	17.421	13.540	**0.144**
RGDP RMSE	1.955	**0.682**	**0.773**	2.818	4.909	7.364	**0.127**
RGDP MASE	2.276	1.410	1.610	5.430	9.932	11.334	**0.186**
IBR RMSE	2.994	**0.500**	**0.500**	4.093	8.023	5.244	**0.116**
IBR MASE	2.993	**0.354**	**0.315**	3.507	8.673	4.291	**0.087**

3 Result

Table 2 shows the relative performance of each model compared to the benchmark model. The values represent the ratio of the model's error metrics (RMSE and MASE) to those of the benchmark. A value less than 1 indicates a performance better than the benchmark, whereas a value greater than 1 indicates poorer performance. Our analysis reveals that the ARIMAX model underperformed across all indicators, with significantly higher RMSE and MASE than the RW benchmark in predictions among the three indicators. The used DL models, including LSTM, CNN-LSTM, and FNN also failed to surpass the benchmark, with the CNN-LSTM showing significantly high RMSE (12.343) and MASE (17.421) for CPI predictions. ML models RF and XGBoost showed better predictive competence, especially in predicting RGDP and IBR, achieving RMSE and MASE values below 1. The ensemble of RF and XGBoost models outperformed all other approaches, with RMSE and MASE values for RGDP at 0.114 and 0.185.

Figure 1 shows a scatter plot of the results and marks the models with both RMSE and MASE ratios. Figures 2, 3 and 4 are feature importance charts for the random forest, XGBoost, and their combined ensemble model for CPI, RGDP and IBR, respectively.

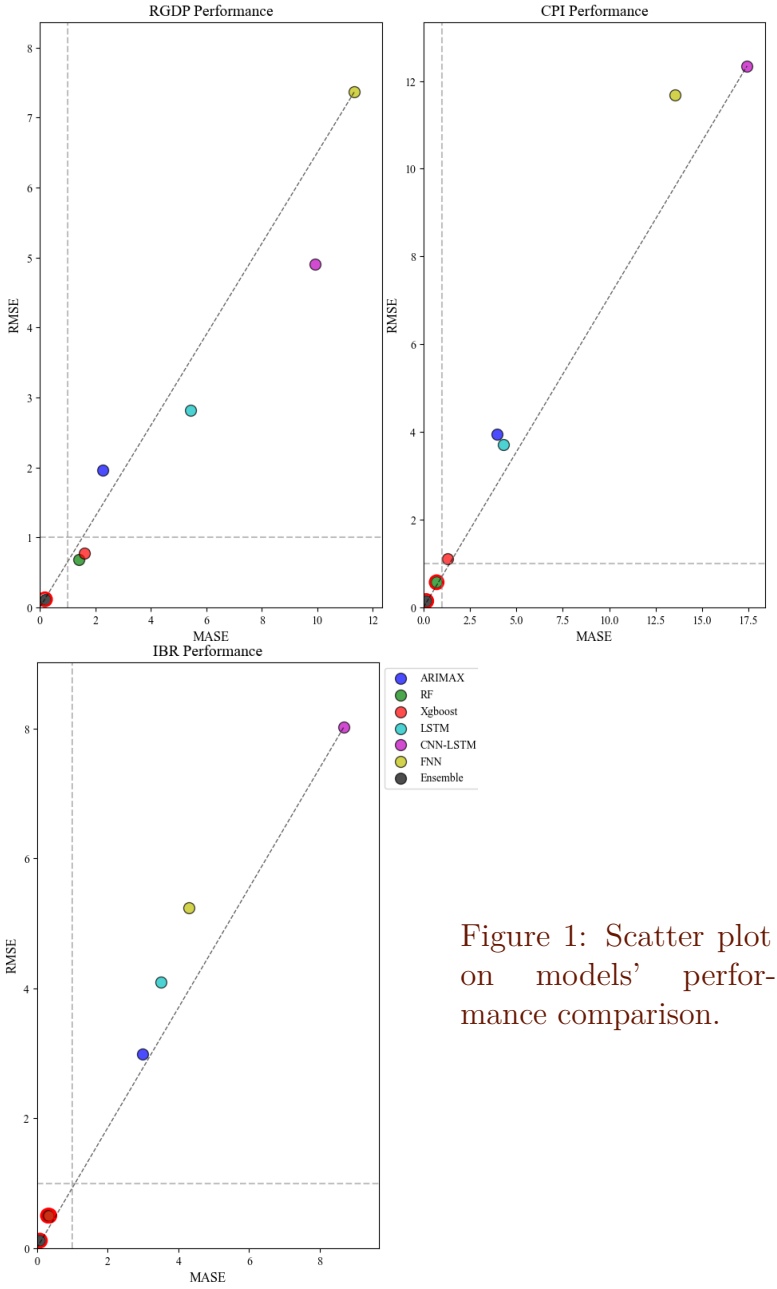

Figure 1: Scatter plot on models' performance comparison.

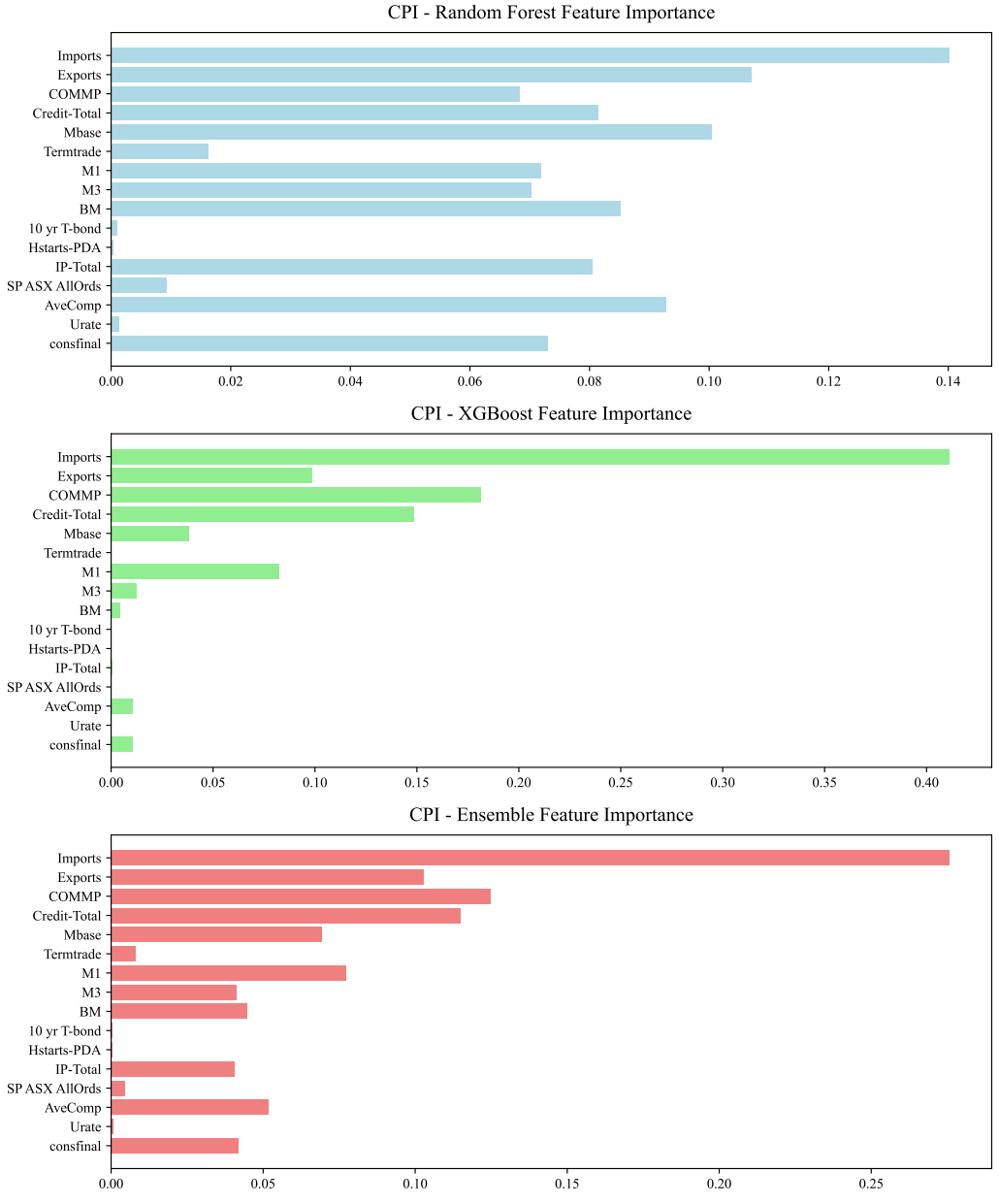

Figure 2: Feature contribution on CPI forecasting.

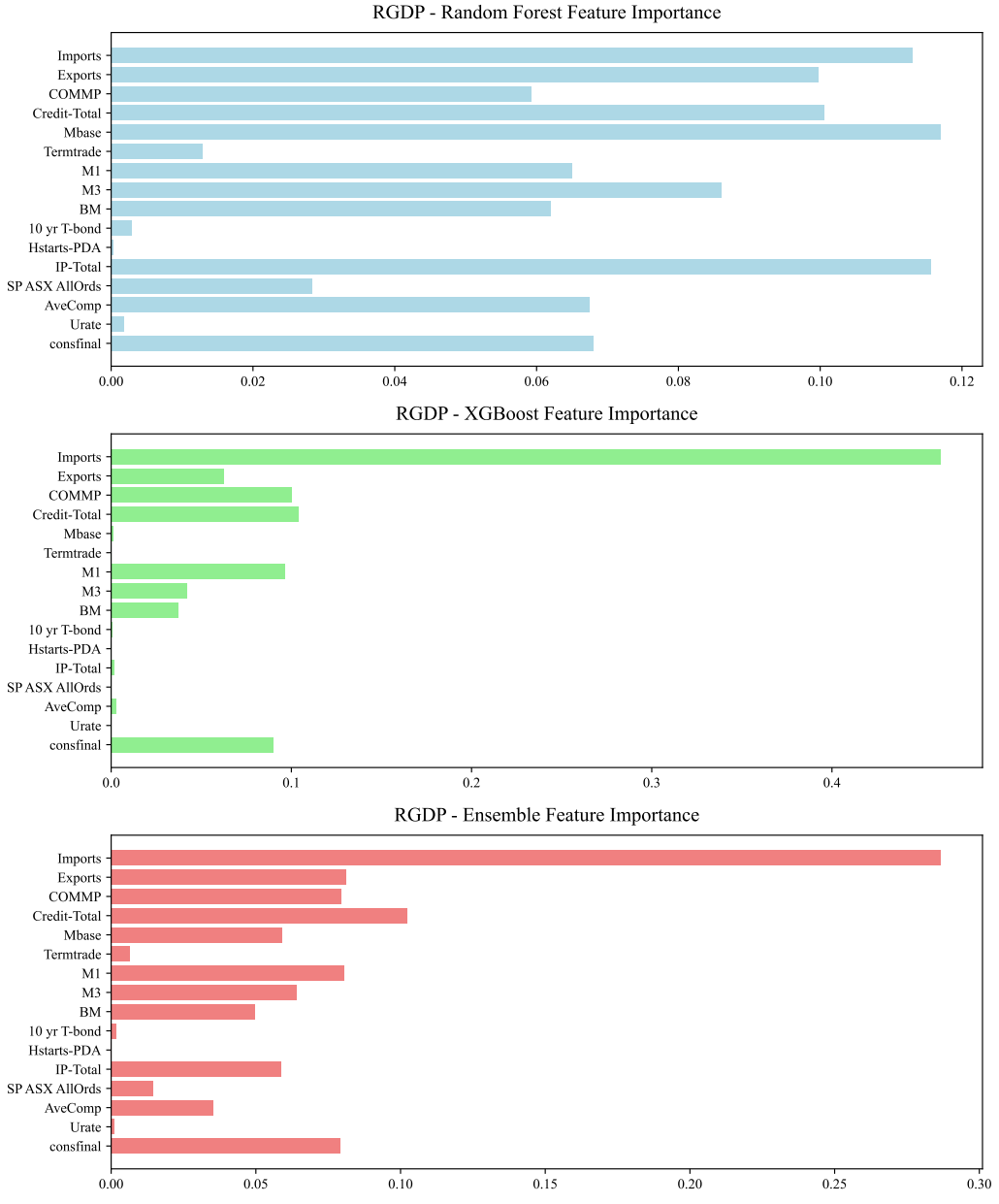

Figure 3: Feature contribution on GDP forecasting.

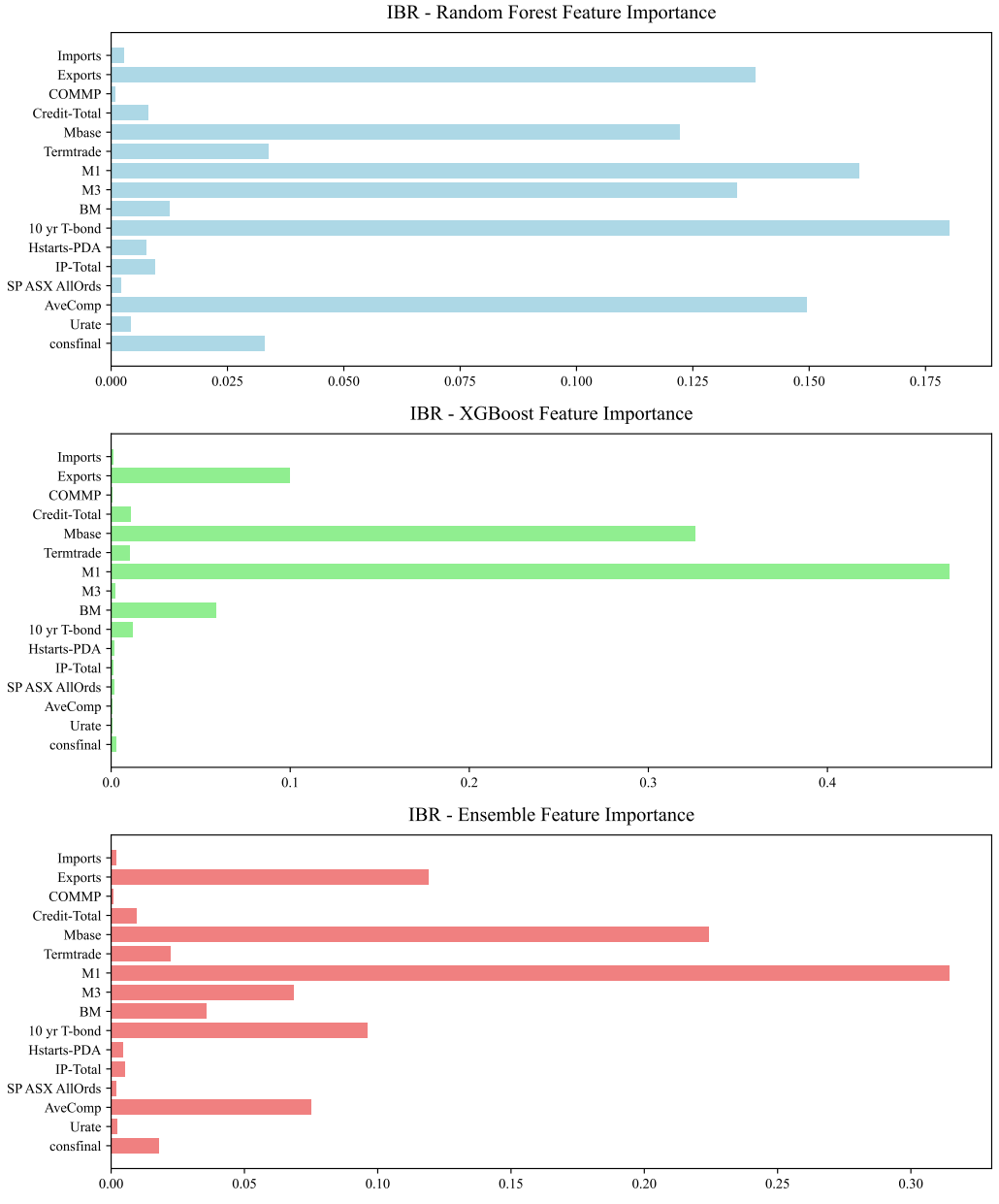

Figure 4: Feature contribution on IBR forecasting.

Table 3: Random Forest parameters for different target variables.

Target	Parameter	Value
CPI	n_estimators	266
	max_depth	11
	min_samples_split	2
	min_samples_leaf	1
RGDP	n_estimators	413
	max_depth	16
	min_samples_split	2
	min_samples_leaf	1
IBR	n_estimators	413
	max_depth	16
	min_samples_split	2
	min_samples_leaf	1

3.1 Model configuration

We implemented hyperparameter tuning for the target variables. The best-performing parameters were recorded for the RF, XGBoost, and Ensemble Model and are shown in Tables 3, 4 and 5, respectively.

4 Discussion

We apply ML and DL models for forecasting macroeconomic increments in Australia, mainly focusing on GDP growth, CPI inflation, and the IBR. By comparing performance with the RW model, we found that RF and XGBoost models have the potential ability for prediction, which also aligns with Yoon's result [29].

Feature importance analysis claims that import prices and quantities play a critical role in predicting CPI, GDP, and the IBR. For CPI, imports appeared as the most influential factor in both the RF and ensemble models, emphasizing

Table 4: XGBoost parameters for different target variables.

Target	Parameter	Value
CPI	colsample_bytree	0.7124
	learning_rate	0.1728
	max_depth	3
	min_child_weight	1
	n_estimators	444
	subsample	0.9948
RGDP	colsample_bytree	0.9637
	learning_rate	0.0876
	max_depth	6
	min_child_weight	2
	n_estimators	489
	subsample	0.6832
IBR	colsample_bytree	0.6053
	learning_rate	0.2927
	max_depth	8
	min_child_weight	2
	n_estimators	364
	subsample	0.6064

the significance of external trade and financial liquidity. In the RGDP forecasts, imports again showed the highest importance, which might suggest that the cost and volume of imported goods play a crucial role in driving domestic inflation. Credit-Total was identified as another key feature across models, demonstrating the importance of credit availability for economic growth. M1 was identified as a critical factor, especially in the XGBoost model. The important role of exports and AveComp (average compensation) indicates the relevance of trade flows and labor market conditions in shaping interest rates.

Table 5: Ensemble model parameters for different target variables (Random Forest and XGBoost).

Target	Model	Parameter	Value
CPI	RF	n_estimators	266
		max_depth	11
		min_samples_split	2
		min_samples_leaf	1
	XGBoost	colsample_bytree	0.9637
		learning_rate	0.0876
		max_depth	6
		min_child_weight	2
		n_estimators	489
		subsample	0.6832
RGDP	RF	n_estimators	266
		max_depth	11
		min_samples_split	2
		min_samples_leaf	1
	XGBoost	colsample_bytree	0.9637
		learning_rate	0.0876
		max_depth	6
		min_child_weight	2
		n_estimators	489
		subsample	0.6832
IBR	RF	n_estimators	266
		max_depth	11
		min_samples_split	2
		min_samples_leaf	1
	XGBoost	colsample_bytree	0.7124
		learning_rate	0.1728
		max_depth	3
		min_child_weight	1
		n_estimators	444
		subsample	0.9948

Table 6: Feature ablation study in prediction task for 1, 5 and 10 features.

	CPI			RGDP			IBR		
	1	5	10	1	5	10	1	5	10
RMSE	0.938	0.170	0.137	0.594	0.149	0.120	0.124	0.114	0.110
MASE	1.012	0.184	0.144	0.975	0.231	0.178	0.084	0.077	0.077

4.1 Ablation attempt

The performance of the ensemble model, which combines the strengths of both RF and XGBoost, is assessed across different target variables—CPI, RGDP (Real GDP), and the IBR. To explore the effect of feature selection on model performance, the analysis is conducted with varying subsets of features: one feature, five features, and ten features, as shown in Table 6. This process of feature ablation aims to identify the optimal set of features that maximizes predictive accuracy, while minimizing overfitting, and to understand the relationship between the number of features used and the model's generalization ability across different target variables.

Table 6 shows that the ensemble model's performance improves significantly as more features are added, and the ensemble also demonstrates strong results when using the full feature set for each target variable. For the CPI, the ensemble model with all features achieves an RMSE of 0.137 and a MASE of 0.144 (Table 2). This performs better compare to cases with fewer features, where RMSE is 0.938 (one feature) and 0.170 (five features). Table 6 shows MASE dropping from 1.012 to 0.184 as more features are added. This shows that the full feature set allows the model to capture more complex patterns and achieve near-optimal predictive accuracy. For RGDP, the model with all features reaches an RMSE of 0.127 and MASE of 0.186 (Table 2). While these values are slightly better than the performance with fewer features (RMSE of 0.149 with five features and 0.594 with one feature, MASE of 0.231 and 0.975, respectively). For the IBR, the ensemble model achieves an RMSE of 0.116 and a MASE of 0.087 when using all features. This improvement is modest compared to the results with five or one feature (RMSE of 0.114 and 0.124,

MASE of 0.077 and 0.084, respectively), suggesting that the model is already performing well with fewer features, and so does not see as significant a boost with the full feature set.

In conclusion, while the ensemble model shows substantial improvement with more features for CPI and RGDP, the performance for IBR plateaus once a sufficient number of features are included. The use of all features allows the model to achieve its best performance, particularly for CPI and RGDP, with diminishing returns for IBR.

Our results reveal the intricate relationships between trade, financial markets, liquidity, and production in macroeconomic forecasting. The ensemble models effectively capture the relation among these complexities by integrating insights of both RF and XGBoost, providing useful guidance for policymakers and analysts in managing the interconnected factors that contribute to macroeconomic stability, supporting informed decision-making in areas such as trade policy, credit control, and monetary regulation [7].

Funding statement

This research was supported by Wenzhou-Kean University, China, through the IRSP project (Grant No. IRSPG202203).

References

[1] F. R. Alharbi and D. Csala. “A seasonal autoregressive integrated moving average with exogenous factors (SARIMAX) forecasting model-based time series approach”. In: *Intentions* 7.4 (2022), p. 94. DOI: 10.3390/inventions7040094 (cit. on p. C173).

[2] A. Alsharef, K. Aggarwal, Sonia, M. Kumar, and A. Mishra. "Review of ML and AutoML solutions to forecast time-series data". In: *Arc. Comput. Meth. Eng.* 29.7 (2022), pp. 5297–5311. DOI: 10.1007/s11831-022-09765-0 (cit. on p. C169).

[3] R. Ashley. "On the usefulness of macroeconomic forecasts as inputs to forecasting models". In: *J. Forecast.* 2.3 (1983), pp. 211–223. DOI: 10.1002/for.3980020304 (cit. on p. C169).

[4] D. Brezak, T. Bacek, D. Majetic, J. Kasac, and B. Novakovic. "A Comparison of Feed-Forward and Recurrent Neural Networks in Time Series Forecasting". In: *IEEE Conf. Comput. Intel. Fin. Eng. Econ.* (2012), pp. 1–6. DOI: 10.1109/CIFEr.2012.6327793 (cit. on p. C177).

[5] T. Chu and F. Qureshi. "Comparing out-of-sample performance of machine learning methods to forecast U.S. GDP growth". In: *Comput. Econ.* 62 (2022), pp. 1567–1609. DOI: 10.1007/s10614-022-10312-z (cit. on p. C169).

[6] Wind Info Co. *Wind Financial*. 2024. URL: https://www.wind.com.cn/ (cit. on p. C171).

[7] A. D'Agostino, L. Gambetti, and D. Giannone. "Macroeconomic forecasting and structural change". In: *J. Appl. Econ.* 28 (2011), pp. 82–101. DOI: 10.1002/jae.1257 (cit. on pp. C168, C188).

[8] S. Demir and E. K. Sahin. "An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost". In: *Neur. Comput. Appl.* 35.4 (2023), pp. 3173–3190. DOI: 10.1007/s00521-022-07856-4 (cit. on p. C175).

[9] F. X. Diebold. "The past, present, and future of macroeconomic forecasting". In: *J. Econ. Perspect.* 12.2 (1998), pp. 175–192. DOI: 10.1257/jep.12.2.175 (cit. on p. C169).

[10] R. Fildes. "Quantitative forecasting—the state of the art: Econometric models". In: *J. Op. Res. Soc.* 36.7 (1985), pp. 549–580. DOI: 10.2307/2582473 (cit. on p. C169).

[11] R. Fildes and H. Stekler. “The state of macroeconomic forecasting”. In: *J. Macroecon.* 24.4 (2002), pp. 435–468. DOI: 10.1016/s0164-0704(02)00055-1 (cit. on p. C169).

[12] H. Ghoddusi, G. G. Creamer, and N. Rafizadeh. “Machine learning in energy economics and finance: A review”. In: *Energy Econ.* 81 (2019), pp. 709–727. DOI: 10.1016/j.eneco.2019.05.006 (cit. on p. C169).

[13] D. Giannone, M. Lenza, and G. E. Primiceri. “Economic predictions with big data: The illusion of sparsity”. In: *Econometrica* 89.5 (2021), pp. 2409–2437. DOI: 10.3982/ECTA17842 (cit. on p. C169).

[14] U. Heilemann and H. Stekler. “Introduction to “The future of macroeconomic forecasting””. In: *Int. J. Forecast.* 23.2 (2007), pp. 159–165. DOI: 10.1016/j.ijforecast.2007.01.001 (cit. on p. C169).

[15] D. F. Hendry. “The econometrics of macroeconomic forecasting”. In: *Econ. J.* 107.444 (1997), pp. 1330–1357. DOI: 10.1111/j.1468-0297.1997.tb00051.x (cit. on p. C169).

[16] S. Hochreiter and J. Schmidhuber. “Long short-term memory”. In: *Neur. Comput.* 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735 (cit. on p. C176).

[17] P. R. Israilevich, G. J. D. Hewings, M. Sonis, and G. R. Schindler. “Forecasting structural change with a regional econometric input-output model”. In: *J. Region. Sci.* 37.4 (1997), pp. 565–590. DOI: 10.1111/0022-4146.00070 (cit. on p. C169).

[18] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang. “A CNN-LSTM-based model to forecast stock prices”. In: *Neur. Comput.* 1, 6622927 (2020), pp. 1–10. DOI: 10.1155/2020/6622927 (cit. on p. C177).

[19] S. Makridakis, E. Spiliotis, and V. Assimakopoulos. “Statistical and machine learning forecasting methods: Concerns and ways forward”. In: *PLoS One* 32.3, e0194889 (2018). DOI: 10.1371/journal.pone.0194889 (cit. on p. C169).

[20] M. C. Medeiros, G. F. Vasconcelos, Á. Veiga, and E. Zilberman. "Forecasting inflation in a data-rich environment: The benefits of machine learning methods". In: *J. Bus. Econ. Stat.* 39.1 (2021), pp. 98–119. DOI: 10.1080/07350015.2019.1637745 (cit. on p. C169).

[21] I. Moosa and K. Burns. "The random walk as a forecasting benchmark: Drift or no drift?" In: *Appl. Econ.* 48.43 (2016), pp. 4131–4142. DOI: 10.1080/00036846.2016.1153788 (cit. on p. C173).

[22] N. Nonejad. "Point forecasts of the price of crude oil: An attempt to 'beat' the end-of-month random-walk benchmark". In: *Empir. Econ.* 67 (2024), pp. 1497–1539. DOI: 10.1007/s00181-024-02599-8 (cit. on p. C173).

[23] A. Panagiotelis, G. Athanasopoulos, R. J. Hyndman, B. Jiang, and F. Vahid. "Macroeconomic forecasting for Australia using a large number of predictors". In: *Int. J. Forecast.* 35.2 (2019), pp. 616–633. DOI: 10.1016/j.ijforecast.2018.12.002 (cit. on pp. C170, C171).

[24] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu. "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019". In: *Appl. Soft Comput.* 90, 106181 (2020). DOI: 10.1016/j.asoc.2020.106181 (cit. on p. C169).

[25] Australian Bureau of Statistics. *Australian economic indicators. accessed 20th November 2024*. 2024. URL: https://www.abs.gov.au/statistics (cit. on p. C171).

[26] Z. I. Taskin, K. Yildirak, and C. H. Aladag. "An enhanced random forest approach using CoClust clustering: MIMIC-III and SMS spam collection application". In: *J. Big Data* 10.1, 38 (2023). DOI: 10.1186/s40537-023-00720-9 (cit. on p. C174).

[27] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso. "Deep learning for time series forecasting: A survey". In: *Big Data* 9.1 (2020), pp. 3–21. DOI: 10.1089/big.2020.0159 (cit. on p. C169).

[28] G. R. West. "Comparison of input-output, input-output + econometric and computable general equilibrium impact models at the regional level". In: *Econ. Sys. Res.* 7.2 (1995), pp. 209–227. DOI: 10.1080/09535319500000021 (cit. on p. C169).

[29] J. Yoon. "Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach". In: *Comput. Econ.* 57.1 (2021), pp. 247–265. DOI: 10.1007/s10614-020-10054-w (cit. on pp. C170, C184).

[30] Y. Zheng, Z. Xu, and A. Xiao. "Deep Learning in economics: A systematic and critical review". In: *A. I. Rev.* 56.9 (2023), pp. 9497–9539. DOI: 10.1007/s10462-022-10272-8 (cit. on p. C169).

Author addresses

1. **Boyuan Li**, College of Science, Mathematics and Technology, Wenzhou-Kean University, Wenzhou, China
2. **Zihui Deng**, College of Science, Mathematics and Technology, Wenzhou-Kean University, Wenzhou, China
3. **Gaurav Gupta**, College of Science, Mathematics and Technology, Wenzhou-Kean University, Wenzhou, China
   mailto:ggupta@kean.edu
4. **Yingxi Chen**, College of Business and Public Management, Wenzhou-Kean University, Wenzhou,China