Economics and Financial Policymaking

J. Econ. & Financ. Policymaking., 2(4) (2025) 1-14
Journal Homepage: https://journalefp.com/

ORGINAL RESEARCH PAPER

A Comparative Performance Evaluation of Econometric and Machine Learning Models in Forecasting Iran's Economic Growth

Ahmad Saradari, Vida Varahrami*

Department of Economics, Faculty of Economics and Political Science, Shahid Beheshti University, Tehran, Iran.
**. Corresponding Author's Email: v_varahrami@sbu.ac.ir*

PAPER INFO

Paper history:
Received: 27 September 2025
Revised: 01 November 2025
Accepted in revised form: 10 November 2025
Published: 21 December 2025

Keywords:
Economic Growth
Forecasting
Machine Learning
Random Forest
Autoregressive Distributed Lag (ARDL)

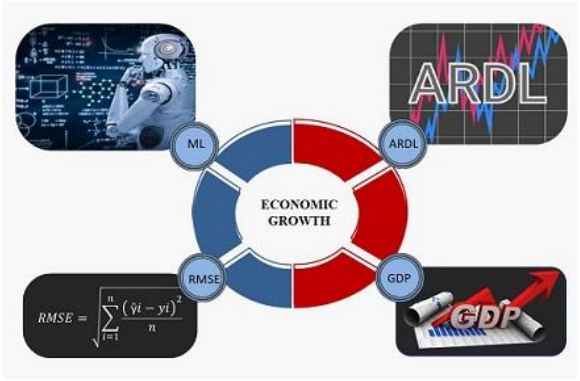
How to cite: Saradari, A., & Varahami, V. (2025). A comparative performance evaluation of econometric and machine learning models in forecasting iran's economic growth. *Economics and Financial Policymaking, 2*(4), 1-14.

©2025 the authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

Forecasting economic growth in Iran's volatile economy has persistently been a challenge. Aiming to identify more effective tools, this study conducts a comparative evaluation of two distinct paradigms: traditional econometrics and modern machine learning. Using annual time-series data from 1991 to 2023, an optimized Autoregressive Distributed Lag (ARDL) model was benchmarked against two advanced algorithms, Random Forest and XGBoost. The models were evaluated on a test set (2016-2023) using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) as performance metrics. The results indicate that machine learning models exhibit superior predictive performance, with Random Forest emerging as the most accurate model. Furthermore, feature importance analysis from this model revealed the key role of variables such as government expenditure, suggesting the presence of significant non-linear relationships overlooked by the linear model. The findings underscore the complementary nature of the two approaches: econometrics for interpretation and machine learning for precise forecasting.

.



1. Introduction

Despite the widespread application of both traditional econometric methods and machine learning techniques in economic forecasting across the world, a critical research gap remains in the specific context of Iran's economy. Existing studies have largely focused on either conventional, theory-driven econometric models or data-driven machine learning approaches in isolation. However, there is a notable absence of a systematic, comprehensive, and evidence-based comparison that directly evaluates the relative performance of these two fundamentally different paradigms in forecasting Iran's economic growth. Given the structural volatility, frequent policy shifts, and high levels of uncertainty characterizing Iran's macroeconomic environment, this gap is particularly consequential for both researchers and policymakers.

This unresolved issue raises an important and policy-relevant question: are traditional and interpretable econometric models—rooted in economic theory and linear relationships—sufficient for accurately forecasting economic growth in such a complex and unstable economy? Or can more flexible, non-linear, and computationally intensive machine learning models provide superior forecasting performance and, consequently, more effective tools for economic policymaking?

The primary innovation and contribution of this study lie precisely in filling this gap in the literature. For the first

time, this paper conducts a direct, fair, and methodologically rigorous comparison between the best-performing traditional econometric model and a set of state-of-the-art machine learning models in the context of Iran's economic growth forecasting. Specifically, the econometric benchmark is an optimized Autoregressive Distributed Lag (ARDL) model, selected through a careful and systematic optimization process to ensure robustness and comparability. This model is then evaluated against advanced machine learning algorithms, namely Random Forest and Extreme Gradient Boosting, which are well known for their ability to capture complex non-linear relationships and interaction effects among variables.

Within this comparative framework, the study seeks to answer several interrelated research questions. First, based on the optimized traditional econometric model, what are the most important linear and theoretically interpretable determinants of Iran's economic growth? Second, how does the out-of-sample forecasting performance of advanced machine learning models compare with that of the traditional econometric approach, and which paradigm yields lower forecasting errors? Finally, do machine learning models uncover additional influential variables or non-linear channels that remain hidden within linear econometric specifications?

Accordingly, the main objective of this research is to provide a comprehensive comparative evaluation of traditional econometric methods and modern machine learning techniques in forecasting Iran's economic growth. By identifying the most accurate and efficient forecasting tool, this study aims not only to contribute to the academic literature but also to offer practical insights for policymakers seeking reliable and timely forecasts in a highly volatile economic environment.

2. Methodology

In this study, annual time-series data for Iran's economy over the period 1991-2023 were used to model economic growth. The dependent variable is real GDP growth, and the explanatory variables were selected based on theoretical foundations of economic growth and previous empirical evidence. These variables include monetary policy indicators (deposit interest rate, monetary base, and inflation rate), fiscal policy indicators (government expenditures and revenues), structural and institutional variables (investment, unemployment rate, and degree of economic freedom), and external sector variables (exchange rate and trade openness).

3. Findings

With the aim of identifying a more efficient and reliable tool for forecasting economic growth in Iran's highly volatile and uncertain economic environment, this research undertakes a systematic comparative evaluation of two fundamentally different modeling paradigms: the traditional econometric approach and the modern machine learning approach. Accurate macroeconomic forecasting has long been recognized as a challenging task, particularly in economies characterized by frequent structural breaks, policy regime changes, external shocks, and high levels of uncertainty. As emphasized by Stock and Watson (2007), these features significantly undermine the predictive power of conventional models and necessitate continuous methodological innovation.

In the context of Iran's economy, these challenges are further intensified by persistent inflationary pressures, fluctuations in oil revenues, international sanctions, and institutional constraints, all of which complicate the dynamics of economic growth. Consequently, relying on a single forecasting framework may lead to biased or inefficient predictions. This study responds to this challenge by juxtaposing a rigorously optimized traditional econometric model with state-of-the-art machine learning techniques, thereby providing a comprehensive assessment of their relative strengths and limitations.

Specifically, an optimized Autoregressive Distributed Lag (ARDL) model is employed as the benchmark econometric framework. The ARDL approach is particularly suitable for macroeconomic analysis due to its flexibility in handling variables with different orders of integration and its strong grounding in economic theory, which facilitates interpretation and policy analysis. The model is carefully optimized through a systematic lag-selection and specification process to ensure robustness, parsimony, and comparability.

This optimized econometric model is then compared with two advanced machine learning models—Random Forest and Extreme Gradient Boosting—which represent powerful, data-driven approaches capable of capturing complex non-linear relationships and interaction effects that may remain hidden in linear specifications. These models have gained increasing attention in the economic forecasting literature due to their superior predictive performance, especially in environments characterized by structural instability and non-linear dynamics.

By conducting a direct and fair comparison between these competing paradigms under a unified empirical framework, this study seeks not only to evaluate their forecasting accuracy but also to shed light on the trade-off between interpretability and predictive performance. In doing so, the research contributes to the growing literature at the intersection of econometrics and machine learning, while offering practical insights into the selection of forecasting tools for policymakers operating in highly volatile economic settings such as Iran.

4. Conclusion and Implications

This study aims to identify a more efficient tool for forecasting economic growth in Iran's highly volatile economy by conducting a comparative evaluation of two modeling paradigms: the traditional econometric approach and the modern machine learning approach. Due to uncertainty and structural changes, macroeconomic forecasting has always been a challenging task (Stock and Watson, 2007). Accordingly, an optimized Autoregressive Distributed Lag (ARDL) model was compared with two advanced machine learning models—Random Forest and Extreme Gradient Boosting.

The analysis of the results led to three main findings. First, the econometric model showed that Iran's economic growth is influenced by interpretable factors such as base

money growth, the degree of economic freedom, and inflation acceleration. Second, in forecasting performance, machine learning models—particularly the Random Forest model—outperformed others by recording the lowest prediction error; this finding is consistent with recent studies in the field (Koop et al., 2021). Finally, a comparison of variable importance revealed that the machine learning model identified "government expenditure" as a key factor, whereas this variable was not statistically significant in the linear model, strongly indicating the presence of nonlinear transmission channels.

Our findings suggest that these two approaches are complementary to each other, as argued by Athey (2018), who emphasizes that modern economics requires new tools to analyze complex relationships. While the econometric model provides a valuable framework for theoretical interpretation, the Random Forest model—by sacrificing some degree of interpretability—achieves much higher predictive accuracy and uncovers hidden nonlinear patterns. Therefore, the final conclusion is that for comprehensive analysis and optimal policymaking, both tools should be used in a combined manner.

Based on the robust findings of the study, the following policy recommendations are proposed. First, given the high importance of the "degree of economic freedom" in both approaches, institutional reforms and improvements in the business environment should be prioritized. Second, the nonlinear importance of "government expenditure" in the machine learning model indicates that fiscal policy should focus on increasing efficiency, transparency, and the optimal allocation of resources toward productive spending. Finally, in light of the econometric model's findings, maintaining macroeconomic stability through sustained inflation control and avoiding abrupt changes in monetary policy should serve as a key anchor for the country's economic policymaking.

Conflict of Interest

The Authors, while observing publication ethics in the referencing, declare the absence of interest of conflict.

نشریه علمی اقتصاد و سیاست‌گذاری مالی

نشریه علمی اقتصاد و سیاست‌گذاری مالی، دوره ۲، شماره ۴، زمستان ۱۴۰۴، صفحات ۱ تا ۱۴
(مقاله پژوهشی)
Journal Homepage: https://journalefp.com/

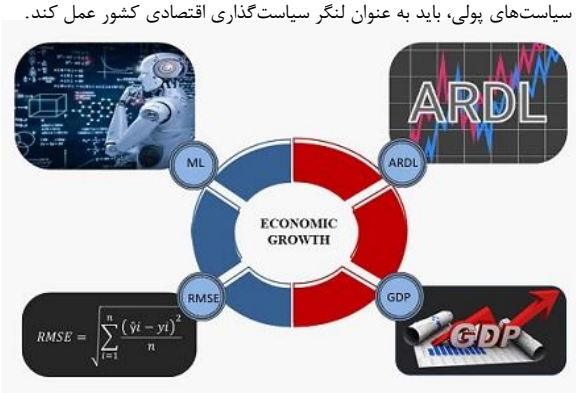
ارزیابی تطبیقی عملکرد مدل‌های اقتصادسنجی و یادگیری ماشین در پیش‌بینی رشد اقتصادی ایران

احمد سراداری ، **ویدا ورهرامی***

گروه اقتصاد، دانشکده علوم اقتصادی و سیاسی، دانشگاه شهید بهشتی، تهران، ایران.
*. ایمیل نویسنده مسئول مکاتبه v_varaharmi@sbu.ac.ir

چکیده

پیش‌بینی رشد اقتصادی در اقتصاد پرنوسان ایران همواره چالش‌برانگیز بوده است. این پژوهش با هدف یافتن ابزاری کارآمدتر، به ارزیابی تطبیقی دو پارادایم اقتصادسنجی سنتی و یادگیری ماشین می‌پردازد. با استفاده از داده‌های سالانه، یک مدل بهینه خودرگرسیون با وقفه‌های گسترده با دو الگوریتم پیشرفته جنگل تصادفی و تقویت گرادیان شدید مقایسه شد. ارزیابی مدل‌ها بر روی داده‌های آزمون با معیارهای ریشه میانگین مربعات خطا و میانگین قدر مطلق خطا صورت گرفت. نتایج نشان داد که مدل‌های یادگیری ماشین عملکرد پیش‌بینی بهتری دارند و جنگل تصادفی دقیق‌ترین مدل است. تحلیل اهمیت ویژگی‌ها در این مدل، نقش کلیدی متغیرهایی مانند مخارج دولت را آشکار ساخت که حاکی از وجود روابط غیرخطی مهمی است که توسط مدل خطی نادیده گرفته شده بود. یافته‌ها بر مکملیت دو رویکرد اقتصادسنجی برای تفسیر و یادگیری ماشین برای پیش‌بینی دقیق تأکید دارند. لذا حفظ ثبات اقتصاد کلان از طریق کنترل پایدار تورم و پرهیز از تغییرات ناگهانی در سیاست‌های پولی، باید به عنوان لنگر سیاست‌گذاری اقتصادی کشور عمل کند.



اطلاعات مقاله

تاریخچه مقاله:
تاریخ دریافت: ۱۴۰۴/۰۷/۰۵
تاریخ بازنگری: ۱۴۰۴/۰۸/۱۰
تاریخ پذیرش: ۱۴۰۴/۰۸/۱۹
تاریخ انتشار: ۱۴۰۴/۰۹/۳۰

کلیدواژگان:
رشد اقتصادی
پیش‌بینی
یادگیری ماشین
جنگل تصادفی
مدل خودرگرسیون با وقفه‌های گسترده

شیوه استناددهی:
سراداری، احمد، و ورهرامی، ویدا. (۲۰۲۵). ارزیابی تطبیقی عملکرد مدل‌های اقتصادسنجی و یادگیری ماشین در پیش‌بینی رشد اقتصادی ایران، *اقتصاد و سیاست‌گذاری مالی*، ۲ (۴)، ۱-۱۴.

© ۱۴۰۴ تمامی حقوق انتشار این مقاله متعلق به نویسنده است. انتشار این مقاله به‌صورت دسترسی آزاد مطابق گواهی (CC BY-NC 4.0) صورت گرفته است.

۱. مقدمه

رشد اقتصادی به عنوان یکی از محوری‌ترین شاخص‌های توسعه و رفاه جوامع، همواره در کانون توجه سیاست‌گذاران و اقتصاددانان قرار داشته است. رشد پایدار اقتصادی نه تنها به بهبود سطح زندگی، کاهش فقر و افزایش اشتغال منجر می‌شود، بلکه ثبات اجتماعی و سیاسی را نیز تقویت می‌کند. در این میان، توانایی پیش‌بینی دقیق مسیر آتی رشد اقتصادی، به عنوان یک ابزار حیاتی برای برنامه‌ریزی‌های کلان، تخصیص بهینه منابع، مدیریت ریسک‌های تجاری و اتخاذ سیاست‌های پولی و مالی پیشگیرانه، از اهمیت فوق‌العاده‌ای برخوردار است.

اقتصاد ایران با ویژگی‌های منحصربه‌فردی همچون وابستگی به درآمدهای نفتی، مواجهه با تحریم‌های بین‌المللی و چالش‌های ساختاری داخلی، همواره نوسانات قابل توجهی را در مسیر رشد اقتصادی خود تجربه کرده است. این پیچیدگی‌ها، وظیفه خطیر پیش‌بینی را برای تحلیلگران دشوارتر ساخته و نیاز به استفاده از مدل‌های دقیق و کارآمد را دوچندان می‌کند. یک پیش‌بینی نادرست می‌تواند به اتخاذ سیاست‌های ناکارآمد و تخصیص نادرست منابع کمیاب منجر شود و هزینه‌های اقتصادی و اجتماعی قابل توجهی را به کشور تحمیل نماید.

به طور سنتی، تحلیل‌ها و پیش‌بینی‌های اقتصادی در ایران عمدتاً بر پایه مدل‌های اقتصادسنجی ساختاری مانند مدل خودرگرسیون با وقفه‌های گسترده[1] استوار بوده است. این مدل‌ها، که توسط اقتصاددانان برجسته‌ای چون پسران، شین و اسمیت[2] (۲۰۰۱) توسعه یافته‌اند، به دلیل پایه‌های نظری قوی و قابلیت تفسیر بالای ضرایب، ابزارهای ارزشمندی محسوب می‌شوند. با این حال، این مدل‌ها اغلب بر فرضیات محدودکننده‌ای مانند خطی بودن روابط بین متغیرها استوار هستند.

در دهه‌های اخیر، با پیشرفت علم داده، پارادایم جدیدی تحت عنوان یادگیری ماشین ظهور کرده است. مدل‌های یادگیری ماشین، مانند جنگل تصادفی[3] که توسط برایمن[4] (۲۰۰۱) معرفی شد، توانایی فوق‌العاده‌ای در کشف الگوهای پیچیده و غیرخطی در داده‌ها بدون نیاز به فرضیات سخت‌گیرانه آماری دارند. همان‌طور که واریان[5] (۲۰۱۴) اشاره می‌کند، این مدل‌ها به خصوص در کاربردهای پیش‌بینی، پتانسیل دستیابی به دقت‌های بسیار بالاتری را نسبت به مدل‌های سنتی دارند.

با وجود کاربرد گسترده هر دو رویکرد در سطح جهان، شکاف تحقیقاتی اصلی در زمینه اقتصاد ایران، فقدان یک مقایسه نظام‌مند، جامع و مبتنی بر شواهد بین عملکرد این دو پارادایم متفاوت در پیش‌بینی رشد اقتصادی است. آیا مدل‌های سنتی و قابل تفسیر اقتصادسنجی برای پیش‌بینی کافی هستند، یا مدل‌های پیچیده‌تر و دقیق‌تر یادگیری ماشین می‌توانند ابزارهای بهتری در اختیار سیاست‌گذار قرار دهند؟ نوآوری اصلی این پژوهش، پر کردن همین شکاف تحقیقاتی است. این مقاله برای اولین بار، یک مقایسه مستقیم و عادلانه بین بهترین مدل اقتصادسنجی سنتی (که پس از یک فرآیند دقیق بهینه‌سازی به دست آمده) و مجموعه‌ای از مدل‌های پیشرفته یادگیری ماشین[6] (جنگل تصادفی و تقویت گرادیان شدید[7]) در پیش‌بینی رشد اقتصادی ایران انجام می‌دهد.

این پژوهش در پی پاسخ به؛ براساس یک مدل بهینه اقتصادسنجی سنتی، مهم‌ترین عوامل خطی و قابل تفسیر مؤثر بر رشد اقتصادی ایران کدام‌اند؟ و عملکرد پیش‌بینی مدل‌های پیشرفته یادگیری ماشین در مقایسه با مدل اقتصادسنجی سنتی چگونه است و کدام رویکرد خطای کمتری دارد؟ می‌باشد. براین اساس، هدف اصلی تحقیق، ارزیابی تطبیقی دو رویکرد اقتصادسنجی سنتی و یادگیری ماشین در پیش‌بینی رشد اقتصادی ایران به منظور شناسایی دقیق‌ترین و کارآمدترین ابزار برای این وظیفه خطیر است.

برای دستیابی به اهداف فوق، مقاله در چهار بخش سازماندهی شده است. پس از این مقدمه، بخش دوم به مرور مبانی نظری و پیشینه تجربی پژوهش در هر دو حوزه اقتصادسنجی و یادگیری ماشین می‌پردازد. بخش سوم به طور کامل روش‌شناسی تحقیق را تشریح می‌کند؛ این بخش شامل معرفی و آماده‌سازی داده‌ها، فرآیند ساخت و بهینه‌سازی مدل اقتصادسنجی و در نهایت، چارچوب آموزش و ارزیابی مدل‌های یادگیری ماشین است و به ارائه و تحلیل نتایج به دست آمده از هر دو رویکرد اختصاص دارد و در بخش پنجم، ضمن مقایسه نهایی مدل‌ها، جمع‌بندی کلی، نتیجه‌گیری و پیشنهادهای سیاستی و پژوهشی ارائه می‌گردد.

۲. ادبیات موضوع

در این بخش، ابتدا مبانی نظری مرتبط با عوامل تعیین‌کننده رشد اقتصادی تشریح شده و سپس، به مرور و تحلیل مطالعات تجربی انجام شده در سطح بین‌المللی و داخلی پرداخته می‌شود تا جایگاه و نوآوری این پژوهش در مقایسه با ادبیات موضوع به روشنی مشخص گردد.

۲-۱. مبانی نظری

نظریه‌های رشد اقتصادی، از دیدگاه نئوکلاسیک تا مدل‌های نوین رشد درون‌زا، با معرفی پیشرانه‌های اصلی رشد، سنگ بنای تحلیلی و چارچوب مفهومی این پژوهش را تشکیل می‌دهند. در ادامه مهم‌ترین دیدگاه‌ها تشریح می‌شود:

۱. دیدگاه نئوکلاسیک: این دیدگاه (سولو[8]، ۱۹۵۶) رشد اقتصادی را تابعی از سه عامل انباشت سرمایه فیزیکی، نیروی کار و پیشرفت فنی برون‌زا می‌داند. در این چارچوب، سرمایه‌گذاری از طریق افزایش سرمایه به ازای هر کارگر، بهره‌وری و تولید را افزایش می‌دهد و به عنوان یک عامل کلیدی در رشد کوتاه‌مدت و میان‌مدت شناخته می‌شود.

۲. دیدگاه رشد درون‌زا: نظریه‌های رشد درون‌زا (رومر[9]، ۱۹۹۰)، برخلاف دیدگاه سنتی، پیشرفت فنی را عاملی درون‌زا و متأثر از سرمایه انسانی، نوآوری و سیاست‌های دولت می‌دانند. در این دیدگاه، سیاست‌های مالی (مخارج و درآمدهای دولت) می‌توانند نقشی دوگانه (مثبت یا منفی) در رشد ایفا کنند. همچنین، ثبات اقتصاد کلان نقشی حیاتی دارد؛ به طوری که تورم پایین و قابل

1. Autoregressive Distributed Lag (ARDL) Model
2. Pesaran, Shin, and Smith
3. Random Forest
4. Breiman
5. Varian
6. Machine Learning
7. XG-Boost
8. Solow
9. Romer

پیش‌بینی، پیش‌نیاز اصلی برای تشویق سرمایه‌گذاری و رشد پایدار است (میشکین[1]، ۲۰۰۷).

۳. نهادها و محیط کسب‌وکار: مکتب نهادگرایی (نورث[2]، ۱۹۹۰) بر نقش بنیادین نهادها یا «قواعد بازی» در عملکرد اقتصادی تأکید دارد. نهادهای کارآمد، مانند تضمین حقوق مالکیت و حاکمیت قانون که در شاخص آزادی اقتصادی تبلور می‌یابند، با کاهش هزینه‌های مبادله و افزایش بازدهی، انگیزه سرمایه‌گذاری و نوآوری را تقویت می‌کنند.

۴. تجارت و تعامل با اقتصاد جهانی: بر اساس نظریه‌های تجارت، تعامل با اقتصاد جهانی از طریق کانال‌هایی چون سرریز فناوری و افزایش رقابت، به رشد شتاب می‌بخشد. نرخ ارز نیز به عنوان یک قیمت کلیدی، بر رقابت‌پذیری تجاری و رشد اقتصادی تأثیرگذار است.

در مجموع، چارچوب مفهومی این پژوهش مبتنی بر این فرضیه است که رشد اقتصادی ایران تابعی از متغیرهای منتخب در حوزه انباشت سرمایه، سیاست‌های پولی و مالی، کیفیت نهادی و تعامل با اقتصاد جهانی است. این چارچوب، مبنای انتخاب متغیرها در بخش روش‌شناسی قرار گرفته است.

۲-۲. مروری بر مطالعات تجربی خارجی

دوپکه و همکاران[3] (۲۰۲۳) در مقاله‌ای با عنوان «پیش‌بینی تولید ناخالص داخلی آلمان با استفاده از شبکه‌های حافظه طولانی کوتاه‌مدت»، به بررسی قابلیت‌های مدل‌های یادگیری عمیق در مقایسه با مدل‌های اقتصادسنجی سری زمانی پرداختند. یافته‌های آن‌ها نشان داد که مدل‌های یادگیری عمیق، به ویژه در شناسایی نقاط عطف چرخه تجاری (شروع رکود یا رونق)، به طور معناداری از مدل‌های خودرگرسیون برداری عملکرد بهتری دارند.

شارپنتیه و همکاران[4] (۲۰۲۲) در یک مقاله مروری جامع با عنوان «یادگیری ماشین در اقتصاد کلان: یک پیمایش»، به بررسی نظام‌مند کاربردهای الگوریتم‌های نوین در پیش‌بینی متغیرهای کلان پرداختند. آن‌ها نتیجه گرفتند که برتری اصلی یادگیری ماشین در مدیریت حجم بالای داده‌ها (ابعاد بالا) و کشف روابط غیرخطی[5] پیچیده است، در حالی که مدل‌های اقتصادسنجی همچنان برای تحلیل‌های ساختاری و تفسیر روابط علی ارزشمند هستند.

مدیروس و واسکونسلوس[6] (۲۰۲۲) در تحقیقی با عنوان «یادگیری ماشین برای پیش‌بینی اقتصاد کلان: پیش‌بینی رکودها»، به بررسی این موضوع پرداختند که آیا می‌توان از این الگوریتم‌ها برای پیش‌بینی رکودهای اقتصادی در ایالات متحده استفاده کرد. آن‌ها نشان دادند که مدل‌های یادگیری ماشین نه تنها در پیش‌بینی زمان وقوع رکود دقیق‌تر از مدل‌های سنتی عمل می‌کنند، بلکه با استفاده از تکنیک‌های تفسیرپذیری، متغیرهای کلیدی هشداردهنده بحران را نیز شناسایی می‌نمایند.

گولت[7] (۲۰۲۱) در یک گزارش کاری برای بانک مرکزی کانادا با عنوان «پیش‌بینی اقتصاد کلان با یادگیری ماشین: یک پیمایش و یک کاربرد برای کانادا»، عملکرد طیف وسیعی از مدل‌ها را مقایسه کرد. این مطالعه نشان داد که مدل‌های جنگل تصادفی و تقویت گرادیان به طور مداوم پیش‌بینی‌های دقیق‌تری برای رشد تولید ناخالص داخلی و تورم نسبت به مدل‌های اقتصادسنجی استاندارد ارائه می‌دهند، به خصوص زمانی که با داده‌های آنی تغذیه شوند.

کوپ و همکاران[8] (۲۰۲۱) در پژوهشی برای بانک مرکزی اروپا با عنوان «پیش‌بینی تولید ناخالص داخلی با یادگیری ماشین: چشم‌انداز منطقه یورو»، به مقایسه گسترده عملکرد مدل‌های اقتصادسنجی و الگوریتم‌های یادگیری ماشین پرداختند. نتایج آن‌ها نشان داد که در افق‌های پیش‌بینی کوتاه‌مدت، مدل‌های یادگیری ماشین به دلیل توانایی در شناسایی روابط غیرخطی در داده‌های با فرکانس بالا، خطای کمتری نسبت به مدل‌های سنتی دارند.

گو و همکاران[9] (۲۰۲۰) در مقاله‌ای بسیار تأثیرگذار با عنوان «قیمت‌گذاری تجربی دارایی از طریق یادگیری ماشین»، نشان دادند که چگونه می‌توان از الگوریتم‌های یادگیری ماشین برای مدیریت و استخراج اطلاعات از مجموعه داده‌های بسیار بزرگ (صدها متغیر) استفاده کرد. اگرچه موضوع اصلی آن‌ها بازارهای مالی است، اما روش‌شناسی آن‌ها در پیش‌بینی‌های اقتصاد کلان نیز به طور گسترده به کار گرفته شده و برتری این رویکرد را در محیط‌های داده-محور نشان می‌دهد.

۲-۳. مروری بر مطالعات تجربی داخلی

فلاحی و همکاران (۱۴۰۲) در مقاله‌ای با عنوان «پیش‌بینی رشد اقتصادی ایران با استفاده از الگوریتم‌های یادگیری ماشین: یک تحلیل مقایسه‌ای»، عملکرد سه مدل ماشین بردار پشتیبان، جنگل تصادفی و شبکه عصبی مصنوعی را با مدل اقتصادسنجی

1. Mishkin
2. North
3. Döpke et al.
4. Charpentier et al.
5. Non-linear Relationships
6. Medeiros & Vasconcelos
7. Goulet
8. Koop et al.
9. Gu et al.

خودرگرسیون برداری مقایسه کردند. نتایج آن‌ها نشان داد که مدل جنگل تصادفی به دلیل قابلیت بالا در مدیریت روابط غیرخطی و انتخاب متغیر، کمترین خطای پیش‌بینی را در دوره آزمون داشته است.

متوسلی و شهنازی (۱۴۰۱) در تحقیقی با عنوان «آیا یادگیری ماشین ابزار بهتری برای پیش‌بینی تورم در ایران است؟»، به مقایسه مدل‌های سنتی سری زمانی (مانند مدل خودرگرسیو میانگین متحرک انباشته) با الگوریتم‌های نوین پرداختند. یافته‌های آن‌ها تأیید کرد که الگوریتم تقویت گرادیان شدید به طور مشخص در پیش‌بینی نقاط عطف تورمی و دوره‌های بی‌ثباتی، برتری معناداری نسبت به مدل‌های خطی اقتصادسنجی دارد.

نادمی و همکاران (۱۴۰۰) در پژوهشی با عنوان «پیش‌بینی قیمت نفت با رویکرد ترکیبی اقتصادسنجی-یادگیری ماشین»، یک مدل دومرحله‌ای را برای پیش‌بینی قیمت نفت خام ایران ارائه دادند. آن‌ها ابتدا با استفاده از یک مدل اقتصادسنجی ناهمسانی واریانس شرطی، نوسانات قیمت را مدل‌سازی کرده و سپس خروجی آن را به عنوان یک ورودی به یک شبکه عصبی حافظه طولانی کوتاه‌مدت دادند. نتایج نشان داد که این مدل ترکیبی، دقت بالاتری نسبت به هر یک از مدل‌ها به تنهایی دارد.

اسلامی و اسماعیلی (۱۳۹۹) در مقاله‌ای با عنوان «انتخاب متغیرهای پیشرو برای رشد اقتصادی ایران با استفاده از مدل‌های لاسو و جنگل تصادفی»، بر یکی از چالش‌های اصلی در مدل‌سازی تأکید کردند. آن‌ها نشان دادند که الگوریتم‌های یادگیری ماشین، به ویژه «رگرسیون لاسو»، می‌توانند به طور مؤثری از میان ده‌ها متغیر بالقوه، مهم‌ترین پیش‌بینی‌کننده‌های رشد اقتصادی را انتخاب کرده و به ساخت مدل‌های اقتصادسنجی ساده‌تر و کارآمدتر کمک کنند.

صادقی و همکاران (۱۳۹۸) در پژوهشی با عنوان «پیش‌بینی بحران‌های بانکی در ایران: رویکرد مقایسه‌ای لاجیت و یادگیری ماشین»، قابلیت مدل‌های مختلف را در پیش‌بینی بحران‌های مالی آزمودند. نتایج تحقیق آن‌ها آشکار ساخت که مدل ماشین بردار پشتیبان به دلیل انعطاف‌پذیری در تفکیک الگوهای پیچیده، با دقت بالاتری نسبت به مدل‌های لاجیت و پروبیت، سیگنال‌های هشدار اولیه بحران بانکی را شناسایی می‌کند.

ابونوری و توکلیان (۱۳۹۷) در تحقیقی با عنوان «مقایسه توان پیش‌بینی مدل‌های آریما و شبکه‌های عصبی مصنوعی در پیش‌بینی نرخ رشد تولید ناخالص داخلی ایران»، از جمله مطالعات پیشگام در این حوزه بودند. یافته‌های آن‌ها نشان داد که مدل شبکه عصبی مصنوعی به دلیل قابلیت بالاتر در شناسایی الگوهای غیرخطی، از نظر معیارهای خطا، عملکرد بهتری نسبت به مدل خطی سری زمانی داشته است.

کمیجانی و همکاران (۱۳۹۶) در مقاله‌ای با عنوان «اثرات نامتقارن شوک‌های نرخ ارز بر رشد اقتصادی ایران»، از یک مدل پیشرفته اقتصادسنجی به نام «مدل خودرگرسیون با وقفه‌های گسترده غیرخطی» استفاده کردند. هرچند این مقاله مستقیماً از یادگیری ماشین استفاده نکرده، اما با نشان دادن وجود روابط غیرخطی و نامتقارن در اقتصاد ایران، لزوم استفاده از مدل‌های انعطاف‌پذیرتر (مانند یادگیری ماشین) را که قادر به کشف چنین روابطی هستند، به طور غیرمستقیم تأیید می‌کند.

مرور ادبیات نظری و تجربی، تصویری چندبعدی از عوامل مؤثر بر رشد اقتصادی ارائه داد. مبانی نظری بر اهمیت متغیرهایی چون انباشت سرمایه، ثبات اقتصاد کلان، کیفیت نهادی و سیاست‌های پولی و مالی تأکید دارند. مطالعات تجربی نیز، چه در سطح بین‌المللی و چه در داخل کشور، ضمن تأیید این روابط، روند فزاینده‌ای را در به‌کارگیری الگوریتم‌های یادگیری ماشین برای دستیابی به دقت پیش‌بینی بالاتر نشان می‌دهند. با این حال، شکاف اصلی در ادبیات داخلی، فقدان یک مقایسه مستقیم و نظام‌مند میان این دو پارادایم در بستر اقتصاد ایران است. بنابراین، نوآوری اصلی این پژوهش، پر کردن همین شکاف از طریق قرار دادن یک مدل بهینه اقتصادسنجی در یک رقابت عادلانه با مدل‌های پیشرفته یادگیری ماشین است تا کارآمدترین ابزار برای پیش‌بینی رشد اقتصادی ایران شناسایی شود.

۳. روش تحقیق

در این بخش، به تشریح دقیق مبانی و مراحل عملی انجام پژوهش پرداخته می‌شود. این بخش شامل معرفی و آماده‌سازی متغیرها، تشریح مدل اقتصادسنجی سنتی و در نهایت، تبیین چارچوب به‌کارگیری و ارزیابی مدل‌های نوین یادگیری ماشین است.

۳-۱. معرفی متغیرها و آماده‌سازی داده‌ها

در این پژوهش از داده‌های سری زمانی سالانه اقتصاد ایران طی دوره ۱۳۷۰ تا ۱۴۰۲برای مدل‌سازی رشد اقتصادی استفاده شده است. متغیر وابسته، رشد تولید ناخالص داخلی حقیقی و متغیرهای توضیحی بر اساس مبانی نظری رشد اقتصادی و شواهد تجربی پیشین انتخاب گردیدند. این متغیرها شامل شاخص‌های سیاست پولی (نرخ سود سپرده، پایه پولی و نرخ تورم)، سیاست مالی (مخارج و درآمدهای دولت)، متغیرهای ساختاری و نهادی (سرمایه‌گذاری، نرخ بیکاری و درجه آزادی

اقتصادی) و متغیرهای بخش خارجی (نرخ ارز و درجه باز بودن تجاری) می‌باشند.

برای جلوگیری از رگرسیون کاذب، که در آن روابط آماری غیرواقعی میان متغیرها گزارش می‌شود، ابتدا آزمون مانایی بر روی تمام متغیرها انجام شد. این آزمون به ما نشان می‌دهد که آیا متغیرها دارای روند تصادفی هستند یا خیر. برای این منظور از آزمون دیکی-فولر تعمیم‌یافته استفاده گردید. نتایج این آزمون که در جدول ۲ ارائه شده، راهنمای ما برای تبدیلات لازم (مانند تفاضل‌گیری) جهت مانا ساختن متغیرها بود. این مرحله با استفاده از کتابخانه‌های پَنداس برای مدیریت داده‌ها و استتس‌مدلز در پایتون اجرا شد.

جدول ۱. خلاصه متغیرهای تحقیق با منابع معتبر جهانی

Table 1. Summary of research variables with reputable global sources

دسته‌بندی Classification	نام متغیر Variable	توضیحات (مبانی نظری انتخاب) Explanations (Theoretical Foundations of Selection)	منبع معتبر جهانی Reliable global source
متغیر وابسته Dependent variable	رشد تولید ناخالص داخلی حقیقی Real Gdp Growth	شاخص اصلی برای سنجش عملکرد و رشد اقتصادی کشور. The main indicator for measuring the country's economic performance and growth	بانک جهانی World Bank
سیاست پولی Monetary policy	نرخ سود سپرده Deposit interest rate	شاخصی از هزینه سرمایه و ابزار سیاست پولی که بر تصمیمات سرمایه‌گذاری تأثیرگذار است. An index of the cost of capital and a monetary policy tool that affects investment decisions.	صندوق بین‌المللی پول IMF
	پایه پولی Monetary base	نماینده مستقیم سیاست‌های پولی بانک مرکزی که بر حجم نقدینگی و فعالیت‌های اقتصادی مؤثر است. A direct representative of the central bank's monetary policies that affects the money supply and economic activities.	صندوق بین‌المللی پول IMF
	تورم Inflation	تورم بالا و بی‌ثبات می‌تواند با ایجاد نااطمینانی، به سرمایه‌گذاری و رشد اقتصادی لطمه بزند. High and unstable inflation can harm investment and economic growth by creating uncertainty.	بانک جهانی World Bank
سیاست مالی Fiscal policy	مخارج و درآمدهای دولت Government expenses and revenues	ابعاد سیاست‌های مالی دولت را نمایندگی می‌کنند و بر تخصیص منابع و رشد اقتصادی تأثیر دارند. They represent the dimensions of the government's fiscal policies and affect resource allocation and economic growth.	صندوق بین‌المللی پول IMF
ساختاری و نهادی Structural and institutional	سرمایه‌گذاری Investment	انباشت سرمایه یکی از موتورهای اصلی رشد اقتصادی بلندمدت است. Capital accumulation is one of the main drivers of long-term economic growth.	بانک جهانی World Bank
	نرخ بیکاری Unemployment rate	شاخص کلیدی سلامت بازار کار و نماینده‌ای از ظرفیت‌های تولیدی استفاده نشده در اقتصاد است. It is a key indicator of labor market health and a representation of unused productive capacities in the economy.	بانک جهانی (بر اساس تخمین ILO) World Bank (ILO)
	درجه آزادی اقتصادی Economic Freedom	شاخصی ترکیبی از حاکمیت قانون، حقوق مالکیت، باز بودن بازارها و کارایی نهادی. An index combining rule of law, property rights, market openness, and institutional efficiency.	بنیاد هریتیج The Heritage Foundation
بخش خارجی Foreign Section	نرخ ارز Exchange rate	یک قیمت کلیدی که بر رقابت‌پذیری تجاری و تصمیمات سرمایه‌گذاری مؤثر است. A key price that affects business competitiveness and investment decisions.	صندوق بین‌المللی پول IMF
	درجه باز بودن تجاری Degree of trade openness	مجموع صادرات و واردات به GDP نشان‌دهنده میزان تعامل یک کشور با اقتصاد جهانی است. The total exports and imports to GDP indicate the level of a country's engagement with the global economy.	بانک جهانی World Bank

جدول ۲. نتایج آزمون مانایی دیکی- فولر تعمیم‌یافته

Table 2. Results of the Augmented Dickey-Fuller Unit Root Test

متغیر Variable	مقدار احتمال (در سطح) Probability value (at level)	مقدار احتمال (تفاضل اول) Probability value (first difference)	مقدار احتمال (تفاضل دوم) Probability value (second difference)	مرتبه نهایی انباشتگی Final order of accumulation
رشد تولید ناخالص داخلی Gross domestic product growth	0.0002	0.0000	0.0000	I(0)
درجه باز بودن تجاری Trade openness	0.4884	0.0000	0.0634	I(1)
نرخ سود سپرده Deposit interest rate	0.8034	0.0129	0.0032	I(1)
تورم Inflation	1.000	0.035	0.0000	I(1)
نرخ ارز Exchange rate	0.0196	0.0000	0.0000	I(0)
پایه پولی Monetary base	0.0001	0.8034	0.0000	I(0)
سرمایه‌گذاری Investment	0.4021	0.1984	0.0002	I(2)
نرخ بیکاری Unemployment rate	0.0049	0.0004	0.0001	I(0)
درجه آزادی اقتصادی Degree of economic freedom	0.6262	0.0002	0.3403	I(1)
مخارج دولت Government expenditure	0.7256	0.0000	0.0000	I(1)
درآمدهای دولت Government revenues	0.3007	0.3455	0.0000	I(2)

۳-۲. معرفی مدل اقتصادسنجی

برای مدل‌سازی رابطه پویا میان متغیرها و ایجاد یک معیار پایه (محک) جهت مقایسه، از چارچوب مدل خودرگرسیون با وقفه‌های گسترده استفاده شد. این مدل به دلیل ماهیت پویا در شناسایی تأثیرات کوتاه‌مدت و بلندمدت و همچنین انعطاف‌پذیری در ورود متغیرهایی با درجات انباشتگی متفاوت (پس از ماناسازی)، گزینه‌ای مناسب برای این پژوهش است. فرم کلی مدل ARDL(p, q) که در آن p تعداد وقفه‌های متغیر وابسته و q تعداد وقفه‌های متغیرهای توضیحی است، به صورت رابطه ریاضی زیر تعریف می‌شود:

$$Y_t = \beta_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{k=1}^{K} \sum_{j=0}^{q_k} \delta_{k,j} X_{k,t-j} + \epsilon_t \qquad (۱)$$

در این رابطه:

Y_t متغیر وابسته (رشد اقتصادی) در زمان t

Y_{t-i} مقادیر گذشته متغیر وابسته با i دوره وقفه

$X_{k,t-j}$ مقدار متغیر توضیحی kام با j دوره وقفه

β_0 عرض از مبدأ یا مقدار ثابت مدل

ϕ_i و $\delta_{k,j}$ ضرایب مدل که میزان تأثیر هر متغیر را نشان می‌دهند

ϵ_t جمله خطای تصادفی که تأثیر عوامل خارج از مدل را پوشش می‌دهد

۳-۳. معرفی مدل‌های یادگیری ماشین و چارچوب ارزیابی

در این پژوهش از دو مدل قدرتمند یادگیری گروهی[1] استفاده می‌شود که در ادامه مبانی نظری آن‌ها تشریح می‌گردد.

۱. جنگل تصادفی: این الگوریتم که توسط بریمن[2] (۲۰۰۱) معرفی شد، یک روش یادگیری گروهی مبتنی بر ساخت تعداد زیادی درخت تصمیم[3] در زمان آموزش است. این مدل با استفاده از روش کیسه‌بندی[4]، از زیرنمونه‌های تصادفی از داده‌های آموزشی برای ساخت هر درخت بهره می‌برد و در هر گره از درخت نیز تنها زیرمجموعه تصادفی از متغیرها را برای تقسیم در نظر می‌گیرد. این دو سطح از تصادفی‌سازی، از بیش‌برازش[5] مدل جلوگیری کرده و قدرت تعمیم آن را افزایش می‌دهد. پیش‌بینی نهایی در یک مسئله رگرسیون، میانگین

1. Ensemble Learning
2. Breiman
3. Decision Tree
4. Bagging (Bootstrap Aggregating)
5. Overfitting

پیش‌بینی تمام درخت‌های منفرد در جنگل است، همان‌طور که در رابطه (۲) نشان داده شده است:

$$\hat{f}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \quad (۲)$$

که در آن $\hat{f}(x)$ پیش‌بینی نهایی، B تعداد کل درختان و $T_b(x)$پیش‌بینی درخت bام است.

۲. تقویت گرادیان شدید: این الگوریتم که توسط چن و گسترین[1] (۲۰۱۶) توسعه داده شد، یک پیاده‌سازی بهینه و مقیاس‌پذیر از روش تقویت گرادیان است. برخلاف جنگل تصادفی که درختان را به صورت موازی می‌سازد، این روش درختان را به صورت متوالی ایجاد می‌کند. هر درخت جدید تلاش می‌کند تا خطاهای باقی‌مانده از مجموع درختان قبلی را اصلاح نماید. ویژگی برجسته این الگوریتم، بهینه‌سازی تابع هدف[2] آن است که شامل تابع زیان[3] و یک عبارت تنظیم‌سازی[4] (شامل جریمه‌های اِل۱ و اِل۲)[5] برای کنترل پیچیدگی مدل و جلوگیری از بیش‌برازش است. این الگوریتم از دو رابطه اصلی پیروی می‌کند.

الف) فرمول ساختار افزایشی مدل: این رابطه نشان می‌دهد که چگونه هر درخت جدید، خطاهای درختان قبلی را جبران می‌کند.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (۳)$$

ب) فرمول تابع هدف کلی مدل: این رابطه، ترکیبی از تابع زیان و عبارت تنظیم‌سازی است که مدل در هر مرحله تلاش می‌کند آن را کمینه کند.

$$\text{Obj}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{j=1}^{t} \Omega(f_j) \quad (۴)$$

برای تضمین یک مقایسه عادلانه، مجموعه داده به دو بخش زمانی تقسیم شد: داده‌های تا سال ۱۳۹۴ برای آموزش مدل‌ها و داده‌های سال‌های ۱۳۹۵ تا ۱۴۰۲ برای آزمون و ارزیابی عملکرد پیش‌بینی آن‌ها. عملکرد مدل‌ها بر روی داده‌های آزمون، با استفاده از دو معیار خطای استاندارد، ریشه میانگین مربعات خطا[6] و میانگین قدر مطلق خطا[7]، سنجیده شد.

1. Chen & Guestrin
2. Objective Function
3. Loss Function
4. Regularization
5. L1 and L2 Penalties
6. Root Mean Squared Error (RMSE)
7. Mean Absolute Error (MAE)

۴. برآورد مدل‌ها و تحلیل نتایج

در این بخش، نتایج حاصل از پیاده‌سازی مدل‌های تشریح‌شده در بخش قبل ارائه و تفسیر می‌شود. ابتدا یافته‌های مدل اقتصادسنجی، سپس عملکرد مدل‌های یادگیری ماشین و در نهایت، یک تحلیل تطبیقی از این دو رویکرد ارائه می‌گردد.

۴-۱. نتایج مدل اقتصادسنجی

پس از فرآیند بهینه‌سازی و حذف متغیرهای بی‌اهمیت، مدل نهایی اقتصادسنجی حاصل شد[8]. برای رسیدن به یک مدل بهینه و جلوگیری از ورود متغیرهای بی‌معنا و مشکل هم‌خطی، از یک مدل عمومی ARDL (1.1) به‌عنوان نقطه شروع استفاده و سپس با روش حذف گام به گام به عقب[9]، متغیرهای بی‌اهمیت حذف شدند تا مدل نهایی به دست آید[10].

جدول ۳ نتایج کامل برآورد این مدل و جدول ۴ خلاصه‌ای از آماره‌های تشخیصی آن را نمایش می‌دهد.

جدول ۳. نتایج نهایی مدل اقتصادسنجی بهینه

Table 3. Final results of the optimized econometric model

متغیر Variable	مقدار احتمال Prob	خطای استاندارد Standard error	ضریب Coefficient
مقدار ثابت Constant	0.129	2.200	-3.460
رشد پایه پولی (وقفه۱) Base money growth (lag 1)	0.004	9.870	31.880
تغییرات نرخ بیکاری (همزمان) Changes in unemployment rate (concurrent)	0.092	0.470	-0.830
تغییرات آزادی اقتصادی (همزمان) Changes in economic freedom (concurrent)	0.000	0.160	0.650
تغییرات آزادی اقتصادی (وقفه ۱) Changes in economic freedom (lag 1)	0.010	0.220	-0.610
شتاب نرخ سود سپرده (وقفه ۱) Deposit interest rate acceleration (lag 1)	0.050	0.480	-1.000
شتاب تورم (همزمان) Inflation acceleration (concurrent)	0.064	4.480	-8.740
شتاب تورم (وقفه ۱) Inflation acceleration (lag 1)	0.090	5.210	-9.230

۸. برآورد این مدل با استفاده از کتابخانه استتس‌مدلز در پایتون انجام گرفت.

9. Backward Stepwise Elimination

۱۰. فرمول‌ها و متغیرهای مورد استفاده در پیوست آمده است.

جدول ۴.خلاصه آماره‌های تشخیصی مدل

Table 4. Summary of the model's diagnostic statistics

آماره Static	مقدار Value
ضریب تعیین تعدیل‌شده Adjusted R-squared	0.542
آماره F و مقدار احتمال آن F-statistic	5.909 (0.000592)
آماره دوربین-واتسون Durbin-Watson Statistic	2.067

تحلیل نتایج نشان‌دهنده عملکرد مطلوب مدل است. ضریب تعیین تعدیل‌شده بیانگر آن است که مدل حاضر توانسته بخش قابل قبولی از نوسانات رشد اقتصادی را توضیح دهد. همچنین، آماره اف معناداری کل مدل را تأیید می‌کند و آماره دوربین-واتسون عدم وجود مشکل خودهمبستگی را نشان می‌دهد. تفسیر ضرایب معنادار حاکی از آن است که رشد پایه پولی با یک وقفه تأثیر مثبت و معناداری بر رشد اقتصادی دارد که با نظریه‌های اقتصادی همخوان است. از سوی دیگر، شتاب تورم و شتاب نرخ سود سپرده با ضرایب منفی خود، اهمیت ثبات اقتصاد کلان را در پیش‌بینی‌پذیر کردن فضای کسب‌وکار برجسته می‌سازند.

۴-۲. نتایج عملکرد مدل‌های یادگیری ماشین

پیاده‌سازی، آموزش و ارزیابی مدل‌های یادگیری ماشین در محیط برنامه‌نویسی پایتون و با استفاده از کتابخانه‌های استاندارد علم داده نظیر سایکیت‌لِرن[1] و ایکس‌جی‌بوست[2] انجام شده است. برای مقایسه عملکرد پیش‌بینی، هر دو مدل بر روی داده‌های آزمون (۱۳۹۵ تا ۱۴۰۲) ارزیابی شدند. جدول ۵ نتایج این مقایسه را بر اساس معیارهای خطا نمایش می‌دهد.

جدول ۵. مقایسه عملکرد مدل‌های یادگیری ماشین بر روی داده‌های آزمون

Table 5. Comparison of machine learning models' performance on test data

مدل Model	MAE	RMSE
جنگل تصادفی Random forest	2.8628	3.6645
XGBoost	3.1202	3.7694

همان‌طور که نتایج جدول ۵ نشان می‌دهد، هر دو مدل یادگیری ماشین عملکرد بهتری نسبت به مدل اقتصادسنجی (که خطای بالاتری داشت) از خود نشان دادند. در این میان، مدل جنگل تصادفی با ثبت کمترین میزان خطا در هر دو معیار، به عنوان «مدل برتر یادگیری ماشین» در این پژوهش انتخاب می‌شود.

۴-۳. تحلیل تطبیقی دو رویکرد

برای درک عمیق‌تر یافته‌ها و مقایسه دو پارادایم، در این بخش به تحلیل اهمیت ویژگی‌ها[3] در مدل برتر جنگل تصادفی پرداخته می‌شود. جدول ۶ متغیرها را بر اساس میزان تأثیرگذاری آن‌ها در کاهش خطای پیش‌بینی مدل رتبه‌بندی می‌کند.

تحلیل تطبیقی نتایج، یافته‌های بسیار مهمی را آشکار می‌سازد. اولاً، برتری چشمگیر مدل جنگل تصادفی در دقت پیش‌بینی، قابلیت بالای این الگوریتم را در کشف الگوهای پیچیده داده‌ها تأیید می‌کند. ثانیاً، برخی متغیرها مانند تغییرات درجه آزادی اقتصادی هم در مدل اقتصادسنجی معنادار بوده و هم در مدل یادگیری ماشین به عنوان یک ویژگی مهم شناسایی شده‌اند که اعتبار بالای این متغیر را نشان می‌دهد.

نکته کلیدی و سوم، تفاوت در شناسایی عوامل مهم است. متغیر تغییرات نسبت مخارج دولت به تولید ناخالص داخلی که در فرآیند بهینه‌سازی از مدل خطی اقتصادسنجی حذف شده بود، در مدل جنگل تصادفی به عنوان یکی از تأثیرگذارترین ویژگی‌ها برای پیش‌بینی شناسایی شده است. این یافته قویاً نشان می‌دهد که سیاست‌های مالی دولت احتمالاً از طریق کانال‌های غیرخطی و پیچیده‌ای بر رشد اقتصادی تأثیر می‌گذارند که مدل‌های سنتی خطی قادر به کشف کامل آن نیستند و این خود، ارزش مکمل بودن دو رویکرد را برای سیاست‌گذار آشکار می‌سازد.

1. Scikit-learn
2. XGBoost (library)
3. Feature Importance

جدول ۶. رتبه‌بندی اهمیت ویژگی‌ها در مدل‌های یادگیری ماشین

Table 6. Ranking the importance of features in machine learning models

اهمیت در مدل جنگل تصادفی Importance in the Random Forest model	اهمیت در مدل تقویت گرادیان Importance in the Gradient Boosting model	میانگین اهمیت Average importance	متغیر Variable
0.134	0.5605	0.3473	تغییرات نسبت مخارج دولت به تولید ناخالص داخلی (همزمان) Changes in the ratio of government expenditures to GDP (concurrent)
0.1556	0.0918	0.1237	تغییرات درجه آزادی اقتصادی (همزمان) Changes in the degree of economic freedom (concurrent)
0.0543	0.1418	0.098	تغییرات نرخ بیکاری (وقفه ۱) Changes in the unemployment rate (lag 1)
0.1259	0.0205	0.0732	شتاب نرخ سود سپرده (وقفه ۱) Acceleration of deposit interest rate (lag 1)
0.096	0.0254	0.0607	نرخ رشد پایه پولی (وقفه ۱) Growth rate of the monetary base (lag 1)
0.0732	0.0468	0.06	نرخ رشد پایه پولی (همزمان) Growth rate of the monetary base (concurrent)
0.0516	0.0351	0.0434	تغییرات نرخ بیکاری (همزمان) Changes in the unemployment rate (concurrent)
0.0508	0.0034	0.0271	نرخ رشد سرمایه‌گذاری (وقفه ۱) nvestment growth rate (lag 1)
0.0375	0.006	0.0217	نرخ رشد سرمایه‌گذاری (همزمان) Investment growth rate (concurrent)
0.0271	0.0155	0.0213	نرخ رشد ارز (وقفه ۱) Exchange rate growth rate (lag 1)
0.0219	0.011	0.0164	نرخ رشد ارز (همزمان) Exchange rate growth rate (concurrent)
0.015	0.0171	0.016	نسبت درآمد دولت به تولید ناخالص داخلی (همزمان) Ratio of government revenue to GDP (concurrent)
0.0301	0	0.015	شتاب تورم (همزمان) Inflation acceleration (concurrent)
0.0272	0	0.0136	شتاب تورم (وقفه ۱) Inflation acceleration (lag 1)
0.0196	0.0053	0.0124	رشد تولید ناخالص داخلی (وقفه ۱) GDP growth (lag 1)
0.0155	0.007	0.0112	درجه باز بودن تجاری (وقفه ۱) Degree of trade openness (lag 1)
0.0073	0.0118	0.0095	درجه باز بودن تجاری (همزمان) Degree of trade openness (concurrent)
0.0165	0	0.0083	نسبت درآمد دولت به تولید ناخالص داخلی (وقفه ۱) Ratio of government revenue to GDP (lag 1)
0.0157	0	0.0079	تغییرات درجه آزادی اقتصادی (وقفه ۱) Changes in the degree of economic freedom (lag 1)
0.0149	0.0001	0.0075	تغییرات نسبت مخارج دولت به تولید ناخالص داخلی (وقفه ۱) Changes in the ratio of government expenditures to GDP (lag1)
0.0107	0.0008	0.0057	شتاب نرخ سود سپرده (همزمان) Acceleration of deposit interest rate (concurrent)

۵. نتیجه‌گیری و ارائه پیشنهادها

این پژوهش با هدف یافتن ابزاری کارآمدتر برای پیش‌بینی رشد اقتصادی در اقتصاد پرنوسان ایران، به ارزیابی تطبیقی دو پارادایم مدل‌سازی پرداخت: رویکرد اقتصادسنجی سنتی و رویکرد نوین یادگیری ماشین. چالش پیش‌بینی اقتصاد کلان، به دلیل وجود عدم قطعیت و تغییرات ساختاری، همواره امری دشوار بوده است (استاک و واتسون[1]، ۲۰۰۷). در این راستا، یک

1. Stock & Watson

مدل بهینه خودرگرسیون با وقفه‌های گسترده با دو مدل پیشرفته یادگیری ماشین (جنگل تصادفی و تقویت گرادیان شدید) مقایسه شد.

تحلیل نتایج به سه یافته اصلی منجر شد. نخست، مدل اقتصادسنجی نشان داد که رشد اقتصادی ایران تحت تأثیر عوامل قابل تفسیری مانند رشد پایه پولی، درجه آزادی اقتصادی و شتاب تورم قرار دارد. دوم، در رقابت پیش‌بینی، مدل‌های یادگیری ماشین و به طور مشخص مدل جنگل تصادفی، با ثبت کمترین خطا عملکرد برتری داشتند؛ این یافته همسو با پژوهش‌های اخیر در این حوزه است (کوپ و همکاران[1]، ۲۰۲۱). و در نهایت، مقایسه اهمیت متغیرها نشان داد که مدل یادگیری ماشین، متغیر «مخارج دولت» را به عنوان یک عامل کلیدی شناسایی کرد، در حالی که این متغیر در مدل خطی معنادار نبود، که این امر قویاً حاکی از وجود کانال‌های تأثیرگذاری غیرخطی است.

یافته‌های ما نشان می‌دهد که این دو رویکرد مکمل یکدیگر هستند، همان‌طور که آتی[2] (۲۰۱۸) استدلال می‌کند، اقتصاد مدرن نیازمند ابزارهای جدیدی برای تحلیل روابط پیچیده است. در حالی که مدل اقتصادسنجی چارچوبی ارزشمند برای تفسیر نظری فراهم می‌کند، مدل جنگل تصادفی با فدا کردن بخشی از تفسیرپذیری، به دقت پیش‌بینی بسیار بالاتری دست یافته و الگوهای غیرخطی پنهان را آشکار می‌سازد. بنابراین، نتیجه‌گیری نهایی این است که برای یک تحلیل جامع و سیاست‌گذاری بهینه، باید از هر دو ابزار به صورت ترکیبی بهره برد.

بر اساس یافته‌های مستحکم پژوهش، پیشنهادهای سیاستی زیر ارائه می‌گردد. اول، با توجه به اهمیت بالای «درجه آزادی اقتصادی» در هر دو رویکرد، تمرکز بر اصلاحات نهادی و بهبود محیط کسب‌وکار باید در اولویت باشد. دوم، اهمیت غیرخطی «مخارج دولت» در مدل یادگیری ماشین نشان می‌دهد که سیاست مالی باید بر افزایش کارایی، شفافیت و تخصیص بهینه منابع به سمت مخارج مولد متمرکز شود و در نهایت، با عنایت به یافته‌های مدل اقتصادسنجی، حفظ ثبات اقتصاد کلان از طریق کنترل پایدار تورم و پرهیز از تغییرات ناگهانی در سیاست‌های پولی، باید به عنوان لنگر سیاست‌گذاری اقتصادی کشور عمل کند.

تضاد منافع

نویسندگان ضمن رعایت اخلاق نشر در ارجاع‌دهی، نبود تضاد منافع را اعلام می‌دارند.

منابع

Abounoori, E., & Tavakolian, H. (2018). Comparing the predictive power of ARIMA models and artificial neural networks in forecasting Iran's GDP growth rate. *Quarterly Journal of Econometric Modeling, 4*(1), 1-24.

Athey, S. (2018). The impact of machine learning on economics. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda*, University of Chicago Press. 507-547.

Breiman, L. (2001). Random forests. *Machine Learning*, *1*(5), 42-56.

Charpentier, A., Gallic, E., & Le Gallo, J. (2022). Machine learning in macroeconomics: A survey. *Journal of Economic Surveys*. *36*(4), 965-925.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

Döpke, J., Fritsche, U., & Pierdzioch, C. (2023). Forecasting German GDP using LSTM networks. *Journal of Forecasting* , *42*(3), 512-527.

Eslami, M., & Esmaeili, A. (2020). Selecting leading variables for Iran's economic growth using LASSO and Random Forest models. *Iranian Economic Research, 25*(85), 47-75.

Falahi, F., Mohammadi, T., & Rezaei, S. (2023). Forecasting Iran's Economic Growth Using Machine Learning Algorithms: A Comparative Analysis. *Quarterly Journal of Monetary and Banking Research, 16*(57), 101-129

Goulet, T. J. T. (2021). *Macroeconomic forecasting with machine learning: A survey and an application to Canada* (Bank of Canada Staff Working Paper 2021-3). Bank of Canada.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223-2273.

Kamijani, A., Eslami, M., & Bagheri, P. (2017). Asymmetric effects of exchange rate shocks on Iran's economic growth. *Economic Research, 52*(3), 591-616.

Koop, G., Onorante, L., & Mitchell, J. (2021) *GDP forecasting with machine learning: A euro area perspective* (ECB Working Paper No. 2592). European Central Bank.

Medeiros, M. C., & Vasconcelos, G. F. (2022). Machine learning for macroeconomic forecasting: Predicting recessions. *International Journal of Forecasting*, *38* (3), 845-869

Mishkin, F. S. (2007). *The economics of money, banking, and financial markets* (8th ed.). Pearson/Addison Wesley.

Motavaseli, M., & Shahnazi, R. (2022). Is machine learning a better tool for predicting inflation in Iran? *Economic Research Journal, 22*(84), 1-34.

Nadimi, Y., Anvari, E., & Hosseini, D. (2021). Predicting oil prices with a combined econometrics-machine learning approach. *Quarterly Journal of Energy Economics Studies, 18*(70), 1-28.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.

1. Koop et al.
2. Athey

Timmermann (Eds.), *Handbook of Economic Forecasting*, 3-50.

The Heritage Foundation. (2025). *2025 Index of Economic Freedom*. The Heritage Foundation. Retrieved August 31, 2025, from https://www.heritage.org/index/

World Bank Group. (2025). *World Development Indicators (WDI)*. Retrieved August 31, 2025, from https://databank.worldbank.org/source/world-development-indicators

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, *98* (5), S71-S102.

Sadeghi, H., Shakeri, A., & Ahmadi, A. (2019). Predicting banking crises in Iran: A comparative approach of logit and machine learning. *Quarterly Journal of Economic Modeling Research*, *10*(37), 1-32.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics, 70*(1), 75-94.

Stock, J. H., & Watson, M. W. (2007). Forecasting in economics. In G. Elliott, C. W. J. Granger, & A.

پیوست

برآورد مدل اقتصاد سنجی

```
import pandas as pd
import numpy as np
import io
from sklearn.ensemble import RandomForestRegressor
from xgboost import XGBRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error
import matplotlib.pyplot as plt
import seaborn as sns

# --- داده‌های خام مستقیماً در کد تعبیه شده‌اند ---
raw_data_csv = """year,GDP_growth,Deposit_Rate,CPI,Exchange_Rate,MoneyBase,Government_expenditure,Government_Revenue,Population,Unemployment,investment,Degree_economic_freedom,Trade_Openness
۱۹۹۱,۳.۲۹۷۹۸۲,۱۵,۳.۴۸۳۶۵۸,۶۷.۵۲۱۷۳,۸۸۳۲,۵۵۲۷۱۳۰۰۰۰۰۰۰,۶۵۴۹.۹,۵۹۹۸۶۹۳۱,۱۱.۱,۲۲۶۸۳۷۸۰۰۰۰۰۰۰,۳۶.۱,۴۴.۱۹۴۱۷۲
۱.۴۷۱۰۱۱,۱۵,۴.۳۸۲۷۱۱,۶۵.۵۶۷۸۴۱,۱۱۰۴۰.۶۹۸۱۰۵۵۰۰۰۰۰۰,۱۰۲۱۱.۵,۶۱۱۸۴۹۸۳,۱۰۶۵۳,۳۰۳۲۵۳۶۰۰۰۰۰۰۰,۳۶.۱,۴۰.۳۹۷۹۲۷-,۱۹۹۲
۱.۶۹۸۲۸۶,۱۵,۵.۳۱۱۹۶,۱۲۶۸.۰۷۸۷۶۱,۱۳۷۹۹,۱۵۱۰۳۷۱۰۰۰۰۰۰۰,۱۶۳۵۰.۱,۶۱۴۰۳۵۴۳,۱۰.۵۶۷,۳۲۲۵۹۴۷۰۰۰۰۰۰۰,۳۶.۱,۴۶.۰۴۰۸۶۴-,۱۹۹۳
۱۹۹۴,۲.۴۰۱۰۴۵,۱۵,۶.۹۸۲۴۱۴,۱۷۴۹.۱۷۳۷۳۶,۱۷۹۳۸.۷,۲۰۰۵۶۴۱۰۰۰۰۰۰۰,۲۲۱۰۸.۴,۶۱۵۲۰۵۹۸,۱۰.۴۴۸,۳۲۵۴۹۹۱۰۰۰۰۰۰۰,۳۶.۱,۴۱.۶۰۹۱۳۶
۱۹۹۵,۶.۳۵۰۵۲۶,۱۵,۱۰.۴۴۹۶۰۱,۱۷۴۸.۳۵۰۲۴۶,۲۴۲۱۷.۳,۲۸۱۲۸۱۸۰۰۰۰۰۰۰,۳۱۲۵۴.۹,۶۲۲۰۵۵۶۷,۹.۸۹۶,۵۹۵۶۶۶۵۰۰۰۰۰۰۰,۳۶.۱,۳۵.۱۴۴۲۴۵
۱۹۹۶,۸.۸۱۵۰۸۷,۱۵,۱۲.۸۷۱۵۷۹,۱۷۴۹.۵,۳۱۴۸۲.۵,۴۱۷۱۶۳۸۰۰۰۰۰۰۰,۴۳۲۵۶.۳,۶۲۸۸۵۹۹۴,۹.۴۵۸,۷۴۶۵۶۴۶۰۰۰۰۰۰۰,۳۸.۶,۳۳.۵۶۵۸۶۳
۱۹۹۷,۳.۰۴,۱۵,۱۵.۱۱۱۴۵۳,۱۷۵۱.۲۷۸۵۴,۴۰۱۴۹.۳,۵۵۸۸۷۲۵۰۰۰۰۰۰۰,۶۵۴۳۲.۸,۶۳۵۶۶۴۲۱,۱۰.۰۵۷,۹۸۷۹۷۹۸۰۰۰۰۰۰۰,۳۸.۶,۳۵.۲۴۹۸۲
۱۹۹۸,۳.۲۲۱۵۳,۱۵,۱۷.۸۵۹۶۶۴,۱۷۵۲.۵,۴۸۱۲۴۶,۶۵۴۳۲۱۰۰۰۰۰۰۰۰,۷۸۹۶۵.۴,۶۴۲۴۶۸۴۸,۱۰.۳۵۸,۱۲۳۴۵۶۰۰۰۰۰۰۰۰,۳۹.۳,۳۱.۷۹۶۳۹۷
۱۹۹۹,۱.۰۰۷۳۸۵,۱۴,۲۱.۴۴۲۸۷,۱۷۵۳.۳۵۴۹۵,۶۰۱۵۵.۸,۵۶۱۲۴۵۴۰۰۰۰۰۰۰,۸۹۸۷۳.۳,۶۵۴۴۵۳۸۹,۱۱.۸۹۷,۱۵۴۵۶۵۰۰۰۰۰۰۰۰,۴۰.۳,۳۷.۸۹۷۶۵۴
۲۰۰۰,۵.۰۸۶۴۳۸,۱۳.۵,۲۵.۸۸۲۱۹۳,۷۹۹۶.۱۸۵۳۶۸,۷۵۱۹۴.۸,۶۵۴۳۲۱۰۰۰۰۰۰۰۰,۱۰۲۳۵۴.۶۶۱۲۸۹۸۷,۱۳.۶۷۹,۲۳۱۲۵۴۰۰۰۰۰۰۰۰,۴۰.۳,۴۲.۵۸۹۶۳۲
۲۰۰۱,۳.۵۸۵۵۷,۱۲.۴,۲۸.۸۴۷۵۲۶,۷۹۲۰.۶۶۹۸۶۶,۹۳۹۹۳.۵,۷۸۹۶۵۴۰۰۰۰۰۰۰۰,۱۲۵۸۹۶,۶۶۸۱۲۵۵۴,۱۲.۷۸۴,۳۲۱۴۵۸۰۰۰۰۰۰۰۰,۴۱.۲,۴۰.۵۳۵۵۶۶
۲۰۰۲,۷.۴۶۴۶۰۱,۱۱.۲,۳۳.۳۷۴۲۴۵,۷۹۶۸.۴۲۱۰۳۴,۱۲۲۱۹۱.۶,۹۸۷۵۶۳۰۰۰۰۰۰۰۰,۱۵۸۹۶۳,۶۷۴۹۶۱۲۱,۱۱.۷۷۲,۴۵۸۹۶۳۰۰۰۰۰۰۰۰,۴۱.۲,۴۸.۷۸۹۶۵۴
۲۰۰۳,۷.۱۸۹۷۳۶,۱۲.۵,۳۸.۵۸۶۱۱۱,۸۱۸۲.۲۶۰۵۲,۱۶۵۰۳۸.۸,۱۲۵۸۹۷۰۰۰۰۰۰۰۰۰,۲۰۱۵۸۹,۶۸۱۷۹۶۸۸,۱۰.۹۷۵,۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۴۱.۵,۵۲.۴۸۹۶۳۲
۲۰۰۴,۴.۹۹۹۷۹۵,۱۳.۸,۴۴.۴۲۶۷۴۳,۸۷۳۳.۸۶۱۰۷۸,۲۱۷۸۵۱,۱۵۴۸۷۹۰۰۰۰۰۰۰۰۰,۲۵۴۸۹۶,۶۸۸۶۳۲۵۵,۱۰.۵۶۹,۸۷۹۶۵۴۰۰۰۰۰۰۰۰۰,۴۱.۵,۵۸.۵۶۵۰۹
۲۰۰۵,۳.۱۸۱۸۷۸,۱۵.۱,۵۰.۸۸۷۲۵,۹۰۲۶.۳۴۷۵۰۹,۲۹۸۴۸۵.۴,۲۶۵۸۹۷۰۰۰۰۰۰۰۰۰,۵۸۹۷۴۱,۷۲۲۲۴۵۳۱,۱۰.۷۷۲,۱۱۲۵۸۷۰۰۰۰۰۰۰۰۰۰,۴۲.۹,۵۶.۳۲۱۴۵۸
۲۰۰۶,۶۳۴۱۲,۱۵.۱,۵۶.۹۶۵۸۷۴,۹۲۳۲.۵,۳۵۸۳۰۰,۳۰۷۷۵۰۲۰۰۰۰۰۰۰۰,۹۳۵۴۴۱.۷,۷۳۷۶۵۴۸۹,۱۰.۴۶۵,۱۲۹۰۱۲۰۰۰۰۰۰۰۰۰۰,۴۳.۲,۴۹.۸۸۷۳۰۷
۲۰۰۷,۷.۷۸۹۴۵۶,۱۵.۱,۶۳.۷۸۶۳۳۲,۹۲۸۱.۱۵۱۸۲۸,۴۵۸۶۰۰,۴۱۹۶۷۲۸۰۰۰۰۰۰۰۰,۱۲۶۸۳۸۳,۷۴۶۰۲۶۴۷,۱۰.۱۵۸,۱۵۸۹۶۳۰۰۰۰۰۰۰۰۰۰,۴۳.۲,۴۷.۸۹۶۵۴۱
۲۰۰۸,۰۵۸۹۶۳۲,۱۵.۱,۷۹.۵۶۸۹۷۴,۹۷۲۵.۵,۵۶۸۹۰۰,۵۶۹۸۴۱۲۰۰۰۰۰۰۰۰,۱۸۹۶۵۴۱,۷۵۴۳۹۸۰۵,۱۰.۰۴۵,۱۹۸۷۵۴۰۰۰۰۰۰۰۰۰,۴۲.۴,۵۱.۲۵۸۹۶۳
۲۰۰۹,۱.۵۸۹۶۳۲,۱۵,۸۷.۳۶۹۸۵۲,۱۰۰۵۱.۷۸۵,۶۸۹۷۰۰,۶۸۴۲۱۳۵۰۰۰۰۰۰۰۰,۲۱۳۴۵۶۸,۷۶۲۷۶۹۶۳,۱۱.۵۸۹,۲۶۵۸۹۷۰۰۰۰۰۰۰۰۰,۴۲.۴,۴۵.۲۱۳۶۹۸
۲۰۱۰,۵.۸۹۶۳۲۱,۱۵,۹۷.۲۵۸۹۶۳,۱۰۲۵۴.۲,۸۵۴۲۰۰,۸۵۴۱۲۳۶۰۰۰۰۰۰۰۰,۲۸۹۶۵۴۱,۷۷۱۱۴۱۲۱,۱۳.۰۴۵,۳۵۸۹۶۴۰۰۰۰۰۰۰۰۰,۴۴.۸,۴۰.۱۲۵۸۹۶
۲۰۱۱,۳.۲۵۸۹۶۳,۱۵,۱۱۸.۵۸۹۶۳۲,۱۲۲۶۰,۱۰۲۵۰۰۰,۱۰۲۵۸۷۴۰۰۰۰۰۰۰۰۰,۳۵۸۹۶۴۱,۷۷۹۵۱۲۷۹,۱۲.۱۱۳,۴۱۹۴۳۱۳۰۰۰۰۰۰۰۰۰,۴۴.۸,۴۹.۵۲۸۷۴۱
۳.۷۴۷۱۷۱,۱۵,۱۵۴.۸۹۶۳۲۱,۱۲۲۶۰,۱۲۳۰۰۰۰,۱۲۳۵۸۷۴۰۰۰۰۰۰۰۰۰,۴۵۸۹۶۳۲,۷۸۷۸۸۴۳۷,۱۲.۰۱۲,۴۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۴۲,۴۸.۹۶۵۴۱۲-,۲۰۱۲
۲۰۱۳,۱.۲۵۸۹۶۳,۱۵,۲۰۸.۵۸۹۶۳۲,۲۴۷۷۷.۵,۱۵۱۸۷۰۰,۱۵۴۷۸۹۶۰۰۰۰۰۰۰۰۰,۵۸۹۶۳۲۱,۷۹۶۲۵۵۹۵,۱۰.۱۴۵,۵۱۲۵۸۷۴۰۰۰۰۰۰۰۰۰,۴۲,۴۶.۳۲۵۸۹۶
۲۰۱۴,۴.۵۸۹۶۳۲,۱۵,۲۴۰.۲۵۸۹۶۳,۲۶۵۱۰.۳,۱۸۴۲۳۰۰,۱۸۸۴۴۳۳۰۰۰۰۰۰۰۰۰,۶۹۸۵۴۱۲,۸۰۴۶۲۷۵۳,۱۰.۲۵۶,۵۵۵۴۵۲۰۰۰۰۰۰۰۰۰۰,۴۱.۸,۴۳.۱۲۵۸۷۴
۱.۵۷۴۲۵,۱۸,۲۷۲.۵۸۹۶۳۲,۲۹۰۱۱.۴۹۱۳۸,۲۲۳۶۰۰۰,۱۸۴۲۲۳۳۰۰۰۰۰۰۰۰۰,۷۵۸۹۶۳۲,۸۱۳۰۰۰۰۰,۱۰.۷۴۱,۴۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۳۵,۴۰.۲۵۸۹۶۳-,۲۰۱۵
۲۰۱۶,۸.۷۴۱۲۵۸,۱۸,۲۸۷.۹۶۴۰۹۴,۳۶۴۴۱.۷۷۴۲۸,۲۶۵۸۰۰۰,۲۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۸۵۴۱۲۵۸,۸۲۶۱۹۳۶۳,۱۲.۰۱۴,۳۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۴۳.۴,۳۸.۵۸۹۶۳۲
۲۰۱۷,۳.۶۹۸۵۲۱,۱۵,۳۱۷.۸۵۲۹۶۳,۴۲۸۹۰.۳,۳۲۱۴۰۰۰,۳۶۹۸۵۲۱۰۰۰۰۰۰۰۰۰,۸۹۸۷۶۵۴,۸۳۹۳۸۷۶۹,۱۱.۴۵۸,۲۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۵۰.۱,۳۵.۲۱۴۵۸۹
۲.۳۶۹۸۵۲,۱۵,۴۴۴.۲۵۸۹۶۳,۴۲۰۰۰,۳۸۵۶۹۰۰,۴۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۸۵۴۱۲۳۶,۸۵۲۵۸۱۷۵,۱۲.۳۶۹,۱۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۵۰.۱,۲۹.۲۲۸۲۲-,۲۰۱۸
۱.۲۵۸۹۶۳,۱۵,۶۲۶.۵۸۹۶۳۲,۴۲۰۰۰,۴۶۲۸۲۸۰,۵۸۹۶۳۲۱۰۰۰۰۰۰۰۰۰,۷۵۸۹۶۳۲,۸۶۵۷۷۵۸۱,۱۰.۳۶۹,۱۲۵۸۹۶۳۰۰۰۰۰۰۰۰۰,۴۹.۸,۳۰.۱۵۸۹۶۳-,۲۰۱۹
۲۰۲۰,۳.۳۲۵۸۹۶,۱۶,۸۵۲.۱۴۵۸۹۶,۴۲۰۰۰,۵۵۵۳۹۳۶,۶۹۸۵۴۱۲۰۰۰۰۰۰۰۰۰,۶۹۸۵۴۱۲,۸۷۹۰۰۰۰۰,۹.۴۵۸,۱۰۲۵۸۹۶۰۰۰۰۰۰۰۰۰,۴۷.۲,۳۲.۵۸۹۶۳۲
۲۰۲۱,۴.۷۱۲۵۸۹,۱۸,۱۲۴۵۸۹۶۳۲۱,۴۲۰۰۰,۶۶۶۴۷۲۳,۸۷۴۱۲۵۸۰۰۰۰۰۰۰۰۰,۶۵۸۹۶۳۲,۸۹۲۲۲۴۱۶,۹.۱۲۵,۱۱۲۵۸۹۷۰۰۰۰۰۰۰۰۰,۴۷.۲,۳۳.۶۹۸۵۲۱
۲۰۲۲,۲.۵۸۹۶۳۲,۲۰.۵,۱۸۵۲.۴۵۸۹۶۳,۴۲۰۰۰,۷۹۹۷۶۶۸,۱۲۵۸۹۶۳۰۰۰۰۰۰۰۰۰,۵۸۹۶۳۲۱,۹۰۰۰۰۰۰,۹.۰۴۱,۱۲۳۵۸۹۶۰۰۰۰۰۰۰۰۰,۵۱.۱,۳۴.۸۹۶۳۲۱
۲۰۲۳,۱.۸۹۶۳۲۵,۲۲.۵,۲۱۴۰.۲۱۹۴۲۹,۴۲۰۰۰,۸۵۹۵۶۰۰,۱۶۹۸۵۲۱۰۰۰۰۰۰۰۰۰,۴۵۸۹۶۳۲,۹۰۶۰۸۷۱۴,۹.۲۳۶,۱۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۵۱.۱,۳۶.۹۸۵۲۱۴
"""
gdp_data_csv = """Data Source,World Development Indicators
Last Updated,12/22/2023

"Year","GDP (current US$)"
۱۹۹۱,۵۲۹۰۰۰۰۰۰۰
۱۹۹۲,۶۲۱۰۰۰۰۰۰۰
۱۹۹۳,۶۶۸۰۰۰۰۰۰۰
۱۹۹۴,۵۲۵۰۰۰۰۰۰۰
۱۹۹۵,۷۴۵۰۰۰۰۰۰۰
۱۹۹۶,۹۳۳۰۰۰۰۰۰۰
```

۱۹۹۷,۹۷۶۰۰۰۰۰۰۰
۱۹۹۸,۹۹۵۰۰۰۰۰۰۰
۱۹۹۹,۹۵۷۰۰۰۰۰۰۰
۲۰۰۰,۱۰۷۰۰۰۰۰۰۰
۲۰۰۱,۱۱۸۰۰۰۰۰۰۰
۲۰۰۲,۱۲۹۰۰۰۰۰۰۰
۲۰۰۳,۱۵۸۰۰۰۰۰۰۰
۲۰۰۴,۱۹۵۰۰۰۰۰۰۰
۲۰۰۵,۲۴۸۰۰۰۰۰۰۰
۲۰۰۶,۲۹۴۰۰۰۰۰۰۰
۲۰۰۷,۳۷۰۰۰۰۰۰۰۰
۲۰۰۸,۴۴۳۰۰۰۰۰۰۰
۲۰۰۹,۴۸۸۰۰۰۰۰۰۰
۲۰۱۰,۵۹۱۰۰۰۰۰۰۰
۲۰۱۱,۶۶۸۰۰۰۰۰۰۰
۲۰۱۲,۶۵۲۰۰۰۰۰۰۰
۲۰۱۳,۵۳۹۰۰۰۰۰۰۰
۲۰۱۴,۴۶۹۰۰۰۰۰۰۰
۲۰۱۵,۳۸۴۰۰۰۰۰۰۰
۲۰۱۶,۴۲۰۰۰۰۰۰۰۰
۲۰۱۷,۴۵۸۰۰۰۰۰۰۰
۲۰۱۸,۲۹۵۰۰۰۰۰۰۰
۲۰۱۹,۲۵۸۰۰۰۰۰۰۰
۲۰۲۰,۲۰۷۰۰۰۰۰۰۰
۲۰۲۱,۲۹۱۰۰۰۰۰۰۰
۲۰۲۲,۳۸۹۰۰۰۰۰۰۰
۲۰۳,۴۱۳۰۰۰۰۰۰۰
"""

```
try:
    # --- ۱. آماده‌سازی دیتافریم‌ها ---
    df_raw = pd.read_csv(io.StringIO(raw_data_csv))
    df_gdp = pd.read_csv(io.StringIO(gdp_data_csv), skiprows=4, names=['year', 'nominal_gdp'])

    # --- ۲. انجام تمام تبدیلات ---
    df_raw['year'] = df_raw['year'].astype(int)
    df_gdp['year'] = df_gdp['year'].astype(int)
    df = pd.merge(df_raw, df_gdp, on='year')
    df.set_index('year', inplace=True)

    model_df = pd.DataFrame(index=df.index)
    model_df['GDP_growth'] = df['GDP_growth']
    model_df['Trade_Openness'] = df['Trade_Openness']
    model_df['Gov_Rev_to_GDP'] = (df['Government_Revenue'] / df['nominal_gdp']) * 100
    model_df['D_Log_Exchange_Rate'] = np.log(df['Exchange_Rate']).diff()
    model_df['D_Log_MoneyBase'] = np.log(df['MoneyBase']).diff()
    model_df['D_Unemployment'] = df['Unemployment'].diff()
    model_df['D_Log_investment'] = np.log(df['investment']).diff()
    model_df['D_Degree_economic_freedom'] = df['Degree_economic_freedom'].diff()
    df['Gov_Exp_to_GDP'] = (df['Government_expenditure'] / df['nominal_gdp']) * 100
    model_df['D_Gov_Exp_to_GDP'] = df['Gov_Exp_to_GDP'].diff()
    model_df['D2_Deposit_Rate'] = df['Deposit_Rate'].diff().diff()
    model_df['D2_Log_CPI'] = np.log(df['CPI']).diff().diff()

    # --- ۳. ساخت مجموعه داده نهایی برای مدل‌سازی ---
    p = 1
    q = 1
    y = model_df['GDP_growth']
    X_exog = model_df.drop('GDP_growth', axis=1)
    final_X_features = pd.DataFrame(index=model_df.index)
    for i in range(1, p + 1):
```

```
    final_X_features[f'GDP_growth_lag_{i}'] = y.shift(i)
for col in X_exog.columns:
    for i in range(q + 1):
        final_X_features[f'{col}_lag_{i}'] = X_exog[col].shift(i)

final_data = pd.concat([y, final_X_features], axis=1).dropna()
y_final = final_data['GDP_growth']
X_final = final_data.drop('GDP_growth', axis=1)

# --- ۴. تقسیم زمانی داده‌ها به آموزش و آزمون ---
train_end_year = 2015
X_train = X_final.loc[:train_end_year]
y_train = y_final.loc[:train_end_year]
X_test = X_final.loc[train_end_year+1:]
y_test = y_final.loc[train_end_year+1:]

# --- ۵. آموزش و ارزیابی مدل‌ها ---
results_list = []
models = {
    'Random Forest': RandomForestRegressor(n_estimators=100, random_state=42),
    'XGBoost': XGBRegressor(n_estimators=100, random_state=42)
}

for name, model in models.items():
    model.fit(X_train, y_train)
    preds = model.predict(X_test)

    results_list.append({
        'Model': name,
        'RMSE': np.sqrt(mean_squared_error(y_test, preds)),
        'MAE': mean_absolute_error(y_test, preds)
    })

results_df = pd.DataFrame(results_list)

# --- ۶. نمایش جدول مقایسه ---
print("--- جدول مقایسه عملکرد مدل‌ها (بر روی داده‌های آزمون) ---")
print(results_df.to_markdown(index=False, floatfmt=".4f"))

# --- ۷. استخراج و نمایش اهمیت ویژگی‌ها ---
print("\n--- اهمیت ویژگی‌ها (Feature Importances) ---")

# استخراج اهمیت ویژگی‌ها از هر دو مدل
rf_model = models['Random Forest']
xgb_model = models['XGBoost']

importances_df = pd.DataFrame({
    'Feature': X_train.columns,
    'Random Forest Importance': rf_model.feature_importances_,
    'XGBoost Importance': xgb_model.feature_importances_
})

# نرمال‌سازی اهمیت‌ها تا جمع آن‌ها ۱ شود
importances_df['Random Forest Importance'] /= importances_df['Random Forest Importance'].sum()
importances_df['XGBoost Importance'] /= importances_df['XGBoost Importance'].sum()

# مرتب‌سازی بر اساس میانگین اهمیت
importances_df['Average Importance'] = importances_df[['Random Forest Importance', 'XGBoost Importance']].mean(axis=1)
importances_df = importances_df.sort_values(by='Average Importance', ascending=False).reset_index(drop=True)

print(importances_df[['Feature', 'Random Forest Importance', 'XGBoost Importance', 'Average Importance']].to_markdown(index=False, floatfmt=".4f"))

# --- ۸. ترسیم نمودار اهمیت ویژگی‌ها ---
plt.figure(figsize=(12, 10))
# انتخاب ۱۰ ویژگی برتر
top_features = importances_df.head(۱۰)
```

```
    # تبدیل داده برای ترسیم کنار هم
    plot_df = pd.melt(top_features, id_vars='Feature','
                value_vars=['Random Forest Importance', 'XGBoost Importance'],
                var_name='Model', value_name='Importance('

    sns.barplot(data=plot_df, x='Importance', y='Feature', hue='Model', palette='viridis')
    plt.title('مقایسه اهمیت ۱۰ ویژگی برتر در مدل‌های جنگل تصادفی و XGBoost')
    plt.xlabel('میزان اهمیت نرمال‌شده')
    plt.ylabel('ویژگی (متغیر)')
    plt.legend(title='مدل')
    plt.tight_layout()
    plt.savefig('feature_importances_comparison.png')
    print('\n' نمودار اهمیت ویژگی‌ها در فایل feature_importances_comparison.png ذخیره شد.')

except Exception as e:
    print(f"یک خطای غیرمنتظره رخ داد: {e}")
```

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
import io

# --- داده‌های خام مستقیماً در کد تعبیه شده‌اند ---
# (داده‌ها مانند کد قبلی هستند و برای اختصار اینجا نمایش داده نمی‌شوند)
raw_data_csv = """year,GDP_growth,Deposit_Rate,CPI,Exchange_Rate,MoneyBase,Government_expenditure,Government_Revenue,Population,Unemployment,investment,Degree_economic_freedom,Trade_Openness
۱۹۹۱,۳.۲۹۷۹۸۲,۱۵,۳.۴۸۳۶۵۸,۶۷.۵۲۱۷۳,۸۸۳۲,۵۵۲۷۱۳۰۰۰۰۰۰۰,۶۵۴۹.۹,۵۹۹۸۶۹۳۱,۱۱.۱,۲۲۶۸۳۷۸۰۰۰۰۰۰۰,۳۶.۱,۴۴.۱۹۴۱۷۲
۱.۴۷۱۰۱۱,۱۵,۴.۳۸۲۷۱۱,۶۵.۵۶۷۸۴۱,۱۱۰۴۰,۶۹۸۱۰۵۵۰۰۰۰۰۰,۱۰۲۱۱.۵,۶۱۱۸۴۹۸۳,۱۰۶۵۳,۳۰۳۲۵۳۶۰۰۰۰۰۰۰,۳۶.۱,۴۰.۳۹۷۹۲۷-,۱۹۹۲
۱.۶۹۸۲۸۶,۱۵,۵.۳۱۱۹۶,۱۲۶۸.۰۷۸۷۶۱,۱۳۷۹۹,۱۵۱۰۳۷۱۰۰۰۰۰۰۰,۱۶۳۵۰.۱,۶۱۴۰۳۵۴۳,۱۰.۵۶۷,۳۲۲۵۹۴۷۰۰۰۰۰۰۰,۳۶.۱,۴۶.۰۴۰۸۶۴-,۱۹۹۳
۱۹۹۴,۲.۴۰۱۰۴۵,۱۵,۶.۹۸۲۴۱۴,۱۷۴۹.۱۷۳۷۳۶,۱۷۹۳۸.۷,۲۰۰۵۶۴۱۰۰۰۰۰۰۰,۲۲۱۰۸.۴,۶۱۵۲۰۵۹۸,۱۰.۴۴۸,۳۲۵۴۹۹۱۰۰۰۰۰۰۰,۳۶.۱,۴۱.۶۰۹۱۳۶
۱۹۹۵,۶.۳۵۰۵۲۶,۱۵,۱۰.۴۴۹۶۰۱,۱۷۴۸.۳۵۰۲۴۶,۲۴۲۱۷.۳,۲۸۱۲۸۱۸۰۰۰۰۰۰۰,۳۱۲۵۴.۹,۶۲۲۰۵۵۶۷,۹.۸۹۶,۵۹۵۶۶۶۵۰۰۰۰۰۰۰,۳۶.۱,۳۵.۱۴۴۲۴۵
۱۹۹۶,۸.۸۱۵۰۸۷,۱۵,۱۲.۸۷۱۵۷۹,۱۷۴۹.۵,۳۱۴۸۲.۵,۴۱۷۱۶۳۸۰۰۰۰۰۰۰,۴۳۲۵۶.۳,۶۲۸۸۵۹۹۴,۹.۴۵۸,۷۴۶۵۶۴۶۰۰۰۰۰۰۰,۳۸.۶,۳۳.۵۶۵۸۶۳
۱۹۹۷,۳.۰۴,۱۵,۱۵.۱۱۱۴۵۳,۱۷۵۱.۲۷۸۵۴,۴۰۱۴۹.۳,۵۵۸۸۷۲۵۰۰۰۰۰۰۰,۶۵۴۳۲.۸,۶۳۵۶۶۴۲۱,۱۰.۰۵۷,۹۸۷۹۷۹۸۰۰۰۰۰۰۰,۳۸.۶,۳۵.۲۴۹۸۲
۱۹۹۸,۳.۲۲۱۵۳,۱۵,۱۷.۸۵۹۶۶۴,۱۷۵۲.۵,۴۸۱۲۴.۶,۶۵۴۳۲۱۰۰۰۰۰۰۰۰,۷۸۹۶۵.۴,۶۴۲۴۶۸۴۸,۱۰.۳۵۸,۱۲۳۴۵۶۰۰۰۰۰۰۰۰,۳۹.۳,۳۱.۷۹۶۳۹۷
۱۹۹۹,۱.۰۰۷۳۸۵,۱۴,۲۱.۴۴۲۸۷,۱۷۵۳.۳۵۴۹۵,۶۰۱۵۵.۸,۵۶۱۲۴۵۴۰۰۰۰۰۰۰,۸۹۸۷۳.۳,۶۵۴۴۵۳۸۹,۱۱.۸۹۷,۱۵۴۵۶۵۰۰۰۰۰۰۰۰,۴۰.۳,۳۷.۸۹۷۶۵۴
۲۰۰۰,۵.۰۸۶۴۳۸,۱۳.۵,۲۵.۸۸۲۱۹۳,۷۹۹۶.۱۸۵۳۶۸,۷۵۱۹۴.۸,۶۵۴۳۲۱۰۰۰۰۰۰۰۰,۱۰۲۳۵۴,۶۶۱۲۸۹۸۷,۱۳.۶۷۹,۲۳۱۲۵۴۰۰۰۰۰۰۰۰,۴۰.۳,۴۲.۵۸۹۶۳۲
۲۰۰۱,۳.۵۸۵۵۷,۱۲.۴,۲۸.۸۴۷۵۲۶,۷۹۲۰.۶۶۹۸۶۶,۹۳۹۹۳.۵,۷۸۹۶۵۴۰۰۰۰۰۰۰۰,۱۲۵۸۹۶,۶۶۸۱۲۵۵۴,۱۲.۷۸۴,۳۲۱۴۵۸۰۰۰۰۰۰۰۰,۴۱.۲,۴۰.۵۳۵۵۶۶
۲۰۰۲,۷.۴۶۴۶۰۱,۱۱.۲,۳۳.۳۷۴۲۴۵,۷۹۶۸.۴۲۱۰۳۴,۱۲۲۱۹۱.۶,۹۸۷۵۶۳۰۰۰۰۰۰۰۰,۱۵۸۹۶۳,۶۷۴۹۶۱۲۱,۱۱.۷۷۲,۴۵۸۹۶۳۰۰۰۰۰۰۰۰,۴۱.۲,۴۸.۷۸۹۶۵۴
۲۰۰۳,۷.۱۸۹۷۳۶,۱۲.۵,۳۸.۵۸۶۱۱۱,۸۱۸۲.۲۶۰۵۲,۱۶۵۰۳۸.۸,۱۲۵۸۹۷۰۰۰۰۰۰۰۰۰,۲۰۱۵۸۹,۶۸۱۷۹۶۸۸,۱۰.۹۷۵,۵۸۹۶۳۲۰۰۰۰۰۰۰۰,۴۱.۵,۵۲.۴۸۹۶۳۲
۲۰۰۴,۴.۹۹۹۷۹۵,۱۳.۸,۴۴.۴۲۶۷۴۳,۸۷۳۳.۸۶۱۰۷۸,۲۱۷۸۵۱,۱۵۴۸۷۹۰۰۰۰۰۰۰۰۰,۲۵۴۸۹۶,۶۸۸۶۳۲۵۵,۱۰.۵۶۹,۸۷۹۶۵۴۰۰۰۰۰۰۰۰,۴۱.۵,۵۸.۵۶۵۰۹
۲۰۰۵,۳.۱۸۱۸۷۸,۱۵.۱,۵۰.۸۸۷۲۵,۹۰۲۶.۳۴۷۵۰۹,۲۹۸۴۸۵.۴,۲۶۵۸۹۷۰۰۰۰۰۰۰۰۰,۵۸۹۷۴۱,۷۲۲۲۴۵۳۱,۱۰.۷۷۲,۱۱۲۵۸۷۰۰۰۰۰۰۰۰۰۰,۴۲.۹,۵۶.۳۲۱۴۵۸
۲۰۰۶,۶.۶۳۴۱۲,۱۵.۱,۵۶.۹۶۵۸۷۴,۹۲۳۲.۵,۳۵۸۳۰۰,۳۰۷۷۵۰۲۰۰۰۰۰۰۰۰,۹۳۵۴۴۱.۷,۷۳۷۶۵۴۸۹,۱۰.۴۶۵,۱۲۹۰۱۲۰۰۰۰۰۰۰۰۰۰,۴۳.۲,۴۹.۸۸۷۳۰۷
۲۰۰۷,۷.۷۸۹۴۵۶,۱۵.۱,۶۳.۷۸۶۳۳۲,۹۲۸۱.۱۵۱۸۲۸,۴۵۸۶۰۰,۴۱۹۶۷۲۸۰۰۰۰۰۰۰۰,۱۲۶۸۳۸۳,۷۴۶۰۲۶۴۷,۱۰.۱۵۸,۱۵۸۹۶۳۰۰۰۰۰۰۰۰۰,۴۳.۲,۴۷.۸۹۶۵۴۱
۲۰۰۸,۰.۵۸۹۶۳۲,۱۵.۱,۷۹.۵۶۸۹۷۴,۹۷۲۵.۵,۵۶۸۹۰۰,۵۶۹۸۴۱۲۰۰۰۰۰۰۰۰,۱۸۹۶۵۴۱,۷۵۴۳۹۸۰۵,۱۰.۰۴۵,۱۹۸۷۵۴۰۰۰۰۰۰۰۰۰۰,۴۲.۴,۵۱.۲۵۸۹۶۳
۲۰۰۹,۱.۵۸۹۶۳۲,۱۵,۸۷.۳۶۹۸۵۲,۱۰۰۵۱.۷۸۵,۶۸۹۷۰۰۶۸۴۲۱۳۵۰۰۰۰۰۰۰۰,۲۱۳۴۵۶۸,۷۶۲۷۶۹۶۳,۱۱.۵۸۹,۲۶۵۸۹۷۰۰۰۰۰۰۰۰۰۰,۴۲.۴,۴۵.۲۱۳۶۹۸
۲۰۱۰,۵.۸۹۶۳۲۱,۱۵,۹۷.۲۵۸۹۶۳,۱۰۲۵۴.۲,۸۵۴۲۰۰,۸۵۴۱۲۳۶۰۰۰۰۰۰۰۰,۲۸۹۶۵۴۱,۷۷۱۱۴۱۲۱,۱۳.۰۴۵,۳۵۸۹۶۴۰۰۰۰۰۰۰۰۰۰,۴۴.۸,۴۰.۱۲۵۸۹۶
۲۰۱۱,۳.۲۵۸۹۶۳,۱۵,۱۱۸.۵۸۹۶۳۲,۱۲۲۶۰,۱۰۲۵۰۰۰,۱۰۲۵۸۷۴۰۰۰۰۰۰۰۰,۳۵۸۹۶۴۱,۷۷۹۵۱۲۷۹,۱۲.۱۱۳,۴۱۹۴۳۱۳۰۰۰۰۰۰۰۰۰,۴۴.۸,۴۹.۵۲۸۷۴۱
۳.۷۴۷۱۷۱,۱۵,۱۵۴.۸۹۶۳۲۱,۱۲۲۶۰,۱۲۳۰۰۰,۱۲۳۵۸۷۴۰۰۰۰۰۰۰۰,۴۵۸۹۶۳۲,۷۸۷۸۸۴۳۷,۱۲.۰۱۲,۴۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۴۲,۴۸.۹۶۵۴۱۲-,۲۰۱۲
۲۰۱۳,۱.۲۵۸۹۶۳,۱۵,۲۰۸.۵۸۹۶۳۲,۲۴۷۷۷.۵,۱۵۱۸۷۰۰,۱۵۴۷۸۹۶۰۰۰۰۰۰۰۰۰,۵۸۹۶۳۲۱,۷۹۶۲۵۵۹۵,۱۰.۱۴۵,۵۱۲۵۸۷۴۰۰۰۰۰۰۰۰۰,۴۲,۴۶.۳۲۵۸۹۶
۲۰۱۴,۴.۵۸۹۶۳۲,۱۵,۲۴۰.۲۵۸۹۶۳,۲۶۵۱۰.۳,۱۸۴۲۳۰۰,۱۸۸۴۴۳۳۰۰۰۰۰۰۰۰۰,۶۹۸۵۴۱۲,۸۰۴۶۲۷۵۳,۱۰.۲۵۶,۵۵۵۴۵۲۰۰۰۰۰۰۰۰۰۰,۴۱.۸,۴۳.۱۲۵۸۷۴
۱.۵۷۴۲۵,۱۸,۲۷۲.۵۸۹۶۳۲,۲۹۰۱۱.۴۹۱۳۸,۲۲۳۶۰۰۰,۱۸۴۲۲۳۳۰۰۰۰۰۰۰۰۰,۷۵۸۹۶۳۲,۸۱۳۰۰۰۰۰,۱۰.۷۴۱,۴۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۳۵,۴۰.۲۵۸۹۶۳-,۲۰۱۵
۲۰۱۶,۸.۷۴۱۲۵۸,۱۸,۲۸۷.۹۶۴۰۹۴,۳۶۴۴۱.۷۷۴۲۸,۲۶۵۸۰۰۰,۲۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۸۵۴۱۲۵۸,۸۲۶۱۹۳۶۳,۱۲.۰۱۴,۳۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۴۳.۴,۳۸.۵۸۹۶۳۲
۲۰۱۷,۳.۶۹۸۵۲۱,۱۵,۳۱۷.۸۵۲۹۶۳,۴۲۸۹۰.۳,۳۲۱۴۰۰۰,۳۶۹۸۵۲۱۰۰۰۰۰۰۰۰۰,۸۹۸۷۶۵۴,۸۳۹۳۸۷۶۹,۱۱.۴۵۸,۲۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۵۰.۱,۳۵.۲۱۴۵۸۹
۲.۳۶۹۸۵۲,۱۵,۴۴۴.۲۵۸۹۶۳,۴۲۰۰۰,۳۸۵۶۹۰۰,۴۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۸۵۴۱۲۳۶,۸۵۲۵۸۱۷۵,۱۲.۳۶۹,۱۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۵۰.۱,۲۹.۲۲۸۲۲-,۲۰۱۸
۱.۲۵۸۹۶۳,۱۵,۶۲۶.۵۸۹۶۳۲,۴۲۰۰۰,۴۶۲۸۲۸۰,۵۸۹۶۳۲۱۰۰۰۰۰۰۰۰۰,۷۵۸۹۶۳۲,۸۶۵۷۷۵۸۱,۱۰.۳۶۹,۱۲۵۸۹۶۳۰۰۰۰۰۰۰۰۰,۴۹.۸,۳۰.۱۵۸۹۶۳-,۲۰۱۹
۲۰۲۰,۳.۳۲۵۸۹۶,۱۶,۸۵۲.۱۴۵۸۹۶,۴۲۰۰۰,۵۵۵۳۹۳۶,۶۹۸۵۴۱۲۰۰۰۰۰۰۰۰۰,۶۹۸۵۴۱۲,۸۷۹۰۰۰۰۰,۹.۴۵۸,۱۰۲۵۸۹۶۰۰۰۰۰۰۰۰۰,۴۷.۲,۳۲.۵۸۹۶۳۲
۲۰۲۱,۴.۷۱۲۵۸۹,۱۸,۱۲۴۵.۸۹۶۳۲۱,۴۲۰۰۰,۶۶۶۴۷۲۳,۸۷۴۱۲۵۸۰۰۰۰۰۰۰۰۰,۶۵۸۹۶۳۲,۸۹۲۲۲۴۱۶,۹.۱۲۵,۱۱۲۵۸۹۷۰۰۰۰۰۰۰۰۰,۴۷.۲,۳۳.۶۹۸۵۲۱
```

۲۰۲۲,۲.۵۸۹۶۳۲,۲۰.۵,۱۸۵۲.۴۵۸۹۶۳,۴۲۰۰۰,۷۹۹۷۶۶۸,۱۲۵۸۹۶۳۰۰۰۰۰۰۰۰۰,۵۸۹۶۳۲۱,۹۰۰۰۰۰۰۰,۹.۰۴۱,۱۲۳۵۸۹۶۰۰۰۰۰۰۰۰۰,۵۱.۱,۳۴.۸۹۶۳۲۱

۲۰۲۳,۱.۸۹۶۳۲۵,۲۲.۵,۲۱۴۰.۲۱۹۴۲۹,۴۲۰۰۰,۸۵۹۵۶۰۰,۱۶۹۸۵۲۱۰۰۰۰۰۰۰۰۰۰,۴۵۸۹۶۳۲,۹۰۶۰۸۷۱۴,۹.۲۳۶,۱۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۵۱.۱,۳۶.۹۸۵۲۱۴

"""

```
gdp_data_csv = """Data Source,World Development Indicators
Last Updated,12/22/2023

"Year","GDP (current US$)"
۱۹۹۱,۵۲۹۰۰۰۰۰۰۰
۱۹۹۲,۶۲۱۰۰۰۰۰۰۰
۱۹۹۳,۶۶۸۰۰۰۰۰۰۰
۱۹۹۴,۵۲۵۰۰۰۰۰۰۰
۱۹۹۵,۷۴۵۰۰۰۰۰۰۰
۱۹۹۶,۹۳۳۰۰۰۰۰۰۰
۱۹۹۷,۹۷۶۰۰۰۰۰۰۰
۱۹۹۸,۹۹۵۰۰۰۰۰۰۰
۱۹۹۹,۹۵۷۰۰۰۰۰۰۰
۲۰۰۰,۱۰۷۰۰۰۰۰۰۰
۲۰۰۱,۱۱۸۰۰۰۰۰۰۰
۲۰۰۲,۱۲۹۰۰۰۰۰۰۰
۲۰۰۳,۱۵۸۰۰۰۰۰۰۰
۲۰۰۴,۱۹۵۰۰۰۰۰۰۰
۲۰۰۵,۲۴۸۰۰۰۰۰۰۰
۲۰۰۶,۲۹۴۰۰۰۰۰۰۰
۲۰۰۷,۳۷۰۰۰۰۰۰۰۰
۲۰۰۸,۴۴۳۰۰۰۰۰۰۰
۲۰۰۹,۴۸۸۰۰۰۰۰۰۰
۲۰۱۰,۵۹۱۰۰۰۰۰۰۰
۲۰۱۱,۶۶۸۰۰۰۰۰۰۰
۲۰۱۲,۶۵۲۰۰۰۰۰۰۰
۲۰۱۳,۵۳۹۰۰۰۰۰۰۰
۲۰۱۴,۴۶۹۰۰۰۰۰۰۰
۲۰۱۵,۳۸۴۰۰۰۰۰۰۰
۲۰۱۶,۴۲۰۰۰۰۰۰۰۰
۲۰۱۷,۴۵۸۰۰۰۰۰۰۰
۲۰۱۸,۲۹۵۰۰۰۰۰۰۰
۲۰۱۹,۲۵۸۰۰۰۰۰۰۰
۲۰۲۰,۲۰۷۰۰۰۰۰۰۰
۲۰۲۱,۲۹۱۰۰۰۰۰۰۰
۲۰۲۲,۳۸۹۰۰۰۰۰۰۰
۲۰۲۳,۴۱۳۰۰۰۰۰۰۰
"""

try:
    # --- ۱. آماده‌سازی دیتافریم‌ها ---
    df_raw = pd.read_csv(io.StringIO(raw_data_csv))
    df_gdp = pd.read_csv(io.StringIO(gdp_data_csv), skiprows=4, names=['year', 'nominal_gdp'])

    # --- ۲. انجام تمام تبدیلات ---
    df_raw['year'] = df_raw['year'].astype(int)
    df_gdp['year'] = df_gdp['year'].astype(int)
    df = pd.merge(df_raw, df_gdp, on='year')
    df.set_index('year', inplace=True)

    model_df = pd.DataFrame(index=df.index)
    model_df['GDP_growth'] = df['GDP_growth']
    model_df['Trade_Openness'] = df['Trade_Openness']
    model_df['Gov_Rev_to_GDP'] = (df['Government_Revenue'] / df['nominal_gdp']) * 100
    model_df['D_Log_Exchange_Rate'] = np.log(df['Exchange_Rate']).diff()

    model_df['D_Log_MoneyBase'] = np.log(df['MoneyBase']).diff()

    model_df['D_Unemployment'] = df['Unemployment'].diff()
```

```
model_df['D_Log_investment'] = np.log(df['investment']).diff()
model_df['D_Degree_economic_freedom'] = df['Degree_economic_freedom'].diff()
df['Gov_Exp_to_GDP'] = (df['Government_expenditure'] / df['nominal_gdp']) * 100
model_df['D_Gov_Exp_to_GDP'] = df['Gov_Exp_to_GDP'].diff()
model_df['D2_Deposit_Rate'] = df['Deposit_Rate'].diff().diff()
model_df['D2_Log_CPI'] = np.log(df['CPI']).diff().diff()

۳. ساخت مدل اولیه # --- ARDL--- (۱ ,۱)
p = 1
q = 1
y = model_df['GDP_growth']
X_exog = model_df.drop('GDP_growth', axis=1)
final_X = pd.DataFrame(index=model_df.index)
for i in range(1, p + 1):
    final_X[f'GDP_growth_lag_{i}'] = y.shift(i)
for col in X_exog.columns:
    for i in range(q + 1):
        final_X[f'{col}_lag_{i}'] = X_exog[col].shift(i)

حذف مقادیر # NaN
final_data = pd.concat([y, final_X], axis=1).dropna()
y_final = final_data['GDP_growth']
X_final = final_data.drop('GDP_growth', axis=1)

--- ۴. اجرای فرآیند حذف گام به گام به عقب # ---
alpha = 0.1
features = list(X_final.columns)

while len(features) > 0:
    X_const = sm.add_constant(X_final[features])
    model = sm.OLS(y_final, X_const).fit()
    p_values = model.pvalues

    p_values_features = p_values.drop('const', errors='ignore')
    if p_values_features.empty:
        break

    feature_to_remove = p_values_features.idxmax()
    max_p_value = p_values_features.max()

    if max_p_value > alpha:
        features.remove(feature_to_remove)
    else:
        break

--- ۵. برآورد و نمایش مدل نهایی و بهینه --- #
print("--- نتایج مدل نهایی بهینه اقتصادسنجی (برای کل دوره) ---")
if features:
    X_final_model = sm.add_constant(X_final[features])
    final_parsimonious_model = sm.OLS(y_final, X_final_model).fit()
    print(final_parsimonious_model.summary())
else:
    print("هیچ متغیر معناداری برای ساخت مدل نهایی یافت نشد.")

except Exception as e:
print)f:داد رخ منتظره غیر خطای یک "{e}(")
```

مدل یادگیری ماشین

```
import pandas as pd
import numpy as np
import io
from sklearn.ensemble import RandomForestRegressor
from xgboost import XGBRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# --- داده‌های خام مستقیماً در کد تعبیه شده‌اند ---
raw_data_csv = """year,GDP_growth,Deposit_Rate,CPI,Exchange_Rate,MoneyBase,Government_expenditure,Government_Revenue,Population,Unemployment,investment,Degree_economic_freedom,Trade_Openness
۱۹۹۱,۳.۲۹۷۹۸۲,۱۵,۳.۴۸۳۶۵۸,۶۷.۵۲۱۷۳,۸۸۳۲,۵۵۲۷۱۳۰۰۰۰۰۰۰,۶۵۴۹.۹,۵۹۹۸۶۹۳۱,۱۱.۱,۲۲۶۸۳۷۸۰۰۰۰۰۰۰,۳۶.۱,۴۴.۱۹۴۱۷۲
۱.۴۷۱۰۱۱,۱۵,۴.۳۸۲۷۱۱,۶۵.۵۶۷۸۴۱,۱۱۰۴۰,۶۹۸۱۰۵۵۰۰۰۰۰۰,۱۰۲۱۱.۵,۶۱۱۸۴۹۸۳,۱۰۶۵۳,۳۰۳۲۵۳۶۰۰۰۰۰۰۰,۳۶.۱,۴۰.۳۹۷۹۲۷-,۱۹۹۲
۱۶۹۸۲۸۶,۱۵,۵.۳۱۱۹۶,۱۲۶۸.۰۷۸۷۶۱,۱۳۷۹۹,۱۵۱۰۳۷۱۰۰۰۰۰۰۰۰,۱۶۳۵۰.۱,۶۱۴۰۳۵۴۳,۱۰.۵۶۷,۳۲۲۵۹۴۷۰۰۰۰۰۰۰,۳۶.۱,۴۶.۰۴۰۸۶۴-,۱۹۹۳
۱۹۹۴,۲.۴۰۱۰۴۵,۱۵,۶.۹۸۲۴۱۴,۱۷۴۹.۱۷۳۷۳۶,۱۷۹۳۸.۷,۲۰۰۵۶۴۱۰۰۰۰۰۰۰۰,۲۲۱۰۸.۴,۶۱۵۲۰۵۹۸,۱۰.۴۴۸,۳۲۵۴۹۹۱۰۰۰۰۰۰۰۰,۳۶.۱,۴۱.۶۰۹۱۳۶
۱۹۹۵,۶.۳۵۰۵۲۶,۱۵,۱۰.۴۴۹۶۰۱,۱۷۴۸.۳۵۰۲۴۶,۲۴۲۱۷.۳,۲۸۱۲۸۱۸۰۰۰۰۰۰۰۰,۳۱۲۵۴.۹,۶۲۲۰۵۵۶۷,۹.۸۹۶,۵۹۵۶۶۶۵۰۰۰۰۰۰۰۰,۳۶.۱,۳۵.۱۴۴۲۴۵
۱۹۹۶,۸.۸۱۵۰۸۷,۱۵,۱۲.۸۷۱۵۷۹,۱۷۴۹.۵,۳۱۴۸۲.۵,۴۱۷۱۶۳۸۰۰۰۰۰۰۰۰,۴۳۲۵۶.۳,۶۲۸۸۵۹۹۴,۹.۴۵۸,۷۴۶۵۶۴۶۰۰۰۰۰۰۰۰,۳۸.۶,۳۳.۵۶۵۸۶۳
۱۹۹۷,۳.۰۴,۱۵,۱۵.۱۱۱۴۵۳,۱۷۵۱.۲۷۸۵۴,۴۰۱۴۹.۳,۵۵۸۸۷۲۵۰۰۰۰۰۰۰۰,۶۵۴۳۲.۸,۶۳۵۶۶۴۲۱,۱۰.۰۵۷,۹۸۷۹۷۹۸۰۰۰۰۰۰۰۰,۳۸.۶,۳۵.۲۴۹۸۲
۱۹۹۸,۳.۲۲۱۵۳,۱۵,۱۷.۸۵۹۶۶۴,۱۷۵۲.۵,۴۸۱۲۴.۶,۶۵۴۳۲۱۰۰۰۰۰۰۰۰,۷۸۹۶۵.۴,۶۴۲۴۶۸۴۸,۱۰.۳۵۸,۱۲۳۴۵۶۰۰۰۰۰۰۰۰۰,۳۹.۳,۳۱.۷۹۶۳۹۷
۱۹۹۹,۱.۰۰۷۳۸۵,۱۴,۲۱.۴۴۲۸۷,۱۷۵۳.۳۵۴۹۵,۶۰۱۵۵.۸,۵۶۱۲۴۵۴۰۰۰۰۰۰۰۰,۸۹۸۷۳.۳,۶۵۴۴۵۳۸۹,۱۱.۸۹۷,۱۵۴۵۶۵۰۰۰۰۰۰۰۰۰,۴۰.۳,۳۷.۸۹۷۶۵۴
۲۰۰۰,۵.۰۸۶۴۳۸,۱۳.۵,۲۵.۸۸۲۱۹۳,۷۹۹۶.۱۸۵۳۶۸,۷۵۱۹۴.۸,۶۵۴۳۲۱۰۰۰۰۰۰۰۰۰,۱۰۲۳۵۴,۶۶۱۲۸۹۸۷,۱۳.۶۷۹,۲۳۱۲۵۴۰۰۰۰۰۰۰۰۰,۴۰.۳,۴۲.۵۸۹۶۳۲
۲۰۰۱,۳.۵۸۵۵۷,۱۲.۴,۲۸.۸۴۷۵۲۶,۷۹۲۰.۶۶۹۸۶۶,۹۳۹۹۳.۵,۷۸۹۶۵۴۰۰۰۰۰۰۰۰۰,۱۲۵۸۹۶,۶۶۸۱۲۵۵۴,۱۲.۷۸۴,۳۲۱۴۵۸۰۰۰۰۰۰۰۰۰,۴۱.۲,۴۰.۵۳۵۵۶۶
۲۰۰۲,۷.۴۶۴۶۰۱,۱۱.۲,۳۳.۳۷۴۲۴۵,۷۹۶۸.۴۲۱۰۳۴,۱۲۲۱۹۱۶,۹۸۷۵۶۳۰۰۰۰۰۰۰۰۰,۱۵۸۹۶۳,۶۷۴۹۶۱۲۱,۱۱.۷۷۲,۴۵۸۹۶۳۰۰۰۰۰۰۰۰۰,۴۱.۲,۴۸.۷۸۹۶۵۴
۲۰۰۳,۷.۱۸۹۷۳۶,۱۲.۵,۳۸.۵۸۶۱۱۱,۸۱۸۲.۲۶۰۵۲,۱۶۵۰۳۸.۸,۱۲۵۸۹۷۰۰۰۰۰۰۰۰۰,۲۰۱۵۸۹,۶۸۱۷۹۶۸۸,۱۰.۹۷۵,۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۴۱.۵,۵۲.۴۸۹۶۳۲
۲۰۰۴,۴.۹۹۹۷۹۵,۱۳.۸,۴۴.۴۲۶۷۴۳,۸۷۳۳.۸۶۱۰۷۸,۲۱۷۸۵۱,۱۵۴۸۷۹۰۰۰۰۰۰۰۰۰,۲۵۴۸۹۶,۶۸۸۶۳۲۵۵,۱۰.۵۶۹,۸۷۹۶۵۴۰۰۰۰۰۰۰۰۰,۴۱.۵,۵۸.۵۶۵۰۹
۲۰۰۵,۳.۱۸۱۸۷۸,۱۵.۱,۵۰.۸۸۷۲۵,۹۰۲۶.۳۴۷۵۰۹,۲۹۸۴۸۵.۴,۲۶۵۸۹۷۰۰۰۰۰۰۰۰۰,۵۸۹۷۴۱,۷۲۲۲۴۵۳۱,۱۰.۷۷۲,۱۱۲۵۸۷۰۰۰۰۰۰۰۰۰,۴۲.۹,۵۶.۳۲۱۴۵۸
۲۰۰۶,۶.۶۳۴۱۲,۱۵.۱,۵۶.۹۶۵۸۷۴,۹۲۳۲.۵,۳۵۸۳۰۰,۳۰۷۷۵۰۲۰۰۰۰۰۰۰۰,۹۳۵۴۴۱.۷,۷۳۷۶۵۴۸۹,۱۰.۴۶۵,۱۲۹۰۱۲۰۰۰۰۰۰۰۰۰,۴۳.۲,۴۹.۸۸۷۳۰۷
۲۰۰۷,۷.۷۸۹۴۵۶,۱۵.۱,۶۳.۷۸۶۳۳۲,۹۲۸۱.۱۵۱۸۲۸,۴۵۸۶۰۰,۴۱۹۶۷۲۸۰۰۰۰۰۰۰۰,۱۲۶۸۳۸۳,۷۴۶۰۲۶۴۷,۱۰.۱۵۸,۱۵۸۹۶۳۰۰۰۰۰۰۰۰۰,۴۳.۲,۴۷.۸۹۶۵۴۱
۲۰۰۸,۰.۵۸۹۶۳۲,۱۵.۱,۷۹.۵۶۸۹۷۴,۹۷۲۵.۵,۵۶۸۹۰۰,۵۶۹۸۴۱۲۰۰۰۰۰۰۰۰,۱۸۹۶۵۴۱,۷۵۴۳۹۸۰۵,۱۰.۰۴۵,۱۹۸۷۵۴۰۰۰۰۰۰۰۰۰,۴۲.۴,۵۱.۲۵۸۹۶۳
۲۰۰۹,۱.۵۸۹۶۳۲,۱۵,۸۷.۳۶۹۸۵۲,۱۰۰۵۱.۷۸۵,۶۸۹۷۰۰,۶۸۴۲۱۳۵۰۰۰۰۰۰۰۰,۲۱۳۴۵۶۸,۷۶۲۷۶۹۶۳,۱۱.۵۸۹,۲۶۵۸۹۷۰۰۰۰۰۰۰۰۰,۴۲.۴,۴۵.۲۱۳۶۹۸
۲۰۱۰,۵.۸۹۶۳۲۱,۱۵,۹۷.۲۵۸۹۶۳,۱۰۲۵۴.۲,۸۵۴۲۰۰,۸۵۴۱۲۳۶۰۰۰۰۰۰۰۰,۲۸۹۶۵۴۱,۷۷۱۱۴۱۲۱,۱۳.۰۴۵,۳۵۸۹۶۴۰۰۰۰۰۰۰۰۰,۴۴.۸,۴۰.۱۲۵۸۹۶
۲۰۱۱,۳.۲۵۸۹۶۳,۱۵,۱۱۸.۵۸۹۶۳۲,۱۲۲۶۰,۱۰۲۵۰۰۰,۱۰۲۵۸۷۴۰۰۰۰۰۰۰۰۰,۳۵۸۹۶۴۱,۷۷۹۵۱۲۷۹,۱۲.۱۱۳,۴۱۹۴۳۱۳۰۰۰۰۰۰۰۰۰,۴۴.۸,۴۹.۵۲۸۷۴۱
۳.۷۴۷۱۷۱,۱۵,۱۵۴.۸۹۶۳۲۱,۱۲۲۶۰,۱۲۳۰۰۰۰,۱۲۳۵۸۷۴۰۰۰۰۰۰۰۰۰,۴۵۸۹۶۳۲,۷۸۷۸۸۴۳۷,۱۲.۰۱۲,۴۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۴۲,۴۸.۹۶۵۴۱۲-,۲۰۱۲
۲۰۱۳,۱.۲۵۸۹۶۳,۱۵,۲۰۸.۵۸۹۶۳۲,۲۴۷۷۷.۵,۱۵۱۸۷۰۰,۱۵۴۷۸۹۶۰۰۰۰۰۰۰۰۰,۵۸۹۶۳۲۱,۷۹۶۲۵۵۹۵,۱۰.۱۴۵,۵۱۲۵۸۷۴۰۰۰۰۰۰۰۰۰,۴۲,۴۶.۳۲۵۸۹۶
۲۰۱۴,۴.۵۸۹۶۳۲,۱۵,۲۴۰.۲۵۸۹۶۳,۲۶۵۱۰.۳,۱۸۴۲۳۰۰,۱۸۸۴۴۳۳۰۰۰۰۰۰۰۰۰,۶۹۸۵۴۱۲,۸۰۴۶۲۷۵۳,۱۰.۲۵۶,۵۵۵۴۵۲۰۰۰۰۰۰۰۰۰,۴۱.۸,۴۳.۱۲۵۸۷۴
۱.۵۷۴۲۵,۱۸,۲۷۲.۵۸۹۶۳۲,۲۹۰۱۱.۴۹۱۳۸,۲۲۳۶۰۰۰,۱۸۴۲۲۳۳۰۰۰۰۰۰۰۰۰,۷۵۸۹۶۳۲,۸۱۳۰۰۰۰۰,۱۰.۷۴۱,۴۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۳۵,۴۰.۲۵۸۹۶۳-,۲۰۱۵
۲۰۱۶,۸.۷۴۱۲۵۸,۱۸,۲۸۷.۹۶۴۰۹۴,۳۶۴۴۱.۷۷۴۲۸,۲۶۵۸۰۰۰,۲۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۸۵۴۱۲۵۸,۸۲۶۱۹۳۶۳,۱۲.۰۱۴,۳۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۴۳.۴,۳۸.۵۸۹۶۳۲
۲۰۱۷,۳۶۹۸۵۲۱,۱۵,۳۱۷.۸۵۲۹۶۳,۴۲۸۹۰.۳,۳۲۱۴۰۰۰,۳۶۹۸۵۲۱۰۰۰۰۰۰۰۰۰,۸۹۸۷۶۵۴,۸۳۹۳۸۷۶۹,۱۱.۴۵۸,۲۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۵۰.۱,۳۵.۲۱۴۵۸۹
۲.۳۶۹۸۵۲,۱۵,۴۴۴.۲۵۸۹۶۳,۴۲۰۰۰,۳۸۵۶۹۰۰,۴۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۸۵۴۱۲۳۶,۸۵۲۵۸۱۷۵,۱۲.۳۶۹,۱۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۵۰.۱,۲۹.۲۲۸۲۲-,۲۰۱۸
۱.۲۵۸۹۶۳,۱۵,۶۲۶.۵۸۹۶۳۲,۴۲۰۰۰,۴۶۲۸۲۸۰,۵۸۹۶۳۲۱۰۰۰۰۰۰۰۰۰,۷۵۸۹۶۳۲,۸۶۵۷۷۵۸۱,۱۰.۳۶۹,۱۲۵۸۹۶۳۰۰۰۰۰۰۰۰۰,۴۹.۸,۳۰.۱۵۸۹۶۳-,۲۰۱۹
۲۰۲۰,۳.۳۲۵۸۹۶,۱۶,۸۵۲.۱۴۵۸۹۶,۴۲۰۰۰,۵۵۵۳۹۳۶,۶۹۸۵۴۱۲۰۰۰۰۰۰۰۰۰,۶۹۸۵۴۱۲,۸۷۹۰۰۰۰۰,۹.۴۵۸,۱۰۲۵۸۹۶۰۰۰۰۰۰۰۰۰,۴۷.۲,۳۲.۵۸۹۶۳۲
۲۰۲۱,۴.۷۱۲۵۸۹,۱۸,۱۲۴۵.۸۹۶۳۲۱,۴۲۰۰۰,۶۶۶۴۷۲۳,۸۷۴۱۲۵۸۰۰۰۰۰۰۰۰۰,۶۵۸۹۶۳۲,۸۹۲۲۲۴۱۶,۹.۱۲۵,۱۱۲۵۸۹۷۰۰۰۰۰۰۰۰۰,۴۷.۲,۳۳.۶۹۸۵۲۱
۲۰۲۲,۲.۵۸۹۶۳۲,۲۰.۵,۱۸۵۲.۴۵۸۹۶۳,۴۲۰۰۰,۷۹۹۷۶۶۸,۱۲۵۸۹۶۳۰۰۰۰۰۰۰۰۰,۵۸۹۶۳۲۱,۹۰۰۰۰۰۰۰,۹.۰۴۱,۱۲۳۵۸۹۶۰۰۰۰۰۰۰۰۰,۵۱.۱,۳۴.۸۹۶۳۲۱
۲۰۲۳,۱.۸۹۶۳۲۵,۲۲.۵,۲۱۴۰.۲۱۹۴۲۹,۴۲۰۰۰,۸۵۹۵۶۰۰,۱۶۹۸۵۲۱۰۰۰۰۰۰۰۰۰,۴۵۸۹۶۳۲,۹۰۶۰۸۷۱۴,۹.۲۳۶,۱۵۸۹۶۳۲۰۰۰۰۰۰۰۰۰,۵۱.۱,۳۶.۹۸۵۲۱۴
"""
gdp_data_csv = """Data Source,World Development Indicators
Last Updated,12/22/2023

"Year","GDP (current US$)"
۱۹۹۱,۵۲۹۰۰۰۰۰۰۰
۱۹۹۲,۶۲۱۰۰۰۰۰۰۰
۱۹۹۳,۶۶۸۰۰۰۰۰۰۰
۱۹۹۴,۵۲۵۰۰۰۰۰۰۰
۱۹۹۵,۷۴۵۰۰۰۰۰۰۰
۱۹۹۶,۹۳۳۰۰۰۰۰۰۰
۱۹۹۷,۹۷۶۰۰۰۰۰۰۰
۱۹۹۸,۹۹۵۰۰۰۰۰۰۰
۱۹۹۹,۹۵۷۰۰۰۰۰۰۰
۲۰۰۰,۱۰۷۰۰۰۰۰۰۰
۲۰۰۱,۱۱۸۰۰۰۰۰۰۰
۲۰۰۲,۱۲۹۰۰۰۰۰۰۰
۲۰۰۳,۱۵۸۰۰۰۰۰۰۰
۲۰۰۴,۱۹۵۰۰۰۰۰۰۰
```

```
۲۰۰۵,۲۴۸۰۰۰۰۰۰۰
۲۰۰۶,۲۹۴۰۰۰۰۰۰۰
۲۰۰۷,۳۷۰۰۰۰۰۰۰۰
۲۰۰۸,۴۴۳۰۰۰۰۰۰۰
۲۰۰۹,۴۸۸۰۰۰۰۰۰۰
۲۰۱۰,۵۹۱۰۰۰۰۰۰۰
۲۰۱۱,۶۶۸۰۰۰۰۰۰۰
۲۰۱۲,۶۵۲۰۰۰۰۰۰۰
۲۰۱۳,۵۳۹۰۰۰۰۰۰۰
۲۰۱۴,۴۶۹۰۰۰۰۰۰۰
۲۰۱۵,۳۸۴۰۰۰۰۰۰۰
۲۰۱۶,۴۲۰۰۰۰۰۰۰۰
۲۰۱۷,۴۵۸۰۰۰۰۰۰۰
۲۰۱۸,۲۹۵۰۰۰۰۰۰۰
۲۰۱۹,۲۵۸۰۰۰۰۰۰۰
۲۰۲۰,۲۰۷۰۰۰۰۰۰۰
۲۰۲۱,۲۹۱۰۰۰۰۰۰۰
۲۰۲۲,۳۸۹۰۰۰۰۰۰۰
۲۰۲۳,۴۱۳۰۰۰۰۰۰۰
"""

try:
    # --- ۱. آماده‌سازی دیتافریم‌ها ---
    df_raw = pd.read_csv(io.StringIO(raw_data_csv))
    df_gdp = pd.read_csv(io.StringIO(gdp_data_csv), skiprows=4, names=['year', 'nominal_gdp'])

    # --- ۲. انجام تمام تبدیلات ---
    df_raw['year'] = df_raw['year'].astype(int)
    df_gdp['year'] = df_gdp['year'].astype(int)
    df = pd.merge(df_raw, df_gdp, on='year')
    df.set_index('year', inplace=True)

    model_df = pd.DataFrame(index=df.index)
    model_df['GDP_growth'] = df['GDP_growth']
    model_df['Trade_Openness'] = df['Trade_Openness']
    model_df['Gov_Rev_to_GDP'] = (df['Government_Revenue'] / df['nominal_gdp']) * 100
    model_df['D_Log_Exchange_Rate'] = np.log(df['Exchange_Rate']).diff()
    model_df['D_Log_MoneyBase'] = np.log(df['MoneyBase']).diff()
    model_df['D_Unemployment'] = df['Unemployment'].diff()
    model_df['D_Log_investment'] = np.log(df['investment']).diff()
    model_df['D_Degree_economic_freedom'] = df['Degree_economic_freedom'].diff()
    df['Gov_Exp_to_GDP'] = (df['Government_expenditure'] / df['nominal_gdp']) * 100
    model_df['D_Gov_Exp_to_GDP'] = df['Gov_Exp_to_GDP'].diff()
    model_df['D2_Deposit_Rate'] = df['Deposit_Rate'].diff().diff()
    model_df['D2_Log_CPI'] = np.log(df['CPI']).diff().diff()

    # --- ۳. ساخت مجموعه داده نهایی برای مدل‌سازی ---
    p = 1
    q = 1
    y = model_df['GDP_growth']
    X_exog = model_df.drop('GDP_growth', axis=1)
    final_X_features = pd.DataFrame(index=model_df.index)
    for i in range(1, p + 1):
        final_X_features[f'GDP_growth_lag_{i}'] = y.shift(i)
    for col in X_exog.columns:
        for i in range(q + 1):
            final_X_features[f'{col}_lag_{i}'] = X_exog[col].shift(i)

    final_data = pd.concat([y, final_X_features], axis=1).dropna()
    y_final = final_data['GDP_growth']
    X_final = final_data.drop('GDP_growth', axis=1)

    # --- ۴. تقسیم زمانی داده‌ها به آموزش و آزمون ---
    train_end_year = 2015
```

```
X_train = X_final.loc[:train_end_year]
y_train = y_final.loc[:train_end_year]
X_test = X_final.loc[train_end_year+1:]
y_test = y_final.loc[train_end_year+1:]

# --- ۵. آموزش و ارزیابی مدل‌ها ---
results_list [] =
models } =
   'Random Forest': RandomForestRegressor(n_estimators=100, random_state=42),
   'XGBoost': XGBRegressor(n_estimators=100, random_state=42)
{

for name, model in models.items:()
   model.fit(X_train, y_train)
   preds = model.predict(X_test)

   results_list.append{ )
      'Model': name,
      'RMSE': np.sqrt(mean_squared_error(y_test, preds)),
      'MAE': mean_absolute_error(y_test, preds)
   ({

results_df = pd.DataFrame(results_list)

# --- ۶. نمایش جدول مقایسه ---
print("--- جدول مقایسه عملکرد مدل‌ها (بر روی داده‌های آزمون) ---")
print(results_df.to_markdown(index=False, floatfmt=".4f"))

# --- ۷. استخراج و نمایش اهمیت ویژگی‌ها ---
print\")nاهمیت ویژگی‌ها --- (Feature Importances) ("---

# استخراج اهمیت ویژگی‌ها از هر دو مدل
rf_model = models['Random Forest']
xgb_model = models['XGBoost']

importances_df = pd.DataFrame{ )
   'Feature': X_train.columns,
   'Random Forest Importance': rf_model.feature_importances_،
   'XGBoost Importance': xgb_model.feature_importances_
({

# نرمال‌سازی اهمیت‌ها تا جمع آن‌ها ۱ شود
importances_df['Random Forest Importance'] /= importances_df['Random Forest Importance'].sum()
importances_df['XGBoost Importance'] /= importances_df['XGBoost Importance'].sum()

# مرتب‌سازی بر اساس میانگین اهمیت
importances_df['Average Importance'] = importances_df[['Random Forest Importance', 'XGBoost Importance']].mean(axis=1)
importances_df = importances_df.sort_values(by='Average Importance', ascending=False).reset_index(drop=True)

print(importances_df[['Feature', 'Random Forest Importance', 'XGBoost Importance', 'Average Importance']].to_markdown(index=False, floatfmt=".4f"))

# --- ۸. ترسیم نمودار اهمیت ویژگی‌ها ---
plt.figure(figsize=(12, 10))
# انتخاب ۱۰ ویژگی برتر
top_features = importances_df.head(۱۰)

# تبدیل داده برای ترسیم کنار هم
plot_df = pd.melt(top_features, id_vars='Feature','
            value_vars=['Random Forest Importance', 'XGBoost Importance'],
            var_name='Model', value_name='Importance('

sns.barplot(data=plot_df, x='Importance', y='Feature', hue='Model', palette='viridis')
plt.title('مقایسه اهمیت ۱۰ ویژگی برتر در مدل‌های جنگل تصادفی و XGBoost(')
plt.xlabel('میزان اهمیت نرمال‌شده')
plt.ylabel('ویژگی (متغیر)')
```

```
    plt.legend)title('مدل'=
    plt.tight_layout()
    plt.savefig('feature_importances_comparison.png')
    print\"n' نمودار اهمیت ویژگی‌ها در فایل feature_importances_comparison.png." ذخیره شد.')

except Exception as e:
    print)f:داد رخ غیرمنتظره خطای یک" {e}")
```