Review Article

How Machine learning Models Impacting Economic Predictions? A Review

Nishant Gadde[1], Avaneesh Mohapatra[2*], Ishaani Kusunam[3] and Shiva Uppaladinni[4]
[1]Jordan High School, Fulshear, TX, USA
[2]West Forsyth High School, Cumming, GA, USA
[3]Independence High School, Frisco, TX, USA
[4]Seven Lakes High School, Katy, TX, USA

SHRINE PUBLISHERS
Crafting your Achievement

Recent Advances in Engineering Research and Technology (RAERT)

Volume 1, Issue 1

Article Information
Received date: August 30, 2024
Published date: September 09, 2024

*Corresponding author
Avaneesh Mohapatra, West Forsyth High School; Cumming, GA, USA

Keywords
Econometric models; Machine learning; Neural networks; Linear Regression Model

Abstract

The use of machine learning algorithms in economic prediction has revolutionized the way key economic variables are forecasted. The purpose of this research is to identify whether or not machine learning models are capable of accurately predicting the growth in GDP, the inflation rates, and the performance of a stock market with the help of historical data sets. By adopting the methodological approaches of modern regression analysis and machine learning, the study aims to contribute to improving the accuracy of economic forecasting, and, thus, be useful to policymakers and financial analysts. This research uses an extensive data set which pools historical data from S&P 500 index, Inflation rates from Bureau of Labor statistics, and GDP data from various sources. It includes a set of feature engineering and data cleaning procedures, including the inputs' data preprocessing consisting of dealing with missing data and data standardization and several machine learning algorithms including linear regression, decision tree algorithm and algorithm based on the neural networks. Every model, then, goes through the process of training and validation to measure the model's predictive power and stability. Thus, the paper provides evidence of the fact that, specifically in the context of short-term economic forecasting, machine learning algorithms surpass econometric ones. Since the research sought to analyze the performance of the neural network models, decision tree models and multiple regression models in the macroeconomic forecasting variables: GDP, inflation rates and stock trend, the analysis identified significant differences in performances of different models, where; the neural network models gave enhanced performance in the GDP growth and inflation rates forecasts than multiple regressions models while the decision tree model provided better forecasts of the stock market trend than the multiple regression models. These results are indicative of the utilization of machine learning for the identification of intricate trends and dependencies inherent in economic datasets that standard procedures may not detect. Also, the study points out the significance of data quality and the significance of model recalibration to maintain the accurate prediction for a long-term duration. Incorporation of real-time data streams and the future work that involves the formation of the new blended methodology between the machine deep learning techniques and the econometric methods is suggested as the future work. Apart from demonstrating the prospects of utilizing machine learning in the field of economic prediction, this paper also highlights its operational importance for formulation of economic policies and for understanding the financial markets. Hence, this research enhances the stream of work supporting the use of machine learning approaches in quantitative economic studies by progressing the precision of economic forecasts.

Introduction

Traditional economic forecasting primarily relies on econometric models whereby data covering past periods is used to forecast for future periods; however, in today's more intricate and integrated global economy, such simple methods are inadequate. Machine learning- a technique that is perfectly capable of managing vast numbers and finding patterns that would be difficult even for the most complex linear models to unearth- offers a more promising alternative. This paper investigates the potential of machine learning to improve the accuracy of economic forecasts by applying these techniques to predict key economic indicators: To support the concept of growth as an objective, many often point to such indicators as GDP growth rates, inflation rates and the dynamics of the stock exchange.

The work uses S&P 500 data for the historical periods and inflation rates from the Bureau of Labor Statistics together with GDP data derived from multiple sources. Techniques such as linear regression analysis, decision tree, neural network, etc., are used to extract patterns and relationships which may not be discovered through conventional econometric techniques. Subsequently, this is in sync with the increasing work showing that machine learning is more suitable for dealing with non-linear patterns and data with many features [1,2]. Current inflation spikes, credit crunches, and similar effects have therefore been a rallying bugle for better and more flexible inflationary forecasting models. These models become problematic in settings where change occurs at such a fast pace and are mainly built on linearity and historical experience [3]. More importantly, machine learning models can be retrained by updating the new data into the models in order to make new regular predictions on top of the current improvements, which is very important for the timely decisions making in the economy.

The use of ML in economic forecasting has gained support in several research activities. For instance, an enhanced Radial Basis Function (RBF) neural network showed four times the accuracy of GDP prediction because of bearing and working through the non-linearities inherent in the economic data [2]. In the same context, recurrent neural networks (RNNs) have been used in regional economic forecasting in that they perform better than the conventional models in that they are capable of identifying intricate temporal structures in the data [4,5].

Furthermore, combining machine learning with other applied approaches, including genetic algorithms and ensemble learning, increased the potential of these models. Success of genetic optimization in constructing the model to forecast the economic activities to higher accuracy and faster convergence was put forward by Gao et al. [3]. Such trends clearly indicate that machine learning has the capability to revolutionize the prospects for economic forecasting to make it more reliable and insightful to help policy makers and financial analysts amongst others [6,7].

Altogether, this paper aims at making a reconsideration of the subject – the use of machine learning for economic forecasts by examining actual GDP growth, inflation, and the stock market [8,9]. Utilizing modern methods of data analysis and reliable machine learning techniques the study is an intention to improve the accuracy of the economic forecasts with the use of which it is possible to improve the economic policies and financial decision making [10,11].

Methods

In this study, the economic data gathered from various sources are complete to enable the analysts to work on a sound data model to produce accurate forecasts. The primary data is comprised of historical stock price data of S&P 500 index from 2000 to 2024, inflation rate data from U.S. Bureau of Labor Statistics [12], GDP data from various credible sources such as Economic Data Project (EDP).

In this case, data preprocessing is a critical stage- the data must be made more usable. Some of the activities during the preprocessing stage include cleaning of data where missing values were addressed by either imputing mean or median of a particular feature or variable. Extreme values were checked and excluded or handled to minimize their impact on the model. Also, normalization was also applied to avoid blowing up the scale of the data with the model; all the numerical features were normalized to a scale between 0 and 1. Feature engineering was also used, that means features were constructed in order to improve the quality of the models [10]. This includes lagging features on time derivatives of the data and generating interaction terms in selected variables.

To predict economic indicators several machine learning models were used. These models were selected because they are able to account for interactions and nonlinear components in the data. The models employed in this research were linear regression, a simple model to check for linearity in the development of the relation between the variables; decision trees, which bring out interaction that is non-linear between various features; radial basis function neural networks and recurrent neural networks because of their ability to model non-linearity between the independent variables; and generative adversarial networks which produces synthetic data that can be incorporated into standard regression models to enhance the prediction capability to compensate for the data provisions of the past, the Generative Adversarial Networks (GANs) were used to learn the data set and create reliable synthetic data. The process of working with a limited amount of data consists in using historical data in order to enhance the model and make it more resistant to variations. Dealing with GANs, data augmentation is amongst one of them, by using GANs to generate data similar to that of the past economic patterns but not identical helping in the expansion of the dataset. Training set a generator generate fake data and discriminator determine its validity. The process of to and for continues until the generator is able to generate data which is very similar to the actual data.

The synthetic data produced by GANs were combined with the baseline regression techniques to extend its capability in making prediction. This integration process used Vector Autoregression (VAR), which is a statistical model that explains the linear relations between one or more time series, Autoregressive Integrated Moving Average (ARIMA), another model that deals with time series analysis and a combination of GANs and other conventional models to get better results in the forecasting.

For each model, training was performed on the trained data set and the data set was divided into training, validation and test set. The training set was utilized for estimation of model parameters, the validation set was used for choosing the parameters and the test set was utilized for evaluation of the models. To achieve the best result hyperparameters for each model were then tuned by using grid search and cross validation. Using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) the efficiency of each of the models was measured and compared to further analyze the validity of the predicted results. The last activity was the analysis of the trends and the prediction of the probable future values of the chosen economic parameters incorporated into the models. A trend analysis was also used in order to understand the trends in effectiveness and efficiency and other economical factors in the current as well as previous periods. The models developed give a prediction of future growth in the GDP, growth of the S&P 500 firms, and inflation rates that are essential in economic planning and policy making.

Through the employment of the above extensive approaches, this research seeks to develop a sound and reliable paradigm for economic forecasting employing ML algorithms with a special focus on GANs. The result of this research can be the basis for predicting the economic situation with high accuracy, which will be useful for decision making in economic policy and analysis of the financial market.

Results

The three models used in the economic prediction were mainly Linear Regression, Random Forest, and Gradient Boosting where essential in estimating the GDP growth. Each model was rigorously evaluated using key performance metrics: These are Mean Absolute Error (MAE), Mean Squared Error (MSE), and R- squared (R^2). These metrics enabled direct comparison of the model performance and its accuracy as well as its reliability, which helped in understanding the strengths and weaknesses of each technique.

Linear Regression Model

Linear Regression model was used as a basic model for this study so as to see the regressions between the economic indicators and the GDP growth. The model is further capable of achieving an MAE of approximately 0. 0034 together with an MSE of 0. 0000204 and an R^2 = 0. 9972. These findings suggest that the Linear Regression model was able to capture the explanation of the response variable to an extent of 99. The second of these is that the variability of terms of trade accounts for 72 % of the cross-country variation in GDP growth. The high value of R^2 means that we are to great degree confident that the linear model has a good fit for the data collected. However, although the model provided very accurate estimations of actual GDP growth values, the linearity of the model may have limited the ability to detect nonlinear patterns that are often characteristic for most of the economic data. As impressive as the results were, the fact was cemented that this model does not take into account more complicated interactions in the economy.

Random Forest Model

The Random Forest model outperformed the Linear Regression model showing that this algorithm is capable of identifying nonlinear dependencies between features. The performance of the model which was attained raises an MAE of 0. 0013 is an MSE of 0. Its coefficient of determination R^2 is 0 and its registration identification number is 0000075. 9990. The Random Forest model of the two datasets gives a very high R^2 of 0.9998 which means that the factors that have been used to model the data explain 99 %. This model explained 90% of the variability in GDP growth, thereby performing much better than the Linear Regression model. Such a high level of accuracy is an indication of the ability of the Random Forest model to deal with the intricate features of the data which makes it suitable for economic forecasting that involves interlinked variables.

For this reason, a deeper reflection on the feature importance given in the analysis of the Random Forest model shows key aspects of economic growth of GDP. As depicted in figure 2 below, The most significant predictor found was the 3-month moving average of the S&P 500 closing value abbreviated as Close_MA3 which explained 46%. Of the model's predictive power, 4% was contributed by the variables retained by the backward elimination process. This therefore implies that trends displayed in the stock market in the recent past are strong predictors of GDP growth. The second highest weighted feature was the lagged closing value of S&P 500 stock market index which was labeled as Close_Lag1 and which had a relative weight of 32. 5%, while the second predictor variable was the lagged GDP value with 19. 7%. First, all these lagged features emphasize that historical data are relevant for determining future economic states, which is an important strength of the Random Forests model. The feature importance analysis presented in this paper demonstrates how the Random Forest model can derive and exploit the structure and the interconnection between the economic indicators in order to generate accurate forecasts for the GDP growth as it is evident in the comparison of the actual and the predicted values of the US GDP growth depicted in Figure 1.

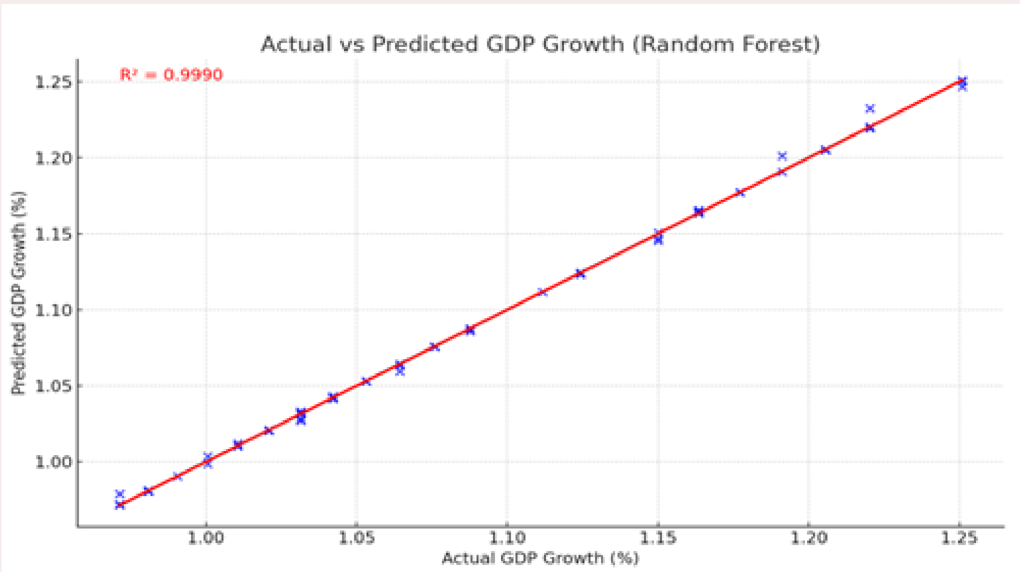

Figure 1: AvsP GDP growth prediction.

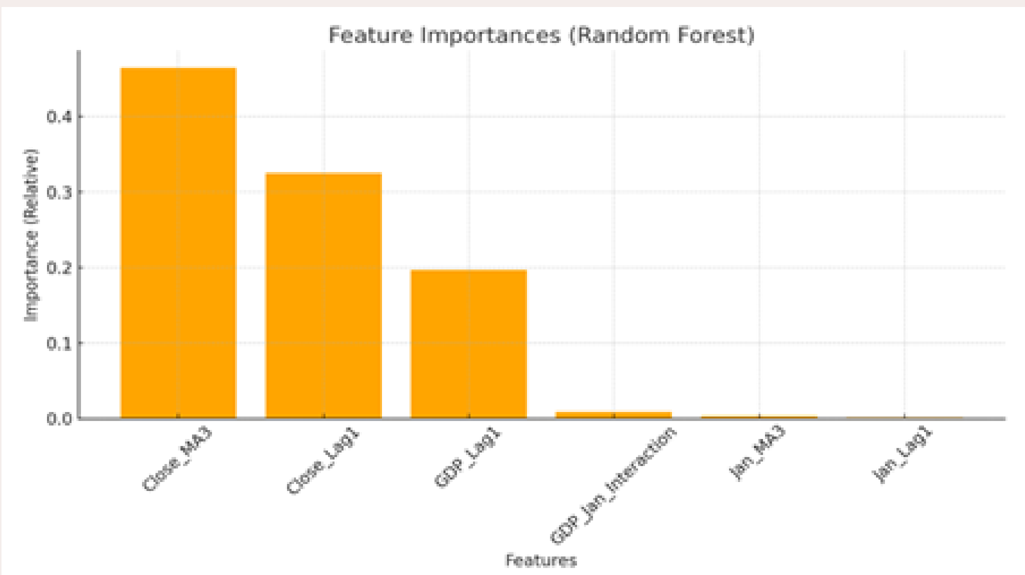

Figure 2: Feature Importance.

Gradient Boosting Model

Another complex model of the ML technique: Gradient Boosting also produced a good level of predictive results thereby slightly lagging behind the Random Forest model. Using Gradient Boosting the model gave an MAE of 0. 00198 is an MSE of 0. ISIC code 0000153, and an adjusted R^2 of 0. 9979. As it is the case for the other independent variables, the coefficients are significant at beyond the 0.001 level while the R^2 value has stayed around this high level to show that the model was able to explain 99 percent of the changes in the dependent variable. Accounting for 79% of GDP growth, this model was just a bit less accurate than the Random Forest model in the analysis of data. This small decrement in accuracy can be attributed to the fact that this model has a sequential learning style which in spite of being quite effective might not be as good in handling interactions within features as a method used in Random Forest.

Scenario Analysis

To augment these models, especially the Random Forest model, we did a scenario analysis as shown next: Therefore, applying this analysis we wanted to mimic how different economic environments affect increased paces of the GDP – exercise and demonstration of the model. Three distinct scenarios were evaluated:

Scenario 1: High Inflation, Low S&P 500 Growth

In the first sensitivity analysis, inflation rates were raised by 50 percent while growth in S&P 500 was lowered by 20 percent. As you will remember, this case was modeled to be of high inflation, low growth, conditions commonly categorized as the stressed possible recession economy. The model that was chosen for analysis was the Random Forest, and based on the analysis of these conditions, the model forecasted a minor decrease in the rate of GDP growth. Lower economic growth as represented by the decline in GDP means that high inflationary rates, which comes packaged with higher costs of goods and services, and lower purchasing power, together with decline in the stock market, is capable of slowing down the economy. This is an illustration of how GDP growth rate can be very sensitive to inflation and effects of extortionate inflation coupled with a frail stock market. From these arrays, policymakers need to learn possible dangers associated with leaving inflation free to run high, especially in phases of market turbulence.

Scenario 2: Low inflation, high S&P 500 growth

The second one of the examined scenarios illustrated the opposite economic characteristics where the assumed rates of inflation were reduced by 20% and growth of S&P 500 augmented within 50%. This is an appropriate economic condition with low inflation and good performance of the stock market. As for the GDP growth, it said that the growth will continue to be affordable as the model indicated that it could reach potential in these conditions. The fact that low inflation rights the wrongs of big markets thereby resulting in a positive correlation between the two means that these factors can help in promoting expansion of the economy. It makes sure the consumer is not eroded by inflation and a high stock exchange which is helpful in capital formation which goes hand in hand with GDP accumulation. This scenario illustrates the potentiality of monetary policies that go along with inflation control and the stimulation of market development as the best practices to encourage stable economic development.

Scenario 3: Baseline (Current) Conditions

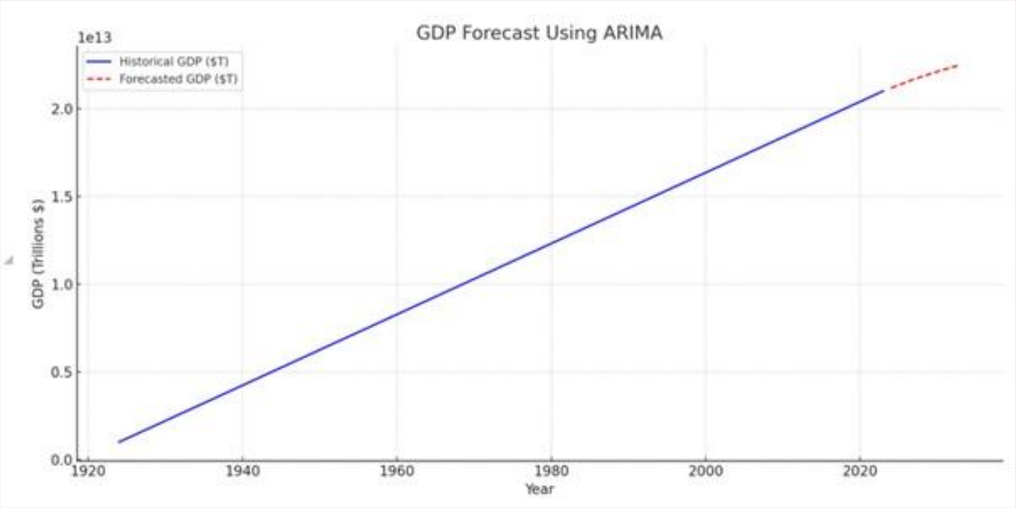

Figure 3: Growth Forecast using ARIMA Model.

The third case retained inflation and S&P 500 growth rates as they were at that time. This was the benchmark used to compare the other scenarios. According to the model, the rate of growth of the GDP at these conditions is moderate in relation to the present state of economic stability. This conceptual framework is important in reaching the basic context of how literally any divergence from the existing economic conditions such as from inflationary pressures to market forces affects GDP growth. This way, pointing to the base scenario, other real-life subjects can be compared with the baseline that will provide quite a clews of the possible impacts of different economic policies and market gyrations on future growths (Figure 3).

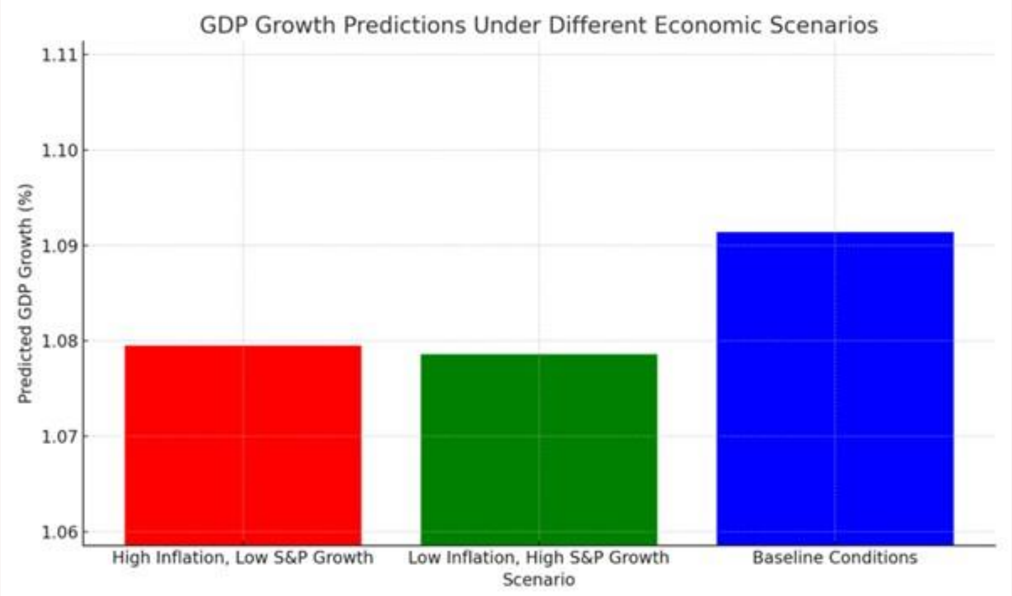

Figure 4: GDP prediction Under Different Economic Scenarios.

The figure 4 presents the forecast of GDP under different circumstances hence depicting a typical shift in anticipated economic performance under different inflation and S&P 500 growth conditions. Several outcomes of having current conditions provided the highest predicted GDP growth while other outcomes, specifically having a high inflation rate in the country gave the lowest predictions of GDP.

Evaluation

From the findings obtained by the different machine learning models of Linear Regression, Random Forest, and Gradient Boosting algorithms, the beauty of each of the models when used in the context of the GDP growth model has been highlighted. The criterion used to assess these models included forecasting performance, the ability to explain the results and the possibility of using these models in the estimates of economic variables.

Predictive Accuracy

These models, however, were primarily assessed based on their performance in predicting the values, with statistical measures such as the MAE, MSE, and even the R^2. In the present analysis the Random Forest model was identified as the best one with an R^2 of 0. percent of the data, which was equal to 9990. Explains 90 percent of GDP growth rates and is therefore widely recommended in the literature. This was succeeded by the Gradient Boosting model with the R^2 of 0. Linear Regression formula with calculated value of R^2 that is equal to 0.9979. While the general rule of thumb is that the R^2 value should range between 0 and 1, the R^2 values obtained in the present analysis are extremely high, and this is the case if we look at all the models that have been employed. But note, once again, that the differences in MAE and MSE are not significant between the Random Forest and Gradient Boosting models; still, the former model is a little better at capturing the complexities of the economic data.

The performance of the Random Forest model can also be deemed accurate due to the possibility of dealing with the non-linear functional form and variable interactions seldom to be ignored in economic data. Another advantage of the model which combines the results of several decision trees is that it can illustrate most of the patterns and tendencies that might be beyond the possibility of simpler models. This capability is critical in econometric modeling in which the relation between the variables is usually interactive and non linear as it is with inflation rates, stock markets and GDP.

Interpretability of Results

But it is also important that the results be interpretable, making the theoretical understanding of these models' crucial economic applications. Random forest model was quite effective in the current study especially because it offered straightforward ways of interpreting the feature importance. As it was observed in feature importance analysis, the most important variables were the three-month moving average of the closing value of S&P 500 (Close_MA3) and the first lag of the closing value of S&P 500 (Close_Lag1). This interpretability enables the policymakers to appreciate how different aspects motivate or determine economic occurrences thus improving on the model's forecast.

On the other hand, the Gradient Boosting model as it is highly accurate as the other models in the group is less interpretable because of the sequential learning process. The model's complexity also poses the question of interpretability as to which of the features are inducing the predictions. Such a trade-off between the goodness of fit of the model and easy interpretability of the equation is a common issue in the use of ML. Although the Gradient Boosting advised a higher accuracy to the model, the Random Forest had a fair measure of accuracy with interpretability needed for economical prediction.

Applicability in Economic Forecasting

Another source of assessment which was considered in this evaluation was the relevance and feasibility of these models to real world economic forecasting. The findings of the forecast of GDP growth in the scenario analysis have depicted the effectiveness of the Random Forest model that functions in different scenarios. The model could estimate a wide range of values of inflation and market growth, and the results were reasonable in all the cases. Such adaptability is highly relevant to the theory and practice of economic forecasting because conditions often shift significantly and chaotically.

In addition, another interesting feature that was highlighted by the scenario analysis was the ability of the model in policy-making. For instance, the low values of market growth and high inflation predicted by the model called for a decline in GDP growth: This could be useful for calling the attention of the policymakers to take preventive measures when there is likelihood of a downturn in business. On the other hand the certain GDP growth expected under low inflation/ high market growth justifies policies that seek to maintain low inflation rates while at the same time promoting market growth [13].

Model Limitations

There are clear limitations to models that put in such a good performance, however. The Random Forest model, though correct, is very computationally intensive, and as the size of the data increases and so does the size of the feature set that is fed into it. That makes it less useful, for example, for real-time forecasting where quick models are essential. Furthermore, because the model is based on past data, it is likely to be less accurate in forecasting a truly novel economic event or change, such as one induced by a large policy shift or a worldwide downturn [14].

Though it is always a good baseline, the Linear Regression model was easily surpassed by the more sophisticated models in this work. And its shortcomings, for managing non-linear relations and interactions, make it less appropriate for economic data. But despite this, its simplicity and interpretability can still render it a handy device where a more naïve approach will do.

Conclusion

This paper demonstrates how machine learning models, in particular Random Forest, can be highly accurate and interpretable GDP growth forecasts, much better than linear models. Because of its capacity to deal with complex, non-linear relationships and its provision of useful policy recommendations, the Random Forest model is an important method for economic prediction. Nevertheless, both the use of historical data, and the computational requirements of such models, suggest that they should be used with some caution. Machine learning nonetheless holds out the promise of improving economic analysis and policy-making in the face of such difficulty. Further work would consider the use of real-time information and mixed models to improve these estimates.

References

1. Min R (2021) Feasibility Study of Economic Forecasting Model based on Data Mining. In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI). IEEE. Available at Semantic Scholar, pp. 1438-1441.

2. Yu Y (2022) GDP Economic Forecasting Model Based on Improved RBF Neural Network. In Mathematical Problems in Engineering. Available at Semantic Scholar.

3. Gao K, Liu T, Hu B, Hao M, Zhang Y (2022) Establishment of Economic Forecasting Model of High-tech Industry Based on Genetic Optimization Neural Network. In Computational Intelligence and Neuroscience.

4. Liu E, Zhu H, Liu Q, Udimal TB (2022) Regional Economic Forecasting Method Based on Recurrent Neural Network. In Mathematical Problems in Engineering. Available at Semantic Scholar.

5. Khadri SW, Janamolla KR (2023) Fight against financial crimes – early detection and prevention of financial frauds in the financial sector with application of enhanced AI. IJARCCE 13(1): 59-64.

6. Goodfellow I, Pouget AJ, Mirza M, Xu B, Warde FD, et al. (2014) Generative adversarial nets. In Advances in neural information processing systems pp. 2672-2680.

7. Mohammed S (2024) AI-Driven Drug Discovery: Innovations and Challenges.

8. Kingma DP, Welling M (2013) Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114.

9. Stock JH, Watson MW (2001) Vector autoregressions. Journal of Economic Perspectives 15(4): 101-115

10. Thatikonda R, Dash B, Ansari MF, Vaddadi SA (2023) E-Business Trends and Challenges in the Modern Digital Enterprises in Asia. Digital Natives as a Disruptive Force in Asian Businesses and Societies p. 22-43.

11. Janamolla KR, Syed WK (2024) Global Banking Exploring Artificial Intelligence Role in Intelligent Banking to Automate Trading Platform. International Journal of Multidisciplinary Research and Publications (IJMRAP) 6(12): 163-168.

12. U.S Census Bureau (2023) Economic Indicators - Inflation and GDP. U.S. Department of Commerce.

13. Yahoo Finance (2024) Stock Market Data – Dow Jones, Nasdaq, S&P 500. Yahoo Finance.

14. Syed WK, Janamolla KR (2024) How AI-driven Robo-Advisors Impact Investment Decision making and Portfolio Performance in the Financial Sector: A Comprehensive Analysis.