Kokand University

RAQAMLI TEXNOLOGIYALARNING YANGI O‘ZBEKISTON RIVOJIGA TA’SIRI

Xalqaro ilmiy-amaliy konferensiyasi to'plami

21 IYUN

2023

kokanduni.uz

RAQAMLI TEXNOLOGIYALARNING YANGI O‘ZBEKISTON RIVOJIGA TA’SIRI

ВЛИЯНИЕ ЦИФРОВЫХ ТЕХНОЛОГИЙ НА РАЗВИТИЕ НОВОГО УЗБЕКИСТАНА

IMPACT OF DIGITAL TECHNOLOGIES ON THE DEVELOPMENT OF NEW UZBEKISTAN

Xalqaro ilmiy-amaliy konferensiyasi maqolalar to'plami

JUNE 21, 2023
KOKAND UNIVERSITY

"O‘zbekiston Respublikasi Oliy ta’lim tizimini 2030 yilgacha rivojlantirish konsepsiyasini tasdiqlash to‘g‘risida" O‘zbekiston Respublika Prezidentining 5847-sonli Farmonida ko‘zda tutilgan vazifalardan biri – ilmiy izlanish yutuklarini amaliyotga joriy etish yo‘li bilan fan sohalarini rivojlantirish, ya’ni xalqaro ilmiy hamjamiyatda e’tirof etilishiga xizmat qilishdir. Shu va boshqa tegishli farmonlarda va qarorlarda belgilangan vazifalarini amalga oshirish maqsadida 2023 yil 21-iyun kuni Qo‘qon universiteti “Raqamli texnologiyalar va matematika” kafedrasi “Raqamli texnologiyalarning Yangi O‘zbekiston rivojiga ta’siri” mavzusidagi xalqaro miqyosida o‘tkaziladigan ilmiy-amaliy konferensiyasi maqolalar to’plamini e‘lon qiladi

Ta'lim sifati yangi O‘zbekiston taraqqiyotini yanada yuksaltirishning muhim omili / Raqamli texnologiyalarning Yangi O‘zbekiston rivojiga ta’siri xalqaro ilmiy-amaliy konferensiyasi to’plami. Kokand university, 2023 yil 21 iyun, - «Innovatsion rivojlanish nashriyot-matbaa uyi» 2023.

FORECASTING GROSS DOMESTIC PRODUCT (GDP) AND GDP GROWTH: AN EXPLORATION OF IMPROVED PREDICTION USING MACHINE LEARNING ALGORITHMS

Azibaev Akhmadkhon Gulomjon ugli

PhD student of Namangan State University (Uzbekistan)

Abstract. This article explores the significance of Gross Domestic Product (GDP) and GDP growth, the importance of accurate forecasting, and the role of machine learning algorithms in improving prediction accuracy. It reviews several studies that highlight the effectiveness of machine learning algorithms, such as random forest regression, linear regression, and autoregressive integrated moving average (ARIMA), in GDP forecasting. These algorithms analyze data, identify patterns, and make accurate forecasts, contributing to enhanced decision-making in economic analysis and planning.

Keywords: Gross Domestic Product (GDP), GDP growth, machine learning algorithms, random forest regression, linear regression, autoregressive integrated moving average (ARIMA), forecasting, economic analysis, decision-making.

Introduction

Gross Domestic Product is the total monetary value of the all the finished goods and services produced within a country in a specific time period. [Dwarakanath G V, (2022)]. Calculating real GDP involves comparing the most recent year’s real GDP with the previous year’s real GDP and then dividing the difference by the prior year’s real GDP. Alternatively, real GDP can be derived by considering the nominal GDP and the prevailing inflation rate. Understanding GDP and its growth is essential for evaluating the strength and trajectory of an economy.

Today, nations around the world are actively striving to increase their GDP and GDP per capita, which is the ratio of total GDP to the population. This metric offers insights into the living standards of a country’s citizens. Consequently, GDP becomes a vital yardstick for policymakers, economists, and researchers when analyzing and predicting future economic levels. Researchers employ various methods, models, and tools to forecast GDP accurately. While traditional methods and models have been widely used, they often fall short of providing satisfactory results. As a result, more sophisticated and precise approaches have emerged. Among these is the utilization of machine learning algorithms (ML), a branch of artificial intelligence (AI) and data science (DS).

Machine learning algorithms offer a promising avenue for GDP prediction due to their ability to analyze vast amounts of data, identify patterns, and make accurate forecasts. By leveraging ML techniques, economists and researchers can incorporate a wide range of economic, financial, and social indicators into their models, enhancing the accuracy of GDP predictions. In this article, we will delve deeper into the significance of GDP and real GDP growth as key factors in assessing the health of an economy.

In their study, Richardson et al. aimed to improve nowcasts of real GDP growth in New Zealand by utilizing machine learning algorithms. They trained various popular ML algorithms using a large real-time dataset comprising approximately 550 New Zealand and international macroeconomic indicators. The authors compared the predictive accuracy of these nowcasts with several benchmarks, including autoregressive models, factor models, a large Bayesian VAR, and statistical models used at the Reserve Bank of New Zealand. Their findings indicated that machine learning algorithms outperformed the statistical benchmarks. Moreover, combining the nowcasts from different ML models further enhanced the overall performance, highlighting the gains in nowcasting accuracy achieved through the use of machine learning methods. [Adam Richardson,(2018)]

Gharte et al. investigated the use of machine learning techniques to improve the accuracy of GDP prediction. Analyzing various social, economic, and cultural parameters from 1970 to 2018, the authors built supervised learning models and compared the performance of three algorithms: Gradient Boosting, Random Forest, and Linear Regression. The study found that Gradient Boosting achieved the best prediction performance, followed by Random Forest and Linear Regression. The authors also developed a web application that estimated and forecasted the GDP of a country based on input attributes, highlighting the potential of machine learning techniques in GDP analysis and prediction. [Tanvi Gharte, (2022)]

Maccarrone et al. compared different models for forecasting the real U.S. GDP. Using quarterly data from 1976 to 2020, they found that the machine learning K-Nearest Neighbour (KNN) model outperformed traditional time series analysis in capturing the self-predictive ability of the U.S. GDP. The authors explored the inclusion of predictors such as the yield curve, its latent factors, and macroeconomic variables to enhance forecasting accuracy, observing improved predictions for longer forecast horizons. The study highlighted the additional guidance provided by machine learning algorithms for data-driven decision-making. [Giovanni Maccarrone, (2021)]

Wang et al. focused on the application of emotion recognition algorithms in analyzing and predicting financial market trends and economic growth. They highlighted the complexity of the financial market and economic growth as a highly intricate system and emphasized the need for accurate prediction results. The authors provided a detailed overview of existing financial development and economic growth forecasting issues, along with an introduction to emotion recognition algorithms. They delved into statistical emotion recognition methods, mixed emotion recognition methods, and emotion recognition methods based on knowledge technology. The study conducted in-depth research on three algorithm models, including the support vector machine algorithm model, artificial neural network algorithm model, and long and short-term memory network algorithm model. [Dahai Wang, (2022)]

Claveria et al. introduced a sentiment construction method based on the evolution of survey-based indicators using genetic algorithms. The study aimed to generate country-specific empirical economic sentiment indicators in the Baltic republics and the European Union. By searching for the optimal non-linear combination of firms' and households' expectations, the authors computed the frequency distribution of survey expectations and examined the lag structure per selected variable. The study evaluated the out-of-sample predictive performance of the generated indicators, demonstrating more accurate estimates of year-on-year GDP growth rates compared to scaled industrial and consumer confidence indicators. The authors further combined the evolved expectations of firms and consumers using non-linear constrained optimization to generate aggregate expectations of year-on-year GDP growth, which outperformed recursive autoregressive predictions of economic growth. [Oscar CLAVERIA, (2021)]

Overall, the reviewed studies emphasize the potential of machine learning algorithms, genetic algorithms, and emotion recognition algorithms in improving GDP prediction and economic forecasting accuracy. These innovative approaches offer new insights and methodologies for researchers and policymakers seeking more precise and reliable predictions of economic growth. By leveraging the power of advanced computational techniques, these studies contribute to the growing body of literature on machine learning-based forecasting methods, highlighting their superiority over traditional statistical models and the potential for enhancing decision-making in economic analysis and planning.

Basic 3 ML algorithms which provide GDP forecasting

Linear Regression: In statistics linear regression is a linier approach for modelling the relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linier regression; for more than one, the process is called multiple linier regression**.** This term is distinct from multivariate linier regression, where multiple corrected dependent variables are predicted, rather than a single scalar variable. In linear regression, the relationships are modeled using linier predictor functions whose unknown model parameters are estimated from the data. [18]

The linear predictor function is represented as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$$

[18] https://en.wikipedia.org/wiki/Talk:Linear_regression

where:

Y – is the dependent variable

$X_1, X_2, \dots, X_n$ – are the independent variables

$\beta_0, \beta_1, \beta_2, \dots, \beta_n$ – are the coefficients that represent the relationship between the variables

ε – is the error term that captures the deviations between the observed values and the predicted values

The goal of linear regression is to estimate the values of the coefficients based on the available data. This estimation is typically done using methods such as ordinary least squares (OLS), which minimizes the sum of squared differences between the observed values and the predicted values. Linear regression is widely used in various fields for tasks such as prediction, forecasting, and understanding the relationship between variables. It provides interpretable coefficients that allow for assessing the impact and significance of each independent variable on the dependent variable.

Random Forest Regression: **RFR** is a supervised learning algorithm that uses **ensemble learning** method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.[19]

In random forest regression, the ensemble is composed of multiple decision trees. Each decision tree is constructed using a subset of the training data and a random selection of features. The random selection of features helps to introduce diversity among the trees and reduces overfitting.

The key steps in building a random forest regression model are as follows:

- Data Preparation: Prepare the training data with a set of independent variables and their corresponding dependent variable.
- Random Sampling: Randomly select subsets of the training data (with replacement) to create multiple subsets, known as bootstrap samples. Each bootstrap sample is used to train an individual decision tree.
- Decision Tree Construction: For each bootstrap sample, construct a decision tree by recursively splitting the data based on the selected features. The splits are determined using a criterion such as the Gini index or information gain.

[19] https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

- Ensemble Generation: Generate an ensemble by combining the predictions from all the individual decision trees. In random forest regression, the predictions are typically averaged or aggregated to obtain the final prediction.
- Prediction: Use the trained random forest model to make predictions on new data by aggregating the predictions from all the individual decision trees.

Random forest regression is commonly used for various applications, including prediction, forecasting, and feature selection. However, it is important to tune the hyperparameters of the random forest model, such as the number of trees, maximum depth, and minimum samples per leaf, to achieve optimal performance.

Auto Regressive Integrated Moving Average: An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. A statistical model is autoregressive if it predicts future values based on past values. For example, an ARIMA model might seek to predict a stock's future prices based on its past performance or forecast a company's earnings based on past periods.[20]

The name "ARIMA" reflects the three components of the model: autoregressive (AR), integrated (I), and moving average (MA).

Autoregressive (AR): The autoregressive component refers to the relationship between the current value of a variable and its past values. It assumes that the future values of the variable can be predicted based on a linear combination of its own past values.

Integrated (I): The integrated component deals with the differencing of the time series data to make it stationary. Stationarity means that the statistical properties of the time series, such as mean and variance, do not change over time. Differencing involves taking the difference between consecutive observations to remove trends or seasonality.

Moving Average (MA): The moving average component considers the dependency between the error term and past errors. It assumes that the future values of the variable can be predicted based on a linear combination of past error terms.

ARIMA models have been widely used in various fields, including finance, economics, and environmental sciences, to analyze and predict time series data. However, it's important to note

20https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average arima.asp#:~:text=An%20autoregressive%20integrated%20moving%20average%2C%20or%20ARIMA%2C%20is%20a%20statistical,values%20based%20on%20past%20values.

that ARIMA assumes linearity, stationarity, and absence of outliers, and the model's performance depends on the quality and characteristics of the data being analyzed.

Abbreviations

- **AI** – Artificial Intelligence
- **ARIMA –** Auto Regressive Integrated Moving Average
- **DS** – Data Science
- **GDP** – Gross Domestic Product
- **ML** – Machine Learning
- **OLS –** Ordinary Least Squares
- **RFR -** Random Forest Regression

REFERENCES:

1. Adam Richardson, T. v. (2018). Nowcasting New Zealand GDP using machine learning algorithms. Big Data for Central Bank Policies / Building Pathways for Policy Making with Big Data. Bali, Indonesia: Irving Fisher Committee on Central Bank Statistics .

2. Dahai Wang, B. L. (2022). Emotion Recognition Algorithm Application Financial Development and Economic Growth Status and Development Trend. Frontiers in Psychology.

3. Dwarakanath G V, S. T. (2022). WORLD'S GDP PREDICTION USING MACHINE LEARNING. Journal of Positive School Psychology, 1039-1047.

4. Giovanni Maccarrone, G. M. (2021). GDP Forecasting: Machine Learning, Linear or Autoregression? Frontiers in Artificial Intellegence.

5. Li, Z. (2022). Neural network economic forecast method based on genetic algorithm. The Institution Engineering and Technology .

6. Oscar CLAVERIA, E. M. (2021). A GENETIC PROGRAMMING APPROACH FOR ESTIMATING ECONOMIC SENTIMENT IN THE BALTIC COUNTRIES AND THE EUROPEAN UNION. Technological and Economic Development of Economy, 262–279.

7. Tanvi Gharte, H. P. (2022). GDP Prediction and Forecasting using Machine Learning . International Research Journal of Engineering and Technology, 2908-2912.