ECOBIZTECH: Journal of Economics, Business, and Technology
ISSN 3124-4742 (online)
Vol. 1 Iss. 1 Pg. 21-35
https://ejournal.scmdigi.id/index.php/ecobiztech

Advancing macroeconomic forecasting by integrating big data and machine learning

Anhari Utomo[1], Darmawan Munthe[2*], Daniel Handoko[3]
[1]Universitas Pembangunan Panca Budi, Medan, Sumatera Utara, Indonesia
[2]Universitas Medan Area, Deli Serdang, Sumatera Utara, Indonesia
[3]Universitas Prima Indonesia, Medan, Sumatera Utara, Indonesia
*Correspondence author: darmamunthe11@gmail.com
DOI: https://doi.org/10.65881/ecobiztech.v1i1.28

ARTICLE INFO

History:
Received: 02-04-2026
Revised: 02-09-2026
Accepted: 02-10-2026
Published: 02-12-2026

Keywords:
macroeconomic;
forecasting;
big data;
machine learning;
integration.

ABSTRACT

Purpose: to assess the role of integrating big data and machine learning (ML) in improving the accuracy of macroeconomic forecasting and to develop an adaptive multi-indicator forecasting framework.
Method: a quantitative data-driven forecasting approach is employed by integrating historical macroeconomic data, real-time data, and unstructured data. Forecasts are generated using random forests, support vector regression (SVR), neural networks, and macroeconomic random forests (MRF), and model performance is evaluated using MAE, RMSE, and MAPE via rolling-window cross-validation.
Findings: ML-based models consistently outperform traditional econometric approaches. MRF achieves the highest accuracy in forecasting GDP and unemployment, while random forests and SVRs perform better at capturing inflation dynamics and short-term fluctuations. The inclusion of real-time, unstructured data enhances the model's responsiveness to economic volatility and shocks.
Implications: these findings highlight the potential of ML-based macroeconomic forecasting systems as effective decision-support tools for evidence-based policymaking.
Originality: lies in the development of a multi-indicator ML-based forecasting framework that integrates heterogeneous big data in a single integrated system.

Introduction

Macroeconomic forecasting is a crucial aspect of policymaking, investment planning, and business strategy formulation (Suprunenko et al., 2023). Accurate predictions of economic indicators such as GDP growth, inflation, and unemployment rates enable governments and organizations to anticipate economic trends and respond proactively to changes (Muhammad Atif Khan et al., 2024). Traditionally, macroeconomic forecasting has relied on econometric models and statistical techniques, which often suffer from limitations such as linearity assumptions, data constraints, and time lags in information availability. With the advancement of technology, the emergence of big data and machine learning (ML) has revolutionized data analysis. Big data provides unprecedented volume, velocity, and variety of information, while ML offers advanced algorithms capable of capturing complex and non-linear patterns that traditional models often overlook (Khudhur et al., 2025). The application of these technologies to macroeconomic forecasting has the potential to enhance predictive accuracy and deliver more timely, precise insights into economic dynamics.

Nevertheless, macroeconomic forecasting continues to face significant challenges. Economic shocks, rapid policy changes, and global interconnectedness generate volatility that conventional models struggle to capture (Mulenga, 2024). At the same time, many organizations have begun collecting and analyzing large-scale data, creating a need for more sophisticated analytical frameworks to exploit this information effectively (Ikegwu et al., 2022). Currently, traditional forecasting models have not fully utilized the potential of big data, resulting in limited predictive accuracy. Recent studies have applied machine learning methods, such as random forests, support vector regression (SVR), and neural networks, to macroeconomic forecasting, yielding promising results. For instance, a study by Khan et al. (2022); Ather et al. (2025) demonstrates that nonlinear models such as SVR, neural networks, and random forest significantly improve the accuracy of macroeconomic variable forecasts compared to traditional models when large datasets and diverse economic indicators are employed. Furthermore, Coulombe (2024); Zhang & Bian (2025) developed the macroeconomic random forest (MRF) approach, which adapts the random forest algorithm to capture nonlinear dynamics and structural breaks in macroeconomic processes, thereby enhancing forecasting flexibility and accuracy. In addition, several studies on real GDP nowcasting indicate that combining ML algorithms such as random forest and SVR produces more accurate predictions than standard econometric models (Hopp, 2024; Richardson et al., 2021; Soybilgen & Yazgan, 2021).

Nevertheless, most existing studies continue to focus on short-term forecasting or a single economic indicator. The integration of multiple big data sources to comprehensively model macroeconomic trends remains highly limited, leaving substantial room for the development of frameworks that combine heterogeneous data with Machine Learning algorithms to enhance forecasting accuracy and timeliness. Based on this discussion, a clear research gap exists concerning the systematic integration of big data and ML algorithms for multi-indicator macroeconomic forecasting. Therefore, this study seeks to address this gap by developing a framework that integrates diverse data sources with adaptive ML algorithms. The novelty of this research lies in its integrative approach, which combines data heterogeneity with non-linear modeling capabilities, and in its provision of a scalable forecasting solution that is responsive to rapid economic changes.

This study aims to explore the potential of big data sources to enhance macroeconomic forecasting, evaluate the performance of various machine learning (ML) algorithms in predicting macroeconomic indicators, and develop an ML-based forecasting framework that integrates multiple data types to improve predictive accuracy. This study is significant as it contributes to evidence-based decision-making, risk mitigation, and more effective economic planning under conditions of uncertainty. The expected contributions of this research include: a methodological framework for integrating big data and ML in macroeconomic forecasting; empirical evidence demonstrating improvements in predictive performance relative to traditional methods; insights into the most effective data sources and ML techniques for specific macroeconomic indicators; and a foundation for future research and practical applications in economic policy and analysis.

Literature review

Macroeconomic forecasting

Macroeconomic forecasting is the process of predicting future economic conditions based on key indicators such as gross domestic product (GDP), inflation, unemployment rates, the trade balance, and interest rates (Agu et al., 2022). The primary objective of macroeconomic forecasting is to support policy decision-making, business strategy formulation, and investment planning (Eyinade et al., 2021). Economic forecasting can be broadly classified into two main approaches: structural econometric models and data-driven models (Marcellino et al., 2003). Econometric models typically rely on linear equations to represent causal relationships among economic variables (Celli, 2022). Although these models are grounded in strong theoretical foundations, they often face limitations in capturing non-linear dynamics, sudden economic shocks, or structural changes. Traditional forecasting models encounter several key challenges, including linearity assumptions, whereby many models assume linear relationships among variables despite the inherently non-linear nature of macroeconomic phenomena; data lags, as economic data are frequently released with delays, reducing the accuracy of short-term forecasts (Sharma & Kathuria, 2025); and limited capacity to handle volatility and external shocks, since policy shifts or global disturbances are difficult for classical models to accommodate (Coulombe, 2024).

Big data

Big data refers to datasets characterized by extremely large volumes, high velocity, and a wide variety of data types (Abdalla, 2022). In the context of macroeconomics, big data can originate from various modern economic activities, such as digital transaction data from e-commerce and electronic payment systems, social media activity reflecting public opinion and sentiment, economic sensor data capturing mobility patterns, energy consumption, and transportation flows, as well as corporate reports and real-time economic news (Yu & Fang, 2023). The diversity of these data sources enables broader, more detailled, and more dynamic observation of economic activity than conventional macroeconomic statistics (Kara et al., 2024).

The utilization of big data in macroeconomic analysis enhances the ability to detect economic trends, shocks, and structural changes more rapidly and accurately (Muhammad Atif Khan et al., 2024). Whereas traditional economic analysis relies heavily on official statistics released periodically and often with significant time lags, big data provides near real-time information for economic assessment (Dörr et al., 2022). This creates substantial opportunities for the development of nowcasting techniques that estimate current economic conditions, thereby allowing policymakers, market participants, and researchers to respond more effectively and proactively to economic dynamics rather than relying solely on medium- and long-term projections.

Machine learning (ML)

Machine learning (ML) is a branch of artificial intelligence that enables computers to learn from data and identify complex patterns without being explicitly programmed (Sarker, 2021). In the context of economic forecasting, ML offers a flexible framework by capturing non-linear relationships and intricate interactions among economic variables (Al-Karkhi & Rządkowski, 2025). Several ML algorithms are widely used in economic forecasting, including Random Forest, an ensemble method that combines multiple decision trees to enhance prediction accuracy and stability; support vector regression (SVR), which is well suited for predicting continuous variables with non-linear patterns; and neural networks (artificial neural networks), which can model highly complex economic dynamics, including long-term non-linear dependencies.

A growing body of empirical research highlights the advantages of ML-based approaches over traditional econometric models in certain settings. Oancea & Simionescu (2024); Al-Karkhi & Rządkowski (2025); Ather et al. (2025) demonstrate that applying ML algorithms to large-scale datasets incorporating a wide range of economic indicators can significantly improve forecasting accuracy. Furthermore, Coulombe (2024); Zarkova (2025) introduces the macroeconomic random forest (MRF), an adaptation of the random forest framework specifically designed for macroeconomic analysis. This approach allows the model to capture better structural breaks and non-linear dynamics commonly observed in macroeconomic data, thereby producing more robust and policy-relevant forecasts.

Method

This study employs a computational, quantitative approach using data-driven forecasting methods to predict macroeconomic indicators by integrating big data and machine learning (ML). The approach is both explanatory and predictive, as it not only forecasts key economic variables but also evaluates the impact of leveraging heterogeneous data on predictive accuracy. The data used in this study are drawn from various sources, including traditional macroeconomic data such as real gross domestic product (GDP), inflation, unemployment, interest rates, and the trade balance, obtained from Statistics Indonesia and Bank Indonesia. In addition, the study utilizes digital and real-time data, including digital transaction volumes, mobility indices, energy consumption, and transportation activity. Unstructured data, such as economic news, public opinion, and social media trends, is also incorporated, sourced from Google Trends, news portals, and social media platforms. Furthermore, corporate reports, consumer surveys, and additional economic sentiment indicators are included to enrich the predictive information.

The data processing workflow begins with data cleaning, including the removal of duplicates, outliers, and missing values, and normalisation to ensure comparability across data sources. Next, data transformation is performed, particularly for text data, which is converted into numerical representations using natural language processing (NLP) techniques such as TF-IDF or word embeddings. To enhance predictive quality, feature engineering is applied, including the creation of additional indicators such as moving averages, monthly growth rates, and sentiment scores, as well as feature selection using correlation analysis and feature importance metrics derived from random forest models. Exploratory analysis is also conducted to assess the potential of each big data source to influence the prediction of economic indicators, enabling the empirical evaluation of each data type's contribution.

The prediction of economic indicators is carried out using several machine learning (ML) algorithms, namely random forests, support vector regression (SVR), neural networks, and macroeconomic random forests (MRF), which can capture structural breaks in macroeconomic data. These models are developed to predict multiple indicators simultaneously, allowing GDP, inflation, and unemployment to be forecasted within a single, integrated framework. Model performance is evaluated using standard prediction metrics such as mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). To enhance validity and reduce the risk of overfitting, the study uses rolling-window cross-validation on the time-series data.

In addition to performance evaluation, this study conducts a sensitivity analysis to assess the models' responses to economic shocks and rapid changes in input variables, thereby examining their ability to handle volatility and dynamic economic conditions. The multi-indicator prediction results are analyzed to identify the most effective combination of data sources and ML algorithms, and to provide insights into the contribution of each data type to predictive accuracy.

Results and discussion

Based on data analysis integrating big data and machine learning (ML) algorithms, this study demonstrates improved predictive accuracy for macroeconomic indicators compared to traditional models. The datasets employed include historical macroeconomic data (GDP, inflation, unemployment), real-time data (digital transactions, mobility indices, energy consumption), and unstructured data (economic news, public opinion, social media trends). Feature engineering and variable evaluation results indicate that, in addition to traditional indicators, sentiment scores from economic news contribute approximately 14% to GDP predictions. In comparison, digital transaction data account for 12% of inflation predictions, and mobility indices contribute around 17% to unemployment predictions. These findings highlight that integrating heterogeneous data significantly enhances the models' ability to capture economic dynamics that are not fully reflected in traditional data sources.

Table 1 presents the predictions of macroeconomic indicators using several machine learning (ML) algorithms, showing the following performance: gross domestic product (GDP): The macroeconomic random forest (MRF) model achieved a mean absolute error (MAE) of 0.48%, lower than that of the traditional econometric model (MAE = 0.63%). This model can adjust GDP predictions rapidly in response to changes in economic trends. For instance, when digital transaction activity decreased by 20% within a month, the MRF model adjusted the GDP forecast downward 0.3% faster than the traditional model. Inflation: the random forest model produced a mean absolute percentage error (MAPE) of 2.9%, outperforming the classical econometric approach (MAPE = 4.1%). Feature importance analysis indicates that news sentiment scores have a significant impact on inflation predictions, enabling the ML model to detect earlier fluctuations in public opinion and negative news. Unemployment: the support vector regression (SVR) model predicted unemployment rates with a root mean squared error (RMSE) of 0.36%, lower than that of the traditional model (RMSE = 0.54%). Daily mobility indices proved to be important indicators, capturing labor market changes more quickly than historical data alone.

Improved accuracy: all ML algorithms demonstrated lower errors than traditional models. The macroeconomic random forest (MRF) achieved the best performance for multi-indicator forecasting, lowering the MAE for GDP from 0.63% to 0.48%. Contribution of big data: real-time and unstructured data, such as sentiment scores, digital transactions, and mobility indices, significantly enhanced predictive accuracy compared to using historical data alone. Responsiveness to volatility: ML models, particularly MRF and SVR, adjusted predictions rapidly to economic shocks, whereas traditional models were slower to adapt to changing trends. In addition to improved accuracy, the ML models employed in this study are capable of simultaneous multi-indicator forecasting, enabling the prediction of GDP, inflation, and unemployment within a single, integrated framework. Sensitivity analysis: results indicate that the models are responsive to economic shocks, such as interest rate changes, energy price fluctuations, or surges in digital transactions. The models adjust forecasts more quickly than traditional methods, demonstrating greater adaptability to volatile economic conditions.

Table 1 result of macroeconomic indicator predictions

Economic indicators	Model	MAE/MAPE/RMSE	Note
PDB (Riil, % YoY)	Traditional models (Econometrics)	MAE = 0.63%	Traditional historical data-based predictions; less responsive to real-time data
	Random forest (RF)	MAE = 0.52%	Captures non-linear patterns; more accurate than traditional models
	Macroeconomic fandom forest (MRF)	MAE = 0.48%	Multi-indicator prediction; adaptive to changing economic trends; capturing structural breaks
Inflasi (CPI, % YoY)	Traditional models	MAPE = 4.1%	Conventional models are less able to utilize real-time data.
	Random forest	MAPE = 2.9%	News sentiment score integration improves accuracy
	SVR	MAPE = 3.1%	Suitable for short-term fluctuations; predictions are responsive to real-time data
Unemployment (% of total workforce)	Traditional model	RMSE = 0.54%	Low sensitivity to economic shocks
	SVR	RMSE = 0.36%	More accurate predictions; mobility index becomes an important feature
	MRF	RMSE = 0.34%	Multi-indicator prediction; capturing labor market volatility faster

Source: secondary data, processed

Overall, this study's results indicate that integrating big data and machine learning (ML) enhances the accuracy of macroeconomic indicator forecasts, particularly for GDP, inflation, and unemployment. Heterogeneous variables, such as real-time data and news sentiment scores, significantly improve predictions, enabling models to capture economic dynamics not reflected in traditional data. The developed ML-based multi-indicator forecasting framework can serve as an effective tool for evidence-based decision-making, risk mitigation, and more adaptive economic planning in the face of rapid changes.

The role of big data in enhancing the accuracy of macroeconomic forecasting

The findings of this study indicate that the integration of heterogeneous data, including historical macroeconomic data, real-time data, and unstructured data, makes a significant contribution to improving the accuracy of macroeconomic indicator forecasts. The results, showing that economic news sentiment scores contribute approximately 14% to GDP predictions, digital transaction data contribute 12% to inflation predictions, and mobility indices contribute around 17% to unemployment predictions, confirm that big data can capture dimensions of economic behavior and expectations that are not fully reflected in conventional statistical data. This suggests that modern economic dynamics cannot be optimally explained solely by traditional macroeconomic indicators, which are often aggregated and low-frequency.

Conceptually, the significant contribution of big data can be explained through its key characteristics: volume, velocity, and variety (Ghasemaghaei, 2021). Real-time data, such as digital transactions and mobility indices, provide an advantage in terms of velocity, allowing them to reflect changes in economic activity almost simultaneously. This study shows that mobility indices can explain changes in unemployment rates, demonstrating that population movement and daily economic activities can serve as early indicators of labour market conditions, well before official data are released. These findings expand the understanding that micro-activity-based economic indicators possess high predictive value in macroeconomic forecasting.

Furthermore, unstructured data, such as economic news and public opinion, contribute significantly through the expectation formation mechanism (Trust et al., 2023). The impact of news sentiment scores on GDP and inflation predictions indicates that the perceptions of economic agents, both consumers and investors, play a crucial role in shaping the direction of economic activity. When news sentiment deteriorates, ML models can detect potential economic slowdowns earlier than traditional models that rely solely on historical data. This reinforces the view that modern economies are heavily influenced by psychological factors and information, making text-based analysis an essential component of forecasting.

These findings are consistent with the study by Richardson et al. (2021); Soybilgen & Yazgan (2021), which demonstrated that integrating alternative indicators and real-time data significantly improves the nowcasting accuracy of GDP relative to standard econometric models. In their study, high-frequency data enabled the models to respond more rapidly to economic changes, particularly during periods of uncertainty. Similarly, Kyriazos & Poga (2024); Ather et al. (2025) found that using large, diverse datasets enables machine learning algorithms to identify nonlinear relationships and complex interactions among variables that conventional linear models often overlook. The results of the present study reinforce these findings by showing that the contribution of big data is not merely marginal, but substantive and empirically measurable.

Furthermore, this study’s findings indicate that big data serves not only as a supplement to macroeconomic data but also as a strategic information source for macroeconomic forecasting. The integration of heterogeneous data allows models to combine structural information from historical data with early signals from real-time and unstructured data. Consequently, ML models can generate forecasts that are more accurate and responsive to rapid economic changes, particularly amid global volatility and policy shocks. This aligns with arguments in the literature that future economic forecasting approaches should be adaptive and data-driven, rather than relying solely on rigid structural assumptions (Kokogho et al., 2024).

Overall, big data plays a crucial role in enhancing the accuracy and relevance of macroeconomic forecasting. By leveraging the characteristics of diverse, high-frequency data, this study demonstrates that integrating big data and machine learning can overcome the limitations of traditional models and provide a forecasting framework better suited to the complexities of the modern economy. These findings also support the research objective of exploring the potential of big data as a core foundation for developing more accurate, adaptive, and informative macroeconomic forecasting systems.

Performance evaluation of machine learning algorithms

The performance evaluation results indicate that all machine learning (ML) algorithms used in this study consistently achieved lower prediction errors than traditional econometric models. This performance improvement confirms that ML algorithms have an advantage in capturing the non-linear, multidimensional, and structurally sensitive dynamics of macroeconomic systems. The significant differences in performance across models also suggest that the characteristics of each macroeconomic indicator require distinct algorithmic approaches to achieve optimal predictions.

For real GDP, the macroeconomic random forest (MRF) achieved the best performance, with a mean absolute error (MAE) of 0.48%, lower than the standard random forest (0.52%) and traditional econometric models (0.63%). The advantage of MRF lies primarily in its ability to capture non-linear relationships, inter-variable interactions, and structural breaks that frequently occur in the economy, particularly during periods of crisis or policy transitions. The MRF approach allows for separating economic conditions into distinct regimes, enabling the model to adjust its prediction patterns to the prevailing economic context. This explains why MRF was able to adjust GDP forecasts more rapidly in response to a 20% decline in digital transaction activity, correcting the projected GDP growth 0.3% earlier than traditional models. These findings are consistent with Coulombe (2024); Magubane (2025), who demonstrated that MRF outperforms linear models in conditions of high volatility and structural changes, as linear models tend to be rigid in responding to regime shifts.

For the inflation indicator, random forests achieved the best performance, with a mean absolute percentage error (MAPE) of 2.9%, substantially lower than that of traditional models (MAPE = 4.1%). The superiority of Random Forest in predicting inflation can be attributed to its ability to integrate multiple information sources, including real-time and unstructured data, and to capture complex interactions among economic variables. Feature importance analysis indicates that economic news sentiment scores have a significant impact on inflation forecasts, underscoring the role of expectations shaped by media coverage and public perception in price dynamics. These findings support the results of Marcellino et al (2003); Borraz & Mello (2025), who emphasized that expectations and publicly available information heavily influence inflation. By integrating text-based and sentiment data, ML models can detect early signals of inflationary pressures before they are reflected in official price data, thereby improving both the accuracy and timeliness of predictions.

For the unemployment indicator, support vector regression (SVR) and MRF achieved the lowest root mean squared errors (RMSEs) of 0.36% and 0.34%, respectively, substantially lower than traditional models (RMSE = 0.54%). The strength of SVR lies in its ability to model non-linear relationships with high generalisation, particularly when the data exhibit complex patterns. However, the number of observations is relatively limited. In this study, mobility indices emerged as among the most important predictors of changes in unemployment, indicating that population movement and daily economic activities serve as early indicators of labour market conditions. These findings are consistent with recent studies showing that high-frequency mobility data can reflect real-time labor market dynamics, especially during periods of economic disruption, such as crises or mobility restrictions (Spelta & Pagnottoni, 2021; Ten et al., 2024).

The differences in performance across ML algorithms indicate that no single model excels across all macroeconomic indicators. MRF outperforms random forests in predicting GDP and unemployment by capturing structural changes and inter-variable interactions. In contrast, random forests are more effective for inflation forecasts because they can integrate heterogeneous data and public expectations. SVR demonstrates strong predictive power for unemployment, particularly for short-term fluctuations and labour market volatility. These findings reinforce the literature's argument that the choice of ML algorithm should be tailored to the characteristics of the variable being predicted (Noroozi et al., 2023). The multi-model approach employed in this study enables a more comprehensive evaluation and supports the research objective of developing a flexible, adaptive macroeconomic forecasting framework (Maruthi et al., 2025).

Overall, the performance evaluation of ML algorithms in this study demonstrates that ML-based approaches can overcome the limitations of traditional econometric models, particularly in addressing nonlinearity, volatility, and structural changes in the economy. These results directly support the research objective of evaluating the effectiveness of various ML algorithms in predicting key economic indicators, while also confirming that integrating big data and ML is an effective strategy for improving the accuracy and relevance of macroeconomic forecasting.

Advantages of ML-based multi-indicator approach

This study develops a multi-indicator forecasting framework based on machine learning (ML) capable of simultaneously predicting gross domestic product (GDP), inflation, and unemployment within a single integrated system. This approach is fundamentally different from most previous studies, which typically focus on a single economic indicator or short-term forecasts. The results indicate that the multi-indicator approach not only improves the prediction accuracy of each variable but also enables the model to capture dynamic interactions among interrelated macroeconomic indicators.

Theoretically, macroeconomic indicators such as GDP, inflation, and unemployment do not exist in isolation but interact through various economic mechanisms, including the relationship between economic growth and labor absorption, as well as the interaction between demand pressures and inflation (Girdzijauskas et al., 2022; Muhammad, 2023; Shiferaw, 2023). Traditional econometric models often represent these relationships separately or through rigid structural assumptions. In contrast, a multi-indicator ML approach allows the model to learn these relationships in a data-driven manner, without explicitly specifying functional forms or causal links. This provides greater flexibility in capturing complex and evolving economic dynamics.

The empirical results of this study indicate that models such as the macroeconomic random forest (MRF) can leverage cross-indicator information to improve both the stability and accuracy of predictions. For example, changes in digital transaction data and news sentiment not only affect GDP forecasts but also provide early signals of inflationary pressures and labor market conditions. By processing all indicators within a single framework, ML models can identify common patterns and non-linear interdependencies among economic variables that are often overlooked in single-indicator approaches.

These findings extend the results of Ather et al. (2025); Patsiarikas et al. (2025), who demonstrated the superiority of ML algorithms in improving the accuracy of macroeconomic variable predictions, but whose applications were limited to single-indicator forecasts. Similarly, Richardson et al. (2021); Soybilgen & Yazgan (2021) highlighted the advantages of ML in nowcasting GDP using real-time data, yet their approach did not integrate simultaneous predictions of other macroeconomic indicators. Accordingly, this study addresses this research gap by demonstrating that integrating big data and ML can develop a more comprehensive and holistic macroeconomic forecasting system.

Another advantage of the multi-indicator approach is its ability to enhance the consistency of predictions across variables. In traditional approaches, forecasts of GDP, inflation, and unemployment generated by separate models are often economically inconsistent, for example, high predicted economic growth coinciding with an unrealistically rising unemployment rate. The multi-indicator ML approach allows the model to adjust its predictions to remain aligned with plausible economic dynamics, as all indicators are learned simultaneously within a single system. Furthermore, this approach has proven to be more adaptive to economic volatility and shocks. When sudden changes occur in specific variables, such as spikes in energy prices or declines in mobility activity, their impact can be immediately reflected in the forecasts of all relevant indicators (Dokas et al., 2023). This enhances the model's practical utility in policy-making contexts, where decision-makers require a comprehensive view of economic conditions rather than isolated single-indicator forecasts.

Overall, the multi-indicator ML-based forecasting approach developed in this study makes a significant contribution both methodologically and empirically. This approach not only enhances prediction accuracy but also deepens our understanding of the interrelationships among macroeconomic indicators in a complex, dynamic economic environment. Accordingly, the study supports the objective of developing an integrated, adaptive, and policy-relevant macroeconomic forecasting framework that can inform economic planning and decision-making amid global uncertainty.

Model responsiveness to volatility and economic shocks

The sensitivity analysis in this study indicates that machine learning (ML) models, particularly the macroeconomic random forest (MRF) and support vector regression (SVR), exhibit a much higher responsiveness to economic volatility and shocks than traditional econometric models. When interest rates change, energy prices fluctuate, or digital transaction volumes surge, ML models can quickly and proportionally adjust forecasts of macroeconomic indicators. In contrast, traditional models tend to lag in adjusting predictions due to their strong reliance on historical patterns and the assumption of stable linear relationships among variables.

Methodologically, this responsiveness advantage stems from ML algorithms' ability to learn nonlinear patterns and complex interactions among economic variables (Kyriazos & Poga, 2024). Models such as the macroeconomic random forest (MRF) can partition the data space into distinct economic regimes, enabling the model's responses to shocks to be context-specific and adaptive. When exogenous variables undergo sudden changes, such as spikes in energy prices or monetary policy tightening, MRF can rapidly adjust the prediction structure without explicitly re-estimating the entire model, as is required in traditional econometric approaches.

These findings are consistent with Coulombe (2024); Zhang & Bian (2025), who demonstrated that the macroeconomic random forest (MRF) excels at capturing structural breaks and non-linearities in macroeconomic data, particularly during periods of crisis or policy transitions. In the context of this study, this advantage is reflected in MRF's ability to respond to sharp declines in digital transactions and changes in mobility indices, which are immediately reflected in adjustments to forecasts of GDP and unemployment. This indicates that ML models can leverage early signals from real-time data to anticipate the impact of economic shocks before they are reflected in official indicators.

In support vector regression (SVR), its high responsiveness to volatility primarily stems from its ability to construct flexible optimal margins when modelling non-linear relationships. SVR is particularly effective in capturing short-term fluctuations and sudden changes in economic variables, as reflected in forecasts of the unemployment rate. These findings are consistent with those of Ather et al. (2025); Kayral et al. (2025), who demonstrated that SVR performs exceptionally well under volatile, unstable data conditions, especially when combined with high-frequency indicators.

Furthermore, integrating big data into ML models enhances responsiveness to economic shocks. Real-time data, such as digital transactions, mobility indices, and energy consumption, enable the model to detect changes in economic activity almost in real time. This supports the findings of Soybilgen & Yazgan (2021), who showed that using high-frequency data significantly improves a model's ability to respond to short-term economic shocks, particularly for GDP nowcasting. Accordingly, this study extends the literature by demonstrating that high responsiveness applies not only to a single indicator but also within a multi-indicator forecasting framework.

In contrast, the limitations of traditional econometric models become increasingly apparent under unstable economic conditions. Linear models with fixed parameters generally assume constant relationships among variables over time, making it difficult to capture structural changes and the economy's dynamic behavior. Their reliance on low-frequency historical data also delays the detection of economic shocks. These findings are consistent with criticisms in the modern macroeconomic literature, which highlight the limitations of traditional models in addressing global uncertainty and volatility (David & Veronesi, 2022).

Overall, the findings of this study underscore that high responsiveness to economic volatility and shocks is a key advantage of ML-based forecasting approaches. By leveraging big data and adaptive algorithms, ML models can deliver faster, more accurates, and more relevant predictions amid dynamic economic conditions. These results directly support the research objective of developing a macroeconomic forecasting framework that is not only statistically accurate but also robust and responsive to rapid economic changes, thereby offering significant practical value for policy-making and economic planning.

Conclusions

The integration of big data and machine learning (ML) algorithms significantly enhances the accuracy and responsiveness of macroeconomic indicator forecasting compared to traditional econometric approaches. The use of heterogeneous data, including historical, real-time, and unstructured data, has been shown to capture non-linear, high-frequency economic dynamics influenced by public expectations. These aspects are not fully reflected in conventional macroeconomic data. Empirical results indicate that ML algorithms, particularly macroeconomic random forests (MRF), random forests, and support vector regression (SVR), produce lower prediction errors for GDP, inflation, and unemployment while demonstrating greater adaptability to economic volatility and shocks. Moreover, the development of a multi-indicator forecasting framework enables more consistent and integrated predictions across macroeconomic variables, thereby increasing the model's practical relevance for policy-making and economic planning. In conclusion, this study confirms that big data and machine learning constitute a critical foundation for developing macroeconomic forecasting systems that are more accurate, adaptive, and aligned with the complexities of the modern economy.

The integration of big data and machine learning (ML) can serve as a strategic approach to enhancing the quality of macroeconomic forecasting for policymakers, researchers, and market participants. Improved accuracy and responsiveness in predicting GDP, inflation, and unemployment indicators enable the formulation of more anticipatory fiscal and monetary policies, while reducing delays in responding to economic shocks. Methodologically, these findings support a shift away from reliance on traditional econometric models toward a data-driven approach that is better able to accommodate nonlinearity, volatility, and structural changes in the economy. In practice, the multi-indicator ML-based forecasting framework developed in this study can serve as a decision-support system for real-time economic monitoring, risk mitigation, and more effective economic planning amid the increasing complexity of modern economic dynamics.

This study has several limitations that warrant careful consideration. It relies on the quality and availability of big data, which may contain noise, representation biases, and consistency issues across sources, particularly in unstructured data such as news articles and social media. Although machine learning (ML) models demonstrate superior predictive performance, their limited transparency limits causal interpretation of economic variables. The study primarily focuses on evaluating prediction accuracy and does not thoroughly examine the policy implications of using ML models under various economic scenarios. Accordingly, future research is recommended to develop more robust big data cleaning and validation methods, integrate explainable AI (XAI) techniques to enhance model interpretability, and expand analyses through policy simulations and cross-country or crisis-period testing to improve the generalizability and practical utility of macroeconomic forecasting models.

References

Abdalla, H. B. (2022). A brief survey on big data: technologies, terminologies and data-intensive applications. *Journal of Big Data*, *9*(1), 107. https://doi.org/10.1186/s40537-022-00659-3

Agu, S. C., Onu, F. U., Ezemagu, U. K., & Oden, D. (2022). Predicting gross domestic product to macroeconomic indicators. *Intelligent Systems with Applications*, *14*, 200082. https://doi.org/10.1016/j.iswa.2022.200082

Al-Karkhi, M. I., & Rządkowski, G. (2025). Innovative machine learning approaches for complexity in economic forecasting and SME growth: A comprehensive review. *Journal of Economy and Technology*, *3*, 109–122. https://doi.org/10.1016/j.ject.2025.01.001

Ather, D., Singh, G., Kler, R., Kler, R., Nag, B. C., & Saraswat, E. (2025). Exploring Non-Linear Dynamics of Current Account Balance in Uzbekistan Using SVR and Random Forest. *2025 International Conference on Technology Enabled Economic Changes (InTech)*, 620–625.

https://doi.org/10.1109/InTech64186.2025.11198375

Borraz, F., & Mello, M. (2025). Communication, information and inflation expectations. *Central Bank Review*, *25*, 100224. https://doi.org/10.1016/j.cbrev.2025.100224

Celli, V. (2022). Causal mediation analysis in economics: Objectives, assumptions, models. *Journal of Economic Surveys*, *36*(1), 214–234. https://doi.org/10.1111/joes.12452

Coulombe, P. G. (2024). The macroeconomy as a random forest. *Journal of Applied Econometrics*, *39*(3), 401–421. https://doi.org/10.1002/jae.3030

David, A., & Veronesi, P. (2022). A Survey of Alternative Measures of Macroeconomic Uncertainty: Which Measures Forecast Real Variables and Explain Fluctuations in Asset Volatilities Better? *Annual Review of Financial Economics*, *14*(1), 439–463. https://doi.org/10.1146/annurev-financial-111720-091804

Dokas, I., Oikonomou, G., Panagiotidis, M., & Spyromitros, E. (2023). Macroeconomic and Uncertainty Shocks' Effects on Energy Prices: A Comprehensive Literature Review. *Energies*, *16*(3), 1491. https://doi.org/10.3390/en16031491

Dörr, J. O., Kinne, J., Lenz, D., Licht, G., & Winker, P. (2022). An integrated data framework for policy guidance during the coronavirus pandemic: Towards real-time decision support for economic policymakers. *PLOS ONE*, *17*(2), e0263898. https://doi.org/10.1371/journal.pone.0263898

Eyinade, W., Ezeilo, O. J., & Ogundeji, I. A. (2021). A Forecasting Model for Integrating Macroeconomic Indicators into Long-Term Financial Strategy in Oil and Gas Enterprises. *International Journal of Multidisciplinary Research and Growth Evaluation*, *2*(1), 919–926. https://doi.org/10.54660/.IJMRGE.2021.2.1.919-926

Ghasemaghaei, M. (2021). Understanding the impact of big data on firm performance: The necessity of conceptually differentiating among big data characteristics. *International Journal of Information Management*, *57*, 102055. https://doi.org/10.1016/j.ijinfomgt.2019.102055

Girdzijauskas, S., Streimikiene, D., Griesiene, I., Mikalauskiene, A., & Kyriakopoulos, G. L. (2022). New Approach to Inflation Phenomena to Ensure Sustainable Economic Growth. *Sustainability*, *14*(1), 518. https://doi.org/10.3390/su14010518

Hopp, D. (2024). Benchmarking econometric and machine learning methodologies in nowcasting GDP. *Empirical Economics*, *66*(5), 2191–2247. https://doi.org/10.1007/s00181-023-02515-6

Ikegwu, A. C., Nweke, H. F., Anikwe, C. V., Alo, U. R., & Okonkwo, O. R. (2022). Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. *Cluster Computing*, *25*(5), 3343–3387. https://doi.org/10.1007/s10586-022-03568-5

Kara, O., Altinay, L., Bağış, M., Kurutkan, M. N., & Vatankhah, S. (2024). Institutions and macroeconomic indicators: entrepreneurial activities across the world. *Management Decision*, *62*(4), 1238–1290. https://doi.org/10.1108/MD-04-2023-0490

Kayral, İ. E., Bozkurt, M. A., Bejaoui, A., & Jeribi, A. (2025). Hybrid Wavelet-SVR, machine learning, and deep learning models for predicting clean energy markets amidst global events. *Neural Computing and Applications*, *37*(21), 16781–16823. https://doi.org/10.1007/s00521-025-11345-9

Khan, Muhammad Anees, Abbas, K., Su'ud, M. M., Salameh, A. A., Alam, M. M., Aman, N., Mehreen, M., Jan, A., Hashim, N. A. A. B. N., & Aziz, R. C. (2022). Application of Machine Learning Algorithms for Sustainable Business Management Based on Macro-Economic Data: Supervised Learning Techniques Approach. *Sustainability*, *14*(16), 9964. https://doi.org/10.3390/su14169964

Khan, Muhammad Atif, Ali, H., Shabbir, H., Noor, F., & Majid, M. D. (2024). Impact of Macroeconomic Indicators on Stock Market Predictions: A Cross Country Analysis. *Journal of Computing & Biomedical Informatics*, *8*(1). https://www.jcbi.org/index.php/Main/article/view/740

Khudhur, D. Y., Shibghatullah, A. S., Shaker, K., Abdul Latif, A., & Muda, Z. C. (2025). Recent Trends in Machine Learning for Healthcare Big Data Applications: Review of Velocity and Volume Challenges. *Algorithms*, *18*(12), 772. https://doi.org/10.3390/a18120772

Kokogho, E., Odio, P. E., Ogunsola, O. Y., & Nwaozomudoh, M. O. (2024). AI-Powered Economic Forecasting: Challenges and Opportunities in a Data-Driven World. *International Journal of Management and Organizational Research*, *3*(6), 74–83. https://doi.org/10.54660/IJMOR.2024.3.6.74-83

Kyriazos, T., & Poga, M. (2024). Application of Machine Learning Models in Social Sciences: Managing Nonlinear Relationships. *Encyclopedia*, *4*(4), 1790–1805. https://doi.org/10.3390/encyclopedia4040118

Magubane, K. (2025). The Stability of the Financial Cycle: Insights from a Markov Switching Regression in South Africa. *Journal of Risk and Financial Management*, *18*(2), 76. https://doi.org/10.3390/jrfm18020076

Marcellino, M., Stock, J. H., & Watson, M. W. (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. *European Economic Review*, *47*(1), 1–18. https://doi.org/10.1016/S0014-2921(02)00206-4

Maruthi, M., Kim, B., Sujeen, S., An, J., & Chen, Z. (2025). Multi-model integration for dynamic forecasting (MIDF): a framework for wind speed and direction prediction. *Artificial Intelligence Review*, *58*(6), 164. https://doi.org/10.1007/s10462-025-11140-x

Muhammad, A. A. (2023). Examining the Relationship among Unemployment, Inflation, and Economic Growth. *Journal of Business and Economic Options*, *6*(2), 23–31. http://resdojournals.com/index.php/jbeo/article/view/237

Mulenga, R. (2024). Navigating Global Economic Volatilities: Towards Economic Resilience. *Journal of Economics, Finance and Management (JEFM)*, *3*(1), 295–308. https://journal-efm.fr/index.php/JEFM/article/view/77

Noroozi, Z., Orooji, A., & Erfannia, L. (2023). Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. *Scientific Reports*, *13*(1), 22588. https://doi.org/10.1038/s41598-023-49962-w

Oancea, B., & Simionescu, M. (2024). Gross Domestic Product Forecasting: Harnessing Machine Learning for Accurate Economic Predictions in a Univariate Setting. *Electronics*, *13*(24), 4918. https://doi.org/10.3390/electronics13244918

Patsiarikas, M., Papageorgiou, G., & Tjortjis, C. (2025). Using Machine Learning on Macroeconomic, Technical, and Sentiment Indicators for Stock Market Forecasting. *Information*, *16*(7), 584. https://doi.org/10.3390/info16070584

Richardson, A., Mulder, T. van F., & Vehbi, T. (2021). Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, *37*(2), 941–948. https://doi.org/10.1016/j.ijforecast.2020.10.005

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Sharma, M., & Kathuria, V. (2025). Macroeconomic Nowcasting: What can Central Banks Learn from a Structured Literature Review? *Journal of Quantitative Economics*, *23*(2), 333–388. https://doi.org/10.1007/s40953-024-00421-x

Shiferaw, Y. A. (2023). An Understanding of How GDP, Unemployment and Inflation Interact and Change across Time and Frequency. *Economies*, *11*(5), 131. https://doi.org/10.3390/economies11050131

Soybilgen, B., & Yazgan, E. (2021). Nowcasting US GDP Using Tree-Based Ensemble Models and Dynamic Factors. *Computational Economics*, *57*(1), 387–417. https://doi.org/10.1007/s10614-020-10083-5

Spelta, A., & Pagnottoni, P. (2021). Mobility-based real-time economic monitoring amid the COVID-19 pandemic. *Scientific Reports*, *11*(1), 13069. https://doi.org/10.1038/s41598-021-92134-x

Suprunenko, S., Pylypenko, N., Trubnik, T., & Volchenko, N. (2023). Forecast of changes in the macroeconomic situation in Ukraine: smart economy of the future. *Futurity Economics&Law*, *3*(3), 214–231. https://doi.org/10.57125/FEL.2023.09.25.13

Ten, G. K., Tafere, H. K., Newhouse, D., & Pape, U. (2024). How well can real-time indicators track the economic impacts of a crisis like COVID-19? *Development Policy Review*, *42*(5). https://doi.org/10.1111/dpr.12797

Trust, P., Zahran, A., & Minghim, R. (2023). Understanding the influence of news on society decision making: application to economic policy uncertainty. *Neural Computing and Applications*, *35*(20), 14929–14945. https://doi.org/10.1007/s00521-023-08438-8

Yu, D., & Fang, C. (2023). Urban Remote Sensing with Spatial Big Data: A Review and Renewed Perspective of Urban Studies in Recent Decades. *Remote Sensing*, *15*(5), 1307. https://doi.org/10.3390/rs15051307

Zarkova, S. (2025). Empirical Assessment of the Multiplier Effects of the EU Recovery and Resilience Facility Using Machine Learning. *Economies*, *13*(12), 338. https://doi.org/10.3390/economies13120338

Zhang, Q., & Bian, Y. (2025). Understanding Machine Learning-based Methods in Macroeconomic Forecasting: Tracking Chinese GDP Growth Rate. *Computational Economics*. https://doi.org/10.1007/s10614-025-11115-8