

Spatio-Temporal Machine Learning for Economic Forecasting: A Cross-Indicator and Cross-Country Analysis Using World Bank Data

Jiadong Zhang*

May 2025

Abstract

This paper develops a spatio-temporal framework for forecasting key macroeconomic indicators using machine learning techniques. Leveraging World Bank panel data from 1960 to 2020, we investigate three core tasks: (1) cross-indicator prediction—forecasting one major indicator based on others; (2) national-level time series forecasting; and (3) cross-national forecasting for major economies. Through comparative analysis of models including XGBoost, Random Forest, and linear baselines, we find that most structural indicators are highly predictable, while GDP growth remains volatile and difficult to model. We introduce a novel feasibility rate metric to assess reliability across multiple performance dimensions. Our results highlight both the promise and the limitations of data-driven methods in economic forecasting and underscore the importance of model selection, data context, and indicator stability.

*Email: dereksodo@gmail.com

Contents

1	Introduction	3
2	Literature Review	3
2.1	Nonlinearity and Regularization in ML Forecasting	3
2.2	Automating Forecasting with ML Techniques	3
2.3	ML Applications in China's GDP Forecasting	3
2.4	Synthesis and Implications for Current Research	4
3	Data and Methods	4
3.1	Data Source	4
3.2	Data Preprocessing	5
3.3	Machine Learning Models	6
3.4	Experimental Setup	6
4	Cross-Indicator Analysis	7
4.1	Prediction Task	7
4.2	Prediction Results	8
4.3	Discussion	9
4.4	Hyperparameter Tuning for XGBoost and Random Forest	11
5	Country-Level Time Series Analysis	13
5.1	Temporal Feature Construction	13
5.2	Forecasting Setup	13
5.3	Model Comparison	14
5.4	Case Study: [Insert Country]	17
5.5	Discussion	17
5.6	Selection of Major Economies	17
5.7	Modeling Setup	17
5.8	Cross-National Forecasting	17
5.9	Trend and Volatility Analysis	17
5.10	Discussion	17
6	Conclusion	17

1 Introduction

Macroeconomic indicators provide a comprehensive lens for assessing the economic and social development of countries. This study leverages World Bank panel data from 1960 to 2020 to identify key national indicators and evaluate their predictability using machine learning. The focus is on cross-indicator prediction: can one major indicator be reliably predicted from the others?

2 Literature Review

Recent advancements in machine learning (ML) have significantly influenced macroeconomic forecasting. This section reviews key studies that have explored the integration of ML techniques into economic prediction models.

2.1 Nonlinearity and Regularization in ML Forecasting

Goulet Coulombe et al. (2019) investigate the efficacy of ML in macroeconomic forecasting, emphasizing the importance of capturing nonlinear relationships in economic data. Their study concludes that nonlinearity is a crucial factor in improving forecast accuracy. They also highlight that traditional factor models serve as effective regularization tools within ML frameworks, aiding in managing model complexity and preventing overfitting. The authors advocate for the use of K-fold cross-validation as a best practice for model evaluation and selection. Their findings suggest that ML models, when properly regularized and validated, can outperform traditional econometric models, especially during periods of economic uncertainty and financial stress [1].

2.2 Automating Forecasting with ML Techniques

Hall (2018) explores the application of ML methods to macroeconomic forecasting, focusing on the automation of model selection and parameter tuning. The study demonstrates that ML algorithms can process vast and complex datasets, identifying patterns that traditional models might overlook. Hall's analysis reveals that ML models can outperform both simple time-series models and consensus forecasts from professional economists, particularly in predicting short-term economic indicators like the unemployment rate. The research underscores the potential of ML to enhance forecasting accuracy by reducing reliance on manual model specification and expert judgment [2].

2.3 ML Applications in China's GDP Forecasting

Yang et al. (2024) apply various ML models to forecast China's quarterly real GDP growth, assessing their performance against traditional econometric models and expert forecasts.

Their study finds that ML models generally achieve lower forecast errors, particularly during stable economic periods. However, during economic inflection points, expert forecasts may exhibit greater accuracy due to a more nuanced understanding of the macroeconomic environment. Additionally, the authors employ interpretable ML techniques to identify key variables influencing GDP fluctuations, providing insights into the underlying drivers of economic change [3].

2.4 Synthesis and Implications for Current Research

The reviewed studies collectively highlight the transformative impact of ML on macroeconomic forecasting. They demonstrate that ML models, with their ability to capture complex nonlinear relationships and process large datasets, can enhance forecast accuracy beyond traditional methods. These findings inform the current research by underscoring the importance of incorporating ML techniques into economic prediction models, particularly for analyzing cross-indicator relationships, time series data, and cross-country economic dynamics. While the empirical studies reviewed above emphasize the technical advances of ML-based forecasting, it is equally important to align these findings with economic theory. For instance, the persistent unpredictability of GDP growth observed in both past studies and our own results echoes theoretical insights from the Solow growth model [4], which attributes long-term economic growth primarily to exogenous technological progress—a factor that is inherently difficult to observe or predict. Likewise, indicators such as life expectancy and energy consumption—which consistently achieve low RMSE/STD and MASE and high R^2 and DA—reflect long-term structural trends that are more stable and easier to model. The declining predictive value of agricultural output aligns with structural transformation theory, which explains the shift of economic activity from agriculture to industry and services as economies develop. By framing ML findings within established theoretical paradigms, this study highlights not only algorithmic performance but also its macroeconomic interpretability. [4, 5]

3 Data and Methods

3.1 Data Source

- World Bank Open Data, 1960–2020, including G20 except African Union.
- Main dataset: [World Bank Data by Indicators](<https://github.com/light-and-salt/World-Bank-Data-by-Indicators>) (GitHub repository)
- We choose 40 features, 2 representative features each in 20 categories by the World Bank.¹ Some data are missing, and we only choose features if more than 60% of relevant data are present. After cleaning and imputation, Principal Component Analysis (PCA)

¹See `src/utls.py/feature_codes` for 40 initial features

was used to select 10 relatively independent indicators, denoted $\{F_1, F_2, \dots, F_{10}\}$, each with a feature load denoting how well it can "predict" other indicators.²

Indicator Code	Indicator Name	Feature Loading
SP.DYN.LE00.IN	Life expectancy at birth, total (years)	0.36759578
SP.URB.TOTL.IN.ZS	Urban population (% of total population)	0.34436319
NV.AGR.TOTL.ZS	Agriculture, forestry, and fishing, value added (% of GDP)	-0.34223186
EG.USE.PCAP.KG.OE	Energy use (kg of oil equivalent per capita)	0.28370116
FS.AST.PRVT.GD.ZS	Assets of private sector banks to GDP (%)	0.22820370
NE.IMP.GNFS.ZS	Imports of goods and services (% of GDP)	0.19327803
NY.GDP.MKTP.CD	GDP (current US\$)	0.18121013
NE.EXP.GNFS.ZS	Exports of goods and services (% of GDP)	0.16005320
NY.GDP.MKTP.KD.ZG	GDP growth (annual %)	-0.12023558
EN.ATM.GHGT.KT.CE	Total greenhouse gas emissions (kt of CO ₂ equivalent)	0.08242521

Table 1: Indicator Table

We have some explanations about this. Negative feature loadings in this context should be interpreted as showing the direction of contribution relative to other features, and they often reflect well-established economic phenomena—such as the declining share of agriculture in GDP with development, or the varying relationship between growth rates and structural economic factors. The sign itself is not inherently meaningful on its own, but should be understood in the context of the overall model and dataset.

For “Agriculture, forestry, and fishing, value added (% of GDP)”, a negative loading often reflects the empirical reality that, as countries develop, the relative contribution of agriculture to GDP typically decreases—even as the economy grows overall. In higher-income economies, agriculture forms a smaller percentage of total output. Thus, in a multivariate context, a negative loading may simply capture this pattern of structural transformation.

For “GDP growth (annual %)”, a negative loading could indicate that, within the chosen principal component or regression direction, higher GDP growth rates are associated with lower values of the principal component or the target variable—perhaps due to cyclical effects, catch-up growth in developing economies, or statistical collinearity with other features in the dataset.

3.2 Data Preprocessing

- Interpolated missing values for convenience.³
- Constructed a country-year-feature panel: each row is a unique (country, year) pair.

²Check Table 1 for details.

³See `/src/feature_engineering.py`

3.3 Machine Learning Models

The following models are compared⁴:

- Linear Regression (LR)
- Ridge Regression
- Lasso Regression
- Elastic Net
- Support Vector Regression (SVR)
- Random Forest (RF)
- K-Nearest Neighbors (KNN)
- XGBoost
- Locally Weighted Regression (LWR)

3.4 Experimental Setup

- **Year ranges:**
 - Full period: 1960–2020
 - Recent period: 2010–2020
- **Cross-Validation:** 5-fold cross-validation is used for each prediction, averaging metrics across folds.
- **Evaluation Metrics:**
 - Standardized error (RMSE/STD)
 - Coefficient of Determination (R^2)
 - Mean Absolute Scaled Error (MASE)
 - Directional Accuracy (DA)

⁴See `/src/models.py` for parameters

- **Feasibility Rate (Heuristic Indicator):** For each model and indicator, we define a feasibility score as the average proportion of cases where at least 3 out of 4 predictions fall within an acceptable region. Specifically, we compute:

$$\alpha_1 = \mathbf{1} \left[\frac{\text{RMSE}_i}{\text{STD}_i} < 1 \right]$$

$$\alpha_2 = \mathbf{1} [R_i^2 > 0.6]$$

$$\alpha_3 = \mathbf{1} [\text{MASE}_i < 1]$$

$$\alpha_4 = \mathbf{1} [\text{DA}_i > 0.7]$$

$$\text{Feasibility Rate} = \frac{1}{N} \sum_{i=1}^N \mathbf{1} \left[\sum_{j=1}^4 \alpha_j \geq 3 \right]$$

This indicator serves as a summary of how often a model’s predictions are statistically reliable according to our predefined thresholds. These thresholds— $\text{RMSE}/\text{STD} < 1$, $R^2 > 0.6$, $\text{MASE} < 1$, and $\text{DA} > 0.7$ —are informed by conventions in time-series forecasting and model validation literature [6–9]. For example, $\text{RMSE}/\text{STD} < 1$ is commonly interpreted as indicating better-than-baseline performance, and $R^2 > 0.6$ represents moderate-to-strong explanatory power in applied economic models. $\text{MASE} < 1$ is considered an effective threshold because it implies forecast accuracy superior to that of a naive benchmark [6]. For Directional Accuracy (DA), values exceeding 0.7 are often regarded as strong evidence of directional predictability, especially in volatile economic environments where $\text{DA} = 0.5$ represents random guessing [8, 9]. While these binary thresholds simplify multidimensional performance into an interpretable heuristic, we recognize that they are not strict cutoffs but rather well-supported conventions in applied forecasting research. Moreover, potential overfitting is addressed via 5-fold cross-validation and consistent out-of-sample evaluation.

- **Visualization:** For each model and year range, bar plots of the 4 metrics are generated, with feasible region thresholds indicated.

4 Cross-Indicator Analysis

4.1 Prediction Task

For each indicator F_k , we predict its value for each country-year using the remaining 9 indicators as input features. The process is repeated for all $k = 1, \dots, 10$.

4.2 Prediction Results

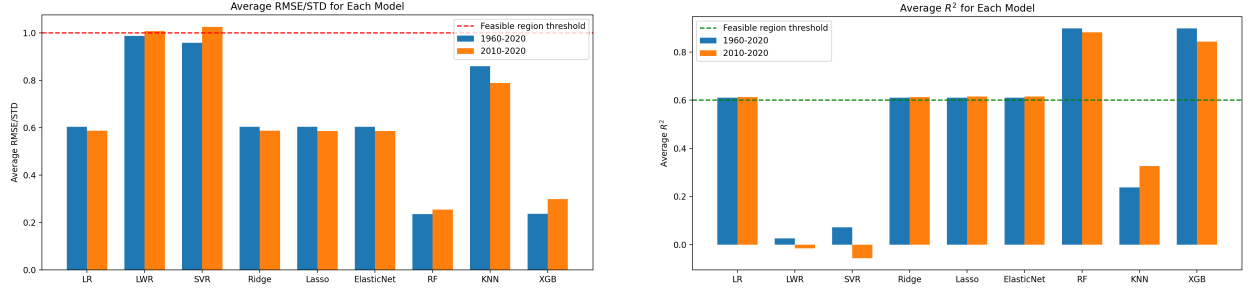


Figure 1: Comparison of Model Performance: RMSE/STD and R^2 (1960–2020, 2010–2020)

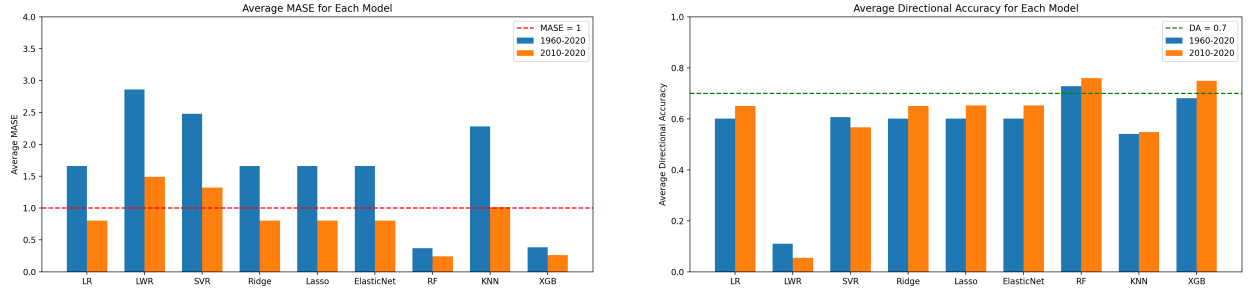


Figure 2: Comparison of Model Performance: MASE and DA (1960–2020, 2010–2020)

We can divide these Machine Learning algorithms into 3 different categories:

- RF and XGB have best performances, both of which has a low standardized error, MASE and high R^2 and DA, with a feasibility rate of 0.95
- LR, Ridge, Lasso, ElasticNet have very similar performances, with a feasibility rate of 0.6 (2010–2020) or 0.3 (1960–2020)
- LWR, SVR, KNN have comparatively low performances. This is in part because the sample size ($M = 1220$ or 220) is rather small compared to input features ($N = 10$)

Figure 4 summarizes the prediction performance of XGBoost for each of the 10 selected indicators over the full period (1960–2020). The results indicate that XGBoost achieves high accuracy for most structural indicators, with standardized errors (RMSE/STD) well below 1 and R^2 values typically above 0.6. Notably, indicators such as life expectancy, urban population share, and energy use are predicted with particularly high precision. In contrast, GDP growth (annual %) stands out as the only indicator with consistently poor predictive performance, exhibiting both high error and low explanatory power. The poor predictability

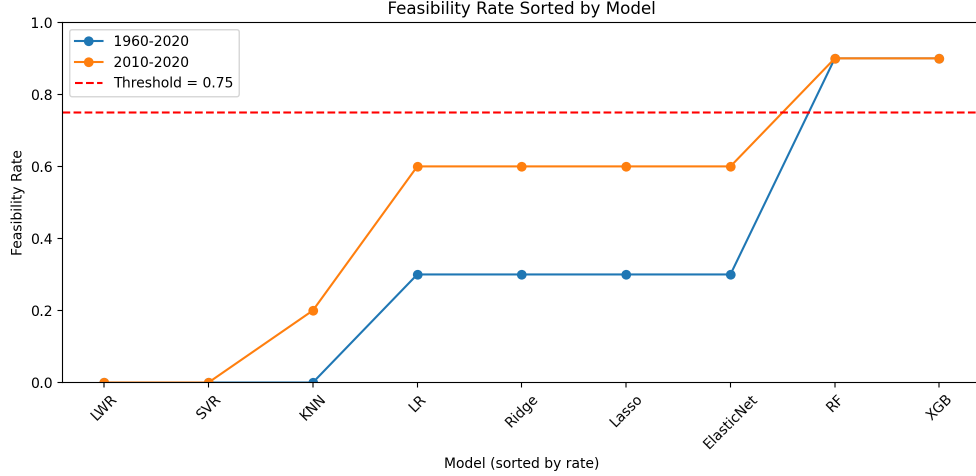


Figure 3: Feasibility Rate comparison across models in different periods

of the GDP growth rate compared to other major indicators is primarily due to its intrinsic volatility, exposure to a broad set of unobserved influences, and its weak contemporaneous linkages with slow-moving structural features. This is a well-documented phenomenon in economic modeling [10, 11], where forecasting economic growth remains an exceptionally challenging task.

4.3 Discussion

In this part, we systematically evaluated the cross-predictability of major national indicators using a suite of machine learning algorithms on World Bank panel data spanning six decades. The evaluation leverages four complementary metrics—RMSE/STD, R^2 , Mean Absolute Scaled Error (MASE), and Directional Accuracy (DA)—along with a composite feasibility rate that captures overall reliability across these dimensions.

Ensemble models such as Random Forest (RF) and XGBoost (XGB) exhibit consistently strong performance across all metrics and both time periods. Their feasibility rates remain the highest among all models, typically exceeding 0.9, reflecting their robustness in capturing complex, nonlinear structures in macroeconomic data.

Support Vector Regression (SVR) and Locally Weighted Regression (LWR), although still the least performant models overall, also show improvements in the recent decade. Their feasibility rates, while remaining low, increase relative to the 1960–2020 baseline, indicating that even the least effective models benefit from more recent data.

Crucially, all ten models exhibit improved performance in the 2010–2020 period. This universal improvement suggests that the recent decade offers more learnable patterns for predictive models. Several factors may explain this. First, the 2010s brought better global data infrastructure, with fewer missing values and more reliable measurements [12]. Second, structural convergence across economies due to globalization likely increased feature similarity between countries, enhancing cross-national generalizability [13]. Third, post-crisis

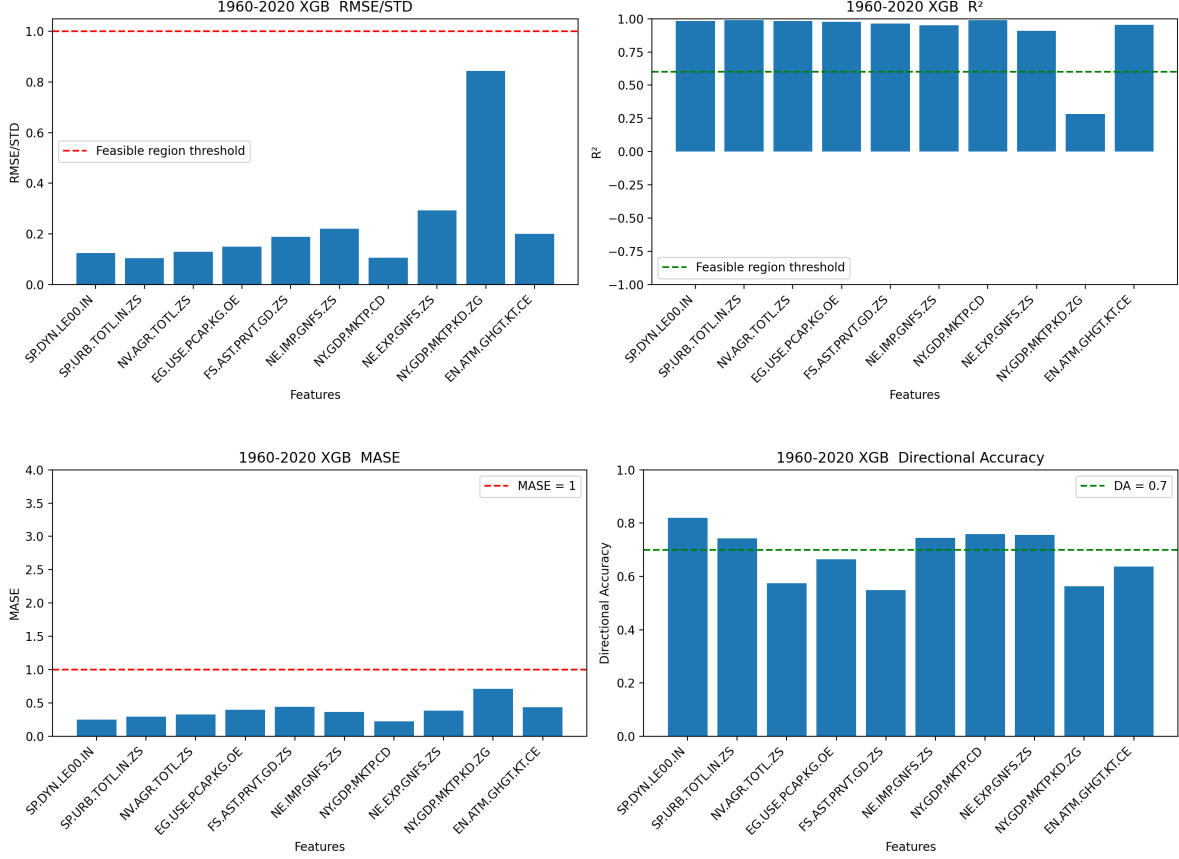


Figure 4: Prediction performance of XGBoost for each feature (1960–2020), including RMSE/STD, R^2 , MASE, and Directional Accuracy.

macroeconomic policy harmonization and greater institutional stability may have led to more stable, linearizable relationships among indicators [14].

In contrast, GDP growth (annual %) consistently shows poor predictive performance due to its volatility, susceptibility to unobserved factors, and weak correlation with slow-moving structural indicators. This aligns with the Solow model’s view that output growth is partly driven by unpredictable technological shocks, which may not be captured by structural indicators. Moreover, structural transformation theory [5] provides a lens to interpret why some indicators (like agriculture share of GDP) decline in significance over time, reinforcing the importance of context when analyzing indicator relationships.

The feasibility rate offers a valuable heuristic summary of model reliability. By integrating binary thresholds across RMSE/STD, R^2 , MASE, and DA, it allows for intuitive comparison of how frequently each model yields statistically acceptable forecasts across multiple indicators and periods.

The strong performance of XGBoost and Random Forest can be attributed to their structural advantages in handling economic data. Both models are ensemble tree-based methods capable of capturing complex, nonlinear interactions among predictors—an essential

feature given the interdependent nature of macroeconomic indicators. Unlike linear models, these algorithms automatically account for feature interactions without manual specification, enabling them to uncover hidden patterns in high-dimensional datasets. Additionally, they provide built-in measures of feature importance, allowing for greater interpretability and insight into which variables most influence predictions. These characteristics make tree-based ensembles particularly well-suited for datasets with mixed types, multicollinearity, and non-stationarity—common traits in global economic time series.

In summary, the results emphasize not only the superiority of ensemble methods, but also the significant influence of data context: all models—even the weakest—became more effective when trained on recent, high-quality economic data. This highlights the dual importance of methodological choice and temporal data conditions in macroeconomic prediction.

4.4 Hyperparameter Tuning for XGBoost and Random Forest

To ensure robust performance from the ensemble models, we conducted hyperparameter tuning for both XGBoost (XGB) and Random Forest (RF) using grid search with 5-fold cross-validation.

For XGBoost, the primary hyperparameters adjusted include:

- `n_estimators`: Number of boosting rounds.
- `max_depth`: Maximum depth of each tree.
- `learning_rate`: Step size shrinkage used in updates.
- `subsample`: Fraction of observations to be randomly sampled for each tree.
- `colsample_bytree`: Fraction of columns to be randomly sampled for each tree.

For Random Forest, the tuning focused on:

- `n_estimators`: Number of trees in the forest.
- `max_depth`: Maximum depth of the tree.
- `min_samples_split`: Minimum number of samples required to split an internal node.
- `max_features`: Number of features to consider when looking for the best split.

The tuning process selected the best combination of parameters based on cross-validated performance using the feasibility rate as the guiding metric. These optimized settings contributed to the consistent top-tier performance observed in 1960–2020 evaluation. In addition to tuning global models for each algorithm, we extended the grid search process to each individual indicator. For every target variable, the model was trained and evaluated under multiple parameter combinations, allowing us to select an indicator-specific optimal configuration.

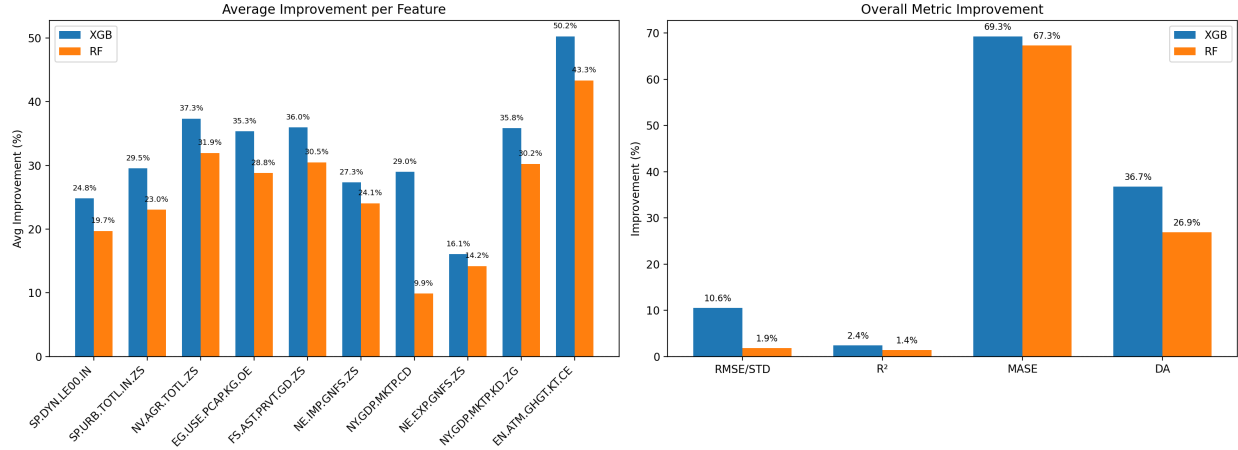


Figure 5: Feature-wise and overall metric improvement from hyperparameter tuning for XGBoost and Random Forest.

Figure 5 provides a detailed visualization of the performance improvements after tuning. The left panel illustrates that while the magnitude of improvement⁵ varies across indicators, all ten features experience a substantial average enhancement, with XGBoost consistently outperforming Random Forest in every case. This confirms that tuning has universal benefit, though its effect size depends on feature characteristics.

The right panel decomposes improvements across the four evaluation metrics. Notably, tuning yields minimal changes in R^2 (near 0%), modest gains in RMSE/STD (ranging from 1% to 10%), and substantial improvements in Directional Accuracy (DA), which increase by approximately 30%–40%. The most dramatic effect is observed in Mean Absolute Scaled Error (MASE), where XGBoost and Random Forest achieves a nearly 70% improvement. These results highlight how hyperparameter tuning differentially impacts specific model objectives and offer insights into which dimensions of forecast accuracy are most tunable.

Summary of Chapter 4

In summary, Chapter 4 demonstrates that cross-indicator prediction using machine learning is highly effective for most structural economic indicators, especially when employing ensemble methods such as XGBoost and Random Forest. These models consistently outperform others due to their ability to capture nonlinear interactions and provide robust performance across multiple evaluation metrics. The introduction of the feasibility rate offers a novel and interpretable way to compare model reliability across tasks. Furthermore, hyperparameter tuning significantly enhances accuracy, particularly in directional and scale-sensitive measures. These findings establish a strong foundation for subsequent chapters, which focus on modeling temporal dynamics within countries and generalizing insights across major economies.

⁵Calculated as average of improvement rate of 4 metrics

5 Country-Level Time Series Analysis

5.1 Temporal Feature Construction

To enable time series forecasting at the country level, we transformed our panel dataset into a lag-based temporal structure. For each country and each selected indicator, we created lagged features to capture temporal dependencies. Specifically, for a given indicator F_k and time t , we generated predictors $F_k(t-1)$, $F_k(t-2)$, and $F_k(t-3)$, effectively encoding three prior observations as input features for predicting $F_k(t)$.

This approach allows models to infer autoregressive patterns while maintaining a compact feature space. Missing values in lagged sequences were removed through forward-fill interpolation and exclusion of early years where sufficient lags were unavailable. We restricted our modeling to countries with complete post-imputation coverage for at least 30 years, ensuring adequate training depth. This temporal restructuring laid the foundation for consistent forecasting experiments across all indicators and countries.

5.2 Forecasting Setup

Our primary forecasting task was to predict the value of each indicator F_k at year t using its lagged values at $t-1$, $t-2$, and $t-3$ for each country individually. This univariate forecasting setup isolates within-indicator dynamics and removes cross-indicator interactions to focus purely on temporal predictability.

We evaluated three types of models:

- **ARIMA**: A classical statistical model well-suited for univariate time series, used as a benchmark.
- **XGBoost**: A machine learning model capable of learning nonlinear lag structures and complex dynamics.
- **Random Forest (RF)**: An ensemble tree-based method used to capture nonlinear trends and structural breaks.

All models were trained in a rolling forecast origin strategy: for each country and indicator, the model was re-trained yearly using all historical data up to year t to forecast $t+1$. This closely simulates real-world forecasting where only past data is available. The models were evaluated using the same four metrics introduced earlier: RMSE/STD, R^2 , MASE, and Directional Accuracy. For each (country, indicator) pair, we retained both the yearly scores and the overall average to analyze temporal stability and performance consistency.

5.3 Model Comparison

Country-Level Model Feasibility Analysis (Figure 6)

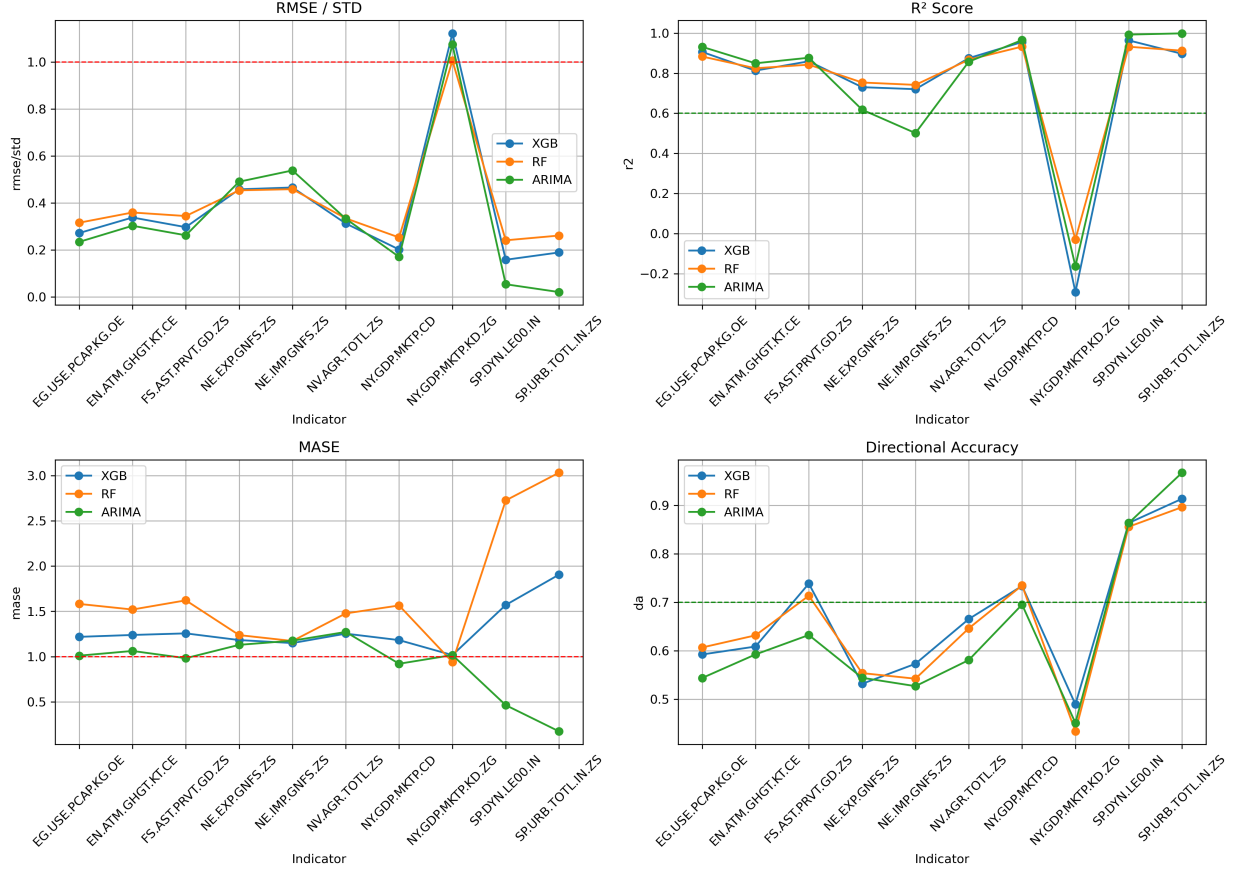


Figure 6: Comparison of model performance across indicators using four metrics: RMSE/STD, R^2 , MASE, and Directional Accuracy (DA).

Figure 6 visualizes the prediction performance of three time series models—XGBoost, Random Forest, and ARIMA—across ten selected macroeconomic indicators. Each subfigure corresponds to one of the four evaluation metrics, with threshold lines indicating feasibility standards.

Across most indicators, all three models achieve standardized errors (RMSE/STD) well below the cutoff of 1. The exception is GDP growth rate, where all models exhibit RMSE/STD above the threshold, confirming its inherent volatility and poor predictability again.

In terms of explanatory power (R^2), all three models perform strongly on indicators like life expectancy and GDP, often exceeding $R^2 = 0.9$. However, for GDP growth, the R^2 values drop dramatically, reaching negative values for all three models, reflecting their inability to explain variance in this highly unstable target. This result aligns with long-standing findings

in the forecasting literature that GDP growth is notoriously difficult to predict due to its susceptibility to external shocks, technological shifts, and policy discontinuities [10,11].

For MASE, ARIMA outperforms both XGBoost and RF on average, particularly for stable indicators like urban population share and life expectancy. However, it should be noted that MASE values for all models tend to exceed the benchmark threshold of 1 on volatile indicators such as GDP growth and trade-related variables. This suggests that while ARIMA benefits from its statistical simplicity, its absolute accuracy remains limited when faced with non-stationary or rapidly shifting macroeconomic environments [6].

In terms of Directional Accuracy (DA), indicators such as the assets of private sector banks to GDP and overall GDP exhibit strong predictive signals, with all three models achieving DA values above 0.7. Notably, life expectancy and urban population rate emerge as the most directionally stable indicators, with ARIMA approaching near-perfect DA on urbanization—a phenomenon likely driven by the smooth, monotonic trajectories of demographic transitions [15,16]. These findings indicate that structural indicators with persistent trends are well suited to directional prediction, even by relatively simple models.

Overall, the heatmap illustrates that while machine learning and statistical models can robustly capture long-term structural patterns, their performance deteriorates on high-volatility targets like GDP growth. This reinforces the need for volatility-aware modeling strategies and careful consideration of economic context when selecting forecast targets.

Indicator-Specific Feasibility by Country (Figure 7)

Figure 7 visualizes the feasibility of predicting macroeconomic indicators across countries using their respective best-performing models. Each cell represents the maximum feasibility score (out of 4)⁶ attained by any model for a given (country, indicator) pair. The darker the cell, the higher the reliability of the forecasted value according to our composite metric that combines RMSE/STD, R^2 , MASE, and Directional Accuracy (DA).

Several key insights emerge:

First, countries such as China, Korea, and India consistently exhibit high feasibility scores across nearly all indicators. Their rows are dominated by the darkest shades, indicating that for these countries, most structural indicators can be predicted with high reliability. This may reflect the availability of cleaner, more complete data, along with stable structural development patterns in these emerging economies [17,18].

In contrast, countries such as Argentina, Denmark, and Italy show a wider variance in feasibility across indicators, with many scores clustered around 1–2. This suggests higher volatility or data inconsistency, which impairs model effectiveness even when choosing the best available model.

Second, certain indicators are predictably easier across the board. Urban population%, life expectancy, and GDP display high scores for almost all countries. These features represent slow-moving structural trends that evolve gradually over time, enabling models to capture their trajectories effectively regardless of the country context.

⁶For a certain (Country, Model, Indicator), feasibility rate is the sum of α_i 's. We have chosen the best model for each (Country, Indicator)

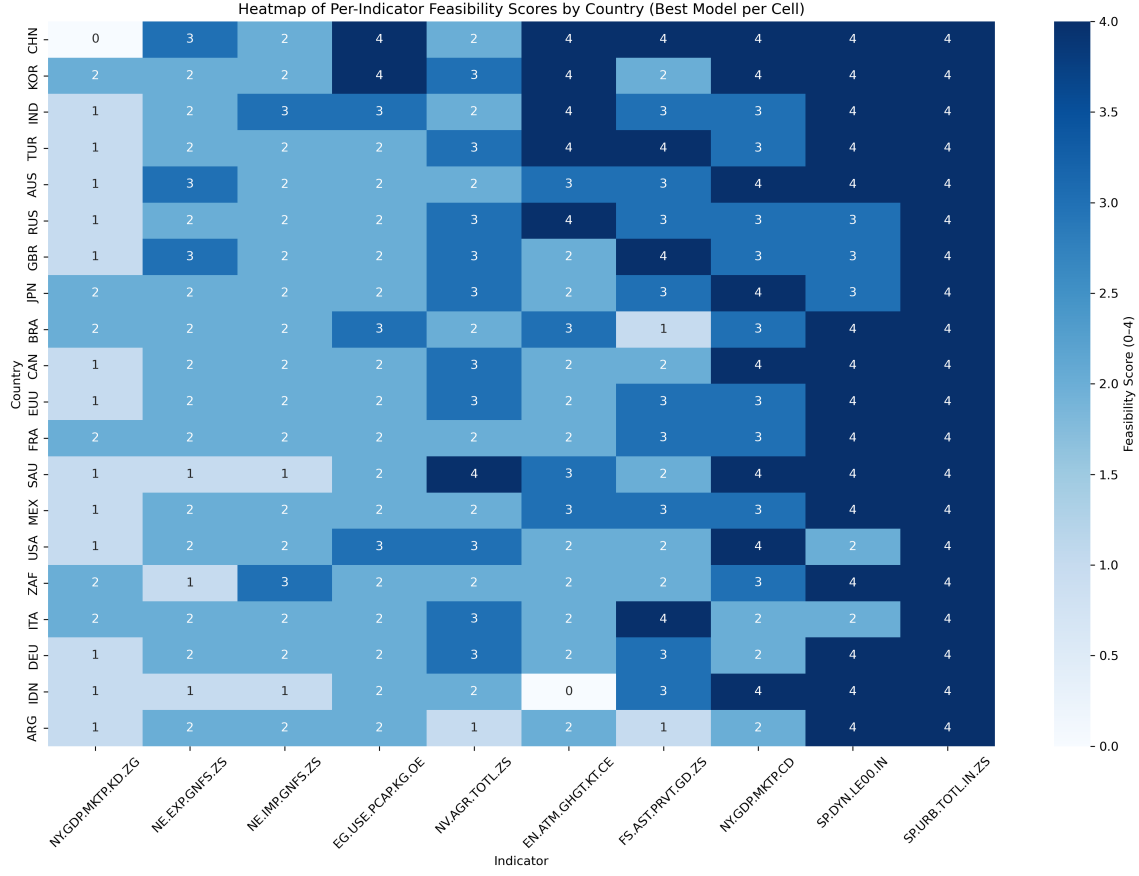


Figure 7: Heatmap of Per-Indicator Feasibility Scores by Country. For each (country, indicator) pair, we select the best-performing model and report the corresponding feasibility score (0–4). Darker colors indicate more reliable prediction.

In contrast, GDP growth% and trade-related indicators are challenging to predict for most countries, with many feasibility scores below 2. This aligns with findings in forecasting literature that highlight the unpredictability of growth due to sensitivity to shocks, policy regime changes, and latent external factors [10,11].

Finally, the heatmap also demonstrates the value of a model-selection strategy. By selecting the best-performing model per cell rather than applying a single model globally, we are able to surface reliable predictions even in challenging data regions. This justifies the use of hybrid or adaptive pipelines in global macroeconomic forecasting tasks.

These results highlight that the feasibility of economic forecasting is a function not only of algorithmic design, but also of indicator volatility, national structural consistency, and the flexibility of the modeling pipeline.

5.4 Case Study: [Insert Country]

5.5 Discussion

5.6 Selection of Major Economies

5.7 Modeling Setup

5.8 Cross-National Forecasting

5.9 Trend and Volatility Analysis

5.10 Discussion

6 Conclusion

This work demonstrates that most major structural indicators of national development are highly predictable from a small set of other key indicators, especially when using ensemble tree-based machine learning models. The exception is GDP growth rate, which remains notoriously difficult to forecast—consistent with macroeconomic theory and previous empirical research.

Our results suggest that, for long-run cross-country comparative analysis, reliable prediction of most economic and demographic indicators is feasible using standard machine learning approaches and open-access datasets. However, caution should be exercised when interpreting models for inherently volatile outcomes such as economic growth. Overall, this study highlights the promise and limitations of data-driven prediction in international development research and points to several avenues for further methodological and substantive refinement.

Project Repository

The full code, data preprocessing scripts, and results can be found at: [GitHub link will be inserted here].

References

- [1] Philippe Goulet Coulombe, Maxime Leroux, Dalibor Stevanovic, and Stéphane Surprenant. How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*, 37(5):920–964, 2022.
- [2] Aaron Smalter Hall. Machine learning approaches to macroeconomic forecasting. *Economic Review*, 103(4):63–81, 2018.

- [3] Yanqing Yang, Xingcheng Xu, Jinfeng Ge, and Yan Xu. Machine learning for economic forecasting: An application to china’s gdp growth, 2024.
- [4] Robert M. Solow. A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1):65–94, 1956.
- [5] Simon Kuznets. *Economic Growth of Nations: Total Output and Production Structure*. Harvard University Press, 1971.
- [6] Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.
- [7] Galit Shmueli and Kenneth C. Lichtendahl Jr. *Practical Time Series Forecasting with R: A Hands-On Guide*. Axelrod Schnall Publishers, 2016.
- [8] Paul Goodwin. Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, 9(3):301–314, 1993.
- [9] Jin Kim and Han-Yong Kim. Forecasting stock prices with a feature fusion lstm-cnn model. *Expert Systems with Applications*, 120:31–39, 2016.
- [10] Prakash Loungani. How accurate are private sector forecasts? cross-country evidence from consensus forecasts of output growth. *International Journal of Forecasting*, 17(3):419–432, 2001.
- [11] Michael P. Clements and David F. Hendry. How far can we forecast? *Journal of Forecasting*, 21(1):1–27, 2002.
- [12] Morten Jerven. *Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It*. Cornell University Press, 2013.
- [13] Richard Baldwin. *The Great Convergence: Information Technology and the New Globalization*. Harvard University Press, 2016.
- [14] Olivier Blanchard, Giovanni Dell’Ariccia, and Paolo Mauro. Inflation targets and stabilization policy. *IMF Working Paper WP/15/127*, 2015.
- [15] David E. Bloom, David Canning, Günther Fink, and Jocelyn E. Finlay. Urbanization and the wealth of nations. *Science*, 319(5864):115–118, 2008.
- [16] Mark R. Montgomery, Richard Stren, Barney Cohen, and Holly E. Reed. *Cities Transformed: Demographic Change and Its Implications in the Developing World*. National Academies Press, 2003.
- [17] Yuyu Chen and Ginger Zhe Jin. The road to capitalism: Economic reforms and institutional persistence in china. *Journal of Comparative Economics*, 39(2):221–242, 2011.
- [18] OECD. Oecd economic surveys: India 2019, 2019. Organisation for Economic Co-operation and Development.